

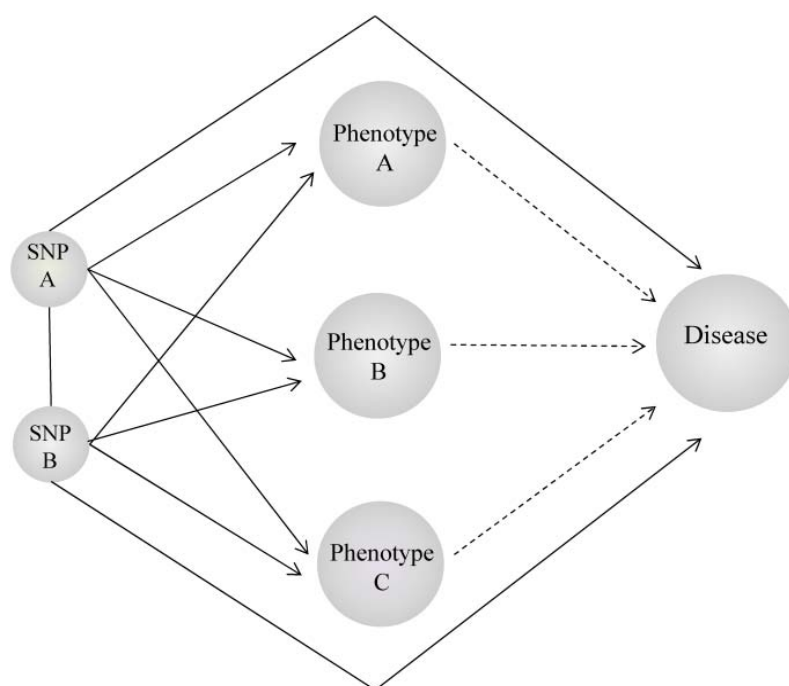


ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ
ΣΤΗ ΒΙΟΪΑΤΡΙΚΗ**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΜΕΘΟΔΟΛΟΓΙΑ ΓΕΝΕΤΙΚΗΣ ΕΠΙΔΗΜΙΟΛΟΓΙΑΣ ΚΑΙ ΜΕΤΑ-ΑΝΑΛΥΣΗΣ ΓΕΝΕΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ



ΝΙΚΗ ΔΗΜΟΥ

ΛΑΜΙΑ 2016

Συμβουλευτική επιτροπή

- | | |
|----------------------------------------------------------|---------------------------------------------------------------------------|
| 1) Αναπληρωτής Καθηγητής
Παντελής Μπάγκος (Επιβλέπων) | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 2) Αναπληρωτής Καθηγητής
Βασίλης Πλαγιανάκος | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 3) Επίκουρη Καθηγήτρια
Μαρία Αδάμ | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |

Εξεταστική επιτροπή

- | | |
|----------------------------------------------------------|------------------------------------------------------------------------------|
| 1) Αναπληρωτής Καθηγητής
Παντελής Μπάγκος (Επιβλέπων) | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 2) Αναπληρωτής Καθηγητής
Βασίλης Πλαγιανάκος | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 3) Επίκουρη Καθηγήτρια
Μαρία Αδάμ | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 4) Αναπληρωτής Καθηγητής
Δημήτριος Ιακωβίδης | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 5) Επίκουρος Καθηγητής
Κωνσταντίνος Δελήμπασης | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 6) Επίκουρος Καθηγητής
Ευάγγελος Ευαγγέλου | Τομέας Υγιεινής και Επιδημιολογίας,
Ιατρική Σχολή, Πανεπιστήμιο Ιωαννίνων |
| 7) Λέκτορας
Ιωάννης Τριανταφύλλου | Τμήμα Πληροφορικής με Εφαρμογές στη
Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |

Αφιερωμένο σε όσους είναι ακόμα δίπλα μου

Πρόλογος

Η παρούσα Διδακτορική Διατριβή πραγματοποιήθηκε στο Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική του Πανεπιστημίου Θεσσαλίας υπό την επίβλεψη του Αναπληρωτή Καθηγητή κ. Παντελή Μπάγκου και μέλη της Τριμελούς Συμβουλευτικής Επιτροπής τον Αναπληρωτή Καθηγητή κ. Βασίλη Πλαγιανάκο και την Επίκουρη Καθηγήτρια κ. Μαρία Αδάμ.

Κατ' αρχήν οφείλω να ευχαριστήσω τον Αναπληρωτή Καθηγητή κ. Παντελή Μπάγκο για τα περίπου δέκα χρόνια που δουλεύουμε μαζί. Αρχικά ως προπτυχιακή φοιτήτρια στο Τμήμα Βιολογίας του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, κατόπιν στο μεταπτυχιακό της Βιοστατιστικής που διοργανώνεται από τα Τμήματα Μαθηματικών και την Ιατρική Σχολή του Πανεπιστημίου Αθηνών έως τώρα στη διδακτορική διατριβή ήταν επιβλέπων καθηγητής μου. Σε κάθε στάδιο, όλα αυτά τα χρόνια, ήταν διαρκώς δίπλα μου και θεωρώ ότι όσα έχω καταφέρει έως τώρα τα οφείλω σε αρκετό βαθμό στη δική του υπομονή και επιμονή.

Επίσης, ευχαριστώ τον Αναπληρωτή Καθηγητή κ. Βασίλη Πλαγιανάκο και την Επίκουρη Καθηγήτρια κ. Μαρία Αδάμ για τη βοήθειά τους σε πρακτικά ζητήματα που προέκυψαν κατά τη διάρκεια της διατριβής με χρήσιμες συμβουλές και παραινέσεις. Η παρουσία και κριτική ανάγνωση της διδακτορικής διατριβής από τα αξιότιμα μέλη της Επταμελούς Εξεταστικής Επιτροπής του Αναπληρωτή Καθηγητή κ. Δημητρίου Ιακωβίδη, του Επίκουρου Καθηγητή κ. Κωνσταντίνου Δελήμπαση και του Λέκτορα κ. Ιωάννη Τριανταφύλλου με τιμά ιδιαίτερα και τους ευχαριστώ θερμά. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον Επίκουρο Καθηγητή κ. Ευάγγελο Ευαγγέλου για τη συνεργασία που είχαμε στα πλαίσια της διδακτορικής διατριβής και για την εμπιστοσύνη που μου επέδειξε να με συμπεριλάβει πλέον στην ερευνητική του ομάδα στο Πανεπιστήμιο Ιωαννίνων. Δεν θα μπορούσα να παραλείψω να ευχαριστήσω τον Ομότιμο Καθηγητή κ. Σταύρο Χαμόδρακα του Τμήματος Βιολογίας του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών που με επέλεξε να είμαι στην ερευνητική του ομάδα για την εκπόνηση της προπτυχιακής διπλωματικής μου εργασίας διδάσκοντάς μου την ακρίβεια και την ορθότητα με την οποία πρέπει να διατυπώνουμε και να γράφουμε στην επιστημονική έρευνα.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω τα μέλη της ερευνητικής ομάδας του Αναπληρωτή Καθηγητή κ. Παντελή Μπάγκου με τα οποία συνεργάστηκα και είχα διαρκή αλληλεπίδραση όλα αυτά τα χρόνια. Πιο συγκεκριμένα, θα ήθελα να ευχαριστήσω την κ. Γκρέτα Μπράλιου, κ. Παυλοπούλου Αθανασία και τον κ. Νίκο Νικολακάκη για τη διαρκή ανταλλαγή ιδεών και απόψεων. Ιδιαίτερα ευχαριστώ την κ. Παναγιώτα Κοντού γιατί εκτός από τη συνεργασία σε επιστημονικά θέματα αντιπροσώπευε πάντα και τις δυο μας σε οποιαδήποτε γραφειοκρατική διαδικασία χρειάστηκε. Επίσης, ευχαριστώ την κ. Κατερίνα Πανταβού όχι μόνο γιατί διαθέτει εξαιρετικό πνεύμα συνεργασίας και προθυμίας αλλά κυρίως γιατί λόγω του χαρακτήρα της με κάνει να μην ξεχνάω ότι ακόμα και στις μέρες υπάρχει ανθρωπιά και έλλειψη υστεροβουλίας.

Ευχαριστίες επίσης πρέπει να απευθύνω στον κ. Γιώργο Νικολόπουλο όχι μόνο για τις εργασίες που δουλέψαμε μαζί αλλά και για την εμπιστοσύνη που μου έδειχνε να με συστήνει σε οποιαδήποτε δουλειά του παρουσιαζόταν. Επίσης, ευχαριστώ την κ. Χρύσα Τσιάρα για την πολύτιμη συνεργασία μας που έχει οδηγήσει ήδη σε εξαιρετικά αποτελέσματα.

Σημαντική ήταν η συνεισφορά των φίλων Κώστα, Εβίτα, Γκάλια, Κατερίνα, Μιχάλη, Κατερίνα, Κατερίνα, Αλεξία, Κώστα που με ενθάρρυναν να συνεχίζω να προσπαθώ και της οικογένειάς μου για την αμέριστη ηθική συμπαράσταση, υπομονή,

αγάπη, κατανόηση και στήριξη σε οτιδήποτε επιχειρώ. Τέλος, θέλω να ευχαριστήσω το σύντροφό μου Κώστα για την υπομονή και αγάπη που μου δείχνει που μπορεί να μη καταλαβαίνει ακριβώς τι είναι το κόκκινο μήνυμα που κοιτάω με απελπισία μπροστά από μια μαύρη οθόνη αλλά πάντα βρίσκει τον τρόπο να μου θυμίζει ότι το μόνο που έχει σημασία είμαι εγώ.

Δημοσιεύσεις

- Κεφάλαιο 1: 1) Dimou NL, Pantavou K, Bagos PG. **Multivariate methods for meta-analysis in genetic association studies: A review.** (σε προετοιμασία)
- Κεφάλαιο 2: 2) Dimou NL, Tsirigos KD, Elofsson A, Bagos PG. **GWAR: Robust Analysis and Meta-Analysis of Genome-Wide Association Studies.** *Bioinformatics* (υπό αναθεώρηση)
- Κεφάλαιο 3: 3) Bagos PG, Dimou NL, Liakopoulos TD, Nikolopoulos GK. **Meta-analysis of Family-based and Case-control Genetic Association Studies That Use the Same Cases.** 2011, *Statistical Applications in Genetics and Molecular Biology*, **10**(1):Article19
4) Pantavou K, Braliou GG, Kontou PI, Dimou NL, Bagos PG. **A meta-analysis of FZD3 polymorphisms and their association with schizophrenia.** *Psychiatric Genetics* (υπό αναθεώρηση)
- Κεφάλαιο 4: 5) Dimou NL and Bagos PG. **A Multivariate Method for Meta-Analysis of Multiple Outcomes in Genetic Association Studies.** (σε προετοιμασία)
6) C. Medina Gomez, J.P Kemp, N. L. Dimou, E. Kreiner-Møller, A. Chesi, D. H. M. Heppe, B. S Zemel, K. Bønnelykke, C. Boer, H. Bisgaard, E. Evangelou, V. W. V. Jaddoe, A. G. Uitterlinden, J. H Tobias, G. Duque, S.F.A Grant, P.G. Bagos, D.M. Evans, F. Rivadeneira. **Variants in regulatory regions of SREBF1 exert pleiotropic effects on BMD and lean mass in children.** (σε προετοιμασία)
7) Dimou NL, Pantavou K, Bagos PG. **Apolipoprotein E polymorphism and left ventricular failure in beta-thalassemia: A meta-analysis.** *Annals of Human Genetics* (υπό αναθεώρηση)
- Κεφάλαιο 5: 8) Tsiara C, Nikolopoulos G, Dimou N, Bagos P, Mensah B, Talias M, Bonovas S, Hatzakis A. **Interleukin Gene Polymorphisms and Susceptibility to HIV-1 Infection: A Systematic Review and Meta-Analysis.** (σε προετοιμασία)
9) Tsiara C, Nikolopoulos GK, Dimou N, Bagos P, Saroglou G, Velonakis E, Hatzakis A. **Effect of Hepatitis C Virus on Immunologic and Virologic Response in HIV Infected Patients Initiating Highly Active Antiretroviral Therapy: An Updated Meta-Analysis** 2013, *Journal of Viral Hepatitis*, **20**(10):715-24
10) Spyridopoulou K, Dimou NL, Hamodrakas SJ, Bagos PG. **Methylene Tetrahydrofolate Reductase gene polymorphisms and their association with methotrexate toxicity: A meta-analysis.** 2012, *Pharmacogenetics and Genomics* **22**(2):117-33
- Κεφάλαιο 6: 11) Dimou NL, Adam M, Bagos PG. **A Multivariate Method for Meta-analysis and Comparison of Diagnostic Tests.** 2016, *Statistics in Medicine*, **35**(20):3509-23

Χρηματοδότηση

- 16/11/2015-30/12/2015: **«Πιλοτικό Πληροφοριακό Σύστημα Καταγραφής Ασθενών»**
Πανεπιστήμιο Θεσσαλίας για λογαριασμό της Abbvie Pharmaceuticals S.A..
- 01/05/2014-28/02/2015: **«Ενοποίηση δεδομένων από διαφορετικές πηγές: μια σύνθεση της επιδημιολογίας με τη βιοπληροφορική, με εφαρμογές στις πολυπαραγοντικές ασθένειες»**
Συγχρηματοδοτήθηκε από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους στο πλαίσιο της πράξης "ΑΡΙΣΤΕΙΑ II" του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Διά Βίου Μάθηση».
- 15/10/2012-04/03/2015: **«Ολοκληρωμένη μελέτη του τρόπου δράσης μεμβρανικών υποδοχέων και της εμπλοκής τους σε ασθένειες με σύγχρονες μεθόδους βιοπληροφορικής»,**
Συγχρηματοδοτήθηκε από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης-ΕΤΠΑ) και από εθνικούς πόρους στο πλαίσιο της δράσης εθνικής εμβέλειας «ΣΥΝΕΡΓΑΣΙΑ»-Πράξη I στα Περιφερειακά Επιχειρησιακά Προγράμματα (ΠΕΠ) Αττικής , Α.Π. 3 «Ενίσχυση της ανταγωνιστικότητας, της καινοτομίας και της Ψηφιακής Σύγκλισης » και (ΠΕΠ) Θεσσαλίας-Στερεάς Ελλάδας-Ηπείρου, Α.Π. 8 «Ψηφιακή Σύγκλιση και Επιχειρηματικότητα Στερεάς Ελλάδας» (Κωδικός 09SYN-13-999)

Περιεχόμενα

Περίληψη	1
Abstract	3
Κεφάλαιο 1: Εισαγωγή	5
1.1. Επιδημιολογία και γενετική επιδημιολογία	5
1.2. Γενετική στη γενετική επιδημιολογία	7
1.2.1. Το γενετικό υλικό	7
1.2.2. Γενετική ποικιλότητα	9
1.3. Μελέτες γενετικής συσχέτισης	11
1.4. Σχεδιασμοί μελετών γενετικής συσχέτισης	13
1.5. Προκαταρκτικοί έλεγχοι σε μελέτες γενετικής συσχέτισης	15
1.5.1. Έλεγχος της ποιότητας αλληλούχισης	15
1.5.2. Ισορροπία Hardy-Weinberg	15
1.5.3. Ελλείπουσες τιμές στους γονοτύπους	16
1.5.4. Πληθυσμιακή διαστρωμάτωση	16
1.6. Μέθοδοι ανάλυσης σε μελέτες γενετικής συσχέτισης	17
1.7. Η μετα-ανάλυση στις μελέτες γενετικής συσχέτισης	19
1.7.1. Μοντέλα σταθερών και τυχαίων επιδράσεων στη μετα-ανάλυση	20
1.7.2. Η πολυμεταβλητή προσέγγιση	22
1.7.3. Η πολυμεταβλητή κανονική κατανομή	24
1.7.4. Η πολυμεταβλητή μετα-ανάλυση	25
1.7.5. Υπολογισμός των εντός των μελετών συνδιακυμάνσεων	27
1.7.6. Μοντέλα μετα-ανάλυσης για τη συσχέτιση ενός γονιδίου με μια ασθένεια	28
1.7.7. Μοντέλα μετα-ανάλυσης για τη συσχέτιση μελετών με απλοτύπους	31
1.7.8. Μοντέλα μετα-ανάλυσης για πολλαπλούς γονιδιακούς πολυμορφισμούς που βρίσκονται σε ανισορροπία σύνδεσης	32
1.7.9. Μοντέλα μετα-ανάλυσης πολλαπλών εκβάσεων	34
1.7.10. Μοντέλα μετα-ανάλυσης Μεντελιανής τυχαιοποίησης	36
1.8. Στόχοι της διδακτορικής διατριβής	38
Κεφάλαιο 2: Robust μέθοδοι ανάλυσης και μετα-ανάλυσης σε μελέτες γενετικής συσχέτισης και σε ευρυγονιδιωματικές μελέτες	41
2.1. Εισαγωγή	41
2.2. Συμβολισμός	41
2.3. Στατιστικοί έλεγχοι	42
2.4. Robust μέθοδοι για την επιλογή γενετικού μοντέλου	44
2.4.1. Οι μέθοδοι MAX και MERT	44
2.4.2. Οι μέθοδοι GMS και GME	46
2.4.3. Η μέθοδος MIN2	47
2.5. Μέθοδοι ανάλυσης ποσοτικών χαρακτηριστικών	48
2.6. Μετα-ανάλυση μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών	49
2.7. Υλοποίηση robust μεθόδων ανάλυσης και μετα-ανάλυσης μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών	53
2.7.1. Υλοποίηση	53
2.7.2. Χρήση	53
2.7.3. Επίδοση	54
2.7.4. Εφαρμογή σε πραγματικά δεδομένα	54
2.8. Συζήτηση	57

Κεφάλαιο 3: Ανάπτυξη μεθοδολογίας μετα-ανάλυσης μελετών γενετικής συσχέτισης σε οικογένειες και ασθενών-μαρτύρων που χρησιμοποιούν τους ίδιους ασθενείς και εφαρμογές	59
Μέρος Α: Μεθοδολογία μετα-ανάλυσης μελετών γενετικής συσχέτισης σε οικογένειες και ασθενών-μαρτύρων που χρησιμοποιούν τους ίδιους ασθενείς	59
3.1. Εισαγωγή	59
3.2. Μεθοδολογία.....	61
3.2.1. Μελέτες ασθενών-μαρτύρων σε πληθυσμούς	61
3.2.2. Μελέτες ασθενών-μαρτύρων σε οικογένειες.....	62
3.2.2.1. Ο έλεγχος transmission disequilibrium test	62
3.2.2.2. Ο έλεγχος haplotype-based haplotype relative risk	63
3.2.2.3. Επεκτάσεις του ελέγχου transmission disequilibrium test.....	64
3.2.3. Μεθοδολογίες για μετα-ανάλυση	65
3.2.4. Υπολογισμός της συνδιακύμανσης.....	68
3.2.5. Υλοποίηση	71
3.3. Αποτελέσματα.....	72
3.4. Συζήτηση.....	79
Μέρος Β: Μετα-ανάλυση πολυμορφισμών του γονιδίου FZD3 και η συσχέτισή τους με την εμφάνιση Σχιζοφρένειας.....	84
3.5. Εισαγωγή	84
3.6. Μεθοδολογία.....	85
3.6.1. Πηγές δεδομένων, κριτήρια επιλεξιμότητας και εξαγωγή δεδομένων	85
3.6.2. Στατιστική ανάλυση.....	86
3.7. Αποτελέσματα.....	87
3.7.1. Ο πολυμορφισμός rs2241802	91
3.7.2. Ο πολυμορφισμός rs2323019	91
3.7.3. Ο πολυμορφισμός rs352203	92
3.7.4. Ο πολυμορφισμός rs3757888	93
3.7.5. Ο πολυμορφισμός rs880481	93
3.7.6. Ο πολυμορφισμός rs960914	93
3.7.7. Ανισορροπία σύνδεσης.....	94
3.8. Συζήτηση.....	99
Κεφάλαιο 4: Ανάπτυξη μεθοδολογίας μετα-ανάλυσης πολλαπλών εκβάσεων σε μελέτες γενετικής συσχέτισης και εφαρμογές	101
Μέρος Α: Πολυμεταβλητή μέθοδος μετα-ανάλυσης πολλαπλών εκβάσεων σε μελέτες γενετικής συσχέτισης.....	101
4.1. Εισαγωγή	101
4.2. Μεθοδολογία.....	103
4.2.1. Το γενικό πολυμεταβλητό μοντέλο μετα-ανάλυσης πολλαπλών διακριτών εκβάσεων	103
4.2.2. Ο υπολογισμός των εντός των μελετών συνδιακυμάνσεων	107
4.2.3. Διαχείριση ελλειπουσών πληροφοριών	110
4.2.4. Το γενικό πολυμεταβλητό μοντέλο μετα-ανάλυσης πολλαπλών συνεχών εκβάσεων	111
4.3. Εφαρμογή της μεθόδου.....	114
4.3.1. Διακριτές εκβάσεις	114
4.3.2. Συνεχείς εκβάσεις	118
4.4. Συζήτηση.....	122

Μέρος Β: Διερεύνηση της συσχέτισης πολυμορφισμών της απολιποπρωτεΐνης Ε με την ανεπάρκεια αριστερής κοιλίας σε ασθενείς με βήτα θαλασσαιμία-μία πολυμεταβλητή μετα-ανάλυση	126
4.5. Εισαγωγή	126
4.6. Μεθοδολογία.....	128
4.6.1. Ανάκτηση δημοσιευμένων μελετών	128
4.6.2. Εξαγωγή δεδομένων	129
4.6.3. Στατιστική ανάλυση.....	129
4.7. Αποτελέσματα.....	134
4.8. Συζήτηση.....	138
Κεφάλαιο 5: Μετα-αναλύσεις μελετών γενετικής συσχέτισης, κλινικών μελετών και μελετών φαρμακογονιδιοματικής.....	141
Μέρος Α: Πολυμορφισμοί των γονιδίων των ιντερλευκινών και προδιάθεση για μόλυνση από τον HIV-1	141
5.1. Εισαγωγή	141
5.2. Μεθοδολογία.....	142
5.2.1. Κριτήρια επιλεξιμότητας, πηγές δεδομένων και στρατηγική αναζήτησης	142
5.2.2. Εξαγωγή δεδομένων και αξιολόγηση ποιότητας	143
5.2.3. Στατιστική ανάλυση.....	143
5.3. Αποτελέσματα.....	144
5.3.1. Περιγραφή των μελετών	144
5.3.2. Προδιάθεση στη μόλυνση από τον HIV	156
5.4. Συζήτηση.....	163
Μέρος Β: Μελέτη της επίδρασης της συλλοίμωξης με τον ιό της Ηπατίτιδας C στην ανοσολογική και ιολογική ανταπόκριση HIV ασθενών μετά την έναρξη αντιρετροϊκής θεραπείας υψηλής δραστηριότητας.....	168
5.5. Εισαγωγή	168
5.6. Μεθοδολογία.....	169
5.6.1. Κριτήρια επιλεξιμότητας, πηγές δεδομένων και διαδικασία αναζήτησης	169
5.6.2. Εξαγωγή δεδομένων	169
5.6.3. Στατιστική ανάλυση.....	170
5.7. Αποτελέσματα.....	171
5.7.1. Περιγραφή των μελετών	171
5.7.2. Ανοσολογική και ιολογική ανταπόκριση.....	178
5.8. Συζήτηση.....	182
Μέρος Γ: Διερεύνηση της συσχέτισης πολυμορφισμών του γονιδίου της αναγωγάσης του μεθυλενοτετραϋδροφυλλικού με την τοξικότητα από μεθοτρεξάτη	185
5.9. Εισαγωγή	185
5.10. Μεθοδολογία.....	186
5.10.1. Επιλογή μελετών.....	186
5.10.2. Κριτήρια επιλεξιμότητας και αποκλεισμού μελετών.....	187
5.10.3. Εξαγωγή δεδομένων	187
5.10.4. Στατιστική ανάλυση.....	188
5.11. Αποτελέσματα.....	189
5.11.1. Ο πολυμορφισμός C677T του γονιδίου της MTHFR (rs1801133)	189
5.11.2. Ο πολυμορφισμός A1298C του γονιδίου της MTHFR (rs1801131).....	197
5.12. Συζήτηση.....	199

Κεφάλαιο 6: Πολυμεταβλητή μεθοδολογία μετα-ανάλυσης και σύγκρισης διαγνωστικών δοκιμασιών	203
6.1. Εισαγωγή	203
6.2. Μεθοδολογία.....	207
6.2.1. Το πολυμεταβλητό μοντέλο.....	207
6.2.2. Υπολογισμός των εντός των μελετών συνδιακυμάνσεων	210
6.2.3. Διαχείριση ελλειπουσών πληροφοριών	211
6.2.4. Η μέθοδος SROC και η AUC	212
6.3. Μελέτη προσομοίωσης	216
6.4. Εφαρμογή της μεθοδολογίας	217
6.5. Συζήτηση.....	221
Παραρτήματα.....	227
Παράρτημα Α.....	227
Παράρτημα Β.....	230
Παράρτημα Γ	234
Παράρτημα Δ.....	238
Παράρτημα Ε	255
Παράρτημα ΣΤ.....	258
Παράρτημα Ζ	262
Παράρτημα Η.....	275
Παράρτημα Θ.....	278
Βιβλιογραφία	281
Δημοσιεύσεις	315

Περίληψη

Η γενετική επιδημιολογία αποτελεί τη συνένωση της κλασσικής επιδημιολογίας με τη γενετική. Σε αυτό το χώρο εφαρμόζονται οι κλασσικές επιδημιολογικές μέθοδοι έτσι ώστε να διαλευκανθεί η πιθανή επίδραση της γενετικής ποικιλομορφίας στην αιτιολογία της νόσου. Ο συνεχώς αυξανόμενος αριθμός δημοσιευμένων μελετών γενετικής συσχέτισης καθιστά επιτακτική τη συλλογή και το συνδυασμό της διαθέσιμης πληροφορίας μέσω μια στατιστικής διαδικασίας γνωστής ως μετα-ανάλυση. Η μετα-ανάλυση είναι μια ποσοτική ανασκόπηση και σύνθεση των αποτελεσμάτων που προκύπτουν από σχετιζόμενες αλλά και ανεξάρτητες μεταξύ τους μελέτες

Σε πρώτη φάση, αναπτύχθηκαν εργαλεία για την robust (στιβαρή) ανάλυση και μετα-ανάλυση ευρυγονιδιωματικών μελετών χρησιμοποιώντας το στατιστικό πακέτο Stata. Οι μεθοδολογίες αυτές, δεν κάνουν καμιά εκ των προτέρων υπόθεση για το μοντέλο κληρονομικότητας και κατά συνέπεια έχουν πάντα μέγιστη ισχύ, με διατήρηση του σφάλματος τύπου I στο 5%. Υλοποιήσαμε τόσο τους κλασσικούς ελέγχους Cochran-Armitage trend tests καθώς επίσης και τις ευρέως χρησιμοποιούμενες robust προσεγγίσεις, MERT, MAX και MIN2. Για τους ελέγχους MAX και MIN2, που παρουσιάζουν αυξημένη ισχύ συγκρινόμενοι με άλλες robust προσεγγίσεις, υλοποιήσαμε ακριβείς και γρήγορες ρουτίνες που στηρίζονται σε αριθμητική ολοκλήρωση για τον υπολογισμό των p-values με μεγάλη ακρίβεια. Η εντολή GWAR υλοποιήθηκε σαν ado file για Stata 12 ή μεγαλύτερη έκδοση αυτής, καθώς επίσης και σαν διαδικτυακή εφαρμογή που βρίσκεται στην ιστοσελίδα www.comp-gen.org/tools/GWAR.

Ακολούθως, προτάθηκε μια απλή μέθοδος μετα-ανάλυσης συνοπτικών δεδομένων για την περίπτωση όπου πραγματοποιείται ταυτόχρονα στον ίδιο πληθυσμό μια μελέτη ασθενών-μαρτύρων σε οικογένειες και σε πληθυσμούς με τους ίδιους ή επικαλυπτόμενους ασθενείς. Το σημείο κλειδί, σε αντίθεση με τις υπάρχουσες μεθοδολογίες που χρησιμοποιούν ατομικά δεδομένα, αποτελεί ο υπολογισμός της συνδιακύμανσης των εκτιμήσεων των μελετών, χρησιμοποιώντας μόνο τα δεδομένα που υπάρχουν στη βιβλιογραφία με τη μορφή ενός 2x2 πίνακα συνάφειας. Ακολούθως, οι μελέτες μπορούν να συνδυαστούν είτε μέσω μιας διαδικασίας δυο βημάτων χρησιμοποιώντας κλασσικές μεθοδολογίες μετα-ανάλυσης είτε με ένα βήμα με ιεραρχικά μοντέλα. Σε κάθε περίπτωση, η μετα-ανάλυση μπορεί να πραγματοποιηθεί με ένα οποιοδήποτε λογισμικό και δεδομένου του αυξημένου μεγέθους δείγματος η ισχύς της μετα-ανάλυσης μεγαλώνει. Η προσέγγιση αυτή επιτρέπει την πραγματοποίηση πολλών διαγνωστικών ελέγχων (συστηματικό σφάλμα δημοσίευσης, αθροιστική μετα-ανάλυση, ανάλυση ευαισθησίας). Η προτεινόμενη μεθοδολογία, εφαρμόστηκε σε ένα σύνολο εννέα μελετών σε οικογένειες και σε πληθυσμούς για τη διερεύνηση της συσχέτισης πολυμορφισμών του γονιδίου FZD3 με την σχιζοφρένεια. Συνολικά εξετάστηκαν 6 πολυμορφισμοί (rs2241802, rs2323019, rs352203, rs3757884, rs880481 και rs960914). Η παρούσα μετα-ανάλυση, υπογράμμισε τον ρόλο του πολυμορφισμού rs352203. Πιο συγκεκριμένα η παρουσία του αλληλομόρφου C έχει προστατευτικό ρόλο για την εμφάνιση της νόσου. Εκτιμήθηκε επίσης ένας αυξημένος κίνδυνος για τους φορείς του αλληλομόρφου A του πολυμορφισμού rs2323019 στους Κινέζους, κάτι που μπορεί να αποδοθεί στο διαφορετικό βαθμό ανισορροπίας σύνδεσης που υπολογίστηκε για τη φυλή αυτή.

Επιπρόσθετα, ένας από τους κύριους άξονες της παρούσας διδακτορικής διατριβής, ήταν η ανάπτυξη μεθοδολογίας για την ταυτόχρονη μετα-ανάλυση μελετών γενετικής συσχέτισης που να περιλαμβάνουν δύο ή περισσότερες εκβάσεις.

Η μέθοδος που προτάθηκε είναι σχετικά απλή αλλά παρουσιάζει αυξημένη ισχύ για την πραγματοποίηση πολυμεταβλητής μετα-ανάλυσης σε μελέτες γενετικής συσχέτισης. Αποτελεί μια απευθείας επέκταση της διμεταβλητής μεθόδου μετα-ανάλυσης μελετών γενετικής συσχέτισης τόσο για συνεχείς όσο και για διακριτές εκβάσεις. Η καινοτομία της μεθόδου έγκειται στο γεγονός ότι παρέχουμε αναλυτικές εκφράσεις για τον υπολογισμό των εντός των μελετών συνδιακυμάνσεων χρησιμοποιώντας την ταξινόμηση των ατόμων που έχουν τις δυο εκβάσεις ανά γονότυπο, η οποία προκύπτει εφαρμόζοντας ένα λογαριθμογραμμικό μοντέλο (log-linear) μέσω του αλγορίθμου iterative proportional fitting υπό τη θεώρηση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης. Η μέθοδος είναι γρήγορη και μπορεί να επεκταθεί για περισσότερες από δύο εκβάσεις. Δεν απαιτούνται ατομικά δεδομένα ή η ταυτόχρονη εκτίμηση παραμέτρων και για τις δυο εκβάσεις. Η μεθοδολογία αυτή εφαρμόστηκε για τις συνεχείς εκβάσεις Total-body lean mass (TB-LM) και Bone mineral density (TB-BMD) έτσι ώστε να διερευνηθούν οι πιθανές πλειοτροπικές επιδράσεις και να διαλευκανθούν τα μονοπάτια που αυτά εμπλέκονται κατά τη διαμόρφωση του σκελετικού συστήματος. Συνολικά, αξιολογήθηκαν περισσότεροι από 2.5 εκατομμύρια γονιδιακοί πολυμορφισμοί τεσσάρων μελετών κοορτής παιδικού πληθυσμού αποκαλύπτοντας 8 γενετικούς τόπους με $p\text{-values} < 5 \times 10^{-8}$.

Επέκταση της προαναφερθείσας μεθοδολογίας, αποτελεί η περίπτωση όπου υπάρχουν μελέτες με περισσότερες από μια εκβάσεις και γονίδια με πολλαπλά αλληλομόρφα και έτσι καλούμαστε να υιοθετήσουμε ένα πιο σύνθετο σχήμα πολυμεταβλητής μεθοδολογίας μετα-ανάλυσης. Μια τέτοια περίπτωση αποτελεί η διερεύνηση της συσχέτισης του γονιδίου της απολιποπρωτεΐνης E (πολλαπλά αλληλομόρφα) με την ανεπάρκεια αριστερής κοιλίας σε ασθενείς με βήτα-θαλασσαιμία (οι ασθενείς ταξινομούνται σε τρεις ομάδες ανάλογα με την σοβαρότητα των συμπτωμάτων) όπου δείξαμε στατιστικά σημαντική συσχέτιση του αλληλομόρφου E4. Επιπλέον, σημαντικά ήταν τα ευρήματα τόσο των πολυμεταβλητών μεθοδολογιών μετα-ανάλυσης όσο και των κλασικών προσεγγίσεων για τη διερεύνηση της συσχέτισης πολυμορφισμών των γονιδίων των ιντερλευκινών με την προδιάθεση για μόλυνση από τον HIV-1, τη μελέτη της επίδρασης της συλλοίμωξης με τον ιό της Ηπατίτιδας C στην ανοσολογική και ιολογική ανταπόκριση HIV ασθενών μετά την έναρξη αντιρετροϊκής θεραπείας υψηλής δραστηριότητας και τη διερεύνηση της συσχέτισης πολυμορφισμών του γονιδίου της αναγωγάσης του μεθυλενοτετραϋδροφυλλικού με την τοξικότητα από μεθοτρεξάτη.

Τέλος, ένα άλλο πεδίο που εφαρμόστηκε η πολυμεταβλητή μεθοδολογία μετα-ανάλυσης, περιλαμβάνει την ανάπτυξη πρωτότυπης μεθοδολογίας και λογισμικού για την ταυτόχρονη διενέργεια μετα-ανάλυσης η οποία συγκρίνει, δύο ή περισσότερες διαγνωστικές δοκιμασίες. Με τη μέθοδο αυτή, επιτρέπεται ο απευθείας υπολογισμός της ευαισθησίας, της ειδικότητας και της επιφάνειας κάτω από την καμπύλη ROC (Area Under the Curve-AUC) των διαγνωστικών δοκιμασιών, καθώς επίσης είναι δυνατές και οι άμεσες συγκρίσεις των παραμέτρων αυτών. Δεν απαιτούνται ατομικά δεδομένα όπως επίσης δεν είναι αναγκαίο σε όλες τις μελέτες να έχουν εξεταστεί ταυτόχρονα και οι δύο δοκιμασίες. Η προτεινόμενη μεθοδολογία αξιολογήθηκε μέσω μελετών προσομοίωσης και βρέθηκε ότι είναι αμερόληπτη αλλά και πιο ισχυρή από την κλασική μέθοδο.

Abstract

Genetic epidemiology is the fusion of traditional epidemiology with genetics, in which epidemiological methods are applied in order to uncover the potential role of genetic variants in the aetiology of disease. The continuously increasing number of published genetic epidemiology studies warrants the need for collecting and synthesizing the available evidence providing a quantitative overall estimate, a procedure known as meta-analysis. Meta-analysis can be broadly defined as the quantitative review and synthesis of the results of related but independent studies.

Firstly, tools for robust analysis and meta-analysis in genome-wide association studies using Stata were developed. These methods do not rely on pre-specifying the genetic model of inheritance and thus they have maximum power and, at the same time, preserve the nominal type I error rate. Both standard Cochran-Armitage trend tests as well as the most widely used robust methods, namely MERT, MAX and MIN2 were implemented. For MAX and MIN2, which are also the most powerful among the robust methods, accurate and fast routines that rely on numerical integration in order to calculate accurately the p-value were implemented. GVAR is implemented as an ado file for Stata 12 or higher, as well as a web-server available in www.compgen.org/tools/GVAR.

Subsequently, a simple and efficient summary-based method for performing a meta-analysis was proposed to account for the case when both a family-based and a population-based case-control study are conducted within the same population, using the same or largely overlapping, set of cases. The key point, contrary to the methods presented earlier that need individual data, is the calculation of the covariance between the study estimates (log-Odds Ratios), using only data derived from the literature in the form of a 2x2 contingency table. Afterwards, the studies can easily be combined either in a two-step procedure using traditional methods for univariate meta-analysis or in a single-step approach using hierarchical models. In any case, the meta-analysis can be performed using standard software and because of the increased sample size the statistical power of the meta-analysis is increased whereas the procedure allows performing several diagnostics (publication bias, cumulative meta-analysis, sensitivity analysis). The proposed approach was applied in nine studies performed under a family-based and a population-based case-control design, to elucidate the potential association of FZD3 polymorphisms with schizophrenia. A total of 6 polymorphisms were assessed (rs2241802, rs2323019, rs352203, rs3757884, rs880481 and rs960914). The present meta-analysis, highlighted the association of rs352203 polymorphism with schizophrenia with allele C presenting a protective role. An increased risk was also estimated for carriers of allele A of rs2323019 polymorphism in Chinese, although this finding could be attributed to different linkage disequilibrium patterns.

One of our major goals was to develop a method for the simultaneous meta-analysis of two or more multiple outcomes in genetic association studies. The proposed method is a simple, yet powerful for performing multivariate meta-analysis of genetic association studies. The model we propose is a direct extension of the bivariate meta-analysis method of genetic association studies to account for both discrete and continuous multiple outcomes. The innovation of the method is that we provide formulas for the calculation of the within-studies covariances using the cross-classification of the genotypes in both outcomes using a log-linear model and the iterative proportional fitting algorithm under the assumption of no three-way interaction. The method is fast and it can be extended for several outcomes. There is

no need for individual patient data or the simultaneous evaluation of both outcomes in all studies. The proposed approach was applied in two continuous outcomes, Total-body lean mass (TB-LM) and Bone mineral density (TB-BMD) to investigate pleiotropic effects on both traits and to elucidate possible biological pathways, which underlie the crosstalk between these two tissues during skeletal development. Overall, more than 2.5 million polymorphisms were assessed in four paediatric cohorts revealing 8 signals with p -values $< 5 \times 10^{-8}$.

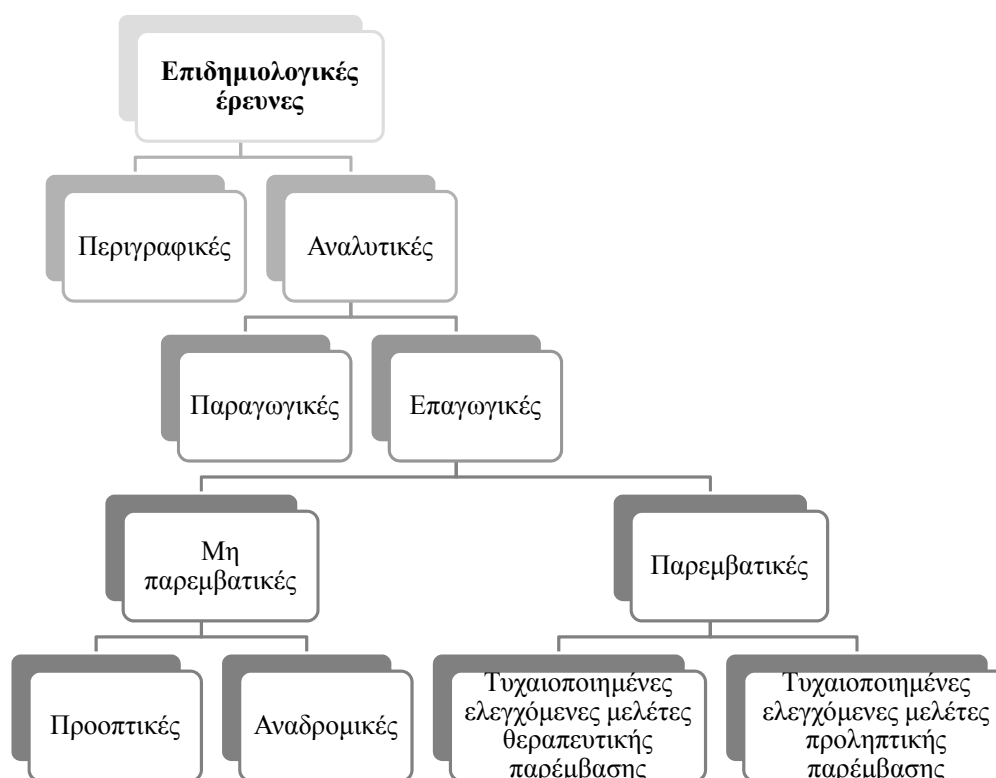
An extension of the aforementioned approach, holds when more than one outcomes are reported in a particular study as well as genes with multiple alleles and thus a more complex multivariate framework of meta-analysis is preferable. Such a scenario involves the investigation of the possible association of Apolipoprotein E gene (multiple alleles) with left ventricular failure in beta thalassaemia (patients are grouped according to disease severity) where a significant association of E4 allele was pointed out. Moreover, promising results were obtained after applying both multivariate methods for meta-analysis and classical approaches for the investigation of interleukin gene polymorphisms and susceptibility to HIV-1 Infection, the effect of Hepatitis C Virus on immunologic and virologic response in HIV infected patients initiating Highly Active Antiretroviral Therapy and Methylene Tetrahydrofolate Reductase gene polymorphisms and their association with methotrexate toxicity.

Finally, another distinct research area that we applied the multivariate methodology of meta-analysis, involves the development of a novel multivariate method and software for meta-analysis and comparison of two or more diagnostic tests. The method allows the direct calculation of sensitivity and specificity, as well as, the area under curve (AUC) and the parameters of the Summary Receiver Operator's Characteristic (SROC) curve, along with the means for a formal comparison of these quantities for different tests. There is no need for individual patient data or the simultaneous evaluation of both diagnostic tests in all studies. The proposed methodology was evaluated with simulation studies and was approximately unbiased with greater power compared with the standard bivariate approach

Κεφάλαιο 1: Εισαγωγή

1.1. Επιδημιολογία και γενετική επιδημιολογία

Επιδημιολογία είναι η μελέτη της κατανομής και της εξέλιξης διαφόρων νοσημάτων ή χαρακτηριστικών στον ανθρώπινο πληθυσμό (περιγραφική επιδημιολογία), και των παραγόντων που τις διαμορφώνουν ή μπορούν να τις επηρεάσουν (αναλυτική επιδημιολογία) (Εικόνα 1.1) (Last 2001). Η ιστορία της επιδημιολογικής μεθοδολογίας βασίζεται στη διαμόρφωση τεσσάρων αρχών: (α) η ανθρώπινη νοσολογία επηρεάζεται περισσότερο από το περιβάλλον παρά από τη γενετική μας υποδομή, (β) οι αριθμητικές μετρήσεις καταστάσεων και γεγονότων που σχετίζονται με την υγεία συμβάλλουν περισσότερο στην κατανόηση της αιτιολογίας και φυσικής ιστορίας των νοσημάτων από τη μη τεκμηριωμένη διατύπωση προσωπικών εμπειριών, (γ) τα καλούμενα «πειράματα της φύσης» μπορούν να συμβάλλουν στην αποκάλυψη της αιτιότητας και (δ) είναι δυνατή η αποκάλυψη αιτιολογικών παραγόντων ή παραγόντων που επηρεάζουν την εξέλιξη νοσημάτων, με τη δημιουργία συνθηκών παρεμβατικής ή μη παρεμβατικής φύσης.



Εικόνα 1.1. Αδρή κατάταξη επιδημιολογικών ερευνών.

Μετά το 1970, και ραγδαία μετά το 1990, οι επιδημιολογικές μέθοδοι εφαρμόστηκαν συστηματικά στην κλινική ιατρική με έμφαση στις τυχαιοποιημένες έρευνες θεραπευτικής παρέμβασης. Τα τελευταία έτη, οι έρευνες αυτές αποτελούν την αιχμή της αξιολόγησης των προτεινόμενων θεραπευτικών, αλλά και διαγνωστικών και προληπτικών, μέτρων, στο πλαίσιο της κλινικής επιδημιολογίας. Η σύγχρονη ανάπτυξη της επιδημιολογίας αντανακλάται στη δημιουργία εξειδικεύσεων ανάλογα με τις κατηγορίες εκθέσεων (π.χ. διατροφική επιδημιολογία, φαρμακοεπιδημιολογία, γενετική επιδημιολογία και λοιπά) ή τη φύση του μελετούμενου αποτελέσματος (π.χ. νευροεπιδημιολογία, περιγεννητική επιδημιολογία). Την τελευταία δεκαετία με τις προόδους που σημειώθηκαν στη μελέτη του ανθρώπινου γονιδιώματος, σημαντική είναι η ανάπτυξη που παρατηρήθηκε στον τομέα της γενετικής επιδημιολογίας.

Η γενετική επιδημιολογία αποτελεί τη συνένωση της κλασσικής επιδημιολογίας με τη γενετική. Σε αυτό το χώρο εφαρμόζονται οι κλασσικές επιδημιολογικές μέθοδοι έτσι ώστε να διαλευκανθεί η πιθανή επίδραση της γενετικής ποικιλομορφίας στην αιτιολογία της νόσου. Αυτός ο αναδυόμενος και ταχύτατα αναπτυσσόμενος κλάδος, μελετά τις γενετικές συνιστώσες των ασθενειών καθώς επίσης και την αλληλεπίδραση των γενετικών και περιβαλλοντικών παραγόντων (Burton, Tobin et al. 2005). Εκτός από τις μελέτες σε οικογένειες, οι οποίες διεξάγονται ευρέως από τους γενετιστές, η σύγχρονη γενετική επιδημιολογία χρησιμοποιεί τις μελέτες ασθενών-μαρτύρων για να ανιχνεύσει συσχετίσεις γονιδίων σε επίπεδο πληθυσμού (Lewis 2002). Αυτές οι δυο διακριτές γραμμές έρευνας αναφέρονται συχνά σαν μελέτες γενετικής συσχέτισης (Genetic Association Studies - GASs) (Cordell and Clayton 2005). Ωστόσο, οι μελέτες γενετικής συσχέτισης ελέγχουν έναν ή λίγους γενετικούς παράγοντες σε κάθε περίπτωση. Συνεπώς, στις πολύπλοκες ασθένειες πολυπαραγοντικής αιτιολογίας, όπου ο φαινότυπος καθορίζεται από την ταυτόχρονη αλληλεπίδραση πολλών γενετικών παραγόντων οι οποίοι ενεργούν από μόνοι τους ή σε συνδυασμό με περιβαλλοντικές παραμέτρους, η μελέτη μιας απλής πολυμορφικής θέσης λαμβάνει υπόψη μικρό μέρος της κληρονομικότητας και η μικρή έως μέτρια επίδρασή της συχνά επισκιάζεται στις αναλύσεις ή μπορεί να θεωρείται ότι δεν έχει κάποια κλινική σημασία. Από το 2005, εμφανίστηκαν οι ευρυγονιδιωματικές μελέτες (Genome Wide Association Studies - GWASs), οι οποίες αναλύουν το σύνολο του γονιδιώματος σε ικανοποιητικό κόστος (Klein, Zeiss et al. 2005). Στις ευρυγονιδιωματικές μελέτες, εκατομμύρια

μονονουκλεοτιδικοί πολυμορφισμοί (Single Nucleotide Polymorphisms - SNPs) διαβάζονται ταυτόχρονα (Manolio 2010). Στις μέρες μας, ο αριθμός των ευρυγονιδιωματικών μελετών αυξάνεται εκθετικά, εισάγοντάς μας στην εποχή των ανακαλύψεων νέων υποκείμενων πολυμορφικών θέσεων που συσχετίζονται με την εμφάνιση πολύπλοκων ασθενειών, πολυπαραγοντικής αιτιολογίας (Hindorff, MacArthur et al. 2009). Έχουν παρουσιαστεί σημαντικές ανασκοπήσεις στατιστικών ζητημάτων που ανακύπτουν στις ευρυγονιδιωματικές μελέτες (Teo 2008; Ziegler, König et al. 2008).

1.2. Γενετική στη γενετική επιδημιολογία

Για την καλύτερη κατανόηση των εννοιών που θα περιγραφούν στις επόμενες παραγράφους ακολουθεί μια σύντομη εισαγωγή βασικών εννοιών κλασσικής γενετικής.

1.2.1. Το γενετικό υλικό

Όλες οι κυτταρικές μορφές έχουν αποθηκευμένες τις γενετικές τους πληροφορίες με τη μορφή μιας γραμμικής αλληλουχίας τεσσάρων χημικών οργανικών αζωτούχων βάσεων, αδερίνης (A), θυμίνης (T), γουανίνης (G) και κυτοσίνης (C), που αποτελούν το μακρομόριο της ζωής, που ονομάζεται DNA [δε(σ)-όξυ-ριβοζονουκλεΐ(νι)κό οξύ]. Η γραμμική διαδοχή των ομοιοπολικά συνδεδεμένων τεσσάρων αζωτούχων βάσεων A, G, C και T συνιστά την κωδικοποίηση των γενετικών πληροφοριών, ενώ η πλευρική συμπληρωματικότητα των αντιπαράλληλων κλώνων (A,T και G,C), μέσω μη ομοιοπολικών αλληλεπιδράσεων υδρογονικής φύσης, καθορίζει τα φαινόμενα τόσο της αντιγραφής του DNA (replication) όσο και της μεταγραφής του σε RNA (transcription). Έτσι διαμορφώνεται η διπλή έλικα του DNA.

Ως γονιδίωμα (genome) νοείται η ποσότητα και το είδος των αλληλουχιών DNA που μεταβιβάζονται κληρονομικά και περιέχουν όλες τις απαραίτητες πληροφορίες για τη μορφολογία και τη φυσιολογία του οργανισμού. Με μόνη εξαίρεση κάποιους ιούς, που το γονιδιώμά τους αποτελείται από μονόκλωνο ή δίκλωνο RNA, το γονιδίωμα όλων των ευκαρυωτικών οργανισμών αποτελείται από δίκλωνο DNA. Στους ευκαρυωτικούς οργανισμούς, οι γενετικές πληροφορίες εδράζονται κατά κύριο λόγο στα χρωμοσώματα, δηλαδή στον πυρήνα των κυττάρων.

Γονίδιο (gene) είναι μια περιοχή του DNA (γενετικός τόπος, locus) υπεύθυνη για την εμφάνιση ενός συγκεκριμένου φαινοτυπικού χαρακτηριστικού. Ένας γενετικός τόπος μπορεί να έχει εναλλακτικές μορφές ενός γονιδίου (αλληλόμορφα, alleles) ενώ, ακόμα και η απουσία του γονιδίου σε περίπτωση εξάλειψης μπορεί να θεωρηθεί αλληλόμορφο εφόσον οδηγεί σε συγκεκριμένο φαινότυπο. Αν και ο γονότυπος (genotype) χρησιμοποιείται μερικές φορές για να περιγράψει τη συνολική γονιδιακή σύσταση ενός ατόμου, οι γενετικοί επιδημιολόγοι χρησιμοποιούν τον όρο για έναν συγκεκριμένο γενετικό τόπο. Έτσι λοιπόν, για ένα γονίδιο με δυο αλληλόμορφα (A, B), με το δεύτερο (B) να θεωρείται παράγοντας κινδύνου για την εμφάνιση μιας νόσου, ο πιθανός γονότυπος ενός ατόμου μπορεί να είναι AA, AB και BB. Ο φαινότυπος είναι ένα μετρήσιμο χαρακτηριστικό ενός ατόμου και μπορεί να αναφέρεται είτε στην παρουσία ή απουσία μια ασθένειας είτε σε ποσοτικά χαρακτηριστικά (συστολική ή διαστολική πίεση). Μερικές ασθένειες εμφανίζονται μόνο από την ταυτόχρονη παρουσία δυο αντιγράφων του μη φυσιολογικού αλληλομόρφου (υπολειπόμενο-recessive αλληλόμορφο). Στην περίπτωση των επικρατών (dominant) αλληλομόρφων αρκεί η παρουσία του ενός μη φυσιολογικού αντιγράφου για την εμφάνιση του φαινοτύπου. Τέλος, μια ενδιάμεση κατάσταση εμφανίζεται για τα συνεπικρατή (codominant) αλληλόμορφα όπου η διεισδυτικότητα της νόσου σε έναν ετεροζυγότη είναι ενδιάμεση από τους αντίστοιχους ομοζυγώτες.

Μικρό μόνο ποσοστό του συνολικού DNA ενός κυττάρου περιέχει λειτουργική πληροφορία, δηλαδή κωδικοποιεί μέσω μεταγραφής σε RNA τον πολυμερισμό των αμινοξέων και τη σύνθεση των πρωτεϊνών. Διάφοροι τύποι αλληλουχιών που δεν κωδικοποιούν πρωτεΐνες ή συνηθισμένα RNA είναι τα εσώνια (introns), το δορυφορικό DNA (satellite DNA), το μινι- και μικρο-δορυφορικό DNA (mini- and micro-satellite DNA), τα SINES (short interspersed elements), τα LINES (long interspersed elements) και τα ψευδογονίδια (pseudogenes). Μετά την «αποκωδικοποίηση» του ανθρώπινου γονιδιώματος προσδιορίστηκε ότι αυτό περιέχει περίπου 30.000 γονίδια. Έτσι αν θεωρήσουμε ότι, κατά μέσο όρο, το μήκος ενός γονιδίου που αντιπροσωπεύεται σε επίπεδο πρωτεΐνης, είναι της τάξης των 1000 νουκλεοτιδικών βάσεων, τότε πολύ λιγότερο από το 1% του DNA μας περιέχει γενετικές πληροφορίες υπεύθυνες για τη φυσιολογία και τη μορφολογία του ανθρώπου.

1.2.2. Γενετική ποικιλότητα

Πολλοί γενετικοί τόποι χαρακτηρίζονται από τη δημιουργία αλληλομόρφων γονιδίων και έτσι τα άτομα ενός πληθυσμού εμφανίζουν διαφορετικούς φαινότυπους. Ο γονιδιακός πολυμορφισμός είναι η εμφάνιση πολλαπλών αλληλομόρφων σε μια γενετική θέση, όπου τουλάχιστον δύο αλληλόμορφα εμφανίζονται με συχνότητα μεγαλύτερη από 1%. Πολυμορφικοί γενετικοί τόποι είναι εκείνοι στους οποίους τουλάχιστον το 2% του πληθυσμού είναι ετερόζυγοι. Ωστόσο, επειδή πολλοί πολυμορφικοί τόποι χαρακτηρίζονται από μεγάλο αριθμό αλληλομόρφων, η αναλογία ετερόζυγων σε ορισμένους γενετικούς τόπους είναι πολύ μεγαλύτερη από αυτή που αναφέρθηκε.

Μέχρι σήμερα, έχει ελεγχθεί ένας μεγάλος αριθμός ενζύμων και πρωτεϊνών του ανθρώπου ως προς την εμφάνιση πολυμορφισμών. Από όλους τους γενετικούς τόπους που έχουν ελεγχθεί, περίπου το 1/3 εμφανίζει ανιχνεύσιμους πολυμορφισμούς σε μία τουλάχιστον φυλή και συχνά σε όλο τον πληθυσμό. Η μεγάλη εξάπλωση των πολυμορφισμών υποδηλώνει ότι κάθε άτομο μπορεί να είναι ετερόζυγο για διάφορες γενετικές θέσεις. Τα μεταλλαγμένα αλληλόμορφα που οδηγούν σε σοβαρές γενετικές ασθένειες είναι συνήθως η πιο εμφανής μορφή γενετικής ποικιλότητας. Διαπιστώθηκε ότι μερικά από τα αυτά τα πολυμορφικά αλληλόμορφα είναι είτε από μόνα τους κλινικώς σημαντικά είτε σε συνδυασμό με σπάνια επιβλαβή αλληλόμορφα. Οι πολυμορφισμοί έχουν μεγάλη σημασία στην ιατρική γενετική όπου χρησιμοποιούνται:

- στη χαρτογράφηση γονιδίων με ανάλυση σύνδεσης.
- στην προσυμπτωματική και προγεννητική διάγνωση γενετικών ασθενειών,
- στην αναγνώριση ετερόζυγων φορέων μιας ασθένειας,
- στον προσδιορισμό ατόμων που βρίσκονται σε ομάδες υψηλού και χαμηλού κινδύνου όσον αφορά στην προδιάθεση για ορισμένες ασθένειες, όπως ο καρκίνος, ο διαβήτης, η στεφανιαία νόσος και λοιπά,
- στον προσδιορισμό της δοκιμασίας πατρότητας και στην ιατροδικαστική και
- στον προσδιορισμό ατόμων που αποτελούν κατάλληλους δότες για μεταμόσχευση οργάνων και ιστών.

Οι κυριότερες δομικές κατηγορίες γονιδιακών πολυμορφισμών είναι οι SNPs που χαρακτηρίζονται από την αντικατάσταση μια βάσης από μια άλλη και οι επαναληπτικές αλληλουχίες (π.χ. με το CA₁₂ υποδηλώνονται 12 επαναλήψεις CA σε μια γραμμή) που εμφανίζονται κυρίως στις μη κωδικές αλληλουχίες του γενετικού

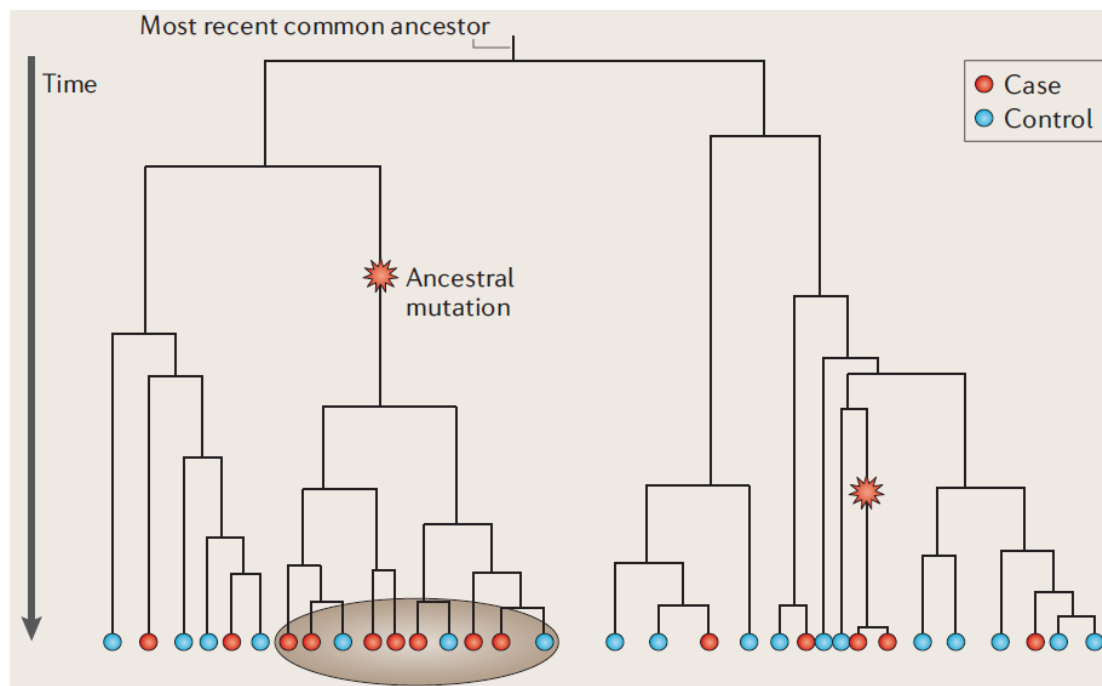
[illegible]

Υπάρχουν τέσσερα σημαντικά πλεονεκτήματα από τη χρήση των SNPs, σε σχέση με άλλους πολυμορφισμούς, για την ανίχνευση γενετικών παραγόντων που συσχετίζονται με πολυπαραγοντικές ασθένειες (Collins, Patrinos et al. 1998; Risch 2000; Palmer and Cookson 2001). Αρχικά, οι SNPs είναι πολυάριθμοι στο γονιδίωμα και εμφανίζονται σε εξώνια, σε εσώνια, σε υποκινητές, σε ενισχυτές και σε διαγονιδιακές περιοχές (Collins, Guyer et al. 1997; Kruglyak 1997) και μερικοί από αυτούς τους πολυμορφισμούς μπορεί να είναι από μόνοι τους λειτουργικοί. Δεύτερον, ομάδες γειτονικών SNPs μπορούν να εμφανίζουν συγκεκριμένα πρότυπα συσχέτισεων έτσι ώστε να επιτρέπουν τη βελτιστοποίηση της χαρτογράφησης γονιδίων (Nickerson, Whitehurst et al. 1992) και τον εντοπισμό θέσεων ανασυνδυασμού (Chakravarti 1998). Τρίτον, οι διαφορές στις συχνότητες των SNPs μεταξύ των πληθυσμών μπορούν να χρησιμοποιηθούν σε μελέτες γενετικής συσχέτισης σε πληθυσμούς (McKeigue 1998; Kuhner, Beerli et al. 2000). Τέλος, οι

SNPs υπόκεινται σε μικρότερο αριθμό μεταλλάξεων σε σχέση με άλλους πολυμορφισμούς παρέχοντας έτσι αξιόπιστες εκτιμήσεις των συσχετίσεων γονιδίων-ασθενειών (Stallings, Ford et al. 1991; Brookes 1999).

1.3. Μελέτες γενετικής συσχέτισης

Οι μελέτες γενετικής συσχέτισης στοχεύουν στην εύρεση συσχετίσεων μεταξύ ενός ή περισσότερων γονιδιακών πολυμορφισμών και ενός χαρακτηριστικού, το οποίο μπορεί να είναι είτε ποσοτικό είτε να αποτελεί μια διχοτομική έκβαση (παρουσία-απουσία μιας νόσου). Η συσχέτιση διαφέρει από τη σύνδεση (linkage) στο ότι το ίδιο αλληλόμορφο συσχετίζεται με ένα χαρακτηριστικό με παρόμοιο τρόπο σε ολόκληρο τον πληθυσμό, ενώ στη σύνδεση επιτρέπονται διαφορετικά αλληλόμορφα να συσχετίζονται με το χαρακτηριστικό σε διαφορετικές οικογένειες. Ωστόσο, οι γενετικές συσχετίσεις στηρίζονται στη γενική αρχή του ότι οι πληθυσμοί προέρχονται από έναν κοινό πρόγονο και έχει διατυπωθεί ότι οι μελέτες συσχέτισης είναι απλώς μια μορφή μελετών σύνδεσης όπου η «οικογένεια» είναι ο ευρύτερος πληθυσμός (Εικόνα 1.3). Η ανάλυση συσχετίσεων έχει μεγαλύτερη ισχύ για την ανίχνευση μικρών επιδράσεων σε σχέση με τις μελέτες σύνδεσης, αλλά απαιτείται η ταυτόχρονη αξιολόγηση πολύ περισσότερων γενετικών δεικτών. Είναι πλέον αποδεκτό ότι για την εμφάνιση πολύπλοκων ασθενειών ευθύνονται πολλά γονίδια που δρουν συνεργατικά, τα περισσότερα από τα οποία έχουν μικρές επιδράσεις. Το γεγονός αυτό, σε συνδυασμό με την ανίχνευση μεγάλου αριθμού SNPs σε όλο το γονιδίωμα με ολοένα και χαμηλότερο κόστος, έχει οδηγήσει στη μεγάλη σημαντικότητα των μελετών γενετικής συσχέτισης στο χώρο της γενετικής επιδημιολογίας (Pokorny, Luan et al. 1976; Livak, Marmaro et al. 1995; Risch and Merikangas 1996).



Εικόνα 1.3. Η λογική των μελετών γενετικής συσχέτισης. Τα αλληλόμορφα εντός του οβάλ σχήματος προέρχονται από ένα προγονικό μεταλλαγμένο αλληλόμορφο το οποίο έχει εμφανιστεί πιθανώς κάποιες εκατοντάδες γενιές στο παρελθόν με συνεπακόλουθη την αυξημένη συχνότητα ασθενών σε αυτή την ομάδα (Balding 2006).

Αν και οι μελέτες σε οικογένειες εξακολουθούν να χρησιμοποιούνται για τη διερεύνηση των συσχετίσεων σε επίπεδο πληθυσμού (μαζί με τις μελέτες σύνδεσης), η έρευνα αυτή παρουσιάζει περισσότερα κοινά σημεία με τις κλασσικές επιδημιολογικές μελέτες περιβαλλοντικών παραγόντων κινδύνου σε σχέση με τις μελέτες σύνδεσης. Η ομοιότητα με την κλασσική επιδημιολογία είναι ξεκάθαρη εάν λάβουμε υπόψη την αιτιολογία της συσχέτισης ενός γενετικού παράγοντα με ένα χαρακτηριστικό. Έτσι, μια σημαντική συσχέτιση μπορεί να είναι είτε (α) άμεση συσχέτιση όπου ο υπό μελέτη γονιδιακός πολυμορφισμός είναι ο πραγματικός αιτιολογικός παράγοντας που συνεισφέρει στην εμφάνιση μιας ασθένειας (β) έμμεση συσχέτιση όπου ο υπό μελέτη γονιδιακός πολυμορφισμός βρίσκεται σε ανισορροπία σύνδεσης (linkage disequilibrium - LD) με το πραγματικό αλληλόμορφο που συσχετίζεται με την ασθένεια και (γ) ένα ψευδώς θετικό αποτέλεσμα που έχει προκύψει είτε λόγω τύχης είτε λόγω κάποιου συγχυτικού παράγοντα (Lewis and Knight 2012). Η διάκριση μεταξύ άμεσης και έμμεσης συσχέτισης δεν είναι πάντα εύκολη και μπορεί να απαιτεί την εκ νέου αλληλούχιση του υπό μελέτη γενετικού τύπου ή/ και όλων των διαθέσιμων πολυμορφισμών ή μελέτες για την ανίχνευση του λειτουργικού του ρόλου.

1.4. Σχεδιασμοί μελετών γενετικής συσχέτισης

Μια σύνοψη των διαφόρων σχεδιασμών μελετών που χρησιμοποιούνται σε μελέτες γενετικής συσχέτισης καθώς και τα πλεονεκτήματα και μειονεκτήματα που συνοδεύουν τον καθένα παρουσιάζονται στον Πίνακα 1.1. Πιο συγκεκριμένα, κλασσικοί επιδημιολογικοί σχεδιασμοί όπως οι μελέτες ασθενών-μαρτύρων ή οι μελέτες κοορτής χρησιμοποιούνται συχνά σε μελέτες γενετικής συσχέτισης. Τα δεδομένα αναλύονται κατά παρόμοιο τρόπο όπως το κάπνισμα, η παχυσαρκία και λοιπά, με τη διαφορά ότι εδώ υπολογίζεται η παρουσία ή απουσία ενός συγκεκριμένου γονιδιακού πολυμορφισμού. Επιπλέον, σχεδιασμοί σε οικογένειες έχουν προταθεί έτσι ώστε να αποφευχθούν συγχυτικοί παράγοντες όπως η πληθυσμιακή διαστρωμάτωση, που παρατηρούνται σε μελέτες ασθενών-μαρτύρων. Στους σχεδιασμούς σε οικογένειες, τα αλληλόμορφα που μεταβιβάζονται σε ασθενείς συγκρίνονται με εκείνα που δεν έχουν μεταβιβαστεί, διαμορφώνοντας έτσι την ομάδα ελέγχου. Ωστόσο, στις μελέτες αυτές απαιτείται μεγάλο μέγεθος δείγματος. Για παράδειγμα, στο σχεδιασμό case-parent triad απαιτείται συνήθως ο ίδιος αριθμός τριάδων (αποτελούμενων από έναν ασθενή και δυο γονείς) με τους ασθενείς σε μια μελέτη ασθενών-μαρτύρων (υποθέτοντας ίδιο αριθμό υγιών) έτσι ώστε η υπολογιζόμενη ισχύς να είναι συγκρίσιμη.

Πίνακας 1.1. Σχεδιασμοί μελετών σε μελέτες γενετικής συσχέτισης (Cordell and Clayton 2005).

Σχεδιασμός μελέτης	Λεπτομέρειες	Πλεονεκτήματα	Μειονεκτήματα	Μέθοδος στατιστικής ανάλυσης
Cross-sectional	Ο γονότυπος και ο φαινότυπος είναι ένα τυχαίο δείγμα του πληθυσμού	Χαμηλό κόστος. Μπορεί να εκτιμηθεί ο επιπολασμός μιας νόσου	Λίγοι ασθενείς στην περίπτωση σπάνιας νόσου	Λογιστική ή γραμμική παλινδρόμηση. χ^2 έλεγχοι συσχέτισης
Cohort	Αλληλούχιση τμήματος του πληθυσμού και εκτίμηση της επίπτωσης μιας νόσου για συγκεκριμένη χρονική περίοδο	Μπορεί να εκτιμηθεί η επίπτωση μιας νόσου	Ακριβή η παρακολούθηση των συμμετεχόντων και προβλήματα από άτομα που διακόπτουν	Μέθοδοι ανάλυσης επιβίωσης
Case-control	Ο γονότυπος προσδιορίζεται για ένα συγκεκριμένο αριθμό ασθενών και υγιών	Δεν απαιτείται η παρακολούθηση των ατόμων. Παρέχει εκτιμήσεις των επιδράσεων της έκθεσης	Απαιτείται προσεκτική επιλογή των υγιών. Πιθανοί συγχυτικοί παράγοντες	Λογιστική παλινδρόμηση. χ^2 έλεγχοι συσχέτισης
Extreme values	Ο γονότυπος προσδιορίζεται για άτομα με ακραίες τιμές ενός ποσοτικού χαρακτηριστικού	Η αλληλούχιση αφορά μόνο τα πιο πληροφοριακά άτομα και έτσι μειώνεται το κόστος	Δεν εκτιμάται το πραγματικό μέτρο επίδρασης του γενετικού παράγοντα	Γραμμική παλινδρόμηση. μη παραμετρικές προσεγγίσεις
Case-parent triads	Ο γονότυπος προσδιορίζεται για ασθενείς και τους γονείς αυτών	Αποφεύγεται η πληθυσμιακή διαστρωμάτωση	Μειωμένη ισχύς σε σχέση με τις μελέτες ασθενών-μαρτύρων	Transmission/disequilibrium test, δεσμευμένη λογιστική παλινδρόμηση ή μοντέλα log-linear
Case-parent-grandparent septets	Ο γονότυπος προσδιορίζεται για ασθενείς και τους γονείς και παππούδες αυτών	Αποφεύγεται η πληθυσμιακή διαστρωμάτωση	Οι παππούδες είναι δύσκολο να βρεθούν	Μοντέλα log-linear
General pedigrees	Ο γονότυπος προσδιορίζεται για ένα τυχαίο δείγμα ή για ένα δείγμα οικογενειών από τον γενικό πληθυσμό με κάποια νόσο	Αυξημένη ισχύς σε μεγάλες οικογένειες. Το δείγμα μπορεί να προϋπάρχει από μελέτες σύνδεσης	Ακριβή η αλληλούχιση. Πολλά ελλείποντα άτομα	Pedigree disequilibrium test, family-based association test, quantitative transmission/disequilibrium test
Case-only	Ο γονότυπος προσδιορίζεται μόνο σε ασθενείς	Παρουσιάζει μέγιστη ισχύ για την ανίχνευση αλληλεπιδράσεων	Μπορούν να εκτιμηθούν μόνο οι αλληλεπιδράσεις.	Λογιστική παλινδρόμηση. χ^2 έλεγχοι συσχέτισης
DNA-pooling	Εφαρμόζεται σε πολλούς από τους σχεδιασμούς που έχουν ήδη αναφερθεί αλλά οι γονότυποι δεν υπολογίζονται σε ατομικό επίπεδο αλλά είναι μια δεξαμενή μεταξύ δυο και 100 ατόμων	Πιθανώς παρουσιάζει χαμηλό κόστος	Δύσκολο να εκτιμηθούν οι διάφορες πηγές μεταβλητότητας λόγω την πειραματικών συνθηκών	Εκτίμηση των συνιστωσών της μεταβλητότητας

1.5. Προκαταρκτικοί έλεγχοι σε μελέτες γενετικής συσχέτισης

Προτού πραγματοποιηθεί η ανάλυση των δεδομένων σε μελέτες γενετικής καθίστανται αναγκαία η διεξαγωγή προκαταρκτικών ελέγχων όπως ο έλεγχος της ποιότητας αλληλούχισης, η ισορροπία Hardy-Weinberg, η ύπαρξη ελλειπουσών παρατηρήσεων και η πληθυσμιακή διαστρωμάτωση.

1.5.1. Έλεγχος της ποιότητας αλληλούχισης

Ένα μειονέκτημα των μελετών ασθενών-μαρτύρων σε σύγκριση με τις μελέτες σε οικογένειες είναι η έλλειψη ελέγχων της ποιότητας αλληλούχισης. Συνήθως στην εργαστηριακή πρακτική για τον περιορισμό των συστηματικών σφαλμάτων, δείγματα τόσο από υγιείς όσο και από ασθενείς τοποθετούνται σε κάθε πλάκα που χρησιμοποιείται για την αλληλούχιση, ελέγχονται οι διαφορές στη συχνότητα των γονοτύπων μεταξύ των πλακών και η αλληλούχιση μπορεί ακολούθως να επαναληφθεί. Ο έλεγχος της ισορροπίας Hardy-Weinberg (Hardy-Weinberg Equilibrium-HWE) στους υγιείς μπορεί επίσης να χρησιμοποιηθεί συμπληρωματικά για να ελεγχθεί η ποιότητα αλληλούχισης.

1.5.2. Ισορροπία Hardy-Weinberg

Υπό τη θεώρηση της ισορροπίας Hardy-Weinberg, τα αλληλόμορφα διαχωρίζονται τυχαία στον πληθυσμό, επιτρέποντας τον υπολογισμό των αναμενόμενων συχνοτήτων των γονοτύπων από τις συχνότητες των αλληλομόρφων.

Αποκλίσεις από την ισορροπία Hardy-Weinberg στους υγιείς μπορεί να οφείλονται σε:

- Σφάλματα στην αλληλούχιση. Σε πολλές πλατφόρμες που χρησιμοποιούνται για την αλληλούχιση, η εύρεση ετερόζυγων ατόμων είναι αρκετά δύσκολη.
- Σύζευξη όμοιων ατόμων. Άτομα με παρόμοιους γονοτύπους ή/και φαινότυπους δίνουν απογόνους πιο συχνά από ότι θα περίμενε κανείς λόγω τύχης.
- Επιλογή. Ο οποιοσδήποτε γονότυπος που μπορεί να οδηγήσει σε πρόωμο θάνατο είναι αρκετά πιθανό να υποαντιπροσωπεύεται.
- Πληθυσμιακή διαστρωμάτωση. Δείγματα υγιών που προκύπτουν από την ένωση γενετικά διακριτών υποπληθυσμών μπορεί να μη βρίσκονται σε ισορροπία Hardy-Weinberg.

- Τύχη. Τα p-values για τον έλεγχο της ισορροπίας Hardy-Weinberg πολλών SNPs θα πρέπει να διορθώνονται κατάλληλα για να αποφεύγεται το σφάλμα πολλαπλών ελέγχων.

Ο έλεγχος για αποκλίσεις από την ισορροπία Hardy-Weinberg μπορεί να γίνει χρησιμοποιώντας τον έλεγχο καλής προσαρμογής χ^2 του Pearson ή τον ακριβή έλεγχο του Fisher για την περίπτωση των γονοτύπων με χαμηλές συχνότητες. Έως τώρα, οι ερευνητές απέρριπταν γενετικούς τόπους με p-values μικρότερα από ένα συγκεκριμένο κατώφλι (συνήθως $\alpha=10^{-3}$ ή 10^{-4}). Ωστόσο, προτείνεται να ελέγχεται εκ των προτέρων αν η απόκλιση από την ισορροπία Hardy-Weinberg οφείλεται σε έλλειψη ή σε διπλασιασμό (Bailey and Eichler 2006; Conrad, Andrews et al. 2006).

1.5.3. Ελλείπουσες τιμές στους γονοτύπους

Για την περίπτωση της συσχέτισης ενός γενετικού τόπου με κάποια ασθένεια οι ελλείπουσες παρατηρήσεις δεν αποτελούν μεγάλο πρόβλημα στην ανάλυση κάτι που όμως δεν ισχύει για την ταυτόχρονη αξιολόγηση πολλών γενετικών τόπων. Έτσι, οι ελλείπουσες τιμές αποτελούν ένα σημαντικό ζήτημα στις ευρυγονιδιωματικές μελέτες. Επίσης, το διαφορετικό κλάσμα ελλειπουσών παρατηρήσεων σε ασθενείς και υγιείς μπορεί να οδηγήσει σε μεροληπτικές εκτιμήσεις. Μια λύση είναι η αντικατάσταση των γονοτύπων που λείπουν χρησιμοποιώντας συχνότητες γειτονικών SNPs κυρίως για την περίπτωση πολύ στενά συνδεδεμένων γενετικών τόπων προσδίδοντας έτσι πιο αξιόπιστες εκτιμήσεις και απλουστεύοντας την ανάλυση. Η αντικατάσταση αυτή μπορεί να γίνει είτε με την εύρεση της πιο πιθανή τιμής (εκτίμηση μέγιστης πιθανοφάνειας) είτε τυχαία μέσω κάποιας κατανομής πιθανότητας. Το πλεονέκτημα της δεύτερης προσέγγισης έγκειται στο ότι η επαναληπτική αυτή διαδικασία μας δίνει ένα μέσο όρο των εκτιμήσεων και μπορεί να διερευνηθεί το κατά πόσο η κάθε τιμή που χρησιμοποιείται μπορεί να επηρεάσει το συνολικό αποτέλεσμα.

1.5.4. Πληθυσμιακή διαστρωμάτωση

Η πιο σημαντική αιτία των πλασματικών συσχετίσεων οφείλεται στη δομή του πληθυσμού της μελέτης. Η πληθυσμιακή διαστρωμάτωση υφίσταται όταν οι υποκείμενοι πληθυσμοί διαφέρουν στη συχνότητα των αλληλομόρφων και αντιπροσωπεύονται με διαφορετικές συχνότητες σε ασθενείς και υγιείς. Η ανίχνευση και η διόρθωση της πληθυσμιακής διαστρωμάτωσης είναι σημαντική, κυρίως στις

ευρυγονιδιωματικές μελέτες, όπου ακόμα και μικρές διαφορές μεταξύ ασθενών και υγιών μπορεί να έχουν πολύ σημαντική επίδραση στην ανάλυση. Έχουν προταθεί διάφορες μέθοδοι για την εκτίμηση της πληθυσμιακής διαστρωμάτωσης όπως η μέθοδος genomic-control, οι έλεγχοι Cochran/Mantel-Haenszel και transmission disequilibrium.

Στη μέθοδο genomic-control θεωρείται ότι η πληθυσμιακή διαστρωμάτωση υπερεκτιμά τα ευρήματα των στατιστικών ελέγχων κατά έναν σταθερό παράγοντα λ , ο οποίος μπορεί να εκτιμηθεί είτε από τη διάμεσο είτε από το μέσο όρο των ελέγχων μιας σειράς μη συνδεδεμένων SNPs ασθενών και υγιών (Devlin and Roeder 1999). Έτσι, οι εκτιμήσεις των στατιστικών ελέγχων διαιρούνται με το λ και συγκρίνονται με μια κατανομή χ^2 ή F (Devlin, Bacanu et al. 2004). Στις ευρυγονιδιωματικές μελέτες, η δομή των επιμέρους πληθυσμών μπορεί να διερευνηθεί με ανάλυση κυρίων συνιστωσών, όπου μοντελοποιείται η προγονική γενετική ποικιλομορφία μεταξύ ασθενών και υγιών και χρησιμοποιείται ως διόρθωση στην ανάλυση (Price, Patterson et al. 2006).

Όταν τα άτομα μπορούν να ταξινομηθούν σε υποομάδες (για παράδειγμα ανάλογα με τον τόπο γέννησης), η ανάλυση μπορεί να πραγματοποιηθεί σε κάθε υποομάδα ξεχωριστά και να συνδυαστούν τα αποτελέσματα χρησιμοποιώντας τον έλεγχο Cochran/Mantel-Haenszel (Clayton, Walker et al. 2005). Γενικώς, η πληθυσμιακή διαστρωμάτωση μπορεί να αποφευχθεί με τη διεξαγωγή μελετών σε οικογένειες. Ο ευρέως χρησιμοποιούμενος έλεγχος είναι ο transmission disequilibrium test (TDT) όπου ελέγχεται η μη-Mεντελιανή μεταβίβαση των αλληλομόρφων από ετερόζυγους γονείς στους ασθενείς απογόνους, με την αυξημένη συχνότητα μεταβίβασης να υποδηλώνει ότι ο υπό μελέτη γονιδιακός πολυμορφισμός συσχετίζεται με αυξημένο κίνδυνο εμφάνισης της νόσου (Spielman, McGinnis et al. 1993).

1.6. Μέθοδοι ανάλυσης σε μελέτες γενετικής συσχέτισης

Στις μελέτες γενετικής συσχέτισης και στις ευρυγονιδιωματικές μελέτες η ανάλυση πραγματοποιείται χρησιμοποιώντας διάφορες στατιστικές μεθόδους (Balding 2006), αλλά το κυριότερο πρόβλημα είναι η έλλειψη γνώσης του μοντέλου κληρονομικότητας (επικρατές, υπολειπόμενο, συνεπικρατές και λοιπά). Κάποιες από τις μεθόδους που χρησιμοποιούνται συχνά δεν υποθέτουν κάποιο μοντέλο

κληρονομικότητας (genetic model-free - GMF), ενώ άλλες κάνουν εκ των προτέρων τέτοιες υποθέσεις. Οι μέθοδοι που υποθέτουν κάποιο μοντέλο κληρονομικότητας έχουν αυξημένη ισχύ υπό τη θεώρηση του ορθού μοντέλου αλλά χαρακτηρίζονται από χαμηλή ισχύ όταν η υπόθεση είναι λανθασμένη. Από την άλλη πλευρά, οι μέθοδοι που δεν κάνουν κάποια υπόθεση για το μοντέλο κληρονομικότητας είναι ανθεκτικές στις παραβιάσεις των προϋποθέσεων (robust) αλλά έχουν μειωμένη ισχύ.

Οι μέθοδοι ανάλυσης που χρησιμοποιούνται στις μελέτες γενετικής συσχέτισης και στις ευρυγονιδιωματικές μελέτες θα περιγραφούν αναλυτικά στην Παράγραφο 2. Συνοπτικά, οι πιο συχνά χρησιμοποιούμενοι έλεγχοι είναι ο κλασσικός έλεγχος του Pearson (χ^2) και ο Cochran-Armitage trend test (CATT) (Sasieni 1997; Balding 2006). Ωστόσο, η μη ορθή θεώρηση του μοντέλου κληρονομικότητας μπορεί να οδηγήσει σε χαμηλή ισχύ για τον CATT. Σε αυτή την περίπτωση, προτείνεται η χρήση των robust ελέγχων, οι οποίοι έχουν αυξημένη ισχύ για όλα τα πιθανά μοντέλα κληρονομικότητας (Freidlin, Zheng et al. 2002). Κάποιες από τις robust μεθόδους βασίζονται στη θεωρία των efficiency robust procedures (Gastwirth 1985; Freidlin, Podgor et al. 1999) όπως οι Maximum Efficiency Robust Test (MERT) και maximum (MAX) test (Freidlin, Zheng et al. 2002). Ο έλεγχος MERT, που αποτελεί έναν γραμμικό συνδυασμό των βέλτιστων CATTs υπό τη θεώρηση του υπολειπόμενου και του επικρατούς μοντέλου κληρονομικότητας, έχει χαμηλότερη ισχύ από τον MAX που υπολογίζεται ως το μέγιστο των απολύτων τιμών των τριών CATTs για το υπολειπόμενο, συνεπικρατές και επικρατές μοντέλο κληρονομικότητας (Freidlin, Zheng et al. 2002). Ωστόσο, η χρήση του MAX εισάγει έναν αυξημένο βαθμό πολυπλοκότητας στην ανάλυση καθώς υπό τη μηδενική υπόθεση της απουσίας συσχέτισης, δεν ακολουθεί την τυποποιημένη κανονική κατανομή, οδηγώντας έτσι σε υπολογιστικά απαιτητικές μεθόδους για τον υπολογισμό των αντίστοιχων p-values. Συνεπώς, παρατηρώντας τη γραμμική εξάρτηση των CATTs, μπορεί να εξαχθεί μια προσεγγιστική κατανομή απλοποιώντας έτσι τον υπολογισμό των p-values (Zang, Fung et al. 2010). Τέλος, στην προσέγγιση MIN2, χρησιμοποιείται ένας συνδυασμός των στατιστικών ελέγχων, υπολογίζοντας το ελάχιστο των p-values για τους CATT (υπό τη θεώρηση του συνεπικρατούς μοντέλου) και χ^2 εξάγοντας παράλληλα μια ασυμπτωτική κατανομή (Joo, Kwak et al. 2009). Έχει δειχθεί ότι ο MAX έχει μεγαλύτερη efficient robustness από τους χ^2 και MIN2 όταν το μοντέλο κληρονομικότητας περιορίζεται αυστηρά σε υπολειπόμενο, συνεπικρατές και επικρατές αλλά όχι απαραίτητα σε διαφορετική περίπτωση (Joo, Kwak et al. 2009).

1.7. Η μετα-ανάλυση στις μελέτες γενετικής συσχέτισης

Ο συνεχώς αυξανόμενος αριθμός δημοσιευμένων μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών (Hirschhorn, Lohmueller et al. 2002; Becker, Barnes et al. 2004) καθιστά επιτακτική τη συλλογή και το συνδυασμό της διαθέσιμης πληροφορίας μέσω μια στατιστικής διαδικασίας γνωστής ως μετα-ανάλυση (Petiti 1994; Normand 1999; Trikalinos, Salanti et al. 2008). Η μετα-ανάλυση είναι μια ποσοτική ανασκόπηση και σύνθεση των αποτελεσμάτων που προκύπτουν από σχετιζόμενες αλλά και ανεξάρτητες μεταξύ τους μελέτες (Glass 1976; Petiti 1994; Greenland 1998; Normand 1999). Έτσι, σε μια μετα-ανάλυση όλη η διαθέσιμη πρωταρχική πληροφορία συνδυάζεται και διερευνάται η πιθανή μεταξύ των μελετών μεταβλητότητα χρησιμοποιώντας επίσημους στατιστικούς ελέγχους. Αν και η μετα-ανάλυση εφαρμόστηκε αρχικά σε τυχαιοποιημένες κλινικές δοκιμές (Chalmers, Berrier et al. 1987; Sacks, Berrier et al. 1987), στις μέρες μας χρησιμοποιείται ευρέως σε μελέτες παρατήρησης (Stroup, Berlin et al. 2000; Nikolopoulos, Bagos et al. 2011) και σε μελέτες γενετικής συσχέτισης και ευρυγονιδιωματικές μελέτες (Minelli, Thompson et al. 2005; Thakkestian, McElduff et al. 2005; Bagos and Nikolopoulos 2007; Evangelou, Maraganore et al. 2007; Bagos 2008; Trikalinos, Salanti et al. 2008; Zeggini and Ioannidis 2009).

Συνήθως, τα δεδομένα μιας μετα-ανάλυσης αναδρομικών μελετών ασθενών-μαρτύρων απεικονίζονται με τη μορφή του Πίνακα 1.2 για διχοτομικά χαρακτηριστικά ή 1.3 για συνεχείς εκβάσεις. Η πλειοψηφία των μετα-αναλύσεων στις μελέτες γενετικής συσχέτισης για έναν γενικό τόπο με δυο αλληλόμορφα (A, B) με έστω B το αλληλόμορφο κινδύνου, διεξάγονται συνενώνοντας τους γονοτύπους σε δυο κατηγορίες θεωρώντας ένα συγκεκριμένο μοντέλο κληρονομικότητας (οι γονότυποι BB+AB έναντι των AA για επικρατές μοντέλο, BB έναντι των AB+AA αντιστοιχεί σε υπολειπόμενο μοντέλο ενώ το B έναντι του A αλληλομόρφου δείχνει συνεπικρατές μοντέλο κληρονομικότητας) και κατόπιν υπολογίζεται για κάθε μελέτη ο σχετικός λόγος συμπληρωματικών πιθανοτήτων (Odds Ratio – OR) για διχοτομικές εκβάσεις ή η μέση διαφορά (επίσης γνωστή ως σταθμισμένη μέση διαφορά (weighted mean difference - WMD)) για συνεχείς εκβάσεις. Έτσι, για παράδειγμα, για το συνεπικρατές μοντέλο σε διχοτομικές εκβάσεις ισχύει ότι:

$$OR = \frac{\text{ποσοστό ασθενών με το αλληλόμορφο B} / \text{ποσοστό ασθενών με το αλληλόμορφο A}}{\text{ποσοστό υγιών με το αλληλόμορφο B} / \text{ποσοστό υγιών με το αλληλόμορφο A}}$$

Πίνακας 1.2. Μια τυπική μορφή των δεδομένων που χρησιμοποιούνται σε μια μετα-ανάλυση μελετών γενετικής συσχέτισης ασθενών-μαρτύρων ενός γενετικού τύπου με δυο αλληλόμορφα με μια διχοτομική έκβαση. Απεικονίζεται η κατανομή των γονοτύπων (AA, AB και BB) για ασθενείς και υγιείς, ενώ οι διαφορετικές μελέτες ($i=1,2,\dots,k$) που περιλαμβάνονται στη μετα-ανάλυση δίνονται σε διαφορετικές γραμμές.

Μελέτη	Ασθενείς			Υγιείς		
1	nAA_{11}	nAB_{11}	nBB_{11}	nAA_{01}	nAB_{01}	nBB_{01}
2	nAA_{12}	nAB_{12}	nBB_{12}	nAA_{02}	nAB_{02}	nBB_{02}
...
k	nAA_{1k}	nAB_{1k}	nBB_{1k}	nAA_{0k}	nAB_{0k}	nBB_{0k}

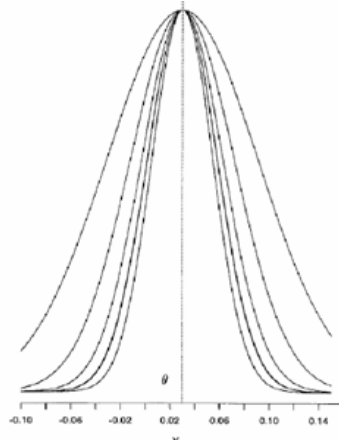
Πίνακας 1.3. Μια τυπική μορφή των δεδομένων που χρησιμοποιούνται σε μια μετα-ανάλυση μελετών γενετικής συσχέτισης ενός γενετικού τύπου με δυο αλληλόμορφα με μια συνεχή έκβαση. Απεικονίζεται η κατανομή των γονοτύπων (AA, AB και BB) για όλα τα άτομα (N), μαζί με το μέσο όρο και την τυπική απόκλιση της συνεχούς έκβασης για κάθε ομάδα. Οι διαφορετικές μελέτες ($i=1, 2 \dots k$) που περιλαμβάνονται στη μετα-ανάλυση δίνονται σε διαφορετικές γραμμές.

Μελέτη	Γονότυποι					
	AA		AB		BB	
	<i>mean (sd)</i>	<i>N</i>	<i>mean (sd)</i>	<i>N</i>	<i>mean (sd)</i>	<i>N</i>
1	$y_{AA1} (s_{AA1})$	AA_1	$y_{AB1} (s_{AB1})$	AB_1	$y_{BB1} (s_{BB1})$	BB_1
2	$y_{AA2} (s_{AA2})$	AA_2	$y_{AB2} (s_{AB2})$	AB_2	$y_{BB2} (s_{BB2})$	BB_2
...
k	$y_{AAk} (s_{AAk})$	AA_k	$y_{ABk} (s_{ABk})$	AB_k	$y_{BBk} (s_{BBk})$	BB_k

1.7.1. Μοντέλα σταθερών και τυχαίων επιδράσεων στη μετα-ανάλυση

Εφόσον υπολογιστεί το κατάλληλο μέτρο επίδρασης (y_i) για κάθε μελέτη ($i=1, 2 \dots k$), ανάλογα με τη μορφή των δεδομένων (Odds Ratios για διακριτές εκβάσεις και μέσες διαφορές για συνεχείς εκβάσεις), προχωρούμε στη διεξαγωγή μιας μονομεταβλητής μετα-ανάλυσης σταθερών ή τυχαίων επιδράσεων (Petiti 1994; Normand 1999). Στο μοντέλο σταθερών επιδράσεων γίνεται η υπόθεση ότι σε κάθε μελέτη μετράται η ίδια υποκείμενη παράμετρος και έτσι δεν υπάρχει μεταξύ των μελετών μεταβλητότητα. Πιο συγκεκριμένα, έστω ότι $i=1,2,\dots,k$ είναι ο αριθμός των μελετών και y_i το συνοπτικό μέτρο επίδρασης από κάθε μελέτη. Θεωρείται ότι το y_i προέρχεται από ένα πληθυσμό εκτιμήσεων των μελετών οι οποίες έχουν ένα κοινό μέσο θ , την παράμετρο που μας ενδιαφέρει. Για μελέτες με μεγάλο μέγεθος δείγματος, κάθε y_i ακολουθεί ασυμπτωτικά την κανονική κατανομή, δηλαδή $y_i \sim N(\theta, s_i^2)$ όπου $s_i^2 = \text{var}(y_i)$. Η εκτίμηση μέγιστης πιθανοφάνειας για το θ είναι:

$$\hat{\theta} = \frac{\sum_{i=1}^k w_i y_i}{\sum_{i=1}^k w_i} \text{ με } w_i = \frac{1}{s_i^2} \quad (1.1)$$



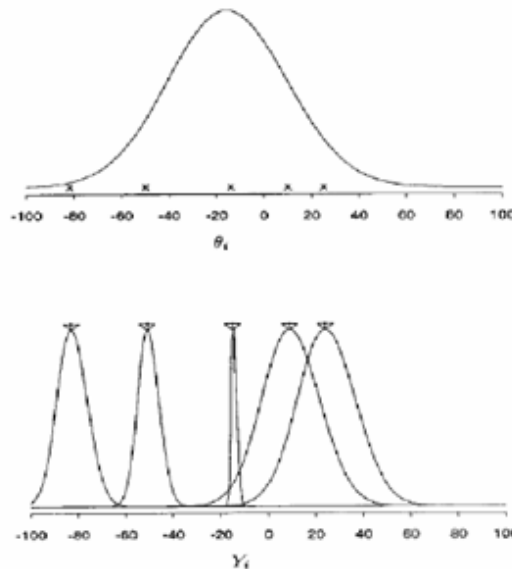
Εικόνα 1.4. Στο μοντέλο σταθερών επιδράσεων, γίνεται η παραδοχή ότι όλες οι μελέτες προέρχονται από έναν κοινό πληθυσμό με όμοια χαρακτηριστικά (ομοιογένεια αποτελεσμάτων). Σε αυτή την περίπτωση, οι τυχόν διαφορές που παρατηρούνται οφείλονται στην τύχη. Έτσι, όλες οι ανεξάρτητες μελέτες ($y_i \sim N(\theta, s_i^2)$ για $i=1,2,\dots,k$) έχουν ένα μέσο κοινό εκτιμητή θ (π.χ. Odds Ratio).

Για το μοντέλο των σταθερών επιδράσεων, ο συνολικός εκτιμητής υπολογίζεται σταθμίζοντας με την αντίστροφη ποσότητα της διασποράς του εκτιμητή της κάθε μελέτης (Inverse Variance Method), ενώ άλλες εναλλακτικές μέθοδοι είναι η μέθοδος του Peto (Yusuf, Peto et al. 1985) ή η μέθοδος των Mantel-Haenszel (Greenland and Robins 1985).

Στο μοντέλο τυχαίων επιδράσεων υποθέτουμε ότι όλες οι μελέτες που συμπεριλαμβάνονται στη μετα-ανάλυση είναι ένα τυχαίο δείγμα κάποιου ολικού πληθυσμού μελετών. Θεωρείται ότι η πραγματική επίδραση δεν είναι σταθερή αλλά ακολουθεί κάποια κατανομή (κανονική). Πιο συγκεκριμένα, το κάθε y_i προέρχεται από ένα πληθυσμό με ειδικό ανά μελέτη θ_i και διασπορά s_i^2 , δηλαδή $y_i | \theta_i, s_i^2 \sim N(\theta_i, s_i^2)$. Τα θ_i θεωρείται ότι προέρχονται από έναν υποπληθυσμό επιδράσεων έτσι ώστε $\theta_i | \theta, \tau^2 \sim N(\theta, \tau^2)$. Έτσι $y_i \sim N(\theta, s_i^2 + \tau^2)$. Το τ^2 είναι δείκτης της ετερογένειας μεταξύ των μελετών και εκτιμάται. Αν η ετερογένεια είναι απύουσα, το τ^2 είναι ίσο με το μηδέν και τα αποτελέσματα που προκύπτουν από τα μοντέλα των τυχαίων και των σταθερών επιδράσεων συμπίπτουν. Για το μοντέλο των τυχαίων επιδράσεων ο πιο απλός εκτιμητής του τ^2 προτάθηκε από τους

DerSimonian και Laird και βασίζεται στη μέθοδο των ροπών (DerSimonian and Laird 1986). Άλλες μέθοδοι έχουν παρουσιαστεί από τον Thompson και Sharp (Thompson and Sharp 1999), και περιλαμβάνουν επαναληπτικές διαδικασίες όπως η μέγιστη πιθανοφάνεια (maximum likelihood-ML), η δεσμευμένη μέγιστη πιθανοφάνεια (restricted maximum likelihood-REML) και εκτιμητές empirical Bayes (EB). Η εκτίμηση μέγιστης πιθανοφάνειας για το θ στην περίπτωση του μοντέλου τυχαίων επιδράσεων είναι:

$$\hat{\theta}(\tau) = \frac{\sum_{i=1}^k w_i(\tau) y_i}{\sum_{i=1}^k w_i(\tau)} \text{ με } w_i(\tau) = \frac{1}{s_i^2 + \tau^2} \quad (1.2)$$



Εικόνα 1.5. Στο μοντέλο τυχαίων επιδράσεων, γίνεται η παραδοχή ότι οι μελέτες προέρχονται από διαφορετικούς πληθυσμούς, οι οποίοι εμφανίζουν διαφορετικές μέσες τιμές και διασπορές (ετερογένεια αποτελεσμάτων). Σε αυτή την περίπτωση, υπάρχουν πραγματικές διαφορές ανάμεσα στις επιμέρους εκτιμήσεις. Έτσι, η καθεμία ανεξάρτητη μελέτη ($y_i | \theta_i, s_i^2 \sim N(\theta_i, s_i^2)$, $\theta_i | \theta, \tau^2 \sim N(\theta, \tau^2)$, $i=1,2,\dots,k$) έχει έναν ξεχωριστό εκτιμητή θ_i (π.χ. Odds Ratio) με κάποια διασπορά (s_i^2). Το κάθε θ_i θεωρείται ότι είναι δείγμα που προέρχεται από ένα σύνολο εκτιμητών με μέση τιμή θ και διασπορά τ^2 . Συνεπώς, σε αυτό το μοντέλο εισάγεται μια νέα μεταβλητή τ^2 που αντιπροσωπεύει την παρατηρούμενη μεταξύ των μελετών ετερογένεια.

1.7.2. Η πολυμεταβλητή προσέγγιση

Η μονομεταβλητή μετα-ανάλυση που παρουσιάστηκε είναι κατάλληλη μόνο για μονοδιάστατες παραμέτρους. Η πολυμεταβλητή μετα-ανάλυση μοντελοποιεί από κοινού τις εκτιμήσεις που προκύπτουν από κάθε μελέτη (van Houwelingen, Arends et

al. 2002; Jackson, Riley et al. 2011; Mavridis and Salanti 2012). Υπό αυτό το πλαίσιο μοντελοποίησης, παρέχονται εκτιμήσεις για όλα τα μέτρα επίδρασης, διευκολύνεται η σύγκριση των αποτελεσμάτων, αποφεύγονται οι πολλαπλοί έλεγχοι και συνεπώς η υπερεκτίμηση του σφάλματος τύπου I. Επιπλέον, λαμβάνεται υπόψη η πιθανή συσχέτιση μεταξύ των παρατηρήσεων οδηγώντας σε αξιόπιστες και ακριβείς εκτιμήσεις ενώ η μονομεταβλητή προσέγγιση αγνοεί αυτή τη δομή (Jackson, Riley et al. 2011). Ο Kirkham και συνεργάτες έδειξαν ένα σημαντικό πλεονέκτημα της πολυμεταβλητής προσέγγισης στην περίπτωση επιλεκτικής παρουσίας των αποτελεσμάτων (Kirkham, Riley et al. 2012). Συνολικά, η πολυμεταβλητή μετα-ανάλυση μπορεί να μοντελοποιήσει μελέτες που παρέχουν εκτιμήσεις μόνο κάποιων και όχι απαραίτητα όλων των παραμέτρων ενδιαφέροντος, δανειζόμενη έτσι «δύναμη» από εξωτερικές μελέτες (Higgins and Whitehead 1996).

Μέχρι τώρα, οι μεθοδολογίες που χρησιμοποιούνται σε μετα-αναλύσεις μελετών γενετικής συσχέτισης και ευρυγονιδωματικών μελετών προσανατολίζονται στην κατεύθυνση της διερεύνησης συσχετίσεων ενός γονιδίου με μια ασθένεια. Πολυμεταβλητές προσεγγίσεις για τη μελέτη των συσχετίσεων αυτής της μορφής, μοντελοποιούν ταυτόχρονα τα Odds Ratios που προκύπτουν από τη σύγκριση των διαφόρων γονοτύπων έτσι ώστε να παρέχεται ταυτόχρονα μια εκτίμηση για το υποκείμενο μοντέλο κληρονομικότητας (Minelli, Thompson et al. 2005; Minelli, Thompson et al. 2005) και έχουν παρουσιαστεί κάτω από ένα ενοποιημένο πλαίσιο (Bagos and Nikolopoulos 2007; Bagos 2008). Ωστόσο, κάποιες μελέτες παρέχουν πληροφορίες για περισσότερες από μια εκβάσεις οι οποίες μπορεί να είναι συσχετισμένες (για παράδειγμα ο διαβήτης με την υπέρταση). Άλλες μελέτες μπορεί να δίνουν δεδομένα για απλότυπους (Bagos 2011) ή για περισσότερους του ενός πολυμορφισμούς, του ίδιου ή γειτονικών γονιδίων που είναι σε ανισορροπία σύνδεσης (linkage disequilibrium - LD) (Bagos and Liakopoulos 2010). Επιπλέον, ένας συνεχώς αυξανόμενος αριθμός μελετών γενετικής συσχέτισης παρέχουν ταυτόχρονα πληροφορίες για συσχετίσεις γονιδίων-ασθενειών και γονιδίων-φαινοτύπων (Μεντελιανή τυχαιοποίηση) (Minelli, Thompson et al. 2004; Thompson, Minelli et al. 2005). Σε αυτές τις περιπτώσεις, κρίνεται απαραίτητη η διεξαγωγή μιας πολυμεταβλητής μετα-ανάλυσης τυχαίων επιδράσεων (van Houwelingen, Arends et al. 2002) κατασκευάζοντας επίσημους στατιστικούς ελέγχους, ενώ οι συνδυασμένες εκτιμήσεις έχουν καλύτερες ιδιότητες (μικρότερο τυπικό σφάλμα και μέσο τετραγωνικό σφάλμα) υπό την υπόθεση ότι οι ελλείπουσες εκτιμήσεις οφείλονται

στην τύχη (Riley, Abrams et al. 2007; Riley, Abrams et al. 2007). Για την καλύτερη κατανόηση της πολυμεταβλητής μετα-ανάλυσης που θα περιγραφεί στην Παράγραφο 1.7.4 κρίνεται σκόπιμη μια σύντομη εισαγωγή στις ιδιότητες της πολυμεταβλητής κανονικής κατανομής.

1.7.3. Η πολυμεταβλητή κανονική κατανομή

Η κανονική κατανομή αποτελεί τη βάση για τις περισσότερες στατιστικές εφαρμογές. Η συνάρτηση πυκνότητας πιθανότητας για τη μονοδιάστατη κανονική κατανομή είναι η ακόλουθη:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right), -\infty < y, \mu < +\infty, \sigma > 0 \quad (1.3)$$

Η πολυμεταβλητή κανονική κατανομή που ορίζεται για το τυχαίο διάνυσμα \mathbf{y} διαστάσεων $p \times 1$ έχει από κοινού συνάρτηση πυκνότητας-πιθανότητας:

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})' \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right) \quad (1.4)$$

όπου $\mathbf{y}' = (y_1, y_2, \dots, y_p)$, $\boldsymbol{\mu}$ είναι ένα διάνυσμα $p \times 1$ και Σ ένας πίνακας $p \times p$.

Είναι ιδιαίτερα ενδιαφέρον ότι η ερμηνεία των παραμέτρων είναι απλή. Το διάνυσμα $\boldsymbol{\mu}$ περιέχει τις αναμενόμενες τιμές καθεμιάς μεταβλητής, είναι δηλαδή το διάνυσμα των μέσων, ενώ ο πίνακας Σ είναι ο πίνακας με τις συνδιακυμάνσεις των μεταβλητών του τυχαίου διανύσματος. Δεδομένου ότι ο πίνακας Σ είναι πίνακας διακύμανσης προκύπτει ότι είναι θετικά ημιορισμένος, συμμετρικός και τα διαγώνια στοιχεία του δεν μπορεί να είναι αρνητικά. Επίσης, η πολυμεταβλητή κανονική κατανομή έχει $p(p+3)/2$ άγνωστες παραμέτρους, εκ των οποίων p είναι οι αναμενόμενες τιμές των τυχαίων μεταβλητών και $p(p+1)/2$ είναι τα διαφορετικά στοιχεία του πίνακα διακύμανσης. Στις ιδιότητες αυτές της πολυμεταβλητής κανονικής κατανομής έχει στηριχτεί η μέθοδος της πολυμεταβλητής μετα-ανάλυσης η οποία θα αναλυθεί εκτενέστερα στην Παράγραφο 1.7.4 (van Houwelingen, Arends et al. 2002).

Ας δούμε την ειδική περίπτωση όπου έχουμε μόνο δύο τυχαίες μεταβλητές Y_1 και Y_2 . Τότε, η από κοινού τους κατανομή έχει συνάρτηση πυκνότητας πιθανότητας:

$$f(y_1, y_2) = \frac{1}{2\pi\sigma_{Y_1}\sigma_{Y_2}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)^2 - 2\rho\left(\frac{y_1-\mu_{Y_1}}{\sigma_{Y_1}}\right)\left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right) + \left(\frac{y_2-\mu_{Y_2}}{\sigma_{Y_2}}\right)^2\right)\right) \quad (1.5)$$

όπου $\rho = \frac{\sigma_{Y_1 Y_2}}{\sigma_{Y_1} \sigma_{Y_2}}$ είναι η συσχέτιση των τυχαίων μεταβλητών Y_1 και Y_2 . Το διάνυσμα

των μέσων είναι $\boldsymbol{\mu} = \begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{bmatrix}$ και ο πίνακας συνδιακύμανσης $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{Y_1} & \sigma_{Y_1 Y_2} \\ \sigma_{Y_1 Y_2} & \sigma_{Y_2} \end{bmatrix}$.

Αν $\rho=0$, τότε η από κοινού συνάρτηση πυκνότητας πιθανότητας της διμεταβλητής κανονικής κατανομής είναι απλά το γινόμενο δύο συναρτήσεων πιθανοτήτων απλών κανονικών κατανομών, δηλαδή οι δύο μεταβλητές είναι ανεξάρτητες.

1.7.4. Η πολυμεταβλητή μετα-ανάλυση

Σε πολλές περιπτώσεις, σε κάθε μελέτη χρειάζεται να μοντελοποιήσουμε περισσότερες της μιας παραμέτρους. Για να προχωρήσουμε σε μια από κοινού ανάλυση, ακολουθώντας το γενικό πολυμεταβλητό πλαίσιο της μετα-ανάλυσης (Berkey, Hoaglin et al. 1998; van Houwelingen, Arends et al. 2002), ορίζουμε ως \mathbf{y}_i το διάνυσμα που περιέχει τις p διαφορετικές εκτιμήσεις και με $\boldsymbol{\beta}$, το διάνυσμα που περιέχει τους συνολικούς μέσους, οι οποίοι δίνονται από:

$$\mathbf{y}_i = \begin{pmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \dots \\ \hat{y}_{pi} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix} \quad (1.6)$$

Στην πολυμεταβλητή μετα-ανάλυση τυχαίων επιδράσεων θεωρούμε ότι το \mathbf{y}_i ακολουθεί μια πολυμεταβλητή κανονική κατανομή (multivariate normal distribution - *MVN*) γύρω από τις συντεταγμένες του διανύσματος $\boldsymbol{\beta}$, σύμφωνα με το περιθώριο μοντέλο (marginal model):

$$\mathbf{y}_i \sim MVN(\boldsymbol{\beta}, \boldsymbol{\Sigma} + \mathbf{C}_i) \quad (1.7)$$

Στο ανωτέρω μοντέλο, ορίζουμε ως \mathbf{C}_i τον εντός των μελετών πίνακα συνδιακύμανσης:

$$C_i = \begin{pmatrix} s_{1i}^2 & \rho_{w12i}s_{1i}s_{2i} & \cdots & \rho_{w1pi}s_{1i}s_{pi} \\ \rho_{w12i}s_{1i}s_{2i} & s_{2i}^2 & \cdots & \rho_{w2pi}s_{2i}s_{pi} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{w1pi}s_{1i}s_{pi} & \rho_{w2pi}s_{2i}s_{pi} & \cdots & s_{pi}^2 \end{pmatrix} \quad (1.8)$$

και με Σ τον μεταξύ των μελετών πίνακα συνδιακύμανσης:

$$\Sigma = \begin{pmatrix} \tau_1^2 & \rho_{B12}\tau_1\tau_2 & \cdots & \rho_{B1p}\tau_1\tau_p \\ \rho_{B12}\tau_1\tau_2 & \tau_2^2 & \cdots & \rho_{B2p}\tau_2\tau_p \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{B1p}\tau_1\tau_p & \rho_{B2p}\tau_2\tau_p & \cdots & \tau_p^2 \end{pmatrix} \quad (1.9)$$

Τα διαγώνια στοιχεία του πίνακα C_i είναι οι ειδικές για κάθε μελέτη εκτιμήσεις τις διασποράς που θεωρούνται γνωστές (van Houwelingen, Arends et al. 2002). Αυτή η προϋπόθεση, σε σχέση με τον αν τα στοιχεία αυτά εκτιμώνται από τα δεδομένα, αποτελεί τη βάση για την κλασσική μετα-ανάλυση και έχει μικρή επίδραση στα αποτελέσματα (Hardy and Thompson 1996). Τα στοιχεία εκτός της διαγωνίου του πίνακα C_i αντιστοιχούν στις ανά δύο εντός των μελετών συνδιακυμάνσεις (για παράδειγμα $\rho_{w12i}s_{1i}s_{2i} = \text{cov}(\hat{y}_{1i}, \hat{y}_{2i})$ βλέπε Παράγραφο 1.7.5). Τα διαγώνια στοιχεία του πίνακα Σ αντιστοιχούν στη μεταξύ των μελετών μεταβλητότητα ενώ τα στοιχεία που βρίσκονται εκτός της διαγωνίου αντιστοιχούν στις μεταξύ των μελετών συνδιακυμάνσεις και εκτιμώνται. Στην περίπτωση όπου είναι μεγάλος ο αριθμός των παραμέτρων που πρέπει να εκτιμηθούν, μπορεί να δοθεί μια δομή για τα στοιχεία του πίνακα Σ .

Το γενικό πολυμεταβλητό μοντέλο της εξίσωσης 1.7 μπορεί να υλοποιηθεί σε οποιοδήποτε στατιστικό πακέτο με μοντέλα τυχαίων επιδράσεων. Η εντολή `mvmeta` στο Stata εκτιμά τις παραμέτρους του μοντέλου χρησιμοποιώντας είτε μέγιστη πιθανοφάνεια είτε δεσμευμένη μέγιστη πιθανοφάνεια με απευθείας μεγιστοποίηση της πιθανοφάνειας μέσω του αλγορίθμου Newton-Raphson (White 2009). Εναλλακτικά, η εντολή `mvmeta` μπορεί να χρησιμοποιηθεί για την υλοποίηση της πολυμεταβλητής εκδοχής της μεθόδου των ροπών των DerSimonian και Laird's (Jackson, White et al. 2010; Jackson, Riley et al. 2011). Η τελευταία επιλογή, όντας μη επαναληπτική, είναι αρκετά ελκυστική για την περίπτωση των ευρυγονιδιωματικών μελετών.

1.7.5. Υπολογισμός των εντός των μελετών συνδιακυμάνσεων

Έχουν προταθεί διάφορες προσεγγίσεις για την αντιμετώπιση του προβλήματος των συσχετισμένων εκτιμήσεων (Gleser and Olkin 1994). Ωστόσο, στη γενική περίπτωση η εντός των μελετών συσχέτιση μπορεί να υπολογιστεί χρησιμοποιώντας ατομικά δεδομένα (Berkey, Hoaglin et al. 1998; van Houwelingen, Arends et al. 2002). Ειδικές περιπτώσεις όπου η συσχέτιση μπορεί να υπολογιστεί αναλυτικά χρησιμοποιώντας αποκλειστικά τα διαθέσιμα συνοπτικά δεδομένα κάθε μελέτης, περιλαμβάνουν μελέτες με πολλαπλές θεραπείες (Higgins and Whitehead 1996) ή πολλαπλές εκβάσεις (Gleser and Olkin 1994), κλινικές δοκιμές με αμοιβαίως αποκλειόμενες εκβάσεις (Trikalinos and Olkin 2008), μοντέλα δόσης-απόκρισης (Greenland and Longnecker 1992; Berlin, Longnecker et al. 1993; Berrington and Cox 2003) και πολυμεταβλητή μοντελοποίηση του λογαρίθμου των Odds Ratios σε μελέτες γενετικής συσχέτισης (Bagos 2008). Πρόσφατα, το πρόβλημα των συσχετισμένων εκτιμήσεων έχει αποκτήσει ιδιαίτερο ενδιαφέρον στις ευρυγονιδιωματικές μελέτες με αλληλοεπικαλυπτόμενα άτομα (Lin and Sullivan 2009; Zaykin and Kozbur 2010), στις εξομοιωμένες ή μη μελέτες ασθενών-μαρτύρων που χρησιμοποιούν τους ίδιους ασθενείς (le Cessie, Nagelkerke et al. 2008) και στην multipoint μετα-ανάλυση μελετών γενετικής συσχέτισης (βλέπε Παράγραφο 1.7.8 για λεπτομέρειες) (Bagos and Liakopoulos 2010). Ωστόσο, σε άλλες πιο γενικές περιπτώσεις, η συσχέτιση δεν μπορεί να υπολογιστεί και έχουν προταθεί τεχνικές προσομοίωσης (Daniels and Hughes 1997; Thompson, Minelli et al. 2005).

Ενσωματώνοντας όλες τις προηγούμενες προσεγγίσεις κάτω από ένα ενοποιημένο πλαίσιο, έχει προταθεί μια γενική μέθοδος για τον υπολογισμό της συνδιακύμανσης του λογαρίθμου δυο συσχετισμένων Odds Ratios χρησιμοποιώντας αποκλειστικά τις παρατηρηθείσες ποσότητες των πινάκων συνάφειας (Πίνακας 1.4) (Bagos 2012). Συνοπτικά, έστω Y_i , X_{1i} και X_{2i} τρεις κατηγορικές τυχαίες μεταβλητές με δυο επίπεδα (δηλαδή 0, 1) που χρησιμοποιούνται για την ταξινόμηση n_i ατόμων για την i μελέτη ($i=1,2,\dots,k$). Στη γενική περίπτωση, τα δεδομένα απεικονίζονται υπό τη μορφή ενός πίνακα συνάφειας τριών διαστάσεων ($2 \times 2 \times 2$) (Πίνακας 1.4) όπου τα άτομα δηλώνονται ως n_{clmi} με $c, l, m \in \{0,1\}$ και ο πίνακας αυτός αναφέρεται σαν μερικός πίνακας συνάφειας. Αποδεικνύεται ότι μια εκτίμηση για τη συνδιακύμανση δίνεται από την ακόλουθη σχέση:

$$\widehat{\text{cov}}(\hat{y}_{1i}, \hat{y}_{2i}) = \sum_c \sum_l \sum_m (-1)^{l-m} \left(\frac{n_{clmi}}{n_{cl+i} n_{c+mi}} \right) \quad (1.10)$$

Η μέθοδος που παρουσιάστηκε είναι πολύ απλή και βασίζεται σε προσεγγίσεις μεγάλου μεγέθους δείγματος (Clayton and Hills 1993; Agresti 2002).

Πίνακας 1.4. Ο πίνακας που περιγράφει την ταξινόμηση των διχοτομικών μεταβλητών Y_i , X_{1i} και X_{2i} για την i μελέτη ($i=1,2,\dots,k$). Δηλώνουμε τις ποσότητες ως n_{clmi} , με το c να αποτελεί το δείκτη για την Y_i , l για την X_{1i} και m για την X_{2i} .

		X_{1i}			
		$X_{1i}=1$		$X_{1i}=0$	
		X_{2i}			
		$X_{2i}=1$	$X_{2i}=0$	$X_{2i}=1$	$X_{2i}=0$
Y_i	$Y_i=1$	n_{111i}	n_{110i}	n_{101i}	n_{100i}
	$Y_i=0$	n_{011i}	n_{010i}	n_{001i}	n_{000i}

1.7.6. Μοντέλα μετα-ανάλυσης για τη συσχέτιση ενός γονιδίου με μια ασθένεια

Όπως έχει ήδη αναφερθεί, στην πλειοψηφία των δημοσιευμένων μετα-αναλύσεων μελετών γενετικής συσχέτισης (Attia, Thakkestian et al. 2003; Ioannidis, Trikalinos et al. 2003; Ioannidis and Trikalinos 2005), τα δεδομένα αναλύονται συνήθως κάτω από ένα πλαίσιο μονομεταβλητής μετα-ανάλυσης αγνοώντας την εντός των μελετών συσχέτιση μεταξύ των συγκρινόμενων γονοτύπων και γίνεται εκ των προτέρων υπόθεση του γενετικού μοντέλου κληρονομικότητας. Ωστόσο, από την Minelli και συνεργάτες έχει προταθεί μια μέθοδος όπου το μοντέλο κληρονομικότητας εκτιμάται από τα δεδομένα (Minelli, Thompson et al. 2005). Πιο συγκεκριμένα, εισήγαν την από κοινού μοντελοποίηση του λογαρίθμου του OR_{BB} , το οποίο είναι το Odds Ratio του γονοτύπου BB έναντι του AA, και του λ που αποτελεί το λόγο του λογαρίθμου του OR_{AB} (Odds Ratio του γονοτύπου AB έναντι του AA) και του λογαρίθμου του OR_{BB} , μια προσέγγιση που λαμβάνει υπόψη το γεγονός ότι τα δυο Odds Ratios είναι συσχετισμένα. Το λ δεν έχει καθορισμένες τιμές, αλλά οι τιμές ίσες με 0, 0.5 και 1 αντιστοιχούν σε υπολειπόμενο, συνεπικρατές και επικρατές μοντέλο αντίστοιχα, ενώ τιμές >1 ή <0 υποδεικνύουν θετική ή αρνητική ετέρωση. Το μοντέλο αυτό υποθέτει

μια μόνο παράμετρο για τη μεταξύ των μελετών μεταβλητότητα τ^2 και αναγνωρίζει το λ ως παράμετρο σταθερών επιδράσεων (Minelli, Thompson et al. 2005; Bagos 2008; Jackson, Riley et al. 2011; Mavridis and Salanti 2012).

Ένα πιο γενικό μοντέλο έχει περιγραφεί από τον Bagos (Bagos 2008). Έτσι μπορεί να πραγματοποιηθεί διμεταβλητή μετα-ανάλυση θεωρώντας ότι οι λογάριθμοι των Odds Ratios για το μεταλλαγμένο αλληλόμορφο (AB έναντι AA και BB έναντι AA) είναι συσχετισμένοι, καθώς εκτιμώνται χρησιμοποιώντας την ίδια κατηγορία αναφοράς, ακολουθώντας το γενικό πολυμεταβλητό πλαίσιο μετα-ανάλυσης για συνοπτικά δεδομένα (van Houwelingen, Zwinderman et al. 1993; Berkey, Hoaglin et al. 1998). Η συσχέτιση αυτή πρέπει να λαμβάνεται υπόψη όταν κάποιος προσπαθεί να εξάγει ταυτόχρονα εκτιμήσεις για τη στατιστική σημαντικότητα των δυο Odds Ratios καθώς επίσης όταν διερευνάται το υποκείμενο μοντέλο κληρονομικότητας. Συνεπώς, το διάνυμα y_i που έχει περιγραφεί στην εξίσωση 1.6 περιέχει δυο διαφορετικές εκτιμήσεις. Η εκτίμηση των παραμέτρων του πολυμεταβλητού μοντέλου που περιγράφεται στην εξίσωση 1.7 απαιτεί τον υπολογισμό των στοιχείων εκτός της διαγωνίου του εντός των μελετών πίνακα συνδιακυμάνσεων 1.8. Δεδομένου ότι οι λογάριθμοι των Odds Ratios για κάθε γονότυπο (AB και BB) συγκρίνονται με την ίδια κατηγορία αναφοράς (AA γονότυπος), η συνδιακύμανση αυτών υπολογίζεται ως το άθροισμα της αντίστροφης ποσότητας των ασθενών και υγιών που φέρουν το γονότυπο AA (Bagos 2008; Bagos 2012).

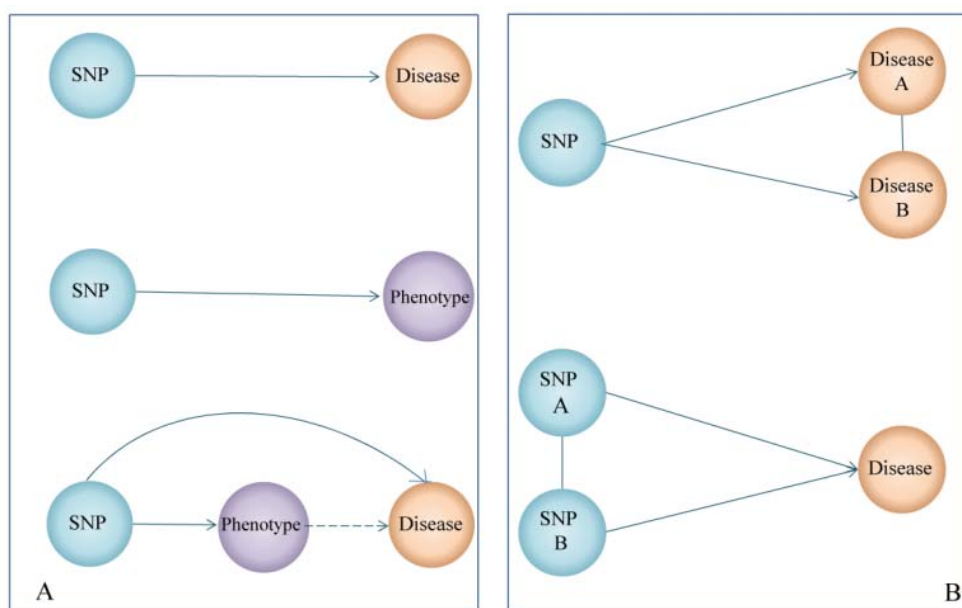
Ένα από τα πλεονεκτήματα της μεθόδου που αναφέρθηκε (Bagos 2008) είναι ότι μπορεί να υλοποιηθεί σε οποιοδήποτε στατιστικό πακέτο για πολυμεταβλητή μετα-ανάλυση (Jackson, Riley et al. 2011; Mavridis and Salanti 2012). Ένα πρόγραμμα στο στατιστικό πακέτο Stata βρίσκεται στην ιστοσελίδα <http://www.compgen.org/tools/multivariate-genetic>. Από την άλλη πλευρά το μοντέλο της Minelli και συνεργατών (Minelli, Thompson et al. 2005), απαιτεί ειδικό κώδικα για την υλοποίησή του. Ωστόσο, σε πολλές πρακτικές εφαρμογές, οι εκτιμήσεις των παραμέτρων των δυο μοντέλων ταυτίζονται. Η Minelli και συνεργάτες παρουσίασαν επίσης μια μπεϋζιανή εκδοχή του μοντέλου τους που χρησιμοποιεί την prospective likelihood (Minelli, Thompson et al. 2005). Αργότερα, ο Bagos έδειξε ότι παρόμοια μοντέλα μπορούν να εκτιμηθούν κάτω από μια frequentist προσέγγιση χρησιμοποιώντας το GLLAMM (Bagos 2011). Τέλος, οι Salanti και Higgins επέκτειναν τη μέθοδο και παρουσίασαν ένα μοντέλο που θα μπορούσε να χρησιμοποιηθεί στις περιπτώσεις όπου κάποιες μελέτες υπέθεταν το μοντέλο κληρονομικότητας και

παρουσίαζαν τα αποτελέσματά τους υπό τη μορφή των collapsed γονοτύπων (Salanti and Higgins 2008).

Οι μέθοδοι που έχουν παρουσιαστεί ως τώρα χρησιμοποιούν συνοπτικά δεδομένα και την προσέγγιση της κανονικής κατανομής. Ωστόσο, θα μπορούσαμε να εκμεταλλευτούμε απευθείας τη διχοτομική φύση των δεδομένων και να πραγματοποιήσουμε την ανάλυση χρησιμοποιώντας μια κλασσική λογιστική παλινδρόμηση. Έτσι, έχουμε δυο εναλλακτικές για τη μοντελοποίηση, χρησιμοποιώντας είτε την prospective ή την retrospective likelihood. Στην πρώτη περίπτωση, θέτουμε ως εξαρτημένη μεταβλητή την παρουσία ή απουσία ασθένειας και τους γονότυπους σαν ανεξάρτητες και πραγματοποιούμε μια λογιστική παλινδρόμηση. Για την περίπτωση αυτή, έχουν παρουσιαστεί μοντέλα τόσο σταθερών όσο τυχαίων επιδράσεων (Bagos and Nikolopoulos 2007). Εναλλακτικά, το μοντέλο μπορεί να παραμετροποιηθεί κάτω από ένα multinomial δειγματοληπτικό πλαίσιο χρησιμοποιώντας την retrospective likelihood. Υπό το μοντέλο αυτό, θεωρούμε τους γονότυπους σαν εξαρτημένες μεταβλητές και την παρουσία/απουσία της νόσου σαν ανεξάρτητη σε μια multinomial (polytomous) λογιστική παλινδρόμηση (McCullagh and Nelder 1989) και ακολούθως να επεκταθεί για την περίπτωση της μετα-ανάλυσης (Skrondal and Rabe-Hesketh 2003). Τα μοντέλα αυτά αποτελούν επεκτάσεις της γενικής μεθόδου πολυμεταβλητής μετα-ανάλυσης που έχει προταθεί από τους van Houwelingen και συνεργάτες (van Houwelingen, Zwinderman et al. 1993). Παρόμοια, οι μέθοδοι μπορούν να θεωρηθούν σαν επεκτάσεις των μονομεταβλητών multilevel μεθόδων για μοντέλα μετα-ανάλυσης τυχαίων επιδράσεων (Thompson and Sharp 1999; Turner, Omar et al. 2000; Higgins, Whitehead et al. 2001). Τα μοντέλα αυτά υπερτερούν σε σχέση με τα μοντέλα που βασίζονται σε συνοπτικά δεδομένα όταν δεν ισχύει η υπόθεση της κανονικότητας. Προγράμματα στο στατιστικό πακέτο Stata για την υλοποίηση των μοντέλων μετα-ανάλυσης χρησιμοποιώντας την prospective ή retrospective likelihood είναι διαθέσιμα στην ιστοσελίδα <http://www.compgen.org/tools/multivariate-genetic>.

Πολυμεταβλητές μέθοδοι μετα-ανάλυσης μπορούν επίσης να εφαρμοστούν για την περίπτωση των συνεχών εκβάσεων (Bagos 2008). Στην περίπτωση αυτή, για τους τρεις γονότυπους (AA, AB and BB) εκτιμάται η μέση τιμή του χαρακτηριστικού (y_{AA} , y_{AB} and y_{BB}) και η τυπική απόκλιση (sd_{AA} , sd_{AB} and sd_{BB}) (Πίνακας 1.3). Η διμεταβλητή μέθοδος μπορεί εύκολα να εφαρμοστεί εκτιμώντας τις διαφορές των εκβάσεων των ατόμων που φέρουν τους δύο γονότυπους κινδύνου (AB, BB) από το

γονότυπο της κατηγορίας αναφοράς AA. Αποδεικνύεται εύκολα ότι οι δυο διαφορές κατανέμονται κανονικά με μέσο ίσο με τη διαφορά των δύο μέσων όρων και διασπορά ίση με το άθροισμα των διασπορών, ενώ εύκολα μπορεί να υπολογιστεί και η αντίστοιχη συνδιακύμανση (Bagos 2008). Συνεπώς, οι διμεταβλητές μέθοδοι που δεν υποθέτουν κάποιο μοντέλο κληρονομικότητας μπορούν να χρησιμοποιηθούν εύκολα χωρίς κάποια αλλαγή. Μια γραφική απεικόνιση των συνηθέστερων κατηγοριών συσχετίσεων στην γενετική επιδημιολογία παρουσιάζεται στην Εικόνα 1.6.



Εικόνα 1.6. Κατηγορίες συσχετίσεων που συναντώνται συχνά σε μελέτες γενετικής επιδημιολογίας. (Α) Συσχέτιση γονιδίου-ασθένειας (επάνω), γονιδίου-φαινοτύπου (μέσο) και Μεντελιανή τυχαιοποίηση (κάτω) με τη διακεκομμένη γραμμή να απεικονίζει την έμμεση συσχέτιση του φαινοτύπου με την υπό μελέτη νόσο. (Β) Ένα γονίδιο που συσχετίζεται με δυο ασθένειες (επάνω) και δυο γονίδια που βρίσκονται σε ανισορροπία σύνδεσης και συσχετίζονται με μια ασθένεια (κάτω).

1.7.7. Μοντέλα μετα-ανάλυσης για τη συσχέτιση μελετών με απλοτύπους

Η ανάλυση των απλοτύπων επιτρέπει την ταυτόχρονη αξιολόγηση πολλών πολυμορφισμών που πιθανά συσχετίζονται με κάποια νόσο. Οι απλότυποι αποτελούν το συνδυασμό στενά συνδεδεμένων αλληλομόρφων σε κάποιο χρωμόσωμα και κατέχουν σημαντική θέση στη διερεύνηση της γενετικής βάσης κάποιας νόσου καθώς ο υπό μελέτη πολυμορφισμός δεν αποτελεί πάντα το αλληλόμορφο κινδύνου. Αντιθέτως, μπορεί να πρόκειται για κάποιον πολυμορφισμό που βρίσκεται σε ανισορροπία σύνδεσης με το πραγματικό αλληλόμορφο κινδύνου (Zondervan and

Cardon 2004). Ένα κύριο πρόβλημα στην ανάλυση των απλοτύπων (και συνεπώς και στις μετα-ανάλυσεις αυτών) είναι ότι απαιτείται η γνώση ή η ανακατασκευή των συχνοτήτων των απλοτύπων, συνήθως χρησιμοποιώντας κάποια προσέγγιση για imputation ελλειπουσών παρατηρήσεων (Niu 2004; Xu, Wu et al. 2004; Marchini, Cutler et al. 2006).

Μια ευρέως χρησιμοποιούμενη προσέγγιση βασίζεται στην κλασσική μέθοδο μετα-ανάλυσης συνοπτικών δεδομένων διχοτομικών χαρακτηριστικών και εκτιμά ξεχωριστές επιδράσεις του j απλοτύπου έναντι των υπολοίπων $j-1$. Ωστόσο, η μέθοδος αυτή έχει αυξημένο σφάλμα τύπου I λόγω των πολλαπλών συγκρίσεων, κυρίως όσο ο αριθμός των απλοτύπων μεγαλώνει (Becker and Knapp 2004; Becker, Cichon et al. 2005). Για το λόγο αυτό, μπορούμε να πραγματοποιήσουμε μια πολυμεταβλητή μετα-ανάλυση μοντελοποιώντας ταυτόχρονα τους λογαρίθμους των Odds Ratios που υπολογίζονται από τη σύγκριση των απλοτύπων $j=2,3,...,r$ σε σχέση με την κατηγορία αναφοράς ($j=1$). Συνεπώς, το διάνυσμα y_i που περιγράφεται στην εξίσωση 1.6 περιλαμβάνει $r-1$ διαφορετικές εκτιμήσεις. Τα στοιχεία του εντός των μελετών πίνακα συνδιακυμάνσεων υπολογίζονται ως το άθροισμα της αντίστροφης ποσότητας των ασθενών και υγιών που φέρουν τον απλότυπο αναφοράς. Η μέθοδος αυτή επιτρέπει έναν συνολικό έλεγχο για τη συσχέτιση των απλοτύπων με αυξημένη ακρίβεια δεδομένου ότι οι συσχετίσεις μιας απλής πολυμορφικής θέσης παρουσιάζουν μειωμένη ισχύ, που εξαρτάται από το βαθμό της ανισορροπίας σύνδεσης και τη συχνότητα των αλληλομόρφων (Kaplan and Morris 2001). Άλλες μέθοδοι που στηρίζονται απευθείας στη διχοτομική φύση των δεδομένων (logistic, multinomial logistic και poisson regression) ή μεθοδολογίες για συνεχή χαρακτηριστικά έχουν περιγραφεί με λεπτομέρεια (Bagos 2011). Η ανάλυση που περιγράφηκε μπορεί να εφαρμοστεί επίσης για την περίπτωση των γονιδίων με πολλαπλά αλληλόμορφα. Πρόγραμμα στο στατιστικό πακέτο Stata που υλοποιεί την προαναφερθείσα μέθοδο βρίσκεται στην ιστοσελίδα <http://www.compgen.org/tools/haplotype>.

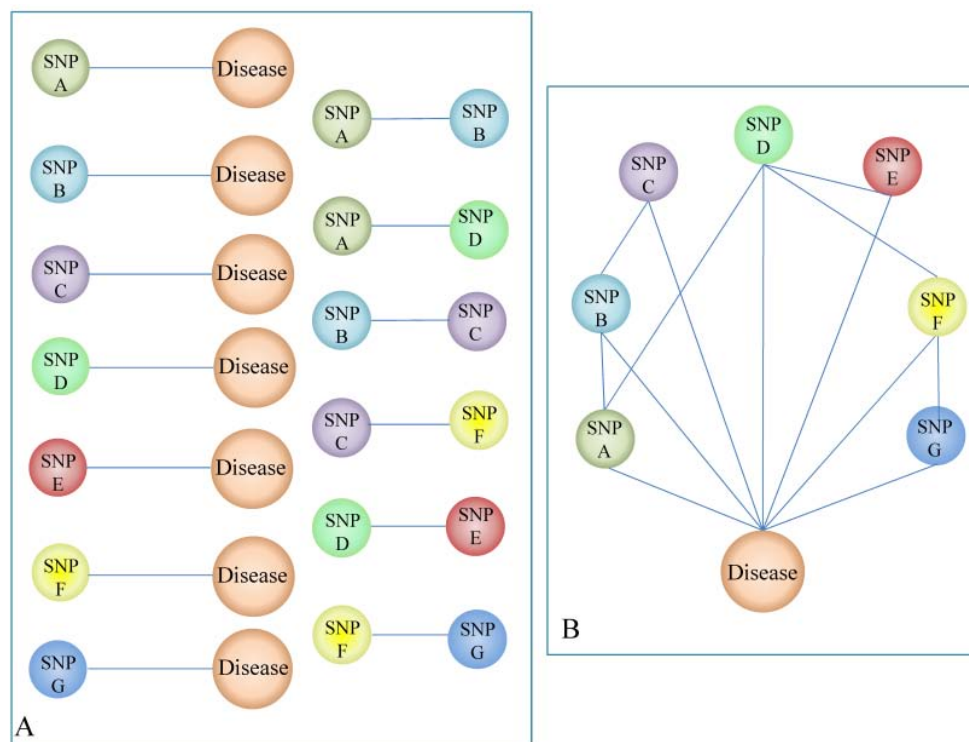
1.7.8. Μοντέλα μετα-ανάλυσης για πολλαπλούς γονιδιακούς πολυμορφισμούς που βρίσκονται σε ανισορροπία σύνδεσης

Ένα αρκετά κοντινό σενάριο με την μετα-ανάλυση των απλοτύπων, συναντάται όταν αξιολογούνται πολλαπλοί συνδεδεμένοι γενετικοί τόποι (πολυμορφισμοί του ίδιου ή γειτονικών γονιδίων). Η περίπτωση αυτή προσεγγίζει τη

συλλογιστική της μετα-ανάλυσης δικτύων (Lumley 2002) (Εικόνα 1.7). Διάφορες μέθοδοι έχουν προταθεί για την multilocus συσχέτιση (Shoemaker, Pungliya et al. 2001; Devlin, Roeder et al. 2003; Marchini, Donnelly et al. 2005). Ένα κύριο μειονέκτημα των μεθοδολογιών αυτών είναι ότι απαιτείται πρόσβαση σε ατομικά δεδομένα έτσι ώστε να γνωρίζουμε την από κοινού κατανομή των γονοτύπων διαφόρων πολυμορφισμών που βρίσκονται σε σύνδεση. Έχουν προταθεί επιπλέον μπεϋζιανές μέθοδοι για τη μετα-ανάλυση δημοσιευμένων μελετών διαφόρων συσχετισμένων γενετικών τόπων, είτε για συνεχείς (Verzilli, Shah et al. 2008), είτε για διχοτομικές εκβάσεις (μελέτες ασθενών-μαρτύρων) (Newcombe, Verzilli et al. 2009). Ωστόσο, στις προσεγγίσεις αυτές γίνεται εκ των προτέρων υπόθεση του γενετικού μοντέλου κληρονομικότητας και εισάγεται ένας αυξημένος βαθμός πολυπλοκότητας στην ανάλυση με τη χρήση πολύπλοκης παραμετροποίησης και αλγορίθμων MCMC. Πρόσφατα, έχει προταθεί μια μπεϋζιανή ημιπαραμετρική προσέγγιση όπου μοντελοποιούνται οι παρατηρηθείσες συχνότητες των γονοτύπων δοθείσης της παρουσίας/απουσίας της νόσου και χρησιμοποιούνται οι εκτιμήσεις της ανισορροπίας σύνδεσης των πολυμορφισμών σαν εκ των προτέρων πληροφορία για την εξαγωγή των εκ των υστέρων εκτιμήσεων (De Iorio, Newcombe et al. 2011). Το μοντέλο επιτρέπει την εκτίμηση του υποκείμενου μοντέλου κληρονομικότητας ενώ για τους αλληλοεπικαλυπτόμενους πολυμορφισμούς μεταξύ των μελετών χρησιμοποιείται ο βαθμός της ανισορροπίας σύνδεσης από την HarMap (HarMap) ή άλλες δημόσιες βάσεις δεδομένων. Το κυριότερο μειονέκτημα του μοντέλου αυτού είναι ότι υποθέτει έναν κοινό βαθμό ανισορροπίας σύνδεσης περιορίζοντας το συνδυασμό πληθυσμών διαφορετικών εθνοτήτων.

Μια πιο γενική μέθοδος πολυμεταβλητής μετα-ανάλυσης πολλαπλών πολυμορφισμών που βρίσκονται σε ανισορροπία σύνδεσης έχει προταθεί κάτω από μια frequentist προσέγγιση (Bagos and Liakopoulos 2010). Το μοντέλο αυτό αποτελεί μια απευθείας επέκταση του κλασσικού διμεταβλητού μοντέλου μετα-ανάλυσης μελετών γενετικής συσχέτισης (Bagos 2008) όπου ο λογάριθμος των Odds Ratios (διακριτές εκβάσεις) ή οι μέσες διαφορές (συνεχείς εκβάσεις) κάθε πολυμορφισμού μοντελοποιούνται ταυτόχρονα κάτω από ένα πολυμεταβλητό πλαίσιο. Οι σχέσεις για τον υπολογισμό των εντός των μελετών συνδιακυμάνσεων απαιτούν τη γνώση των από κοινού γονοτύπων, κάτι που σπανίως είναι διαθέσιμο. Έτσι, έχει προταθεί ο υπολογισμός των πιθανοτήτων των από κοινού γονοτύπων με τη χρήση μιας εκτίμησης του βαθμού της ανισορροπίας σύνδεσης υπό τη θεώρηση ότι υπάρχει

ισορροπία Hardy-Weinberg. Ωστόσο, ακόμα και όταν η υπόθεση αυτή παραβιάζεται χρησιμοποιείται μόνο για τον υπολογισμό των συνδιακυμάνσεων ενώ μικρή είναι η επίδραση στις συνολικές εκτιμήσεις ακόμα και όταν παρατηρούνται αποκλίσεις από την ισορροπία Hardy-Weinberg. Τέλος, η υπόθεση ότι ο βαθμός της ανισορροπίας σύνδεσης στους πληθυσμούς της μελέτης ταυτίζεται με αυτόν που περιγράφεται στη βιβλιογραφία ή στο HarMap (HarMap) μπορεί να ελεγχθεί συγκρίνοντας τις αντίστοιχες συχνότητες (Verzilli, Shah et al. 2008; Newcombe, Verzilli et al. 2009).



Εικόνα 1.7. (Α) Συσχέτιση πολλαπλών πολυμορφισμών που βρίσκονται σε ανισορροπία σύνδεσης με μια νόσο. (Β) Δίκτυο των ίδιων δεδομένων.

1.7.9. Μοντέλα μετα-ανάλυσης πολλαπλών εκβάσεων

Αρκετές μελέτες περιέχουν πληροφορίες για πολλαπλές συσχετισμένες εκβάσεις με τα ίδια ή επικαλυπτόμενα άτομα. Για παράδειγμα, σε μια μελέτη υπέρτασης συχνά μετράται ταυτόχρονα η συστολική και διαστολική πίεση (Ehret, Munroe et al. 2011; Franceschini, Fox et al. 2013). Επίσης, μια αρνητική συσχέτιση ισχύει για την περίπτωση των αμοιβαίως αποκλειόμενων εκβάσεων (Trikalinos and Olkin 2008). Συσχετισμένες εκτιμήσεις προκύπτουν ακόμα στις ευρυγονιδιωματικές μελέτες που χρησιμοποιούν τους ίδιους υγιείς (Zaykin and Kozbur 2010).

Έχουν προταθεί μονομεταβλητές μέθοδοι μετα-ανάλυσης όπου τα συνοπτικά δεδομένα από κάθε μελέτη χρησιμοποιούνται για τον υπολογισμό ενός στατιστικού

που εκτιμά τις συσχετίσεις γενετικών τόπων με πολλαπλούς φαινότυπους, οι οποίες είναι γνωστές ως cross-phenotype (CP) associations. Ο υπολογιζόμενος εκτιμητής για την Cross Phenotype Meta-Analysis (CPMA) (Cotsapas, Voight et al. 2011) δεν αποκαλύπτει ποιοι φαινότυποι συσχετίζονται με το τον υπό μελέτη πολυμορφισμό και συνεπώς είναι απαραίτητη μια clustering ανάλυση. Στην CPMA θεωρείται ότι τα p-values για το κάθε χαρακτηριστικό προέρχονται από μη επικαλυπτόμενες μελέτες κοορτής. Μια συνάρτηση στην R (cpma) περιλαμβάνεται στο πακέτο «omics» για τη διεξαγωγή της ανάλυσης. Η subset-based προσέγγιση αποτελεί τη μόνη διαθέσιμη μέθοδο που ανιχνεύει ένα υποσύνολο χαρακτηριστικών τα οποία επηρεάζονται από συγκεκριμένους πολυμορφισμούς αλλά υπεισέρχεται αυξημένο σφάλμα τύπου I λόγω των πολλαπλών ελέγχων (Bhattacharjee, Rajaraman et al. 2012). Η μέθοδος αυτή έχει υλοποιηθεί στην R ως πακέτο (ASSET). Ποικίλες παραλλαγές του ελέγχου O'Brien linear combination test (O'Brien 1984) έχουν προταθεί οι οποίες λαμβάνουν υπόψη την ετερογένεια των συσχετισμένων εκβάσεων που μετρούνται στα ίδια άτομα (Xu, Tian et al. 2003; Yang, Wu et al. 2010). Επιπλέον, έχει προταθεί μια αρκετά συναφής μεθοδολογία όπου επιτρέπεται η χρήση του μεγέθους δείγματος ως βάρος και της ετερογένειας, χρησιμοποιώντας μόνο τις συνοπτικές εκτιμήσεις (Zhu, Feng et al. 2015). Το λογισμικό CPASSOC (Cross Phenotype Association) για την πραγματοποίηση της ανάλυσης διατίθεται δωρεάν στην ιστοσελίδα <http://hal.case.edu/zhu-web/>. Η προσέγγιση αυτή μπορεί να εμπλουτιστεί χρησιμοποιώντας μια ενδιαφέρουσα μεθοδολογία που παρουσιάστηκε από τον Huang και συνεργάτες (Huang, Johnson et al. 2011). Οι ερευνητές αυτοί ανέπτυξαν το λογισμικό PRIME (<https://www.framinghamheartstudy.org/researchers/gwas-pleiotropic-tool.php>) για τον υπολογισμό του Pleiotropy Index (PI) ενός συνόλου πολυμορφισμών. Ωστόσο, η μεθοδολογία αυτή εντοπίζει γενικώς πολυμορφισμούς που είναι συστηματικά σημαντικοί μεταξύ των πολλαπλών ανεξάρτητων ελέγχων για κάθε χαρακτηριστικό για τα ίδια δεδομένα ή μέσω μετα-ανάλυσης των ανεξάρτητων μελετών. Η μέθοδος TATES (Trait-based Association Test όπου χρησιμοποιείται η διαδικασία Extended Simes) απαιτεί μόνο τα p-values μιας κλασσικής μονομεταβλητής ευρυγονιδιωματικής μελέτης για τον υπολογισμό ενός trait-based p-value διορθώνοντας παράλληλα για τις συσχετίσεις μεταξύ των συνιστωσών αλλά έως τώρα δεν έχει εφαρμοστεί για μετα-ανάλυση (van der Sluis, Posthuma et al. 2013). Η TATES είναι διαθέσιμη ως ένα δωρεάν λογισμικό στην ιστοσελίδα <http://ctg.cncr.nl/software/tates>. Τέλος, έχει προταθεί μια εναλλακτική μέθοδος

(Conneely and Boehnke 2007; Conneely and Boehnke 2010) η οποία βασίζεται σε διόρθωση λόγω των πολλαπλών ελέγχων και λαμβάνει υπόψη τη συσχέτιση μεταξύ των χαρακτηριστικών.

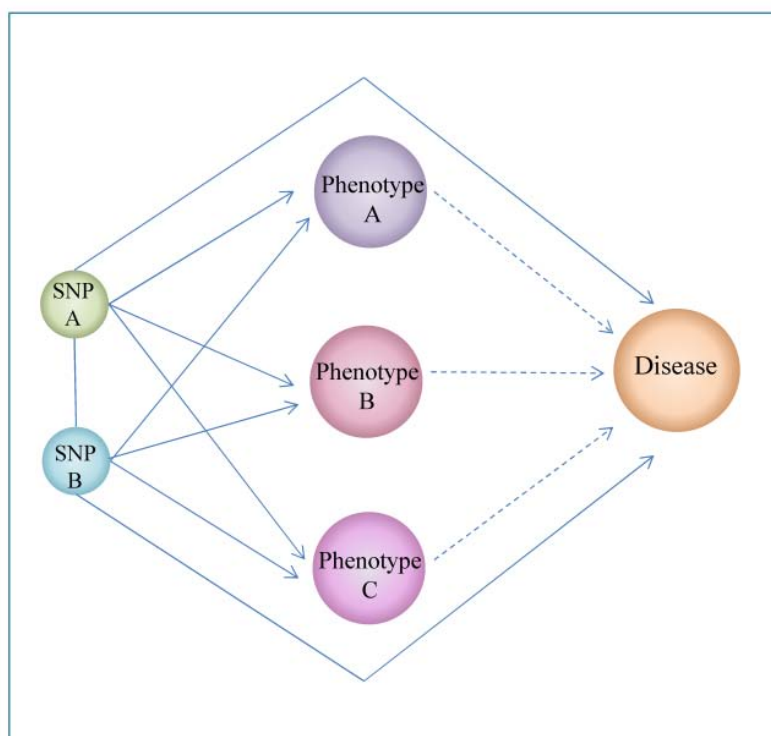
Όπως έχει γίνει αντιληπτό, οι συσχετίσεις μεταξύ των συσχετισμένων πολλαπλών εκβάσεων δεν λαμβάνονται υπόψη στην περίπτωση όπου διεξάγονται ξεχωριστές μονομεταβλητές αναλύσεις με συνεπακόλουθη μείωση της ισχύος για την ανίχνευση συσχετίσεων και έτσι έχουν προταθεί διάφορες πολυμεταβλητές μέθοδοι (Shriner 2012; Yang and Wang 2012; Solovieff, Cotsapas et al. 2013). Μια ελκυστική πολυμεταβλητή μεθοδολογία μετα-ανάλυσης, τόσο για συνεχείς όσο και για διακριτές εκβάσεις θα παρουσιαστεί αναλυτικά στην Παράγραφο 4. Συνοπτικά, επεκτείνοντας την κλασσική μέθοδο διμεταβλητής μετα-ανάλυσης που παρουσιάστηκε στην Παράγραφο 1.7.6 (Bagos 2008) εκτιμώνται δυο λογάριθμοι των Odds Ratios ή μέσες διαφορές (AB έναντι του AA και BB έναντι του AA) για κάθε έκβαση, αν δεν υποθέσουμε κάποιο μοντέλο κληρονομικότητας εκ των προτέρων, που μπορούν να μοντελοποιηθούν ταυτόχρονα κάτω από ένα πολυμεταβλητό πλαίσιο. Οι εντός των μελετών συνδιακυμάνσεις για τις διακριτές εκβάσεις υπολογίζονται χρησιμοποιώντας την από κοινού ταξινόμηση των γονοτύπων για τις εξεταζόμενες εκβάσεις, η οποία μπορεί να εκτιμηθεί χρησιμοποιώντας ένα log-linear μοντέλο και τον iterative proportional fitting algorithm κάνοντας την υπόθεση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης. Δεδομένου ότι το μοντέλο βασίζεται στο γενικό μοντέλο πολυμεταβλητής μετα-ανάλυσης (van Houwelingen, Arends et al. 2002; Mavridis and Salanti 2012), μπορεί να επεκταθεί εύκολα για πολλαπλές εκβάσεις, να υλοποιηθεί χρησιμοποιώντας κάποιο κλασσικό στατιστικό πακέτο πολυμεταβλητής μετα-ανάλυσης (Jackson, Riley et al. 2011; Mavridis and Salanti 2012) και τελικά, επιτρέπει την απευθείας σύγκριση των εκτιμώμενων παραμέτρων χρησιμοποιώντας επίσημους στατιστικούς ελέγχους. Προγράμματα στο στατιστικό πακέτο Stata τόσο για συνεχείς όσο και για διακριτές εκβάσεις μπορούν να βρεθούν στην ιστοσελίδα <http://www.compgen.org/tools/multiple-outcomes>.

1.7.10. Μοντέλα μετα-ανάλυσης Μεντελιανής τυχαιοποίησης

Ο ολοένα αυξανόμενος αριθμός μελετών που περιλαμβάνουν ταυτόχρονα συσχετίσεις γονιδίων με ασθένειες και γονιδίων με φαινότυπους οι οποίοι θεωρούνται ενδιάμεσοι στο αιτιολογικό μονοπάτι των ασθενειών καθιστά αναγκαία την ενοποίηση της

υπάρχουσας γνώσης με τη διενέργεια μετα-ανάλυσης. Η μέθοδος της «Μεντελιανής τυχαιοποίησης» έχει προταθεί έτσι ώστε να λαμβάνονται υπόψη οι συσχετίσεις μεταξύ γονοτύπου-φαινοτύπου και γονοτύπου-ασθένειας, εξάγοντας μια έμμεση εκτίμηση για τη σχέση φαινοτύπου-ασθένειας στην οποία δεν υπεισέρχονται συγχυτικοί παράγοντες (Minelli, Thompson et al. 2004; Thompson, Minelli et al. 2005). Το γράφημα του λογαρίθμου του Odds Ratio του γονοτύπου-ασθένειας έναντι της μέσης διαφοράς γονοτύπου-φαινοτύπου χρησιμεύει διερευνητικά για την ανίχνευση αποκλίσεων από τις υποθέσεις της Μεντελιανής τυχαιοποίησης. Το γράφημα αυτό αναμένεται να παρουσιάζει μια σταθερή τάση εάν ο φαινότυπος λειτουργεί ως ενδιάμεσο στο αιτιολογικό μονοπάτι της νόσου και η γραμμή διέρχεται από την αρχή των αξόνων.

Υπό το πλαίσιο της Μεντελιανής τυχαιοποίησης, μπορεί να εκτιμηθεί η συσχέτιση μεταξύ γονοτύπου-ασθένειας και γονοτύπου-φαινοτύπου χρησιμοποιώντας μια διμεταβλητή μετα-ανάλυση όπου κάποιες μελέτες και όχι απαραίτητα όλες δίνουν πληροφορίες και για τις δυο συσχετίσεις (Εικόνα 1.8). Συνεπώς, για κάθε μελέτη μπορεί να υπολογιστεί ο λογάριθμος του Odds Ratio της συσχέτισης γονοτύπου-ασθένειας και της μέσης διαφοράς στο φαινότυπο. Η εντός των μελετών συνδιακύμανση θεωρείται αμελητέα στις περισσότερες μελέτες Μεντελιανής τυχαιοποίησης. Η μεταξύ των μελετών συνδιακύμανση μπορεί να εκτιμηθεί χρησιμοποιώντας την κλασσική προσέγγιση που περιγράφηκε στην Παράγραφο 1.7.4 ή θεωρώντας ότι η ετερογένεια γονοτύπου-ασθένειας και γονοτύπου-φαινοτύπου είναι ανεξάρτητη.



Εικόνα 1.8. Γραφική απεικόνιση της επέκτασης της μεθόδου της Μεντελιανής τυχαιοποίησης. Η προσέγγιση αυτή ενσωματώνει μέτρα συσχέτισης του πολυμορφισμού με περισσότερους από έναν φαινότυπους καθώς και με την υπό μελέτη νόσο.

1.8. Στόχοι της διδακτορικής διατριβής

Αρχικός στόχος της διδακτορικής διατριβής ήταν η ανάπτυξη μεθοδολογίας και λογισμικού για την robust ανάλυση και μετα-ανάλυση ευρυγονιδιωματικών μελετών. Οι μεθοδολογίες αυτές, δεν κάνουν καμιά εκ των προτέρων υπόθεση για το μοντέλο κληρονομικότητας και κατά συνέπεια έχουν πάντα μέγιστη ισχύ, με διατήρηση του σφάλματος τύπου I στο 5%. Ακολούθως, στοχεύσαμε στην κατασκευή μιας μεθόδου μετα-ανάλυσης συνοπτικών δεδομένων για την από κοινού μοντελοποίηση δεδομένων που προέρχονται από μια μελέτη ασθενών-μαρτύρων σε οικογένειες και σε πληθυσμούς με τους ίδιους ή επικαλυπτόμενους ασθενείς επιτυγχάνοντας μεγαλύτερη ακρίβεια στις εκτιμήσεις των παραμέτρων. Επιπλέον, επικεντρωθήκαμε στην ανάπτυξη πολυμεταβλητής μεθοδολογίας μετα-ανάλυσης μελετών γενετικής συσχέτισης που περιλαμβάνουν δύο ή περισσότερες συνεχείς ή διακριτές εκβάσεις χρησιμοποιώντας μόνο τα συνοπτικά δεδομένα που δίνονται από τις δημοσιευμένες εργασίες. Μια πρακτική εφαρμογή που περιλαμβάνει τη διερεύνηση της συσχέτισης του γονιδίου της απολιποπρωτεΐνης E (πολλαπλά αλληλομόρφα) με την ανεπάρκεια

αριστερής κοιλίας σε ασθενείς με βήτα-θαλασσαιμία (οι ασθενείς ταξινομούνται σε τρεις ομάδες ανάλογα με την σοβαρότητα των συμπτωμάτων) οδηγεί στην αναγκαιότητα διαμόρφωσης μιας σύνθετης πολυμεταβλητής μεθοδολογίας μετα-ανάλυσης για την περίπτωση των πολλαπλών εκβάσεων και πολλαπλών αλληλομόρφων. Επιπλέον, προσπαθήσαμε να δώσουμε απάντηση σε σημαντικά ζητήματα στο χώρο των κλινικών μελετών και των μελετών φαρμακογονιδιοματικής όπως η επίδραση της συλλοίμωξης με τον ιό της Ηπατίτιδας C στην ανοσολογική και ιολογική ανταπόκριση HIV ασθενών μετά την έναρξη αντιρετροϊκής θεραπείας υψηλής δραστηριότητας και η διερεύνηση της συσχέτισης πολυμορφισμών του γονιδίου της αναγωγής του μεθυλενοτετραϋδροφυλλικού με την τοξικότητα από μεθοτρεξάτη. Τέλος, στοχεύσαμε στην ανάπτυξη μιας πολυμεταβλητής μεθοδολογίας για την ταυτόχρονη διενέργεια μετα-ανάλυσης και σύγκρισης δύο ή περισσότερων διαγνωστικών δοκιμασιών υπολογίζοντας αμερόληπτες και πιο ακριβείς εκτιμήσεις σε σχέση με τις υπάρχουσες μεθοδολογίες.

Κεφάλαιο 2: Robust μέθοδοι ανάλυσης και μετα-ανάλυσης σε μελέτες γενετικής συσχέτισης και σε ευρυγονιδιωματικές μελέτες

2.1. Εισαγωγή

Στο πλαίσιο των μελετών γενετικής συσχέτισης και των ευρυγονιδιωματικών μελετών υπάρχουν πολλές μεθοδολογίες για την πραγματοποίηση της ανάλυσης αλλά το κυριότερο πρόβλημα που ανακύπτει είναι η έλλειψη γνώσης του μοντέλου κληρονομικότητας. Στις περιπτώσεις αυτές, ο κλασσικός έλεγχος CATT δεν επαρκεί και καταφεύγουμε στη χρήση των robust προσεγγίσεων. Θα ξεκινήσουμε με περιγραφή των προσεγγίσεων αυτών καθώς επίσης και το πώς μπορούν να επεκταθούν για μετα-ανάλυση κάτι που δεν έχει παρουσιαστεί έως τώρα. Ο απώτερος σκοπός ήταν η υλοποίηση του ελέγχου CATT για υπολειπόμενο, επικρατές και συνεπικρατές μοντέλο κληρονομικότητας και, κυρίως, των robust προσεγγίσεων, MERT (Freidlin, Zheng et al. 2002), MAX (Zang, Fung et al. 2010) και MIN2 (Joo, Kwak et al. 2009) με την κατασκευή εντολής στο στατιστικό πρόγραμμα Stata καθώς και διαδικτυακής εφαρμογής (<http://www.compgen.org/tools/GWAR>). Όλες οι προαναφερθείσες προσεγγίσεις επεκτάθηκαν για μετα-ανάλυση σταθερών ή τυχαίων επιδράσεων συνοπτικών δεδομένων (Zhou, Shi et al. 2011).

2.2. Συμβολισμός

Έστω μια μελέτη ασθενών-μαρτύρων που εξετάζει ένα γενετικό τόπο με δυο αλληλόμορφα A και B με συχνότητες αλληλομόρφων P_A και P_B αντίστοιχα. Οι τρεις πιθανοί γονότυποι δίνονται ως $g_0=AA$, $g_1=AB$ και $g_2=BB$ με αντίστοιχες συχνότητες $\pi_j=P(g_j)$ για $j=0,1,2$. Επιπλέον, η διεισδυτικότητα της νόσου ορίζεται ως $f_j=P(\text{case}|g_j)$ και ο επιπολασμός $k=P(\text{case})=\sum f_j\pi_j$.

Πίνακας 2.1. Ο συνήθης 2x3 πίνακας συνάφειας για την κατανομή των ασθενών και υγιών σε μελέτες γενετικής συσχέτισης και σε ευρυγονιδιωματικές μελέτες για έναν γενετικό τόπο με δυο αλληλόμορφα.

	<i>AA</i> (g_0)	<i>AB</i> (g_1)	<i>BB</i> (g_2)	Total
Ασθενείς	$r_0= nAA_1$	$r_1= nAB_1$	$r_2= nBB_1$	r
Υγιείς	$s_0= nAA_0$	$s_1= nAB_0$	$s_2= nBB_0$	s
Σύνολο	$n_0= nAA_+$	$n_1= nAB_+$	$n_2= nBB_+$	n

Σε μια μελέτη ασθενών-μαρτύρων πραγματοποιείται ανεξάρτητη δειγματοληψία r ασθενών και s υγιών και, έτσι, οι γονότυποι (r_0, r_1, r_2) και (s_0, s_1, s_2) ακολουθούν multinomial κατανομή, $\text{mul}(r; p_0, p_1, p_2)$ και $\text{mul}(s; q_0, q_1, q_2)$ αντίστοιχα, με $p_j = P(g_j | \text{case}) = f_j \pi_j / k$ και $q_j = P(g_j | \text{control}) = (1 - f_j) \pi_j / (1 - k)$ για $j=0,1,2$. Η μηδενική υπόθεση ορίζεται ως $H_0: f_0 = f_1 = f_2 = k$. Αν υποθέσουμε ότι το B είναι το αλληλόμορφο κινδύνου, οι σχετικοί κίνδυνοι των γονοτύπων (genotype relative risks-GRR) ορίζονται ως $\gamma_j = f_j / f_0$ για $j=1,2$ και $f_0 > 0$. Έτσι, η μηδενική υπόθεση μπορεί να οριστεί ως $H_0: \gamma_2 = \gamma_1 = 1$ έναντι $H_1: \gamma_2 \geq \gamma_1 \geq 1$ με τουλάχιστον μια ανισότητα. Συνεπώς, το γενετικό μοντέλο είναι επικρατές, υπολειπόμενο και συνεπικρατές αν $\gamma_2 = \gamma_1 > 1$, $\gamma_2 > \gamma_1 = 1$, $\gamma_2 = 2\gamma_1 - 1 > 1$ και $\gamma_2 = \gamma_1^2 > 1$, αντίστοιχα. Ένα λιγότερο πιθανό σενάριο παρατηρείται όταν $\gamma_1 > \gamma_2 = 1$, όπου το γενετικό μοντέλο είναι overdominant.

2.3. Στατιστικοί έλεγχοι

Υπάρχει μια μεγάλη ποικιλία μεθόδων ανάλυσης των δεδομένων μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών (Balding 2006; Langefeld and Fingerlin 2007). Ίσως ο πιο προφανής τρόπος ανάλυσης των γονοτύπων σε μια μελέτη ασθενών-μαρτύρων (χωρίς να υποθέσουμε κάποιο μοντέλο κληρονομικότητας) είναι η εξέταση της συσχέτισης μεταξύ γραμμών και στηλών ενός 2×3 πίνακα συνάφειας (Πίνακας 2.1). Αυτό μπορεί να πραγματοποιηθεί χρησιμοποιώντας τον κλασσικό χ^2 έλεγχο του Pearson (χ^2) (Balding 2006; Langefeld and Fingerlin 2007):

$$T_{\chi^2}^2 = \sum_{j=0}^2 \frac{(r_j - n_j r / n)^2}{n_j r / n} + \sum_{j=0}^2 \frac{(s_j - n_j s / n)^2}{n_j s / n} \quad (2.1)$$

Κατά παρόμοιο τρόπο, μπορεί κάποιος να εφαρμόσει τον ακριβή έλεγχο του Fisher, κάτι που όμως είναι υπολογιστικά απαιτητικό. Μια άλλη επιλογή αποτελεί ο υπολογισμός των Odds Ratios που αντιστοιχούν στη σύγκριση των γονοτύπων που φέρουν το αλληλόμορφο κινδύνου (δηλαδή g_1 και g_2) έναντι του g_0 . Αυτοί οι έλεγχοι μπορούν να πραγματοποιηθούν κάτω από ένα πλαίσιο λογιστικής παλινδρόμησης χρησιμοποιώντας την παρουσία/απουσία μιας νόσου σαν εξαρτημένη μεταβλητή (Balding 2006). Σύμφωνα με την προσέγγιση αυτή, μοντελοποιείται η πιθανότητα κάποιο άτομο με γονότυπο g_j ($j=1, 2$) να έχει τη νόσο ($P(\text{case} | g_j)$); ο γονότυπος g_0 θεωρείται σαν κατηγορία αναφοράς και δημιουργούνται δυο δείκτριες μεταβλητές με τιμές $x_j=1$ για g_j και 0 αλλιώς:

$$\text{logit}\left[P(\text{case} | g_j)\right] = \alpha + \beta_1 x_1 + \beta_2 x_2 \quad (2.2)$$

Μετά την εκτίμηση των παραμέτρων του μοντέλου, μπορούμε να πραγματοποιήσουμε τον έλεγχο του Wald για $\beta_1=\beta_2=0$ με 2 βαθμούς ελευθερίας. Εναλλακτικά, μπορεί κάποιος να υπολογίσει τα Odds Ratios χρησιμοποιώντας τα συνοπτικά δεδομένα του πίνακα συνάφειας (Chen and Chatterjee 2007). Ασυμπτωτικά, αυτές οι μέθοδοι είναι ταυτόσημες με το χ^2_2 , αλλά το βασικό τους μειονέκτημα είναι η χαμηλή ισχύς.

Από την άλλη πλευρά, υπάρχουν μέθοδοι που υποθέτουν κάποιο γενετικό μοντέλο εκ των προτέρων. Αυτές οι μέθοδοι περιλαμβάνουν κάποια μορφή ανακατάταξης των δεδομένων ή μια διαφορετική βαθμολόγηση των γονοτύπων (Sasieni 1997; Zheng, Freidlin et al. 2003). Για παράδειγμα, εάν ανακατατάξουμε τα δεδομένα του Πίνακα 2.1 ομαδοποιώντας τους γονότυπους g_1 και g_2 και πραγματοποιήσουμε έναν χ^2_1 , υποθέτουμε επικρατές μοντέλο κληρονομικότητας; εάν ομαδοποιήσουμε τους γονότυπους g_0 και g_1 και τους συγκρίνουμε με τον g_2 ελέγχουμε για υπολειπόμενο μοντέλο κληρονομικότητας ενώ η σύγκριση των αλληλομόρφων στηρίζεται σε θεώρηση ενός συνεπικρατούς μοντέλου υπό τη θεώρηση ότι υπάρχει ισορροπία Hardy-Weinberg. Ωστόσο, ο πιο συνηθισμένος έλεγχος είναι ο CATT (Cochran 1954; Armitage 1955). Πρόκειται για ένα score test της γραμμικής τάσης των ποσοστών ενός πίνακα συνάφειας. Έτσι, θεωρώντας μια γενική μορφή σκορ $x_i=(x_0, x_1, x_2)$ για τους γονότυπους (g_0, g_1, g_2), μπορεί να διαμορφωθεί ένας έλεγχος Z ως εξής:

$$Z_{CATT(x)} = \frac{U_x}{\sqrt{\text{var}_{H_0}(U_x)}} = \frac{\sqrt{n} \sum_{i=0}^2 x_i (sr_i - rs_i)}{\sqrt{rsn \left[n \sum_{i=0}^2 x_i^2 n_i - \left(\sum_{i=0}^2 x_i n_i \right)^2 \right]}} \sim N(0,1) \quad (2.3)$$

όπου το U_x είναι το στατιστικό των σκορ (score test). Ο υπολογισμός τους βασίζεται στην πρώτη παράγωγο της συνάρτησης πιθανοφάνειας και παρουσιάζουν μεγαλύτερη ισχύ όταν η πραγματική τιμή της παραμέτρου που μας ενδιαφέρει είναι κοντά με τη θεωρητική. Με $\text{var}_{H_0}(U_x)$ ορίζεται η διασπορά του U_x υπό την H_0 . Τα σκορ αυτά συνήθως ορίζονται ως $x_i=(0, 1/2, 1)$, υποθέτοντας συνεπικρατές μοντέλο κληρονομικότητας ($Z_{CATT(1/2)}$). Αξίζει να αναφερθεί ότι ο έλεγχος CATT δεν επηρεάζεται από γραμμικούς μετασχηματισμούς [δηλαδή η κωδικοποίηση του $x_i=(0, 1, 2)$ είναι ισοδύναμη] και δεν εξαρτάται από την υπόθεση της ισορροπίας Hardy-Weinberg.

Ο ίδιος έλεγχος μπορεί να κατασκευαστεί εκτιμώντας ένα μοντέλο λογιστικής παλινδρόμησης παρόμοια με αυτό που περιγράφεται στην εξίσωση 2.2 χρησιμοποιώντας την ίδια κωδικοποίηση για τους γονότυπους $x_i = (0, \frac{1}{2}, 1)$. Ωστόσο, η κωδικοποίηση $x_i = (0, 1, 1)$, υποδηλώνει επικρατές μοντέλο ($Z_{CATT(1)}$), ενώ $x_i = (0, 0, 1)$ αντιστοιχεί σε υπολειπόμενο μοντέλο κληρονομικότητας ($Z_{CATT(0)}$). Συνεπώς, η λογιστική παλινδρόμηση δεν είναι, από μόνη της, μια μέθοδος που δεν υποθέτει κάποιο μοντέλο κληρονομικότητας αλλά εξαρτάται από την κωδικοποίηση των γονοτύπων που χρησιμοποιείται. Αξίζει να αναφερθεί ότι ο έλεγχος CATT είναι ο πιο ισχυρός στατιστικός έλεγχος υπό τη θεώρηση επικρατούς, συνεπικρατούς και υπολειπόμενου μοντέλου κληρονομικότητας κάτι που εξηγεί την ευρεία χρήση του στις ευρυγονιδιωματικές μελέτες (Zheng, Freidlin et al. 2003).

2.4. Robust μέθοδοι για την επιλογή γενετικού μοντέλου

Διάφορες προσεγγίσεις έχουν προταθεί για την εύρεση robust μεθόδων που επιλέγουν το σωστό μοντέλο κληρονομικότητας και πραγματοποιούν ταυτόχρονα την ανάλυση των δεδομένων μεγιστοποιώντας την ισχύ και διατηρώντας σταθερό το σφάλμα τύπου I στο 5%. Κάποιες από αυτές τις μεθόδους στηρίζονται στη θεωρία των efficiency robust procedures (Gastwirth 1985; Freidlin, Podgor et al. 1999), κάποιες άλλες χρησιμοποιούν αποκλίσεις από την ισορροπία Hardy-Weinberg για την επιλογή του υποκείμενου μοντέλου κληρονομικότητας (Zheng and Ng 2008; Joo, Kwak et al. 2010) ενώ άλλες χρησιμοποιούν έναν συνδυασμό στατιστικών ελέγχων (Joo, Kwak et al. 2009).

2.4.1. Οι μέθοδοι MAX και MERT

Οι πιο βασικές μέθοδοι στηρίζονται στη θεωρία των efficiency robust procedures (Gastwirth 1985; Freidlin, Podgor et al. 1999), και είναι ο έλεγχος MERT και MAX (Freidlin, Zheng et al. 2002). Από μια οικογένεια βέλτιστων ελέγχων, ο MERT είναι ο γραμμικός συνδυασμός δυο βέλτιστων ελέγχων για τα ακραία μέλη της οικογένειας, δηλαδή για τους ελέγχους εκείνους με την ελάχιστη συσχέτιση (ρ). Δυο συνθήκες πρέπει να ικανοποιούνται (Gastwirth 1985; Freidlin, Podgor et al. 1999) (i) $\rho \geq 0$, και (ii) το $\rho+1$ πρέπει να είναι μικρότερο από το άθροισμα των δυο άλλων συσχετίσεων. Ο Freidlin και συνεργάτες έδειξαν ότι ο MERT είναι ένας γραμμικός συνδυασμός του

βέλτιστου CATT για το υπολειπόμενο και επικρατές μοντέλο και ακολουθεί ασυμπτωτικά την τυποποιημένη κανονική κατανομή (Freidlin, Zheng et al. 2002):

$$Z_{MERT} = \frac{Z_{CATT(0)} + Z_{CATT(1)}}{\sqrt{2(1 + \rho_{CATT(0,1)})}} \sim N(0,1) \quad (2.4)$$

Ο έλεγχος MAX (αρχικά ονομάστηκε MAX3) στηρίζεται στην απλή ιδέα του ελέγχου όλων των τριών πιθανών μοντέλων και κατόπιν επιλέγεται εκείνο με το μεγαλύτερη τιμή του στατιστικού (χρησιμοποιείται η απόλυτη τιμή για την περίπτωση όπου το αλληλόμορφο κινδύνου είναι το A):

$$Z_{MAX} = \max(|Z_{CATT(0)}|, |Z_{CATT(1/2)}|, |Z_{CATT(1)}|) \quad (2.5)$$

Ωστόσο, για τον υπολογισμό του σωστού p-value (δηλαδή $P(|Z_{MAX}| > t)$) χρειάζεται να λάβουμε υπόψη τις συσχετίσεις μεταξύ των στατιστικών ελέγχων έτσι ώστε να μην οδηγηθούμε σε υπερεκτίμηση του σφάλματος τύπου I. Ο Freidlin και συνεργάτες υπολόγισαν τις συσχετίσεις κάτω από την H_0 (Freidlin, Zheng et al. 2002), αλλά ακόμα και σε αυτή την περίπτωση, απαιτούνται προσομοιώσεις ή αριθμητική ολοκλήρωση καθιστώντας τη μέθοδο υπολογιστικά δύσκολη. Προσομοιώσεις έχουν δείξει ότι ο έλεγχος MAX παρουσιάζει αυξημένη ισχύ, αλλά ο MERT είναι πιο εύχρηστος καθώς για τον υπολογισμό της στατιστικής σημαντικότητας χρησιμοποιείται η τυποποιημένη κανονική κατανομή (Freidlin, Zheng et al. 2002). Η ασυμπτωτική κατανομή του ελέγχου MAX μπορεί να υπολογιστεί χρησιμοποιώντας τη μέθοδο Δέλτα (Gonzalez, Carrasco et al. 2008), καθώς επίσης έχουν περιγραφεί και λιγότερο υπολογιστικά δύσκολες μέθοδοι (rhombus formula) (Li, Zheng et al. 2008).

Οι Hothorn και Hothorn έδειξαν ότι ο δεσμευμένος έλεγχος MAX είναι μια ειδική περίπτωση γραμμικών ελέγχων, των οποίων η δεσμευμένη κατανομή μπορεί να υπολογιστεί εύκολα (Hothorn and Hothorn 2009). Έτσι, ο έλεγχος MAX επανατοποθετήθηκε σε ένα πιο ευέλικτο πλαίσιο δεσμευμένων ανεξάρτητων ελέγχων (Strasser and Weber 1999) με αρκετά πλεονεκτήματα όπως η στρωματοποιημένη ανάλυση (η οποία μπορεί να έχει ιδιαίτερο ενδιαφέρον και στο χώρο της μετα-ανάλυσης).

Πρόσφατα, ο Zang και συνεργάτες μελέτησαν τη γραμμική εξάρτηση των $Z_{CATT(0)}$, $Z_{CATT(1/2)}$ και $Z_{CATT(1)}$, και ανέπτυξαν πιο γρήγορους αλγόριθμους για τον

υπολογισμό της στατιστικής σημαντικότητας του MAX (Zang, Fung et al. 2010). Συνεπώς, το p-value του MAX υπολογίζεται ως εξής:

$$\begin{aligned}
 P(Z_{MAX} < t) = & 2 \int_0^{t(1-\omega_1)/\omega_0} \Phi \left(\frac{t - \rho_{CATT(0,1)} z_0}{\sqrt{1 - \rho_{CATT(0,1)}^2}} \right) \phi(z_0) dz_0 \\
 & + 2 \int_{t(1-\omega_1)/\omega_0}^t \Phi \left(\frac{(t - \omega_0 z_0)/\omega_1 - \rho_{CATT(0,1)} z_0}{\sqrt{1 - \rho_{CATT(0,1)}^2}} \right) \phi(z_0) dz_0 \\
 & - 2 \int_0^t \Phi \left(\frac{-t - \rho_{CATT(0,1)} z_0}{\sqrt{1 - \rho_{CATT(0,1)}^2}} \right) \phi(z_0) dz_0
 \end{aligned} \quad (2.6)$$

όπου:

$$\omega_0 = \frac{\rho_{CATT(0,1/2)} - \rho_{CATT(0,1)} \rho_{CATT(1/2,1)}}{1 - \rho_{CATT(0,1)}^2} \quad (2.7)$$

$$\omega_1 = \frac{\rho_{CATT(1/2,1)} - \rho_{CATT(0,1)} \rho_{CATT(0,1/2)}}{1 - \rho_{CATT(0,1)}^2} \quad (2.8)$$

Αξίζει να αναφερθούν διάφορες αντιστοιχίες με υπάρχοντες ελέγχους. Για παράδειγμα, ο έλεγχος χ^2 μπορεί να γραφεί ως ο συνδυασμός των CATTs για το επικρατές και και συνεπικρατές μοντέλο (Zheng, Freidlin et al. 2006). Τέλος, ο Zheng και συνεργάτες έδειξαν ότι ο CATT, ο MAX και ο χ^2 , είναι όλοι έλεγχοι τάσης αλλά με διαφορετικά σκορ; η διαφορά έγκειται στο αν τα σκορ είναι προκαθορισμένα ή εξαρτώνται από τα δεδομένα, και στον αν τα σκορ παρουσιάζουν κάποια διαβάθμιση (Zheng, Joo et al. 2009).

2.4.2. Οι μέθοδοι GMS και GME

Μια σημαντική κατηγορία robust μεθόδων χρησιμοποιεί αποκλίσεις από την ισορροπία Hardy-Weinberg για την επιλογή του υποκείμενου μοντέλου κληρονομικότητας. Η θεωρητική αιτιολόγηση των μεθόδων αυτών έχει παρουσιαστεί στην εργασία του Wittke-Thompson και συνεργατών, οι οποίοι μελέτησαν τις αποκλίσεις από την ισορροπία Hardy-Weinberg σε ασθενείς και υγιείς σε μελέτες γενετικής συσχέτισης (Wittke-Thompson, Pluzhnikov et al. 2005). Στη μέθοδο Genetic Model Selection (GMS), ο Zheng και Ng έδειξαν ότι ο έλεγχος Transmission Disequilibrium Trend Test (Z_{HWDTT}), ο οποίος προτάθηκε από τον Song και Elston (Song and Elston 2006), έχει θετικό πρόσημο όταν το μοντέλο κληρονομικότητας

είναι υπολειπόμενο και αρνητικό όταν είναι επικρατές (Zheng and Ng 2008). Συνεπώς, πρότειναν τη χρήση του $Z_{CATT(0)}$ εάν $Z_{HWDTT} > c$, $Z_{CATT(1)}$ εάν $Z_{HWDTT} < -c$ και $Z_{CATT(1/2)}$ διαφορετικά, όπου το c είναι ένα προκαθορισμένο κατώφλι.

Ωστόσο, όπως έδειξαν ο Joo και συνεργάτες αργότερα, εάν το αλληλόμορφο κινδύνου είναι άγνωστο, κάτι που είναι αρκετά πιθανό στις μελέτες γενετικής συσχέτισης και στις ευρυγονιδιωματικές μελέτες, η μέθοδος GMS μπορεί να αποτυγχάνει (Joo, Kwak et al. 2010). Για παράδειγμα, εάν το A είναι το αλληλόμορφο κινδύνου τα $Z_{CATT(0)}$, $Z_{CATT(1/2)}$ και $Z_{CATT(1)}$ είναι βέλτιστα για το επικρατές, συνεπικρατές και υπολειπόμενο μοντέλο αντίστοιχα. Έτσι, έχει προταθεί μια τροποποιημένη προσέγγιση της μεθόδου GMS, όπου αρχικά χρησιμοποιείται το $Z_{CATT(1/2)}$ για να ανιχνεύσει το αλληλόμορφο κινδύνου και κατόπιν εφαρμόζεται η μέθοδος GMS. Δεδομένου ότι πραγματοποιούνται πολλαπλοί έλεγχοι, η συνολική σημαντικότητα πρέπει να σταθμιστεί επιτρέποντας τη συσχέτιση μεταξύ των δυο ελέγχων (Joo, Kwak et al. 2010). Συνάμα, ο Joo και συνεργάτες, στηριζόμενοι σε παλαιότερη εργασία (Lee 2003), υπογράμμισαν ότι όταν οι GRRs είναι μικροί, η πιθανότητα επιλογής του πραγματικού γενετικού μοντέλου χρησιμοποιώντας το Z_{HWDTT} μικραίνει, κυρίως λόγω της μειωμένης ισχύος να ανιχνεύονται προσθετικές επιδράσεις. Από την άλλη πλευρά, η πιθανότητα να αποκλείονται ορθά τα λιγότερο πιθανά μοντέλα παραμένει υψηλή. Έτσι, πρότειναν τη μέθοδο Genetic Model Exclusion (GME), όπου το επικρατές ή υπολειπόμενο μοντέλο κληρονομικότητας αποκλείεται εάν $Z_{HWDTT} > 0$ ή $Z_{HWDTT} < 0$ αντίστοιχα, ενώ, όταν $Z_{HWDTT} = 0$, επιλέγεται το συνεπικρατές μοντέλο (Joo, Kwak et al. 2010). Τόσο η μέθοδος GMS όσο και η GME είναι robust με ισχύ συγκρίσιμη με τον έλεγχο MAX. Ο Zang και συνεργάτες απέδειξαν ότι τα $Z_{CATT(0)}$, Z_{HWDTT} και $Z_{CATT(1)}$ είναι γραμμικώς εξαρτημένα, ένα εύρημα που τους επέτρεψε να κατασκευάσουν πιο γρήγορους αλγόριθμους για τη μέθοδο GMS (Zang, Fung et al. 2010).

2.4.3. Η μέθοδος MIN2

Μια επιπλέον ενδιαφέρουσα robust προσέγγιση υιοθετήθηκε από ερευνητές του Wellcome Trust Case-Control Consortium (WTCCC 2007). Εφάρμοσαν τον έλεγχο χ^2 μαζί με τον CATT(1/2) και ακολούθως, επέλεξαν το ελάχιστο των p-values:

$$MIN2 = \min \left(P_{T^2}, P_{Z_{CATT(1/2)}} \right) \quad (2.9)$$

Η χρήση της μεθόδου MIN2 δικαιολογείται από προσομοιώσεις που δείχνουν ότι ο χ^2 έχει 5% χαμηλότερη ισχύ σε σχέση με τον MAX ενώ είναι καλύτερος σε σχέση με τον MERT, εκτός και αν ισχύει το συνεπικρατές μοντέλο (Zheng, Freidlin et al. 2006). Ωστόσο, η μέθοδος MIN2 δεν έχει σωστό p-value μιας και τα στατιστικά είναι συσχετισμένα και πραγματοποιούνται ταυτόχρονα πολλαπλοί έλεγχοι. Ο Joo και συνεργάτες υπολόγισαν την από κοινού κατανομή η οποία απαιτείται για τον υπολογισμό του p-value (Joo, Kwak et al. 2009):

$$P\left(Z_{CATT(1/2)}^2 < t_1, T_{\chi^2}^2 < t_2\right) = \begin{cases} 1 - \frac{1}{2}e^{-\frac{t_1}{2}} - \frac{1}{2}e^{-\frac{t_2}{2}} + \frac{1}{2\pi} \int_{t_1}^{t_2} e^{-\frac{u}{2}} \arcsin\left(\frac{2t_1}{u} - 1\right) du, & t_1 < t_2 \\ 1 - e^{-\frac{t_2}{2}}, & t_1 > t_2 \end{cases} \quad (2.10)$$

Αξίζει να σημειωθούν δυο ιδιότητες της από κοινού κατανομής. Αρχικά, σε αντίθεση με τον έλεγχο MAX, η μέθοδος MIN2 είναι ανεξάρτητη από τη συχνότητα των αλληλομόρφων. Δεύτερον, το p-value της MIN2 μεγαλώνει πάντα και έτσι μπορεί να χρησιμοποιηθεί ευθέως για την κατάταξη όλων των SNPs όπως πραγματοποιήθηκε στο WTCCC (2007). Αποτελέσματα προσομοιώσεων δείχνουν ότι η MIN2 έχει χαμηλότερη ισχύ (1-4%) σε σχέση με τον έλεγχο MAX για το υπολειπόμενο μοντέλο, αλλά είναι καλύτερη σε σχέση με τον MAX για την περίπτωση του συνεπικρατούς μοντέλου (κατά ~2%), ενώ για το επικρατές μοντέλο και οι δυο έλεγχοι έχουν παρόμοια αποτελέσματα.

2.5. Μέθοδοι ανάλυσης ποσοτικών χαρακτηριστικών

Όλες οι μέθοδοι που έχουν περιγραφεί ως τώρα, περιλαμβάνουν ανάλυση δεδομένων του παραδοσιακού σχεδιασμού μελετών ασθενών-μαρτύρων. Ωστόσο, υπάρχει ολοένα και μεγαλύτερο ενδιαφέρον στις μελέτες γενετικής συσχέτισης και στις ευρυγονιδιωματικές μελέτες για τη μελέτη ποσοτικών χαρακτηριστικών όπως τα επίπεδα χοληστερόλης, η συστολική πίεση και λοιπά. Σε αυτή την περίπτωση αν ορίσουμε ως X_j την τιμή του χαρακτηριστικού για το γονότυπο g_j (και υποθέτοντας ότι το B είναι το αλληλόμορφο που προκαλεί αύξηση στο X) τότε το πρόβλημα ανάγεται σε ένα πρόβλημα γραμμικής παλινδρόμησης με το X να είναι η εξαρτημένη μεταβλητή και οι γονότυποι οι επεξηγηματικές μεταβλητές (Balding 2006). Κατά παρόμοιο τρόπο με τη λογιστική παλινδρόμηση για διχοτομικά χαρακτηριστικά, η κωδικοποίηση των γονοτύπων μπορεί να καθορίσει το υποκείμενο μοντέλο κληρονομικότητας. Για παράδειγμα, χρησιμοποιώντας μια κωδικοποίηση της μορφής

$x_i=(0, \frac{1}{2}, 1)$, $x_i=(0, 1, 1)$ και $x_i=(0, 0, 1)$ για τους γονότυπους ελέγχουμε για συνεπικρατές, επικρατές και υπολειπόμενο μοντέλο αντίστοιχα. Επιπλέον, μπορούμε να μην υποθέσουμε εκ των προτέρων κάποιο μοντέλο κληρονομικότητας χρησιμοποιώντας δύο δείκτριες $(0, 1, 0)$ και $(0, 0, 1)$ για τον g_1 και g_2 γονότυπο αντίστοιχα.

Καθίσταται συνεπώς προφανές ότι παρόμοιες στρατηγικές για την επιλογή γενετικού μοντέλου μπορούν επίσης να εφαρμοστούν σε ποσοτικά χαρακτηριστικά υπό το πρίσμα της γραμμικής παλινδρόμησης (Lettre, Lange et al. 2007). Συνεπώς, μερικές από τις robust προσεγγίσεις που συζητήθηκαν ως τώρα όπως ο έλεγχος MAX έχουν παρουσιαστεί σε ένα τέτοιο πλαίσιο (So and Sham 2011). Επιπλέον, έχειδειχθεί εύκολα ότι ο έλεγχος MERT μπορεί να χρησιμοποιηθεί για συνεχείς εκβάσεις. Ωστόσο, αυτό δεν ισχύει για άλλους ελέγχους όπως ο GMS, GME και MIN2, που αφορούν αποκλειστικά μελέτες ασθενών-μαρτύρων.

2.6. Μετα-ανάλυση μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών

Όπως έχει ήδη παρουσιαστεί στην εισαγωγή, η μετα-ανάλυση αποτελεί μια μεθοδολογία για την αύξηση της ισχύος έτσι ώστε να ανιχνεύονται ασθενείς επιδράσεις γονιδίων, να εκτιμώνται τα μέτρα επίδρασης της κάθε μελέτης με μεγαλύτερη ακρίβεια και να διερευνάται η μεταξύ των μελετών μεταβλητότητα. Η μετα-ανάλυση των k ανεξάρτητων μελετών μπορεί να πραγματοποιηθεί είτε χρησιμοποιώντας ατομικά δεδομένα είτε ενσωματώνοντας συνοπτικά δεδομένα, όπως οι λογάριθμοι των Odds Ratios ή τα Z-values από κάθε μελέτη. Γενικώς, οι μεθοδολογίες που χρησιμοποιούν ατομικά δεδομένα βασίζονται σε multilevel models και συνεπώς είναι υπολογιστικά δύσκολες στην υλοποίησή τους (Turner, Omar et al. 2000; Higgins, Whitehead et al. 2001).

Ωστόσο, προηγούμενες εργασίες έχουν δείξει ότι, τουλάχιστον σε συγκεκριμένες περιπτώσεις, οι μεθοδολογίες που χρησιμοποιούν συνοπτικά δεδομένα ταυτίζονται με εκείνες για ατομικά δεδομένα (Olkin and Sampson 1998; Mathew and Nordstrom 1999). Τα αποτελέσματα αυτά πρόσφατα γενικεύτηκαν και φάνηκε ότι η χρήση ατομικών δεδομένων για μετα-ανάλυση ευρυγονιδιωματικών μελετών δεν οδηγεί απαραίτητα σε μεγαλύτερη ακρίβεια (Lin and Zeng 2010; Lin and Zeng 2010). Εμπειρικά ευρήματα ενισχύουν τα αποτελέσματα αυτά (Janssens, Gonzalez-Zuloeta

Ladd et al. 2009) και συνεπώς, μεθοδολογίες ανάλυσης συνοπτικών δεδομένων χρησιμοποιούνται κυρίως σε μετα-ανάλυσεις ευρυγονιδιωματικών μελετών (de Bakker, Ferreira et al. 2008; Magi and Morris 2010; Willer, Li et al. 2010).

Κάποιοι από τους ελέγχους που συζητήθηκαν στις προηγούμενες παραγράφους, όπως ο έλεγχος CATT, μπορούν να διαμορφωθούν ως ένα score test:

$$Z_{score(i)} = \frac{U_{\beta_i}}{\sqrt{\text{var}_{H_0}(U_{\beta_i})}} = \frac{U_{\beta_i}}{\sqrt{\hat{I}(\beta_i)}} \quad (2.11)$$

όπου U_{β_i} αποτελεί το score statistic για κάθε i μελέτη, και $\hat{I}(\beta_i)$ η εκτίμηση του πίνακα πληροφορίας υπό την H_0 (δηλαδή όταν $\hat{\beta}_i=0$). Διάφοροι άλλοι έλεγχοι (όπως εκείνοι που βασίζονται στους λογάριθμους των Odds Ratios) μπορούν να διαμορφωθούν ως ένα Wald statistic:

$$Z_{Wald(i)} = \frac{\hat{\beta}_i}{\sqrt{\text{var}(\hat{\beta}_i)}} = \frac{\hat{\beta}_i}{\sqrt{1/\hat{I}(\beta_i)}} = \hat{\beta}_i \sqrt{\hat{I}(\beta_i)} \quad (2.12)$$

Στην εξίσωση 2.12 το $\hat{\beta}_i$ αποτελεί την εκτίμηση μέγιστης πιθανοφάνειας και $\text{var}(\hat{\beta}_i)$ η διασπορά αυτού. Όταν και οι δυο έλεγχοι εξετάζουν την ίδια μηδενική υπόθεση και επιπλέον όταν το $\hat{\beta}_i$ είναι μικρό κατά απόλυτη τιμή, τα score και Wald statistics είναι ασυμπτωτικά ισοδύναμα υπό την H_0 ($\hat{\beta}_i=0$) (Agresti 2002).

Μολονότι μέθοδοι για το συνδυασμό δεδομένων ανεξάρτητων μελετών οι οποίες χρησιμοποιούν το Wald statistic, έχουν περιγραφεί στη βιβλιογραφία εδώ και αρκετά χρόνια (Petiti 1994; Normand 1999), η χρήση των βέλτιστων βαρών κάθε μελέτης παρουσιάστηκε πρόσφατα (Zhou, Shi et al. 2011). Ας θεωρήσουμε μια μετα-ανάλυση k μελετών χρησιμοποιώντας μοντέλο σταθερών επιδράσεων όπου η πραγματική επίδραση είναι κοινή ανάμεσα στις k μελέτες (δηλαδή $\beta_i \sim N(\beta, s_i^2)$). Έτσι, το συνδυασμένο Z υπολογίζεται ως ένας σταθμισμένος μέσος όρος:

$$\bar{Z} = \frac{\sum_{i=1}^k \sqrt{w_i} Z_i}{\sqrt{\sum_{i=1}^k w_i}} \quad (2.13)$$

Ο Zhou και οι συνεργάτες έδειξαν ότι σε κάθε περίπτωση τα βέλτιστα βάρη υπολογίζονται ως $w_i = \hat{I}(\beta_i)$ (Zhou, Shi et al. 2011). Πρέπει να τονιστεί ότι όταν συνδυάζουμε Z-values, ανεξάρτητα από τη μέθοδο από την οποία προκύπτουν, το

$\hat{I}(\beta_i)$ δεν είναι συνήθως διαθέσιμο στις ανεξάρτητες μελέτες. Συνεπώς, μια προσέγγιση που χρησιμοποιείται συνήθως στη βιβλιογραφία είναι η στάθμιση των επιμέρους μελετών ανάλογα με το μέγεθος δείγματος ($w_i = n_i$) (de Bakker, Ferreira et al. 2008; Magi and Morris 2010; Willer, Li et al. 2010). Ωστόσο, ακόμα και αν η προσέγγιση αυτή είναι ικανοποιητική για αρκετές περιπτώσεις, ο Zhou και οι συνεργάτες έχουν δείξει ότι δεν είναι βέλτιστη. Χρησιμοποιώντας γενικευμένα γραμμικά μοντέλα, μελέτησαν μια ποικιλία εκβάσεων, και υποθέτοντας για παράδειγμα συνεπικρατές μοντέλο κληρονομικότητας με κωδικοποίηση $x_i = (0, 1, 2)$, έδειξαν ότι τα βέλτιστα βάρη δίνονται από την ακόλουθη σχέση:

$$w_i = \hat{I}(\beta_i) = 2 \frac{r_i s_i}{n_i} P_A P_B \quad (2.14)$$

Η σχέση αυτή φαίνεται να προσεγγίζει τη γνωστή διασπορά του ελέγχου CATT. Παρόμοιες εκφράσεις μπορούν να δοθούν επίσης και για τα υπόλοιπα μοντέλα κληρονομικότητας. Στηριζόμενοι σε αυτό το αποτέλεσμα, ο Zhou και οι συνεργάτες έδειξαν ότι μια καλή επιλογή για τα βάρη είναι $w_i = r_i s_i / n_i = 1 / (1/r_i + 1/s_i)$ δεδομένου ότι η ποσότητα αυτή είναι ανάλογη με τα πραγματικά βάρη που δίνονται στην εξίσωση 2.14.

Η εξίσωση 2.13 αποτελεί μια γενική σχέση και στην περίπτωση που συνδυάζονται τα score statistics, λαμβάνει τη μορφή:

$$\bar{Z}_{score} = \frac{\sum_{i=1}^k U_{\beta_i}}{\sqrt{\sum_{i=1}^k \hat{I}(\beta_i)}} \quad (2.15)$$

Στην περίπτωση που συνδυάζουμε τα Wald statistics, η εξίσωση 2.13 αντιστοιχεί στον υπολογισμό μιας συνολικής εκτίμησης για το β :

$$\hat{\beta} = \frac{\sum_{i=1}^k w_i \beta_i}{\sum_{i=1}^k w_i} \quad (2.16)$$

Η εκτίμηση αυτή αποτελεί τη γνωστή inverse-variance εκτίμηση που χρησιμοποιείται στη μετα-ανάλυση με βάρη $w_i = 1/s_i^2$ (Petiti 1994; Normand 1999).

Όλες οι μεθοδολογίες που έχουν περιγραφεί υποθέτουν ομοιογένεια των εκτιμήσεων μεταξύ των μελετών. Στην περίπτωση της ύπαρξης της μεταξύ των μελετών ετερογένειας, υποθέτουμε ότι το πραγματικό μέτρο επίδρασης διαφοροποιείται μεταξύ των μελετών $\beta_i \sim N(\beta, s_i^2 + \tau^2)$ και εκτιμάται μια επιπλέον

συνιστώσα, η μεταξύ των μελετών μεταβλητότητα (τ^2) (μοντέλα τυχαίων επιδράσεων). Η συνήθης χρησιμοποιούμενη μεθοδολογία για την εκτίμηση του τ^2 είναι η μη επαναληπτική μέθοδος των ροπών που έχει προταθεί από τους DerSimonian και Laird (DerSimonian and Laird 1986), αν και υπάρχουν πολλές εναλλακτικές οι οποίες περιλαμβάνουν επαναληπτικές διαδικασίες (Thompson and Sharp 1999). Στην περίπτωση όπου το $\tau^2=0$, τα μοντέλα σταθερών και τυχαίων επιδράσεων συγκλίνουν. Ακολούθως, τα βάρη μπορούν να υπολογιστούν ως:

$$w_i = \left(\tau^2 + \hat{I}(\beta_i)^{-1} \right)^{-1} \quad (2.17)$$

και να χρησιμοποιηθεί η εξίσωση 2.16 έτσι ώστε να υπολογιστεί μια συνολική εκτίμηση. Το μοντέλο των τυχαίων επιδράσεων για το συνδυασμό των score statistics (ή Z-values) δεν μπορούσε να υλοποιηθεί μέχρι ο Zhou και οι συνεργάτες αυτού να αποδείξουν μια χρήσιμη προσέγγιση. Έδειξαν ότι όταν το β_i είναι μικρό κατά απόλυτη τιμή, ο μέσος του U_{β_i} μπορεί να προσεγγιστεί από το $\beta_i \hat{I}(\beta_i)$. Το αποτέλεσμα αυτό υποδεικνύει ότι το μέτρο επίδρασης μπορεί να εκτιμηθεί από την ακόλουθη σχέση:

$$\hat{\beta}_i = U_{\beta_i} / \hat{I}(\beta_i) \quad (2.18)$$

με διασπορά ίση με:

$$\text{var}(\hat{\beta}_i) = 1 / \hat{I}(\beta_i) \quad (2.19)$$

Το αποτέλεσμα αυτό υπονοεί μια απλή μέθοδο για την πραγματοποίηση της μετα-ανάλυσης χρησιμοποιώντας τις robust προσεγγίσεις MAX, MERT και MIN2. Συνεπώς, υπολογίζοντας τα κατάλληλα βάρη ως $w_i = (1/r_i + 1/s_i)^{-1}$, ακολουθεί η μετατροπή του score test σε Wald test, λαμβάνοντας υπόψη την κατεύθυνση της συσχέτισης. Κατόπιν μπορεί να πραγματοποιηθεί μια κλασσική μετα-ανάλυση τυχαίων επιδράσεων. Πρέπει να λάβει κανείς υπόψη του ότι τα πραγματικά μέτρα επίδρασης που υπολογίζονται ως $\hat{\beta}_i = U_{\beta_i} / \hat{I}(\beta_i)$ καθώς και από τη μετα-ανάλυση δεν είναι ορθά (Zhou, Shi et al. 2011). Συνεπώς, η συμπερασματολογία στηρίζεται στο συνολικό Z-score, το οποίο είναι ορθό σε κάθε περίπτωση. Στο σημείο αυτό, αξίζει να τονιστεί ότι οι μεθοδολογίες που περιγράφηκαν ως τώρα μπορούν να εφαρμοστούν και στην περίπτωση των ποσοτικών χαρακτηριστικών (Thakkinstian, McElduff et al. 2005; Bagos 2008). Επιπλέον, η προσέγγιση του Zhou και

συνεργατών παρουσιάζει ως ειδική περίπτωση των ανάλυση ποσοτικών χαρακτηριστικών (Zhou, Shi et al. 2011).

2.7. Υλοποίηση robust μεθόδων ανάλυσης και μετα-ανάλυσης μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών

2.7.1. Υλοποίηση

Οι μεθοδολογίες που περιγράφηκαν για την ανάλυση και μετα-ανάλυση δεδομένων μελετών γενετικής συσχέτισης και ευρυγονιδιωματικών μελετών υλοποιήθηκαν στο στατιστικό πρόγραμμα Stata (StataCorp) ενώ κατασκευάστηκε απευθείας εντολή (GWAR). Για τους ελέγχους MAX και MIN2, υλοποιήσαμε τις μεθοδολογίες που περιγράφονται στις εξισώσεις 2.6 και 2.10, αντίστοιχα, που απαιτούν τον υπολογισμό ολοκληρωμάτων. Για το σκοπό αυτό, πραγματοποιούμε αριθμητική ολοκλήρωση με Gauss-Legendre quadrature με τη χρήση της εντολής *integrate* που οδηγεί σε μείωση του υπολογιστικού χρόνου για την ολοκλήρωση της ανάλυσης. Για το κομμάτι της μετα-ανάλυσης, χρησιμοποιούμε την εντολή *metan* στο Stata. Λόγω της χρήσης του *integrate*, η GWAR απαιτεί την έκδοση Stata v12 ή μεταγενέστερης. Η εντολή GWAR υλοποιήθηκε ως ado file καθώς και ως διαδικτυακή εφαρμογή που είναι διαθέσιμη στην ιστοσελίδα www.compgen.org/tools/GWAR.

2.7.2. Χρήση

Υπάρχουν δυο εντολές: *gwar* και *min2gwar*. Η σύνταξη για την *gwar* είναι: *gwar varlist, Method(string) [SNP(varlist min=1 max=1) EFfect(string)]*. Ως *varlist* περιλαμβάνονται οι μεταβλητές που αντιστοιχούν στους γονοτύπους για ασθενείς και υγιείς. Οι τρεις πρώτες μεταβλητές αντιστοιχούν στους γονότυπους για τους υγιείς (δηλαδή aa0, ab0, bb0), ενώ οι υπόλοιπες για τους ασθενείς (δηλαδή aa1, ab1, bb1). Ως αλληλόμορφο κινδύνου θεωρείται το b. Στην απλή ανάλυση δεδομένων, ο χρήστης πρέπει να ορίσει μόνο τους γονότυπους και τη μέθοδο της ανάλυσης (max, min2, mert καθώς επίσης και τις τρεις παραλλαγές του CATT). Για την περίπτωση της μετα-ανάλυσης, ο χρήστης πρέπει να καθορίσει επίσης τη μεταβλητή που δηλώνει τον πολυμορφισμό και την απαιτούμενη μέθοδο μετα-ανάλυσης (σταθερών ή τυχαίων επιδράσεων).

Η εντολή `min2gwar` αποτελεί την εκδοχή όπου χρησιμοποιούνται απευθείας τα *p-values* των ελέγχων του Pearson και του CATT(1/2). Στην περίπτωση της απλής ανάλυσης, η σύνταξή της είναι απλή (`min2gwar p1 p2`). Στην μετα-ανάλυση, ο χρήστης πρέπει να ορίσει τρεις επιπλέον μεταβλητές, το συνολικό αριθμό ασθενών και υγιών και μια μεταβλητή (-1, +1) για την κατεύθυνση της συσχέτισης. Κατά παρόμοιο τρόπο, η μετα-ανάλυση πραγματοποιείται ομαδοποιώντας τις παρατηρήσεις του ίδιου πολυμορφισμού, και συνεπώς είναι ευθύνη του χρήστη να είναι τα δεδομένα στην κατάλληλη μορφή όταν καλείται το πρόγραμμα.

Στη διαδικτυακή εφαρμογή, παρέχεται η δυνατότητα να δίνεται ολόκληρο αρχείο, ή μπορεί να χρησιμοποιηθεί σαν ένα απλό εργαλείο όπου ο χρήστης συμπληρώνει τους γονότυπους και παίρνει τα αποτελέσματα της ανάλυσης. Για μεγάλο μέγεθος αρχείων επιτρέπουμε το ανέβασμα συμπιεσμένων αρχείων στα περισσότερα γνωστά formats.

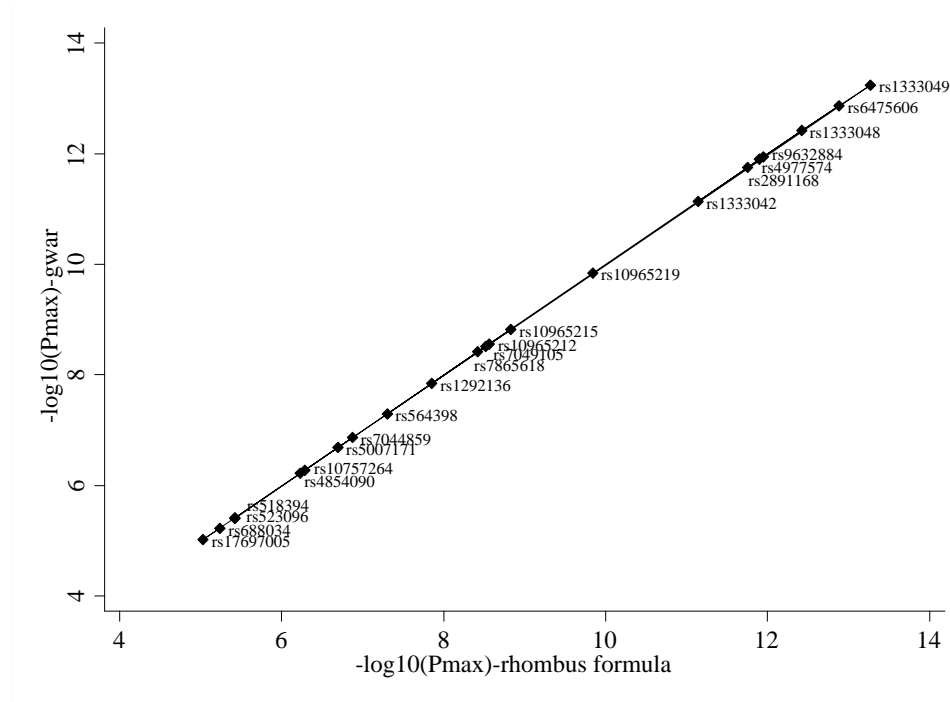
2.7.3. Επίδοση

Η εντολή `GWAR` είναι γρήγορη, αξιόπιστη και κατέχει χαρακτηριστικά που δεν υπάρχουν ως τώρα σε άλλα παρόμοια πακέτα. Μια απλή GWAS με 100,000 SNPs μπορεί να αναλυθεί σε έναν κλασσικό υπολογιστή εντός λίγων ωρών ακόμα και όταν υλοποιούνται οι πιο χρονοβόροι υπολογιστικά έλεγχοι `MAX` ή `MIN2`. Η ακρίβεια μπορεί να φανεί μέσω σύγκρισης γνωστών αποτελεσμάτων από τη βιβλιογραφία. Αξίζει να αναφερθεί ότι λόγω της προσεγγιστικής σχέσης που χρησιμοποιείται, με την εντολή `GWAR` μπορούν να υπολογιστούν *p-values* έως και 10^{-18} για τον έλεγχο `MAX` και 10^{-52} για τον `MIN2`, αν και ξεπερνούν τα εύρος μιας οποιασδήποτε λογικής μεθόδου προσομοίωσης.

2.7.4. Εφαρμογή σε πραγματικά δεδομένα

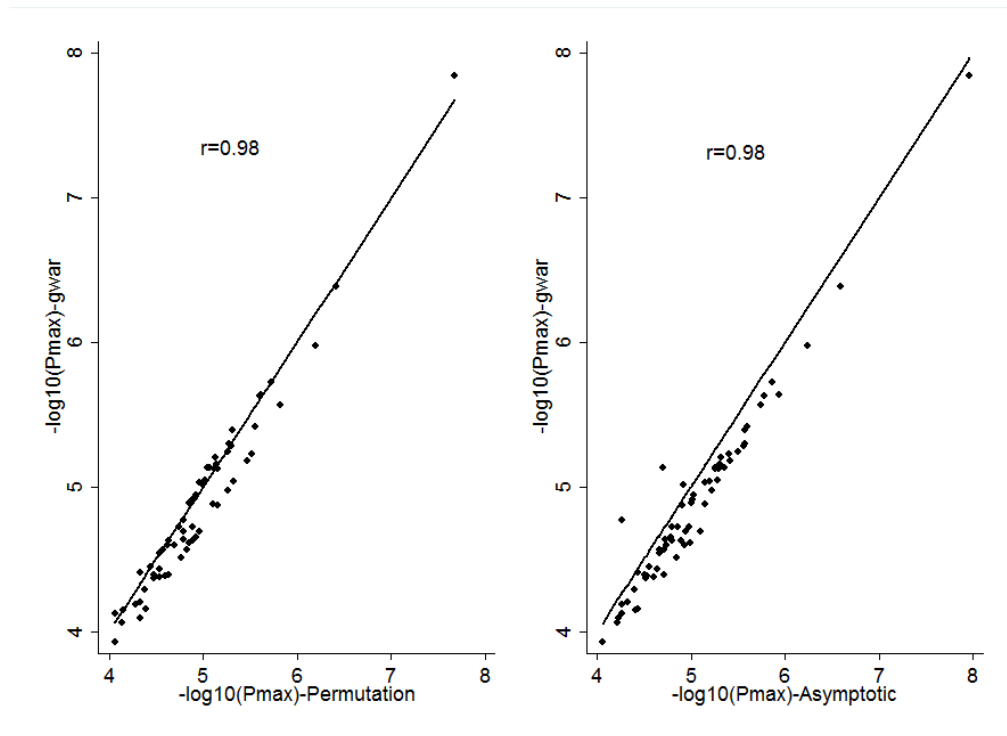
Αναλύθηκαν ξανά τα δεδομένα 22 SNPs με *p-values* (χρησιμοποιώντας την rhombus formula για τον προσεγγιστικό υπολογισμό των *p-values* του ελέγχου `MAX` (Li, Zheng et al. 2008)) κάτω του 10^{-5} μιας ευρυγονιδιωματικής μελέτης για τη στεφανιαία νόσο (Coronary Artery Disease - CAD) (WTCCC 2007). Τα αποτελέσματα του ελέγχου `MAX` με την εντολή `GWAR` είναι σχεδόν ταυτόσημα με εκείνα που προκύπτουν από τη χρήση της rhombus formula (Εικόνα 2.1). Πρέπει επίσης να τονιστεί ότι και οι δυο προσεγγίσεις μπορούν να χρησιμοποιηθούν για την

εκτίμηση πολύ χαμηλών p-values (κάτω του 10^{-7}) όπου οι resampling procedures δεν επαρκούν.



Εικόνα 2.1. Η σύγκριση των p-values που προέρχονται από τον έλεγχο MAX με τη χρήση της εντολής GVAR έναντι αυτών που προκύπτουν από την rhombus formula για 22 επιλεγμένους SNPs μιας ευρυγονιδιοματικής μελέτης για τη στεφανιαία νόσο.

Επιπλέον, αναλύθηκαν ξανά τα δεδομένα μιας ευρυγονιδιοματικής μελέτης που στόχευε να ανιχνεύσει γονιδιακούς πολυμορφισμούς που συσχετίζονται με την εμφάνιση διαβήτη τύπου 2 (Sladek, Rocheleau et al. 2007). Υπολογίστηκαν τα p-values από τον έλεγχο MAX μέσω της εντολής GVAR και συγκρίθηκαν με εκείνα που προέρχονται από μια permutation approach καθώς και χρησιμοποιώντας μια προσεγγιστική ασυμπτωτική κατανομή του ελέγχου MAX (Gonzalez, Carrasco et al. 2008). Παρατηρούμε μια σχεδόν τέλεια συμφωνία μεταξύ των αποτελεσμάτων με σχεδόν πάντα ελαφρώς μεγαλύτερα p-values χρησιμοποιώντας την εντολή GVAR σε σχέση με την ασυμπτωτική προσέγγιση (Εικόνα 2.2).



Εικόνα 2.2. Η σύγκριση των p-values που προέρχονται από τη χρήση του ελέγχου MAX με την εντολή GWAR έναντι αυτών που προκύπτουν από την permutation approach και την ασυμπτωτική κατανομή του ελέγχου MAX για 67 SNPs από την ευρυγονιδιοματική μελέτη για το διαβήτη τύπου 2 (οι πολυμορφισμοί rs7903146, rs12255372, rs10885409 και rs7904519 αναλύθηκαν επίσης, με εκτιμώμενα p-values < 3.30E-10 για την permutation approach, 6.36E-19, 9.33E-15, 2.47E-11 και 5.60E-11 για την ασυμπτωτική προσέγγιση και 3.77E-15, 1.89E-14, 4.21E-11 και 6.44E-11 για την GWAR αντίστοιχα.

Τέλος, αναλύθηκαν 6 SNPs μιας ευρυγονιδιοματικής μελέτης για την υπέρταση (WTCCC 2007). Υπολογίστηκαν τα p-values του ελέγχου MAX και MIN2 με την εντολή GWAR και συγκρίθηκαν με εκείνα που αναφέρονται από τους (Joo, Kwak et al. 2009). Τα αποτελέσματα είναι ταυτόσημα για τον έλεγχο MIN2 με κάποιες πολύ μικρές αποκλίσεις για τον έλεγχο MAX (Πίνακας 2.2).

Πίνακας 2.2. Η εφαρμογή των ελέγχων MIN2 και MAX σε επιλεγμένους SNPs που παρουσιάζουν συσχέτιση με την υπέρταση στην μελέτη του WTCCC (2007).

	MIN2 reported in (Joo, Kwak et al. 2009)	MIN2 in gwar	MAX reported in (Joo, Kwak et al. 2009)	MAX in gwar
rs2820037	1.31E-06	1.31E-06	3.16E-06	3.23E-06
rs6997709	1.34E-05	1.34E-05	2.06E-05	2.07E-05
rs7961152	1.25E-05	1.25E-05	1.80E-05	2.01E-05
rs11110912	1.56E-05	1.56E-05	8.16E-06	8.15E-06
rs1937506	1.56E-05	1.56E-05	2.46E-05	2.43E-05
rs2398162	9.64E-06	9.64E-06	2.50E-06	2.42E-06

2.8. Συζήτηση

Η εντολή GVAR, αποτελεί ένα γρήγορο και αξιόπιστο εργαλείο για την ανάλυση και μετα-ανάλυση μελετών γενετικής συσχέτισης επιλεγμένων γονιδίων και ευρυγονιδιωματικών μελετών. Υλοποιήσαμε τόσο τους κλασσικούς ελέγχους CATTs καθώς επίσης και τις ευρέως χρησιμοποιούμενες robust προσεγγίσεις, MERT, MAX και MIN2. Για τους ελέγχους MAX και MIN2, που παρουσιάζουν αυξημένη ισχύ συγκρινόμενοι με άλλες robust προσεγγίσεις, υλοποιήσαμε ακριβείς και γρήγορες ρουτίνες που στηρίζονται σε αριθμητική ολοκλήρωση για τον υπολογισμό των p-values με μεγάλη ακρίβεια. Η εντολή GVAR υλοποιήθηκε σαν ado file για Stata 12 ή μεγαλύτερη έκδοση αυτής, καθώς επίσης και σαν διαδικτυακή εφαρμογή που βρίσκεται στην ιστοσελίδα www.compgen.org/tools/GVAR.

Η εντολή GVAR προσφέρει λειτουργικότητες που δεν υπάρχουν σε άλλα παρόμοια εργαλεία. Αρχικά, υλοποιείται ο έλεγχος MIN2 που δεν υπάρχει έως τώρα σε οποιοδήποτε άλλο στατιστικό πακέτο. Επιπλέον, η υλοποίηση του ελέγχου MAX είναι από τις πιο γρήγορες που υπάρχουν διαθέσιμες αυτή τη στιγμή. Η εντολή GVAR πραγματοποιεί επίσης μετα-ανάλυση χρησιμοποιώντας όλους τους προαναφερόμενους στατιστικούς ελέγχους. Παρόμοιες μεθοδολογίες δεν υπάρχουν σε άλλα πακέτα μετα-ανάλυσης ευρυγονιδιωματικών μελετών (METAL (Willer, Li et al. 2010), GWAMA (Magi and Morris 2010), PLINK (Purcell, Neale et al. 2007) και MetABEL (Aulchenko, Ripke et al. 2007)). Επιπλέον, ακόμα και στην περίπτωση

μιας κλασσικής μετα-ανάλυσης όπου γίνεται εξαρχής υπόθεση του γενετικού μοντέλου, η GWA είναι μοναδική καθώς παρέχει την επιλογή για μετα-ανάλυση σταθερών και τυχαίων επιδράσεων. Για λόγους σύγκρισης, πρέπει να αναφέρουμε ότι μόνο το GWAMA παρέχει τη δυνατότητα για μοντέλα τυχαίων επιδράσεων, ενώ τα PLINK, MetABEL και METAL επιτρέπουν μόνο σταθερές επιδράσεις.

Η διαδικτυακή εφαρμογή που υλοποιήσαμε είναι επίσης μοναδική στο είδος της και επιτρέπει στους ερευνητές να ανεβάζουν αρχεία και να παίρνουν τα αποτελέσματα χωρίς να απαιτείται να χρησιμοποιήσουν τοπικά το λογισμικό. Το εργαλείο απαιτεί μόνο συνοπτικά δεδομένα για τους γονότυπους, και έτσι εγγυάται η ασφάλεια των δεδομένων (σε κάθε περίπτωση μπορεί ο χρήστης να μη δώσει την πραγματική λίστα με τα SNP ids). Ωστόσο, υπάρχουν θεωρητικές αποδείξεις ότι τα συνοπτικά στατιστικά παρέχουν την ίδια ακρίβεια στις εκτιμήσεις (Lin and Zeng 2010; Lin and Zeng 2010) και η μετα-ανάλυση των ευρυγονιδιωματικών μελετών πραγματοποιείται συνήθως χρησιμοποιώντας συνοπτικές εκτιμήσεις από τις μελέτες που συμπεριλαμβάνονται (de Bakker, Ferreira et al. 2008; Magi and Morris 2010; Willer, Li et al. 2010; Begum, Ghosh et al. 2012). Ωστόσο, ο έλεγχος MIN2 μπορεί να πραγματοποιηθεί χωρίς καν να παρέχει ο χρήστης τις συχνότητες των γονοτύπων αλλά μόνο τα p-values από τον έλεγχο του Pearson και τον κλασσικό έλεγχο CATT (για την περίπτωση της απλής ανάλυσης), ή επιπρόσθετα το συνολικό αριθμό των ασθενών και υγιών καθώς και την κατεύθυνση της συσχέτισης (για την περίπτωση της μετα-ανάλυσης). Με την επιλογή αυτή, ο αναλυτής πρέπει να πραγματοποιήσει κάποιους επιπλέον υπολογισμούς, αλλά αποφεύγονται πιθανά προβλήματα εμπιστευτικότητας, κυρίως στην περίπτωση μεγάλων consortia μετα-αναλύσεων ευρυγονιδιωματικών μελετών.

Με την εντολή GWA επιτρέπεται μόνο η ανάλυση μελετών ασθενών-μαρτύρων. Πιθανές μελλοντικές βελτιστοποιήσεις της GWA περιλαμβάνουν την ανάλυση ποσοτικών χαρακτηριστικών αν και μόνο οι έλεγχοι MERT και MAX μπορούν πιθανά να υλοποιηθούν. Επιπλέον, μια άλλη πιθανή επέκταση αφορά την υλοποίηση σαν standalone version (για παράδειγμα σε C, Java and so on), έτσι ώστε να μεγιστοποιηθεί η ταχύτητα και ελαχιστοποιηθεί η εξάρτηση από εξωτερικό λογισμικό.

Κεφάλαιο 3: Ανάπτυξη μεθοδολογίας μετα-ανάλυσης μελετών γενετικής συσχέτισης σε οικογένειες και ασθενών-μαρτύρων που χρησιμοποιούν τους ίδιους ασθενείς και εφαρμογές

Μέρος Α: Μεθοδολογία μετα-ανάλυσης μελετών γενετικής συσχέτισης σε οικογένειες και ασθενών-μαρτύρων που χρησιμοποιούν τους ίδιους ασθενείς

3.1. Εισαγωγή

Όπως έχει αναφερθεί στην Παράγραφο 1.2, οι μελέτες γενετικής συσχέτισης μπορούν να διακριθούν σε μελέτες σε οικογένειες και σε πληθυσμούς (family-based και population-based). Έχουν καταβληθεί διάφορες προσπάθειες για την εκτίμηση ενός κοινού μέτρου επίδρασης ώστε να επιτρέπεται η απευθείας σύγκριση των αποτελεσμάτων των προσεγγίσεων αυτών (Mitchell 2000). Επιπλέον, έχει περιγραφεί ο συνδυασμός των εκτιμήσεων μελετών σε οικογένειες και σε πληθυσμούς χρησιμοποιώντας συνοπτικά δεδομένα στο πλαίσιο της μετα-ανάλυσης (Kazeem and Farrall 2005). Εμπειρικές αξιολογήσεις έδειξαν ότι οι δυο σχεδιασμοί είναι συγκρίσιμοι (Evangelou, Trikalinos et al. 2006) και έτσι, η ενοποίηση των δεδομένων στη μετα-ανάλυση θα μπορούσε να αυξήσει την ισχύ για να ανιχνευτούν ασθενείς επιδράσεις γονιδίων.

Σε πολλές περιπτώσεις, οι ερευνητές προσπαθώντας να εξαλείψουν τους συγχυτικούς παράγοντες και ταυτόχρονα να αυξήσουν την ισχύ, πραγματοποιούν ταυτόχρονα στον ίδιο πληθυσμό μελέτες ασθενών-μαρτύρων σε οικογένειες και σε πληθυσμούς, χρησιμοποιώντας έτσι τους ίδιους ή μεγάλο αριθμό επικαλυπτόμενων ασθενών (Ackerman, Usen et al. 2005; Pomp, Van Stralen et al. 2010). Συνεπώς, οι διαφορετικές εκτιμήσεις κάθε μελέτης είναι εξαρτημένες και ο συνδυασμός τους χρειάζεται ειδική μεθοδολογία. Τα τελευταία χρόνια, έχουν προταθεί διάφορες μέθοδοι σε αυτήν την κατεύθυνση, επιτρέποντας μια συνδυασμένη ανάλυση (Nagelkerke, Hoebee et al. 2004; Epstein, Veal et al. 2005; Joo, Tian et al. 2007; Chen and Lin 2008; Pfeiffer, Pee et al. 2008; Glaser and Holmans 2009; Guo, Lunetta et al. 2009; Hsu, Starr et al. 2009; Gebregziabher, Guimaraes et al. 2010). Ωστόσο, αυτές οι μέθοδοι εξαρτώνται από τη διαθεσιμότητα ατομικών δεδομένων και έτσι δεν

επαρκούν για τις μετα-αναλύσεις δημοσιευμένων μελετών όπου ο ερευνητής έχει πρόσβαση μόνο σε συνοπτικά δεδομένα (συχνότητες γονοτύπων και αλληλομόρφων). Οι διάφορες προσεγγίσεις έχουν παρουσιαστεί και συγκριθεί, τονίζοντας τα πλεονεκτήματα και τα μειονεκτήματα που παρουσιάζουν (Infante-Rivard, Mirea et al. 2009). Αξίζει να τονιστεί ότι η διάκριση μεταξύ μελετών σε πληθυσμούς και σε οικογένειες, μοιάζει με την διάκριση στο χώρο της επιδημιολογίας μεταξύ εξομοιωμένων ή μη μελετών ασθενών-μαρτύρων, των οποίων τα πλεονεκτήματα και μειονεκτήματα έχουν περιγραφεί (Thompson, Kelsey et al. 1982; Feinstein 1987; Hamajima, Hirose et al. 1994). Μολονότι, μεθοδολογίες για την ενοποίηση των δεδομένων από τέτοιες μελέτες έχουν προταθεί εδώ και αρκετά χρόνια, χρησιμοποιούν ατομικά δεδομένα και απαιτούν την ανεξαρτησία των πληθυσμών (Duffy, Rohan et al. 1989; Moreno, Martin et al. 1996).

Πίνακας 3.1. Ο 2x2 πίνακας συνάφειας που περιγράφει τα δεδομένα μιας μελέτης ασθενών-μαρτύρων όπου το αλληλόμορφο B θεωρείται το αλληλόμορφο κινδύνου. Ο συνολικός αριθμός των αλληλομόρφων B και A συγκρίνεται μεταξύ των ασθενών και υγιών. Για συντομία ορίζουμε ως $n_{01}=2nBB_0+nAB_0$, $n_{00}=2nAA_0+nAB_0$, $n_{10}=2nAA_1+nAB_1$ και $n_{11}=2nBB_1+nAB_1$. Ο συνολικός αριθμός των αλληλομόρφων των ασθενών είναι n_1 και των υγιών n_0 .

		Αλληλόμορφο		Σύνολο
		B	A	
Κατάσταση	Ασθενείς	n_{11}	n_{10}	n_1
	Υγιείς	n_{01}	n_{00}	n_0
	Σύνολο			$n_0 + n_1$

Έχει παρουσιαστεί μια γενική μέθοδος, που χρησιμοποιεί τον sandwich variance estimator ή την μέθοδο bootstrap (le Cessie, Nagelkerke et al. 2008), και μπορεί να εφαρμοστεί και στις δυο περιπτώσεις (μελέτες γενετικής συσχέτισης και κλασσικές επιδημιολογικές αναλύσεις). Η μέθοδος αυτή εφαρμόστηκε επίσης για το συνδυασμό των αποτελεσμάτων από μελέτες σε οικογένειες και σε πληθυσμούς (Janssen, Bont et al. 2007). Ωστόσο, απαιτείται η γνώση ατομικών δεδομένων κάτι που σπανίως είναι διαθέσιμο στη μετα-ανάλυση.

Στις παραγράφους που ακολουθούν, θα περιγραφεί μια απλή μέθοδος ενοποίησης συνοπτικών δεδομένων μελετών γενετικής συσχέτισης, οι οποίες πραγματοποιούνται ταυτόχρονα σε οικογένειες και σε πληθυσμούς και χρησιμοποιούν επικαλυπτόμενους ασθενείς. Το σημείο κλειδί της μεθόδου είναι ο υπολογισμός της συνδιακύμανσης των εκτιμήσεων που προκύπτουν από τους δυο

σχεδιασμούς χρησιμοποιώντας μόνο τα συνοπτικά δεδομένα που δίνονται στη βιβλιογραφία. Ακολούθως, οι μελέτες μπορούν να συνδυαστούν χρησιμοποιώντας είτε μια διαδικασία δυο βημάτων που στηρίζεται σε κλασσικές μεθόδους μονομεταβλητής μετα-ανάλυσης είτε με ένα μόνο βήμα με ιεραρχικά μοντέλα. Σε κάθε περίπτωση, η μετα-ανάλυση μπορεί να πραγματοποιηθεί με οποιοδήποτε στατιστικό πακέτο με συνεπακόλουθη αύξηση της ισχύος λόγω του αυξανόμενου μεγέθους δείγματος.

Η μέθοδος εφαρμόστηκε σε τρία σύνολα δεδομένων. Αρχικά, αναλύθηκαν τα δεδομένα 356 πολυμορφισμών (SNPs) που παρουσιάστηκαν από τον Janssen και συνεργάτες (2007) έτσι ώστε να συγκρίνουμε την ακρίβεια στην εκτίμηση της συνδιακύμανσης. Ακολούθως, αναλύσαμε μια ήδη δημοσιευμένη μετα-ανάλυση για τη διερεύνηση της συσχέτισης του 10-Repeat Allele ενός πολυμορφισμού VNTR στην 3' αμετάφραστη περιοχή του γονιδίου του μεταφορέα της ντοπαμίνης (Dopamine Transporter Gene) με τη Διαταραχή Ελλειμματικής Προσοχής - Υπερκινητικότητας (ΔΕΠ-Υ) (Attention Deficit Hyperactivity Disorder-ADHD) (Yang, Chan et al. 2007), καθώς επίσης και δεδομένα μιας ευρυγονιδιοματικής μελέτης για την ανίχνευση γονιδιακών πολυμορφισμών που συσχετίζονται με τη Σκλήρυνση κατά Πλάκας (Hafler, Compston et al. 2007).

3.2. Μεθοδολογία

3.2.1. Μελέτες ασθενών-μαρτύρων σε πληθυσμούς

Έστω ένας γενετικός τόπος με δυο αλληλόμορφα (A, B) με το B να αποτελεί το αλληλόμορφο κινδύνου. Σε έναν παραδοσιακό σχεδιασμό μελετών ασθενών-μαρτύρων σε πληθυσμούς, πραγματοποιείται ανεξάρτητη δειγματοληψία ασθενών ($j=1$) και υγιών ($j=0$) και οι γονότυποι αυτών (AA_j, AB_j, BB_j) συγκρίνονται σε έναν 3×2 πίνακα συνάφειας (Sasieni 1997). Όπως έχει ήδη αναφερθεί, η ταξινόμηση των ατόμων σε έναν 2×2 πίνακα συνάφειας, οδηγεί στην υπόθεση διαφορετικών μοντέλων κληρονομικότητας (επικρατές, υπολειπόμενο, συνεπικρατές) (Sasieni 1997); ωστόσο, για λόγους που θα γίνουν κατανοητοί στη συνέχεια, θα εστιάσουμε την προσοχή μας στο συνεπικρατές μοντέλο κληρονομικότητας. Σύμφωνα με το μοντέλο αυτό, ο συνολικός αριθμός των αλληλομόρφων B ($2nBB_j + nAB_j$) και A ($2nAA_j + nAB_j$) συγκρίνεται μεταξύ ασθενών και υγιών. Τα δεδομένα απεικονίζονται με τη μορφή του

Πίνακα 3.1, όπου για συντομία ορίζουμε ως $n_{01}=2nBB_0+nAB_0$, $n_{00}=2nAA_0+nAB_0$, $n_{10}=2nAA_1+nAB_1$ και $n_{11}=2nBB_1+nAB_1$. Σε μια μετα-ανάλυση k μελετών ασθενών-μαρτύρων ο λογάριθμος του Odds Ratio για την i μελέτη ($i=1,2,\dots,k$), ορίζεται ως y_i^{CC} και δίνεται από την ακόλουθη σχέση:

$$y_i^{CC} = \log \left(\frac{n_{11}n_{00}}{n_{01}n_{10}} \right) \quad (3.1)$$

Ο λογάριθμος του Odds Ratio ακολουθεί προσεγγιστικά την κανονική κατανομή (Woolf 1955) με διασπορά ίση με:

$$\text{var}(y_i^{CC}) = \frac{1}{n_{11}} + \frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} \quad (3.2)$$

Η διασπορά μπορεί να υπολογιστεί εύκολα θεωρώντας ότι οι ποσότητες των κελιών του 2x2 πίνακα συνάφειας είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson (U), με $E[U]=\text{var}[U]=u$. Συνεπώς, χρησιμοποιώντας τη μέθοδο Δέλτα (Agresti 2002) ισχύει ότι:

$$\text{var}[\log(U)] \approx \text{var}(U) \left(\frac{\partial \log(E[U])}{\partial (E[U])} \right)^2 = u \left(\frac{1}{u} \frac{\partial U}{\partial U} \right)^2 = \frac{1}{u} \quad (3.3)$$

3.2.2. Μελέτες ασθενών-μαρτύρων σε οικογένειες

Αν και υπάρχουν πολλοί σχεδιασμοί μελετών που πραγματοποιούνται σε οικογένειες (Πίνακας 1.1), η πιο συχνά χρησιμοποιούμενη προσέγγιση στις μελέτες γενετικής συσχέτισης είναι η family-trio, στην οποία περιλαμβάνεται ο ασθενής απόγονος και οι υγιείς γονείς σαν μάρτυρες. Σε αυτό το πλαίσιο, οι πιο συχνά χρησιμοποιούμενοι έλεγχοι είναι οι transmission disequilibrium test (Spielman, McGinnis et al. 1993; Spielman and Ewens 1996) και haplotype-based haplotype relative risk (Terwilliger and Ott 1992).

3.2.2.1. Ο έλεγχος transmission disequilibrium test

Στον έλεγχο transmission disequilibrium test (TDT) υποθέτουμε ότι κάτω από τη μηδενική υπόθεση της απουσίας συσχέτισης, οι δυο γονότυποι των γονέων είναι ανεξάρτητοι. Σε αυτή την περίπτωση, κάθε ένας από τους n_1 γονείς των ασθενών μπορεί να ταξινομηθεί ανάλογα με το αλληλόμορφο έχει μεταβιβαστεί ή όχι. Τα δεδομένα απεικονίζονται με τη μορφή του Πίνακα 3.2 όπου είναι προφανές ότι μόνο

τα μη συμφωνούντα ζεύγη συνεισφέρουν στην ανάλυση. Έτσι, ο έλεγχος TDT είναι άλλη μια αναδιατύπωση του ελέγχου McNemar για ζευγαρωτές παρατηρήσεις. Το σκορ (A) υπολογίζεται χρησιμοποιώντας τα δεδομένα του Πίνακα 3.2 και δίνεται ως ακολούθως:

$$A = (b - c)/2 \quad (3.4)$$

με διασπορά (V) ίση με:

$$V = (b + c)/4 \quad (3.5)$$

Έτσι, η στατιστική σημαντικότητα μπορεί να ελεγχθεί εκτιμώντας το $z = A/\sqrt{V} \sim N(0,1)$.

Πίνακας 3.2. Παρουσίαση των δεδομένων μιας μελέτης σε οικογένειες χρησιμοποιώντας τον έλεγχο Transmission Disequilibrium Test (TDT). Τα αλληλόμορφα που μεταβιβάζονται συγκρίνονται με εκείνα που δεν μεταβιβάζονται και το Odds Ratio δίνεται χρησιμοποιώντας το λόγο των μη συμφωνούντων ζευγών (b/c). Για σύγκριση με τον Πίνακα 3.1 ορίζουμε ως $a+b=w=n_{11}$ και $c+d=x=n_{10}$.

	<i>Αλληλόμορφο που μεταβιβαστεί</i>	<i>που δεν έχει</i>	
<i>Αλληλόμορφο που έχει B</i>	<i>B</i>	<i>A</i>	Σύνολο
<i>μεταβιβαστεί A</i>	<i>a</i>	<i>b</i>	<i>w</i>
	<i>c</i>	<i>d</i>	<i>x</i>
Σύνολο	<i>y</i>	<i>z</i>	<i>n₁</i>

Για τη μετα-ανάλυση ωστόσο, είναι πιο εύχρηστο να υπολογίζεται το κατάλληλο Odds Ratio που προέρχεται από τον Πίνακα 3.2, ο λογάριθμος του οποίου δίνεται από την ακόλουθη σχέση:

$$y_i^{TDT} = \log\left(\frac{b}{c}\right) \quad (3.6)$$

Κατά αντιστοιχία με την εξίσωση 3.2 μια ασυμπτωτική εκτίμηση της διασποράς δίνεται από την ακόλουθη σχέση:

$$\text{var}\left(y_i^{TDT}\right) = \frac{1}{b} + \frac{1}{c} \quad (3.7)$$

3.2.2.2. Ο έλεγχος haplotype-based haplotype relative risk

Ο έλεγχος haplotype-based haplotype relative risk (HHRR) αποτελεί μια προσπάθεια αύξησης της ισχύος (με άλλα λόγια μείωση της διασποράς). Εδώ χρησιμοποιείται η unmatched εκδοχή του Πίνακα 3.2, μιας και, υπό τη μηδενική υπόθεση, τα δυο αλληλόμορφα κάθε γονέα είναι ανεξάρτητα. Η μετάβαση στην unmatched ανάλυση παρουσιάζεται στον Πίνακα 3.3.

Ο έλεγχος haplotype-based haplotype relative risk παρουσιάστηκε αρχικά σαν ένα χ^2 στατιστικό αλλά κατά παρόμοιο τρόπο με τον έλεγχο TDT κατασκευάστηκε το αντίστοιχο Odds Ratio χρησιμοποιώντας τα δεδομένα του Πίνακα 3.3:

$$y_i^{HHRR} = \log\left(\frac{wz}{xy}\right) \quad (3.8)$$

Παρόμοια, μια ασυμπτωτική εκτίμηση της διασποράς δίνεται από την ακόλουθη σχέση:

$$\text{var}(y_i^{HHRR}) = \frac{1}{w} + \frac{1}{z} + \frac{1}{x} + \frac{1}{y} \quad (3.9)$$

Πίνακας 3.3. Παρουσίαση των δεδομένων μιας μελέτης σε οικογένειες χρησιμοποιώντας τον έλεγχο Haplotype-based Haplotype Relative Risk (HHRR). Τα αλληλόμορφα των γονέων που διαμορφώνουν ένα πληθυσμό “pseudo-control” που μεταβιβάζονται συγκρίνονται με εκείνα που δεν μεταβιβάζονται. Το Odds Ratio δίνεται από τη σχέση wz/xy . Η σύνδεση με τα δεδομένα του Πίνακα 3.1 μπορεί να γίνει αν παρατηρήσουμε ότι οι πρώτες γραμμές των πινάκων είναι ταυτόσημες ($n_{11}=w$ και $n_{10}=x$).

	Αλληλόμορφο		
	<i>B</i>	<i>A</i>	Σύνολο
Αλληλόμορφο που έχει μεταβιβαστεί	<i>w</i>	<i>x</i>	<i>n₁</i>
Αλληλόμορφο που δεν έχει μεταβιβαστεί	<i>y</i>	<i>z</i>	<i>n₁</i>
Σύνολο	<i>w+y</i>	<i>x+z</i>	<i>2n₁</i>

Από ιστορικής άποψης, αξίζει να αναφερθεί ότι οι Falk και Rubinstein ήταν οι πρώτοι που πρότειναν τη χρήση των αλληλομόρφων που δεν έχουν μεταβιβαστεί για τη διαμόρφωση του pseudo-control γονοτύπου (Falk and Rubinstein 1987). Αργότερα, οι Terwilliger και Ott επέκτειναν την ιδέα αυτή και πρότειναν, για πρώτη φορά, τη χρήση του ελέγχου McNemar, αν και συμπέραναν ότι είχε χαμηλότερη ισχύ σε σχέση με την unmatched ανάλυση που αντιστοιχεί στον έλεγχο HHRR (Terwilliger and Ott 1992). Κατόπιν, ο έλεγχος McNemar επαναδιατυπώθηκε και κατασκευάστηκε ο ευρέως χρησιμοποιούμενος έλεγχος TDT (Spielman, McGinnis et al. 1993; Spielman and Ewens 1996).

3.2.2.3. Επεκτάσεις του ελέγχου transmission disequilibrium test

Ο έλεγχος TDT αποτελεί την κυριότερη μέθοδο για την ανάλυση δεδομένων σε οικογένειες και έχουν προταθεί διάφορες επεκτάσεις για να καλυφθούν κάποιες ειδικές περιπτώσεις. Αυτές οι επεκτάσεις είναι πιο κατάλληλες για ασθένειες που εκδηλώνονται κατά την ενηλικίωση με ελλιπείς πληροφορίες για τους γονείς και

περιλαμβάνουν περιπτώσεις όπου μόνο ο ένας γονέας (Sun, Flanders et al. 1999), ή μόνο τα αδέλφια είναι διαθέσιμα (Curtis 1997; Spielman and Ewens 1998), καταστάσεις όπου χρησιμοποιούνται τα «μη συμφωνούντα» αδέλφια (Horvath and Laird 1998; Deng, Chen et al. 2002) και τέλος σχεδιασμοί case-spouse και case-offspring (Lee 2003).

Αν και οι έλεγχοι αυτοί δεν προκύπτουν συνήθως απευθείας από έναν αντίστοιχο 2x2 πίνακα συνάφειας (όπως ο Πίνακας 3.2), όλοι χρησιμοποιούν μια παραλλαγή του score test που έχει δοθεί στις εξισώσεις 3.4-5. Συνεπώς, σε κάθε περίπτωση μπορούν να υπολογιστούν οι αντίστοιχοι έλεγχοι και να ανακατασκευάσουμε έναν παρόμοιο 2x2 πίνακα. Κατόπιν, μπορούμε να υπολογίσουμε προσεγγιστικά τις εκτιμήσεις των \tilde{b} και \tilde{c} που θα μπορούσαν να προέρχονται από έναν τέτοιο πίνακα λύνοντας το σύστημα των εξισώσεων 3.4 και 3.5. Έτσι διαμορφώνονται οι ακόλουθες σχέσεις:

$$\tilde{b} = 2V + A \quad (3.10)$$

και

$$\tilde{c} = 2V - A \quad (3.11)$$

Οι ποσότητες αυτές μπορούν να χρησιμοποιηθούν για τον υπολογισμό του λογαρίθμου των Odds Ratios με την αντίστοιχη διασπορά αυτών μέσω των εξισώσεων 3.6-7 και ακολούθως να πραγματοποιήσουμε την ανάλυση σαν να είχε χρησιμοποιηθεί εξ' αρχής ο έλεγχος TDT. Σε κάθε περίπτωση, πρέπει να λάβουμε υπόψη ότι οι εξισώσεις 3.10 και 3.11 αποτελούν προσεγγίσεις και δεν μπορούν να χρησιμοποιηθούν όταν οι αντίστοιχοι έλεγχοι περιλαμβάνουν συμμεταβλητές, κυρίως όταν αυτές συσχετίζονται τόσο με την παρουσία ή απουσία της νόσου όσο και με το αλληλόμορφο.

3.2.3. Μεθοδολογίες για μετα-ανάλυση

Εάν οι πρωταρχικές μελέτες μιας μετα-ανάλυσης παρέχουν εκτιμήσεις για μόνο ένα από τα Odds Ratios που παρουσιάστηκαν στις προηγούμενες ενότητες, μπορεί κανείς να προχωρήσει απευθείας στη διεξαγωγή μιας μονομεταβλητής μετα-ανάλυσης (Normand 1999) όπως έχει προταθεί από τους Kazeem και Farrall (Kazeem and Farrall 2005). Πρέπει να τονιστεί ότι ανάμεσα στους διαφόρους ελέγχους που μπορούν να πραγματοποιηθούν σε μια μελέτη ασθενών-μαρτύρων σε πληθυσμούς, μόνο η εξαγωγή του συνοπτικού κινδύνου υπό τη θεώρηση συνεπικρατούς μοντέλου

κληρονομικότητας είναι συγκρίσιμη με αντίστοιχες εκτιμήσεις από μελέτες σε οικογένειες. Σε μια τέτοια μετα-ανάλυση η οποία περιλαμβάνει μια συνοπτική εκτίμηση y_i ($i=1,2,\dots,k$) από k μελέτες, το κλασσικό μοντέλο τυχαίων επιδράσεων (van Houwelingen, Arends et al. 2002) μπορεί να γραφτεί με τη μορφή ενός γραμμικού μοντέλου με τυχαίο σταθερό όρο (random intercept):

$$\begin{aligned} y_i &= \beta + \beta_i + \varepsilon_i \\ \beta_i &\sim N(0, \tau^2), \varepsilon_i \sim N(0, \Sigma_i) \end{aligned} \quad (3.12)$$

Σε αυτό το μοντέλο, με το β ορίζεται η συνολική μέση επίδραση η οποία πρέπει να εκτιμηθεί, το β_i αποτελεί την τυχαία επίδραση (τυχαίος σταθερός όρος) για κάθε i μελέτη, το τ^2 αποτελεί τη μεταξύ των μελετών μεταβλητότητα ενώ με ε_i υποδηλώνονται τα τυχαία σφάλματα τα οποία θεωρούνται γνωστά:

$$\Sigma_i = \begin{bmatrix} \text{var}(y_1) & 0 & \dots & 0 \\ 0 & \text{var}(y_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \text{var}(y_k) \end{bmatrix} \quad (3.13)$$

Ωστόσο, εάν συνδυάσουμε τα αποτελέσματα από μελέτες σε οικογένειες και σε πληθυσμούς οι οποίες χρησιμοποιούν τους ίδιους ή επικαλυπτόμενους ασθενείς (Pomp, Van Stralen et al. 2010), προκύπτουν $j=1,2$ συσχετισμένες εκτιμήσεις για κάθε μια από τις i μελέτες. Μια προσέγγιση που μπορεί να εφαρμοστεί εύκολα, περιλαμβάνει τη διεξαγωγή μιας ανάλυσης σε δυο στάδια όπου μπορούν να συνδυαστούν τα αποτελέσματα έτσι ώστε να υπολογιστεί μια συνολική, σταθμισμένη εκτίμηση για κάθε i μελέτη (le Cessie, Nagelkerke et al. 2008):

$$y_i^{pooled} = \lambda_i y_i^{CC} + (1 - \lambda_i) y_i^{TDT} \quad (3.14)$$

Το βέλτιστο βάρος λ είναι εκείνο που ελαχιστοποιεί τη συνολική διασπορά:

$$\begin{aligned} \text{var}(y_i^{pooled}) &= \lambda_i^2 \text{var}(y_i^{CC}) + (1 - \lambda_i)^2 \text{var}(y_i^{TDT}) \\ &\quad + 2\lambda_i(1 - \lambda_i) \text{cov}(y_i^{CC}, y_i^{TDT}) \end{aligned} \quad (3.15)$$

Ελαχιστοποιώντας την εξίσωση 3.15 οδηγούμαστε στη ακόλουθη σχέση:

$$\lambda_i = \frac{\text{var}(y_i^{TDT}) - \text{cov}(y_i^{CC}, y_i^{TDT})}{\text{var}(y_i^{CC}) + \text{var}(y_i^{TDT}) - 2\text{cov}(y_i^{CC}, y_i^{TDT})} \quad (3.16)$$

Ακολουθώντας την προσέγγιση αυτή για κάθε i μελέτη μιας μετα-ανάλυσης υπολογίζουμε μια μονοδιάστατη εκτίμηση για το Odds Ratio των αλληλομόρφων και προχωρούμε στη διεξαγωγή μιας κλασσικής μονομεταβλητής μετα-ανάλυσης (Normand 1999).

Εναλλακτικά, μπορούμε να εισάγουμε απευθείας στον πίνακα διακυμάνσεων-συνδιακυμάνσεων του μοντέλου τις συνδιακυμάνσεις που εκτιμώνται λόγω των επικαλυπτόμενων ασθενών:

$$\begin{aligned} y_{ij} &= \beta + \beta_i + \varepsilon_{ij} \\ \beta_i &\sim N(0, \tau^2), \varepsilon_{ij} \sim N(0, \Sigma_{ij}) \end{aligned} \quad (3.17)$$

Έτσι, ο πίνακας διακυμάνσεων-συνδιακυμάνσεων θα είναι:

$$\Sigma_{ij} = \begin{bmatrix} \text{var}(y_1^{CC}) & \text{cov}(y_1^{CC}, y_1^{TDT}) & \dots & 0 & 0 \\ \text{cov}(y_1^{CC}, y_1^{TDT}) & \text{var}(y_1^{TDT}) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \text{var}(y_k^{CC}) & \text{cov}(y_k^{CC}, y_k^{TDT}) \\ 0 & 0 & \dots & \text{cov}(y_k^{CC}, y_k^{TDT}) & \text{var}(y_k^{TDT}) \end{bmatrix} \quad (3.18)$$

Επιπλέον, αντί να μοντελοποιούμε απευθείας τις συσχετίσεις, μια διαφορετική προσέγγιση για να λάβουμε υπόψη τη συσχέτιση μεταξύ των εκτιμήσεων περιλαμβάνει τη χρήση του sandwich-type εκτιμητή της διασποράς (Williams 2000). Η προσέγγιση αυτή ομοιάζει αρκετά (σε επίπεδο μετα-ανάλυσης) με εκείνη που προτάθηκε από τον le Cessie και συνεργάτες (2008), για την ενοποίηση των αποτελεσμάτων μιας μελέτης.

Τέλος, θα μπορούσαμε να χρησιμοποιήσουμε ιεραρχικά μοντέλα που εντάσσονται στο γενικό πλαίσιο των γενικευμένων γραμμικών μεικτών μοντέλων (Stram 1996; Breslow, Leroux et al. 1998; Platt, Leroux et al. 1999; Thompson, Turner et al. 2001). Έτσι, διαμορφώνονται (*j*) υποομάδες nested εντός των (*i*) μελετών με τους αντίστοιχους τυχαίους όρους:

$$\begin{aligned} y_{ij} &= \beta + \beta_i + \beta_{ij} + \varepsilon_{ij} \\ \beta_i &\sim N(0, \tau_s^2), \beta_{ij} \sim N(0, \tau_g^2), \varepsilon_{ij} \sim N(0, \Sigma_{ij}) \end{aligned} \quad (3.19)$$

Η μεταξύ των μελετών μεταβλητότητα ορίζεται ως τ_s^2 ενώ με τ_g^2 ορίζεται η διασπορά μεταξύ των ομάδων μιας μελέτης. Κατά παρόμοιο τρόπο με το μοντέλο της εξίσωσης 3.17, ο πίνακας διακυμάνσεων-συνδιακυμάνσεων Σ_{ij} δίνεται με τη μορφή της εξίσωσης 3.13, όπου χρησιμοποιείται ο sandwich-type εκτιμητής της διασποράς (Williams 2000) για να υπολογιστεί η συσχέτιση των εκτιμήσεων; εναλλακτικά η συνδιακύμανση μπορεί να μοντελοποιηθεί απευθείας μέσω της εξίσωσης 3.18.

3.2.4. Υπολογισμός της συνδιακύμανσης

Μια απαραίτητη συνθήκη (για την περίπτωση όπου δεν χρησιμοποιείται ο sandwich-type εκτιμητής της διασποράς) αποτελεί ο υπολογισμός της εντός των μελετών μεταβλητότητας της εξίσωσης 3.18. Εν συντομία, πρέπει να υπολογιστεί αναλυτικά η συνδιακύμανση του λογαρίθμου των Odds Ratios, για παράδειγμα μιας μελέτης ασθενών-μαρτύρων και μιας μελέτης σε οικογένειες όπου εκτιμάται ο HHRR:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HHRR}) &= \text{cov}\left[\log\left(\frac{n_{11}n_{00}}{n_{01}n_{10}}\right), \log\left(\frac{wz}{xy}\right)\right] \\ &= \text{cov}\left[\log(n_{11}n_{00}) - \log(n_{01}n_{10}), \log(wz) - \log(xy)\right]\end{aligned}\quad (3.20)$$

η οποία κατόπιν υπολογισμών (Παράρτημα Α), συνοψίζεται στην ακόλουθη σχέση:

$$\text{cov}(y_i^{CC}, y_i^{HHRR}) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \quad (3.21)$$

Στην περίπτωση όπου παρουσιάζονται τα αποτελέσματα μιας μελέτης ασθενών-μαρτύρων και ταυτόχρονα περιλαμβάνεται μια ανάλυση με τον έλεγχο TDT η συνδιακύμανση υπολογίζεται επίσης από την ακόλουθη σχέση (Παράρτημα Α):

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \quad (3.22)$$

Πρέπει να τονιστεί ότι οι εξισώσεις 3.21 και 3.22 ισχύουν επίσης σε περιπτώσεις όπου κάποιοι ασθενείς (n_{11-} και n_{10-}) δεν συμμετέχουν για κάποιο λόγο στην ανάλυση των δεδομένων σε οικογένειες (για παράδειγμα όταν λείπουν οι γονείς ή απέτυχε η αλληλούχιση). Σε αυτή την περίπτωση, ισχύει ότι $n_{11}=n_{11*}+n_{11-}=w+n_{11-}$ και $n_{10}=n_{10*}+n_{10-}=x+n_{10-}$ και συνεπώς οι εξισώσεις 3.21 και 3.22 δεν τροποποιούνται. Παρόμοια είναι η αντιμετώπιση της αντίστροφης περίπτωσης όπου κάποιοι ασθενείς συμμετέχουν μόνο στις μελέτες σε οικογένειες και όχι στην ανάλυση ασθενών-μαρτύρων. Στη γενική περίπτωση, όπου υπάρχει μόνο μερική επικάλυψη των ασθενών, οι εξισώσεις πρέπει να τροποποιηθούν. Για παράδειγμα, εάν θεωρήσουμε ότι οι δυο σχεδιασμοί μοιράζονται $m/2$ ασθενείς που συνεισφέρουν m αλληλόμορφα (με $m < n_1$) οι αντίστοιχες ποσότητες της εξίσωσης 3.21 πρέπει να διαμεριστούν σε δυο υποσύνολα που δηλώνονται με τους δείκτες s και d (s :shared και d :distinct). Έτσι, η εξίσωση 3.21 διαμορφώνεται ως εξής:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HHRR}) &= \text{cov}\left[\log(n_{11s} + n_{11d}), \log(w_s + w_d)\right] \\ &\quad + \text{cov}\left[\log(n_{10s} + n_{10d}), \log(x_s + x_d)\right]\end{aligned}\quad (3.23)$$

Στο παράρτημα Α αποδεικνύεται ότι η συνδιακύμανση υπολογίζεται τελικά από την ακόλουθη σχέση:

$$\text{cov}(y_i^{CC}, y_i^{HHRR}) = \frac{n_{11s}}{n_{11}w} + \frac{n_{10s}}{n_{10}x} \quad (3.24)$$

Παρόμοια (Παράρτημα Α), η γενική μορφή της εξίσωσης (3.22) γίνεται:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \frac{b_s}{b} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) \quad (3.25)$$

Ένα πρακτικό πρόβλημα που ανακύπτει με τη χρήση της εξίσωσης 3.24 είναι το ότι δεν αναμένεται σε κάθε μελέτη να αναφέρονται όλες τις ποσότητες που χρειάζονται για τον υπολογισμό της (δηλαδή n_{11s} , n_{11d} και λοιπά). Σε αυτές τις περιπτώσεις, μπορεί να χρησιμοποιηθεί η αρχική εκτίμηση της εξίσωσης 3.21 που υποθέτει επικάλυψη για όλα τα άτομα οδηγώντας έτσι σε μεγαλύτερη συνδιακύμανση. Μια άλλη εναλλακτική είναι να θεωρήσουμε ότι η επικάλυψη είναι σταθερή (δηλαδή $n_{11s}/n_{11} = n_{10s}/n_{10}$) και να οδηγηθούμε σε μια προσεγγιστική εκτίμηση της συνδιακύμανσης:

$$\text{cov}(y_i^{CC}, y_i^{HHRR}) \approx \frac{m}{n_1} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) = \frac{w_s + x_s}{w + x} \left(\frac{1}{w} + \frac{1}{x} \right) \quad (3.26)$$

Κατά παρόμοιο τρόπο η συνδιακύμανση της εξίσωσης 3.25 διαμορφώνεται ως εξής:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) \approx \frac{m}{n_1} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) \quad (3.27)$$

Προφανώς, οι δυο εκτιμήσεις μπορούν να χρησιμοποιηθούν από κοινού, με τις εκτιμήσεις των εξισώσεων (3.21) και (3.22) να αποτελούν το άνω όριο για την εκτίμηση της συνδιακύμανσης. Αξίζει να αναφερθεί ότι παρόμοιες εξισώσεις έχουν παρουσιαστεί κάτω από ένα διαφορετικό πλαίσιο, για παράδειγμα, για τη σύγκριση συνεχών μεταβλητών έναντι μιας κοινής ομάδας ελέγχου (Dunnett 1955) καθώς επίσης και σε μια ευρυγονιδιωματική μελέτη (Zaykin and Kozbur 2010).

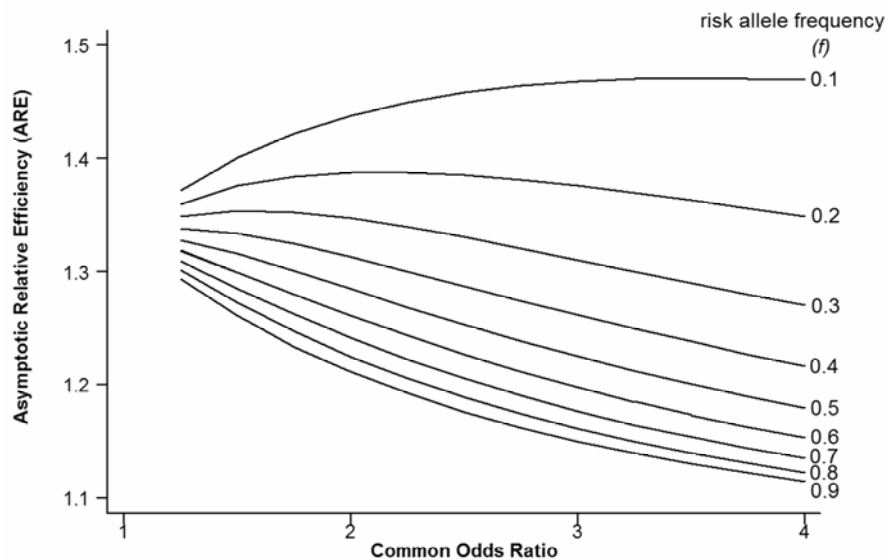
Αντικαθιστώντας τις εκτιμήσεις της συνδιακύμανσης που υπολογίζεται από τις εξισώσεις 3.21 και 3.22, στην έκφραση της διασποράς για τη συνολική εκτίμηση (Εξίσωση 3.15), προκύπτει μια έκφραση όπου χρησιμοποιούνται αποκλειστικά οι παρατηρούμενες συχνότητες. Συνεπώς, επιβεβαιώνεται το γενικό αποτέλεσμα (Keller and Olkin 2004), ότι δηλαδή ο συνδυασμός αυξάνει την ισχύ μειώνοντας τη διασπορά:

$$\text{var}(y_i^{\text{pooled}}) \leq \min \{ \text{var}(y_i^{CC}), \text{var}(y_i^{TDT}) \} \quad (3.28)$$

Στη συγκεκριμένη περίπτωση, η κλειστή έκφραση που δίνεται για τη συνδιακύμανση υποδηλώνει ότι ισχύει μόνο το σύμβολο της ανισότητας και αν υποθέσουμε ότι όλοι οι έλεγχοι ακολουθούν ασυμπτωτικά την κανονική κατανομή, μπορούμε εύκολα να συμπεράνουμε ότι η Asymptotic Relative Efficiency (*ARE*) της συνδυασμένης εκτίμησης σε σύγκριση με εκείνη με τη μικρότερη διασπορά είναι πάντα μεγαλύτερη της μονάδας:

$$ARE = \frac{\min\{\text{var}(y_i^{CC}), \text{var}(y_i^{TDT})\}}{\text{var}(y_i^{pooled})} > 1 \quad (3.29)$$

Θεωρώντας ίσο αριθμό ασθενών και μαρτύρων και ένα κοινό Odds Ratio για την ανάλυση ασθενών-μαρτύρων και TDT, μπορεί ναδειχθεί ότι η *ARE* είναι μια συνάρτηση τριών μόνο παραμέτρων: του κοινού Odds Ratio, της συχνότητας του αλληλομόρφου κινδύνου στους υγιείς (*f*) και του συνολικού μεγέθους δείγματος (*n*). Στην Εικόνα 3.1, απεικονίζεται η *ARE* για ένα εύρος πιθανών τιμών των *f* και Odds Ratios (η επίδραση του *n* είναι αμελητέα για *n*>50). Τα αποτελέσματα δείχνουν ότι η συνδυασμένη εκτίμηση είναι κατά 12-47% πιο αποτελεσματική (efficient) συγκρινόμενη με τον συνήθη εκτιμητή ασθενών-μαρτύρων. Παρόμοια αποτελέσματα μπορούμε επίσης να έχουμε συνενώνοντας τα αποτελέσματα του ελέγχου HHRR και μελετών ασθενών-μαρτύρων.



Εικόνα 3.1. Η *ARE* της συνδυασμένης εκτίμησης σε σύγκριση με την εκτίμηση από τις μελέτες ασθενών-μαρτύρων, απεικονίζεται για ένα εύρος πιθανών τιμών των *f* και Odds Ratio. Για τις πιο συχνά αναμενόμενες τιμές του Odds Ratio (1.25 έως 2), η *ARE* κυμαίνεται από 1.25 έως 1.4. Ακόμα και για την ακραία περίπτωση όπου *f*=0.9 και *OR*=4, η *ARE* είναι μεγαλύτερη από 1.12 υποδηλώνοντας ότι η συνολική εκτίμηση είναι κατά 12% πιο αποτελεσματική (efficient) (θα έχει κατά 11% μικρότερη διασπορά). Η επίδραση του *n* είναι αμελητέα για *n*>50.

Τέλος, αξίζει να αναφερθεί, ότι οι εκτιμήσεις των μελετών που δίνουν αποτελέσματα ταυτόχρονα για τους ελέγχους TDT και HHRR είναι επαναλαμβανόμενες και μόνο η μια εκ των δυο πρέπει να χρησιμοποιείται. Για παράδειγμα, η συνδιακύμανση μεταξύ των y_i^{TDT} και y_i^{HHRR} , μπορεί να δειχθεί εύκολα ότι ισούται με $\text{var}(y_i^{HHRR})$ και συνεπώς, η εξίσωση 3.15 ελαχιστοποιείται όταν $y_i^{pooled} = y_i^{HHRR}$. Αυτό μπορεί να γίνει εύκολα κατανοητό, αν θεωρήσουμε ότι οι δυο έλεγχοι χρησιμοποιούν τα ίδια δεδομένα και ο HHRR έχει μικρότερη διασπορά σε σχέση με τον TDT.

3.2.5. Υλοποίηση

Οι προσεγγίσεις που έχουν περιγραφεί ως τώρα παρουσιάζουν έναν κλιμακούμενο βαθμό δυσκολίας κατά την υλοποίηση. Για παράδειγμα, η προσέγγιση των δυο σταδίων μπορεί να υλοποιηθεί ακόμα και σε ένα απλό λογιστικό φύλλο καθώς δεν περιλαμβάνει επαναληπτικές διαδικασίες. Επιπλέον, με την προσέγγιση αυτή επιτρέπεται η πραγματοποίηση πολλών γνωστών πληροφοριακών διαγνωστικών ελέγχων που χρησιμοποιούνται στη μετα-ανάλυση. Έτσι μπορεί να πραγματοποιηθεί έλεγχος για την ύπαρξη συστηματικού σφάλματος δημοσίευσης (Begg and Mazumdar 1994; Egger, Davey Smith et al. 1997), αθροιστικής μετα-ανάλυσης (Lau, Antman et al. 1992; Bago and Nikolopoulos 2009) καθώς και άλλες αναλύσεις ευαισθησίας ή αναλύσεις υποομάδων. Από την άλλη πλευρά, οι προσεγγίσεις που χρησιμοποιούν γραμμικά μεικτά μοντέλα απαιτούν τη χρήση ειδικών στατιστικών πακέτων όπως το Stata, R, ή SAS.

Για την προσέγγιση των δυο σταδίων υλοποιήθηκε ένα πρόγραμμα στο στατιστικό πακέτο Stata χρησιμοποιώντας τις εκτιμήσεις των συνδιακυμάνσεων των εξισώσεων 3.21 και 3.27 (<http://www.compgen.org/tools/meta-tdt>). Σε αυτό το πρόγραμμα, παρουσιάζουμε την υλοποίηση της κλασσικής μετα-ανάλυσης θέτοντας ως βάρη την αντίστροφη ποσότητα της διασποράς και χρησιμοποιώντας την μη επαναληπτική μέθοδο των ροπών των DerSimonian και Laird (DerSimonian and Laird 1986); ωστόσο, παρουσιάζουμε την εφαρμογή άλλων μεθόδων εκτίμησης όπως η μέγιστη πιθανοφάνεια και η δεσμευμένη μέγιστη πιθανοφάνεια (Thompson and Sharp 1999). Επιπλέον, παρουσιάζουμε την εκτίμηση των μοντέλων των εξισώσεων 3.17 και 3.19 χρησιμοποιώντας τον sandwich-type εκτιμητή της διασποράς (Williams

2000) με την εντολή `gllamm` στο Stata (Rabe-Hesketh, Skrondal et al. 2002; Rabe-Hesketh, Skrondal et al. 2005).

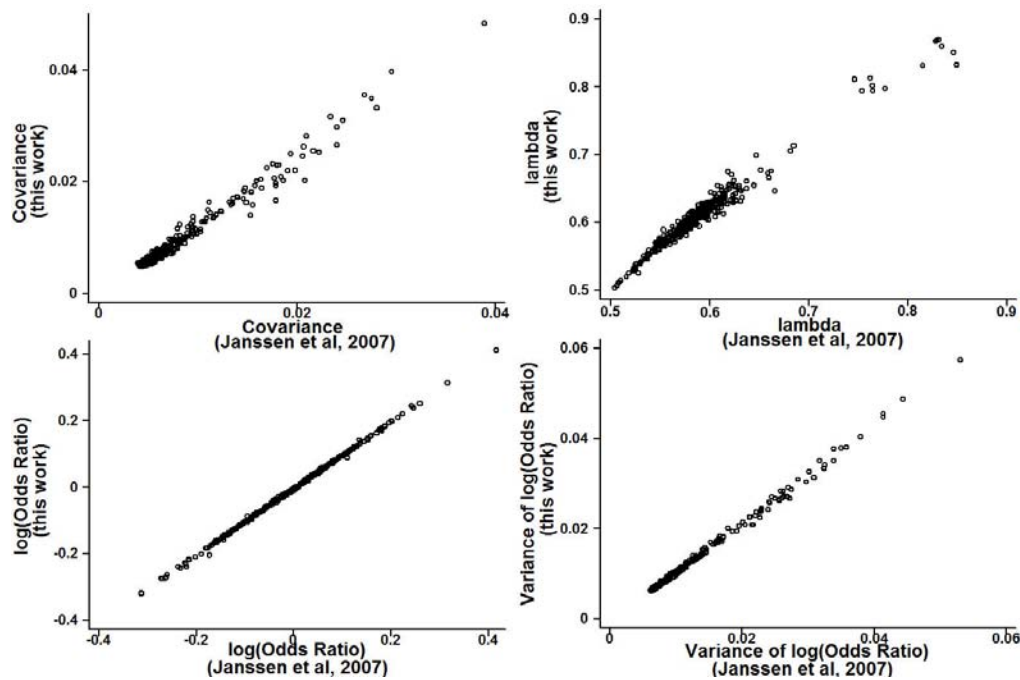
3.3. Αποτελέσματα

Στην ενότητα αυτή γίνεται εφαρμογή της μεθοδολογίας που παρουσιάστηκε σε τρία σύνολα δεδομένων. Αρχικά, χρησιμοποιήθηκαν δεδομένα 356 SNPs, που αξιολογούσαν την πιθανή συσχέτιση ενός συγκεκριμένου ιού (Respiratory Syncytial Virus-RSV) με τον κίνδυνο εμφάνισης Βρογχολίτιδας (Janssen, Bont et al. 2007). Στη συγκεκριμένη μελέτη, τα δεδομένα 470 παιδιών που νοσηλεύτηκαν για τη συγκεκριμένη νόσο, των γονιών τους, και 1008 ανεξάρτητων υγιών αναλύθηκαν χρησιμοποιώντας τη μέθοδο που περιγράφηκε από τους (le Cessie, Nagelkerke et al. 2008). Τα συγκεκριμένα δεδομένα επιλέχθηκαν δεδομένου του μεγάλου αριθμού αποτελεσμάτων για την εμπειρική αξιολόγηση της προτεινόμενης μεθοδολογίας σε σχέση με την προϋπάρχουσα μεθοδολογία (le Cessie, Nagelkerke et al. 2008), όπου χρησιμοποιούνται ατομικά δεδομένα.

Η μέθοδος εφαρμόστηκε επίσης σε μια ευρυγονιδιωματική μελέτη που στόχευε να ανιχνεύσει πιθανούς γονιδιακούς πολυμορφισμούς που συσχετίζονται με την εμφάνιση Σκλήρυνσης κατά Πλάκας. Στη συγκεκριμένη μελέτη, πραγματοποιήθηκε αλληλούχιση 931 family trios, τα δεδομένα των οποίων αναλύθηκαν χρησιμοποιώντας τον έλεγχο TDT και ακολούθως, οι ίδιοι ασθενείς συγκρίθηκαν έναντι 2,431 ανεξάρτητων υγιών (Hafler, Compston et al. 2007). Συνήθως στις ευρυγονιδιωματικές μελέτες, οι ερευνητές ενδιαφέρονται να αναπαράγουν τα αρχικώς στατιστικά σημαντικά αποτελέσματα και έχειδειχθεί ότι το συνδυασμός των σταδίων screening και replication αυξάνει την ισχύ των εκτιμήσεων (Skol, Scott et al. 2006).

Τέλος, η μεθοδολογία εφαρμόστηκε για να διερευνήσει τη συσχέτιση του 10-Repeat Allele ενός πολυμορφισμού VNTR στην 3' αμετάφραστη περιοχή του γονιδίου του μεταφορέα της ντοπαμίνης (DAT1) με τη ΔΕΠ-Υ (Yang, Chan et al. 2007). Η συγκεκριμένη μετα-ανάλυση περιλαμβάνει 26 μελέτες, από τις οποίες δυο είχαν δεδομένα τόσο για τον έλεγχο TDT όσο και για τη ανάλυση ασθενών-μαρτύρων χρησιμοποιώντας τους ίδιους ασθενείς ενώ σε μια μελέτη υπήρχαν πληροφορίες για τον έλεγχο HHRR και την ανάλυση ασθενών-μαρτύρων. Μια μελέτη, είχε δεδομένα τόσο για τον HHRR όσο και για τον TDT, αλλά όπως έχει ήδη αναφερθεί μόνο ένας

από τους δυο μπορεί να χρησιμοποιηθεί (επιλέξαμε την χρήση του ελέγχου TDT). Η μετα-ανάλυση περιλαμβάνει επίσης 15 μελέτες που μπορούσε να εκτιμηθεί αποκλειστικά ο έλεγχος TDT, 3 μελέτες που περιελάμβαναν μόνο το σχεδιασμό ασθενών-μαρτύρων, και 5 μελέτες μόνο για τον υπολογισμό του ελέγχου HHRR. Στην αρχική δημοσίευση, τα δεδομένα αναλύθηκαν ξεχωριστά, αλλά κατέστη δυνατή η πραγματοποίηση της ανάλυσης ενσωματώνοντας όλη τη διαθέσιμη πληροφορία.

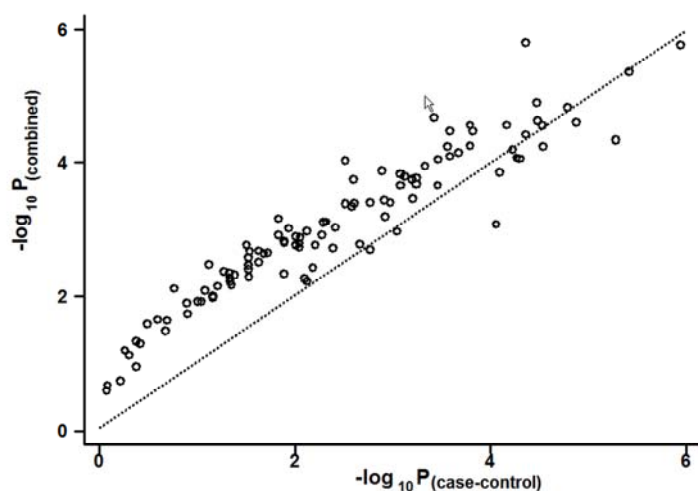


Εικόνα 3.2. Σύγκριση των αποτελεσμάτων χρησιμοποιώντας τη μεθοδολογία που παρουσιάστηκε έναντι των εκτιμήσεων που προκύπτουν από τη χρήση του sandwich estimator για τα δεδομένα 356 SNPs που αναφέρονται από τους Janssen και συνεργάτες, (2007). Παρουσιάζονται τα αποτελέσματα για τη συνδιακύμανση, την εκτίμηση του lambda, το συνδυασμένο λογάριθμο του Odds Ratio και τη διασπορά αυτού. Χρησιμοποιήθηκαν reconstructed δεδομένα για τους 2x2 πίνακες συνάφειας και σε όλες τις περιπτώσεις το R^2 ήταν μεγαλύτερο από 0.96.

Στην ανάλυση των δεδομένων σχετικά με την πιθανή συσχέτιση πολυμορφισμών με τη Βρογχιολίτιδα (Janssen, Bont et al. 2007) και την ευρυγονιδιωματική μελέτη για τη Σκλήρυνση κατά Πλάκας (Hafler, Compston et al. 2007), μια επιπλέον δυσκολία πηγάζει από το γεγονός ότι οι 2x2 πίνακες συνάφειας δεν ήταν διαθέσιμοι στις αρχικές δημοσιεύσεις. Έτσι, εφαρμόσαμε μια μέθοδο που έχει προταθεί (Di Pietrantonj 2006) για να υπολογίσουμε τα δεδομένα που απαιτούνται για την ανάλυση ασθενών-μαρτύρων (τα δεδομένα που απαιτούνται για τον υπολογισμό του ελέγχου TDT μπορούν εύκολα να υπολογιστούν καθώς απαιτούνται μόνο δυο ποσότητες). Μιας και η μέθοδος υπολογισμού των δεδομένων

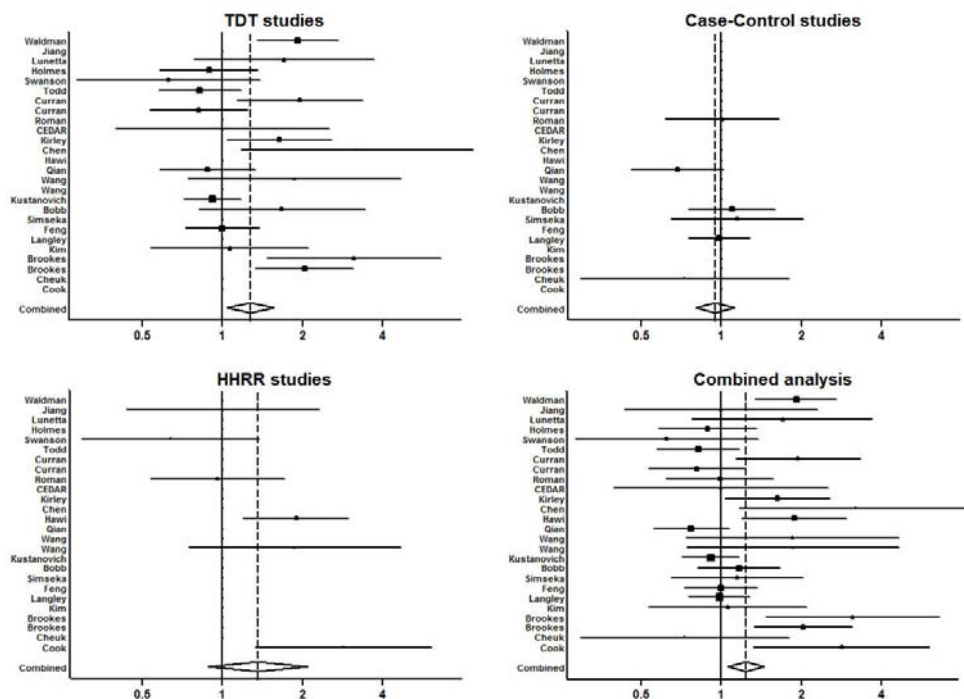
δίνει δυο πιθανές λύσεις (Di Pietrantonj 2006), κρατήσαμε μόνο εκείνη που μας δίνει συχνότητες αλληλομόρφων κοντά σε εκείνες που αναφέρονται στο Harmap για Ευρωπαίους. Για τα δεδομένα της μελέτης των Janssen και συνεργατών, έξι από τους 356 SNPs που αναλύθηκαν αρχικά, αποκλείστηκαν καθώς έδιναν πολύ μικρές συχνότητες αλληλομόρφων και το ίδιο ακριβώς συνέβη για τρεις από τους 110 SNPs στην ευρυγονιδιωματική μελέτη για τη Σκλήρυνση κατά Πλάκας.

Όταν χρησιμοποιήθηκε η εξίσωση (3.22) για τα δεδομένα των 356 SNPs που αναφέρθηκαν από τον Janssen και συνεργάτες, υπήρχε μεγάλο ποσοστό συμφωνίας με τα αρχικά αποτελέσματα που προκύπτουν από τη χρήση του sandwich estimator. Τα αποτελέσματα δείχνουν ξεκάθαρα ότι τα δεδομένα που υπολογίσαμε ήταν αρκετά κοντά στα πραγματικά και έτσι οι υπολογιζόμενες συνδιακυμάνσεις προσεγγίζουν αρκετά ικανοποιητικά εκείνες που προκύπτουν από τη χρήση του sandwich estimator. Παρόμοια είναι τα ευρήματα και για την παράμετρο λ που εκτιμήθηκε χρησιμοποιώντας την εξίσωση 3.16. Το πιο σημαντικό εύρημα είναι ότι η συνολική εκτίμηση του συνδυασμένου λογαρίθμου του Odds Ratio και η διασπορά αυτού είναι πολύ κοντά με εκείνες που δίνονται στη δημοσιευμένη εργασία (Εικόνα 3.2). Η μέθοδος που παρουσιάστηκε, δίνει οριακά μεγαλύτερες εκτιμήσεις της συνδιακύμανσης κάτι που οδηγεί σε πιο συντηρητικές εκτιμήσεις σε σχέση με τον sandwich estimator όπου χρησιμοποιούνται ατομικά δεδομένα. Μιας και η προτεινόμενη μεθοδολογία στηρίζεται στην προσεγγιστικό υπολογισμό των συχνοτήτων του πίνακα συνάφειας, είναι προφανές ότι θα αποκλίνει περισσότερο όταν οι τιμές των κελιών του πίνακα είναι μικρές ή ακόμα και μηδέν. Μια λύση στο πρόβλημα αυτό αποτελεί η χρήση μιας μικρής σταθεράς (δηλαδή $c=1/2$) στις παρατηρούμενες ποσότητες (Anscombe 1956; Haldane 1956; Gart and Zweifel 1967).



Εικόνα 3.3. Εφαρμογή της μεθόδου στην ευρυγονιδιωματική μελέτη για τη Σκλήρυνση κατά Πλάκας (Hafler et al, 2007). Απεικονίζεται ο αρνητικός λογάριθμος (με βάση το 10) των p -values από τη συνδυασμένη ανάλυση, έναντι των αντίστοιχων τιμών από την ανάλυση μόνο των ασθενών-μαρτύρων.

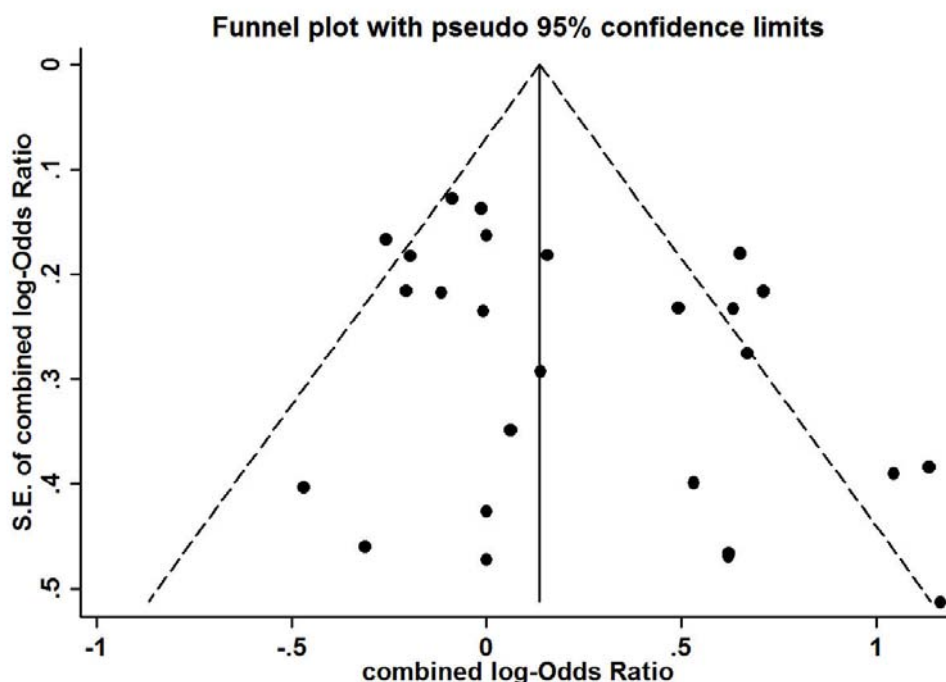
Όταν η μέθοδος εφαρμόστηκε στους πιο σημαντικούς SNPs (συνολικά 110) που περιγράφονται στην ευρυγονιδιωματική μελέτη για τη Σκλήρυνση κατά Πλάκας (Hafler, Compston et al. 2007), τα αποτελέσματα έδειξαν ξεκάθαρα τη χρησιμότητα της προτεινόμενης μεθοδολογίας (Εικόνα 3.3). Σε σχέση με την ανάλυση των ασθενών-μαρτύρων, στη συνδυασμένη ανάλυση εκτιμώνται χαμηλότερα p -values για τους περισσότερους SNPs, υποδεικνύοντας αυξημένη ισχύ για να ανιχνευτούν ασθενείς επιδράσεις γονιδίων. Μολονότι κανένας πολυμορφισμός δεν έφτασε το κοινώς αποδεκτό όριο του $p < 5 \times 10^{-8}$, η συνδυασμένη ανάλυση έδωσε 9 επιπλέον SNPs με $p < 10^{-4}$, σε σύγκριση με την αρχική ανάλυση που χρησιμοποιεί τα δεδομένα μόνο ασθενών-μαρτύρων.



Εικόνα 3.4. Τα forest plots που περιγράφουν τις τρεις επιμέρους μετα-αναλύσεις (TDT, Case-Control και HHRR μελέτες) καθώς επίσης και τη συνδυασμένη ανάλυση όπου χρησιμοποιείται η προσέγγιση των δυο σταδίων. Τα αποτελέσματα των ξεχωριστών αναλύσεων είναι σχεδόν ταυτόσημα με εκείνα που παρουσιάστηκαν στην αρχική δημοσίευση (Yang et al, 2007), ενώ η από κοινού ανάλυση είναι εφικτή μόνο με την προτεινόμενη μεθοδολογία.

Στη μετα-ανάλυση για την διερεύνηση της συσχέτισης του αλληλομόρφου 10-Repeat Allele ενός πολυμορφισμού VNTR στην 3' αμετάφραστη περιοχή του γονιδίου του μεταφορέα της ντοπαμίνης με τη ΔΕΠ-Y (Yang, Chan et al. 2007), οι επιμέρους αναλύσεις έδωσαν παρόμοια αποτελέσματα με εκείνα των πρωταρχικών αναλύσεων εκτιμώντας στατιστικά σημαντικές συσχετίσεις μόνο για τις 17 TDT μελέτες με Odds Ratio ίσο με 1.27 (95% C.I.: 1.04, 1.56). Αναλύοντας τα διαθέσιμα δεδομένα για τις 6 μελέτες ασθενών-μαρτύρων το Odds Ratio ήταν 0.95 (95% C.I.: 0.80, 1.12), ενώ στην ανάλυση των 6 HHRR μελετών ήταν 1.36 (95% C.I.: 0.88, 2.09). Ξεκάθαρα, τα αυτά αποτελέσματα δείχνουν ότι υπάρχουν διαφορές μεταξύ των αναλύσεων σε οικογένειες (TDT και HHRR) και σε πληθυσμούς, αλλά οι εκτιμήσεις δεν μπορούν να συγκριθούν άμεσα μιας και η απλή ανάλυση δεν λαμβάνει υπόψη τη συσχέτιση εντός των μελετών. Εάν, προσπαθήσουμε να πραγματοποιήσουμε μια κλασσική μετα-ανάλυση χρησιμοποιώντας από κάθε μελέτη με εκτιμήσεις για οικογένειες και πληθυσμούς μόνο εκείνες που αναφέρονται στο δεύτερο σχεδιασμό, τα αποτελέσματα καταδεικνύουν μια στατιστικά σημαντική συνολική συσχέτιση (OR=1.23; 95% C.I.: 1.05, 1.45). Όταν η ίδια διαδικασία επαναλήφθηκε μόνο για τις

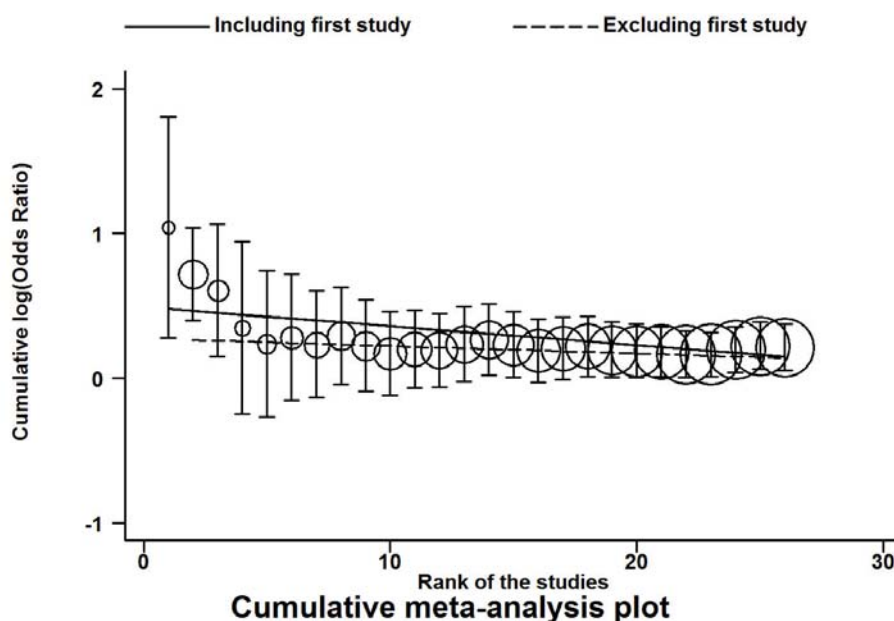
αναλύσεις σε οικογένειες, τα αποτελέσματα ήταν παρόμοια αν και με ελαφρώς μεγαλύτερα διαστήματα εμπιστοσύνης (OR=1.27, 95% C.I.: 1.08, 1.49). Ωστόσο, στην αρχική δημοσίευση δεν πραγματοποιήθηκε κάποια τέτοια ανάλυση.



Εικόνα 3.5. Το funnel plot για τη διερεύνηση της ασυμμετρίας της συνδυασμένης εκτίμησης. Υπάρχει μόνο μια μικρή ένδειξη για συστηματικό σφάλμα δημοσίευσης. Οι έλεγχοι του Begg και Mazumbar είναι μη στατιστικά σημαντικοί αλλά άλλοι έλεγχοι (ο έλεγχος του Egger, της παλινδρόμησης τυχαίων επιδράσεων και ο “trim and fill”) παρέχουν ένδειξη για την ύπαρξη ετερογένειας.

Από την ανάλυση των δυο σταδίων εκτιμήθηκε ένα Odds Ratio ίσο με 1.24 (95% C.I.: 1.06, 1.45). Η αυτή ανάλυση έδωσε στενότερα διαστήματα εμπιστοσύνης σε σχέση με όλες τις προηγούμενες προσεγγίσεις. Αν και στο συγκεκριμένο παράδειγμα οι διαφορές είναι αμελητέες από πρακτικής άποψης, γίνεται εμφανές ότι αν αυξανόταν ο αριθμός των μελετών που έδιναν πληροφορίες για τις διάφορες υποομάδες, η διαφορά θα πιο ήταν εμφανής. Στην Εικόνα 3.4 παρουσιάζονται τα αποτελέσματα των τριών επιμέρους αναλύσεων, καθώς επίσης και η συνδυασμένη ανάλυση. Η μεταξύ των μελετών μεταβλητότητα της από κοινού ανάλυσης είναι αρκετά μεγάλη με τ^2 ίσο με 0.0922 ($I^2=62.7\%$), αλλά το εύρημα αυτό δεν μπορεί να αποδοθεί μόνο στο σχεδιασμό της μελέτης ή τη φυλή (για παράδειγμα, το I^2 των μελετών που υπολογίζεται ο έλεγχος TDT είναι ίσο με 66.1%). Τα ιεραρχικά μοντέλα των εξισώσεων 3.17 και 3.19 με μέγιστη πιθανοφάνεια και με τον sandwich estimator της διασποράς δίνουν πολύ κοντινές εκτιμήσεις, μιας και το συνολικό Odds Ratio ήταν ίσο με 1.23 (95% C.I.: 1.03, 1.46) και στις δυο περιπτώσεις. Το τελευταίο

αποτέλεσμα δείχνει ότι η μεταξύ των ομάδων μεταβλητότητα εντός μια μελέτης είναι αμελητέα ($\tau_g^2 < 10^{-8}$), αλλά αυτό μπορεί να οφείλεται στον μικρό αριθμό μελετών που παρείχαν δεδομένα για αυτές τις υποομάδες. Η μεταξύ των μελετών μεταβλητότητα τ^2 που εκτιμάται από τα μοντέλα των εξισώσεων 3.17 και 3.19 ισούται με 0.0994, κάτι που δείχνει επίσης ότι οι προσεγγίσεις των δυο και του ενός σταδίου οδηγούν σε σχεδόν ταυτόσημα αποτελέσματα.



Εικόνα 3.6. Γράφημα αθροιστικής μετα-ανάλυσης για την ανίχνευση ύπαρξης διαχρονικής τάσης. Οι γραμμές παλινδρόμησης οι οποίες κατασκευάζονται συμπεριλαμβάνοντας ή αποκλείοντας την πρώτη μελέτη βασίζονται στον έλεγχο παλινδρόμησης γενικευμένων ελαχίστων τετραγώνων (Bagos and Nikolopoulos, 2009). Τόσο ο έλεγχος που βασίζεται στην παλινδρόμηση όσο και ο έλεγχος της πρώτης έναντι των υπολοίπων μελετών καταδεικνύουν την παρουσία του φαινομένου του Πρωτέα μιας και στην πρώτη δημοσιευμένη μελέτη εκτιμάται ένα αρκετά μεγαλύτερο Odds Ratio συγκρινόμενη με τις υπόλοιπες. Εάν η πρώτη μελέτη αφαιρεθεί από την ανάλυση (στικτή γραμμή) η διαχρονική τάση εξαφανίζεται αλλά η συνολική σημαντικότητα δεν αλλάζει.

Η συνδυασμένη ανάλυση μας επιτρέπει να χρησιμοποιήσουμε όλους τους γνωστούς διαγνωστικούς ελέγχους αξιοποιώντας όλα τα διαθέσιμα δεδομένα. Για παράδειγμα, είναι εφικτό να πραγματοποιηθεί μια σειρά ελέγχων για το συστηματικό σφάλμα δημοσίευσης (Macaskill, Walter et al. 2001). Στα συγκεκριμένα δεδομένα, η μέθοδος των Begg και Mazumbar (Begg and Mazumdar 1994) αποκλείει την πιθανότητα ύπαρξης συστηματικού σφάλματος δημοσίευσης, αλλά η μέθοδος του Egger (Egger, Davey Smith et al. 1997) και το ανάλογο αυτής τυχαίων επιδράσεων (Thompson and Sharp 1999), παρέχει μια μέτρια ένδειξη για τέτοιου είδους σφάλμα (p -values ίσα με 0.042 και 0.060, αντίστοιχα). Παρόμοια ήταν τα ευρήματα από την μη παραμετρική μέθοδο trim and fill (Duval and Tweedie 2000). Ο οπτικός έλεγχος

του funnel plot ωστόσο, δείχνει μια μικρή ένδειξη ασυμμετρίας των εκτιμήσεων (Εικόνα 3.5). Ωστόσο, η μεγάλη παρατηρούμενη ετερογένεια μπορεί να οφείλεται για τα αποτελέσματα των ελέγχων αυτών.

Το γράφημα αθροιστικής μετα-ανάλυσης (Lau, Antman et al. 1992), αποκάλυψε ότι η πρώτη μελέτη για το συγκεκριμένο ερευνητικό ερώτημα, έχει ένα αρκετά μεγάλο Odds Ratio σε σύγκριση με τις υπόλοιπες δημοσιευμένες μελέτες (Εικόνα 3.6). Οι επίσημοι στατιστικοί έλεγχοι δηλαδή ο έλεγχος παλινδρόμησης που βασίζεται σε γενικευμένα ελάχιστα τετράγωνα (Bagos and Nikolopoulos 2009) και η σύγκριση της πρώτης έναντι των υπολοίπων μελετών (Ioannidis and Trikalinos 2005), οδήγησαν σε παρόμοια αποτελέσματα (p -values ίσα με 0.031 και 0.001 αντίστοιχα). Αν αποκλειστεί η πρώτη δημοσιευμένη μελέτη, η συνολική εκτίμηση παραμένει στατιστικά σημαντική αν και μειώνεται το εκτιμώμενο μέτρο επίδρασης (OR=1.21; 95% C.I.: 1.03, 1.41) και η διαχρονική τάση σχεδόν εξαφανίζεται. Έτσι, φαίνεται ότι η πρώτη μελέτη υπερεκτιμούσε τον πραγματικό κίνδυνο του αλληλομόρφου εκτιμώντας μόνο το HHRR στατιστικό (OR=2.84; 95% C.I.: 1.32, 6.09). Αυτά τα αποτελέσματα δείχνουν την παρουσία του φαινομένου του Πρωτέα (Ioannidis and Trikalinos 2005), το οποίο ενδέχεται να επηρεάσει την επαναληψιμότητα των αποτελεσμάτων στις μελέτες γενετικής συσχέτισης (Ioannidis, Ntzani et al. 2001). Ξεκάθαρα, η προτεινόμενη μέθοδος επέτρεψε την διερεύνηση τέτοιου είδους μεροληψίας κάτι που δεν ήταν εφικτό στην πρωταρχική δημοσίευση.

3.4. Συζήτηση

Τα τελευταία χρόνια, υπάρχει ολοένα και μεγαλύτερο ενδιαφέρον για την ενοποίηση μελετών γενετικής συσχέτισης σε οικογένειες και σε πληθυσμούς οι οποίες περιλαμβάνουν τους ίδιους ασθενείς. Διάφορες μέθοδοι έχουν προταθεί για το συγκεκριμένο πρόβλημα επιτρέποντας την από κοινού ανάλυση (Nagelkerke, Hoebee et al. 2004; Epstein, Veal et al. 2005; Joo, Tian et al. 2007; Pfeiffer, Pee et al. 2008; Glaser and Holmans 2009; Guo, Lunetta et al. 2009; Hsu, Starr et al. 2009; Gebregziabher, Guimaraes et al. 2010; Lasky-Su, Won et al. 2010). Ωστόσο, αυτές οι μέθοδοι απαιτούν ατομικά δεδομένα και έτσι δεν είναι κατάλληλες για τη μετα-ανάλυση δημοσιευμένων μελετών όπου ο ερευνητής έχει πρόσβαση μόνο σε συνοπτικά δεδομένα.

Η μέθοδος που παρουσιάστηκε υλοποιείται εύκολα και απαιτεί μόνο τα συνοπτικά δεδομένα που είναι διαθέσιμα στη βιβλιογραφία. Η προσέγγιση στηρίζεται στη γενική μεθοδολογία για συσχετισμένες παρατηρήσεις (Gleser and Olkin 1994). Άλλες κατηγορίες μελετών όπου η συσχέτιση μπορεί να υπολογιστεί αναλυτικά αφορούν μελέτες πολλαπλών θεραπειών (Higgins and Whitehead 1996; Lu and Ades 2004), κλινικές δοκιμές με αμοιβαίως αποκλειόμενες εκβάσεις (Trikalinos and Olkin 2008), μοντέλα δόσης-απόκρισης (Greenland and Longnecker 1992; Berrington and Cox 2003), πολυμεταβλητή μοντελοποίηση του λογαρίθμου των Odds Ratios σε μελέτες γενετικής συσχέτισης (Bagos 2008), μελέτες γενετικής συσχέτισης με επικαλυπτόμενους μάρτυρες (Lin and Sullivan 2009) και μετα-ανάλυση πολλαπλών γονιδιακών πολυμορφισμών που βρίσκονται σε ανισορροπία σύνδεσης (Bagos and Liakopoulos 2010).

Η θεωρία στην οποία βασίζεται η προτεινόμενη μεθοδολογία είναι αρκετά απλή και αφορά προσεγγίσεις μεγάλου μεγέθους δείγματος (Clayton and Hills 1993; Agresti 2002), οι οποίες χρησιμοποιούνται συνήθως σε μετα-αναλύσεις συνοπτικών δεδομένων. Συγκρίνοντας τα αποτελέσματα με εκείνα που προκύπτουν από τη γενική μέθοδο ατομικών δεδομένων του le Cessie και συνεργατών (2008), δείξαμε ότι η μέθοδος είναι αρκετά ακριβής. Η ακρίβεια είναι ακόμα μεγαλύτερη αν λάβουμε υπόψη ότι δεν χρησιμοποιήσαμε τα πραγματικά δεδομένα αλλά εκτιμήσεις αυτών χρησιμοποιώντας ειδική μεθοδολογία. Πρέπει να τονιστεί ότι η συνδιακύμανση που υπολογίζεται από την εξίσωση (3.21) είναι η προσεγγιστική εκτίμηση της διασποράς λογαριθμισμένων ποσοτήτων των ασθενών. Έχειδειχθεί ότι αυτές οι προσεγγίσεις είναι αρκετά ικανοποιητικές ακόμα και για μικρά μεγέθη δείγματος (Agresti 1999) κάτι που δικαιολογεί την ευρεία χρήση τους. Θεωρητικά, αναμένεται η μέθοδος να οδηγεί σε μεροληπτικές εκτιμήσεις όταν το μέγεθος δείγματος είναι πολύ μικρό ή όταν υπάρχουν κελιά του πίνακα συνάφειας με λίγα ή καθόλου άτομα. Ωστόσο, στην πλειοψηφία πρακτικών εφαρμογών μετα-αναλύσεων συνοπτικών δεδομένων οι συγκεκριμένες ποσότητες που απαιτούνται για τον υπολογισμό των συνδιακυμάνσεων πρέπει να είναι εξ' αρχής αρκετά μεγάλες. Για παράδειγμα, όταν η συχνότητα των αλληλομόρφων κινδύνου είναι πολύ μικρή (<1%) ή πολύ μεγάλη (>99%) η εκτίμηση της συνδιακύμανσης είναι μεροληπτική. Ωστόσο, πρέπει να αναφερθεί ότι η μικρή μεροληψία που εμφανίστηκε στην ανάλυση που παρουσιάστηκε σε σχέση με εκείνη που προκύπτει από τη χρήση ατομικών δεδομένων, είναι γνωστή στη βιβλιογραφία και η ελαχιστοποίηση αυτής είναι δυνατή

με τη χρήση μιας μικρής σταθεράς (δηλαδή $c=1/2$) στις παρατηρούμενες ποσότητες (Anscombe 1956; Haldane 1956; Gart and Zweifel 1967).

Αξίζει να αναφερθούν κάποιοι περιορισμοί της μετα-ανάλυσης συνοπτικών δεδομένων. Αρχικά, όπως έχει ήδη αναφερθεί, στην προτεινόμενη μεθοδολογία υπολογίζονται συνολικές-συνδυασμένες εκτιμήσεις υποθέτοντας συνεπικρατές μοντέλο κληρονομικότητας. Αν και διάφορα μοντέλα κληρονομικότητας αναφέρονται συχνά σε μελέτες γενετικής συσχέτισης ασθενών-μαρτύρων (Bagos 2008), ο σχεδιασμός μελετών σε οικογένειες όπου χρησιμοποιούνται οι προσεγγίσεις TDT/HHRR επιτρέπει μόνο τη σύγκριση των αλληλομόρφων. Προσεγγίσεις που επιτρέπουν διαφορετικά μοντέλα κληρονομικότητας, αν και υπάρχουν, δεν χρησιμοποιούνται συχνά (Schaid and Sommer 1994). Ένα επιπλέον μειονέκτημα είναι το ότι η σύγκριση των αλληλομόρφων υποθέτει ότι ο πληθυσμός βρίσκεται σε ισορροπία Hardy-Weinberg. Η ισορροπία αυτή αποτελεί θέμα προς συζήτηση στις μελέτες γενετικής συσχέτισης (Salanti, Amountza et al. 2005; Trikalinos, Salanti et al. 2006; Minelli, Thompson et al. 2008) και μπορεί να διατυπωθεί ότι οι παρατηρούμενες αποκλίσεις μπορεί να οδηγήσουν σε μεροληπτικές εκτιμήσεις της διασποράς του λογαρίθμου του Odds Ratio από τη σύγκριση των αλληλομόρφων (Sato, Suganami et al. 2006; Trikalinos, Salanti et al. 2006).

Πρέπει να τονιστεί ωστόσο, ότι η προτεινόμενη μέθοδος είναι η μόνη ως τώρα διαθέσιμη μεθοδολογία που επιτρέπει το συνδυασμό των εκτιμήσεων των αποτελεσμάτων χρησιμοποιώντας συνοπτικά δεδομένα και έτσι αναμένεται να είναι εξαιρετικά χρήσιμη όταν δεν υπάρχουν ατομικά δεδομένα ή είναι δύσκολο να ανακτηθούν. Δείξαμε ότι ακόμα και όταν δεν έχουμε πρόσβαση σε πρωταρχικά δεδομένα μπορούμε να τα εκτιμήσουμε χρησιμοποιώντας μια απλή τεχνική (Di Pietrantonj 2006). Το μόνο πρόβλημα που μπορεί πιθανώς να προκύψει αποτελεί η περίπτωση όπου οι εκτιμήσεις των odds ratios που δίνονται στη βιβλιογραφία έχουν σταθμιστεί για πιθανούς συγχυτικούς παράγοντες, κυρίως όταν αυτοί συσχετίζονται τόσο με τη ασθένεια όσο και με τον πολυμορφισμό.

Δείξαμε επίσης μέσω εφαρμογής σε πραγματικά δεδομένα, ότι η συγκεκριμένη μεθοδολογία, η οποία παρέχει μια κλειστή μορφή για τον υπολογισμό της συνδιακύμανσης, μπορεί να χρησιμοποιηθεί και για την ανάλυση ευρυγονιδιωματικών μελετών όπου εφαρμόζεται επίσης ο σχεδιασμός TDT (Murphy, Weiss et al. 2008; Benyamin, Visscher et al. 2009; Murphy, S et al. 2010). Η μετα-ανάλυση των ευρυγονιδιωματικών μελετών αποτελεί μια συνήθη προσέγγιση για την

αύξηση της στατιστικής ισχύος και ανίχνευσης ασθενών επιδράσεων γονιδίων (Neale, Medland et al. 2010) και έτσι η μεθοδολογία που προτείνουμε θα μπορούσε να αποτελεί μια εναλλακτική προσέγγιση για το συνδυασμό αποτελεσμάτων σε οικογένειες και σε πληθυσμούς (Lasky-Su, Won et al. 2010) ή για την πραγματοποίηση μελετών επαλήθευσης χρησιμοποιώντας τους ίδιους ασθενείς (Murphy, Weiss et al. 2008; Murphy, S et al. 2010). Επιπλέον, μελέτες υποδεικνύουν ότι οι χρονοβόρες μέθοδοι μετα-ανάλυσης που χρησιμοποιούν ατομικά δεδομένα δεν οδηγούν απαραίτητα σε καλύτερες εκτιμήσεις (Lin and Zeng 2010), καθιστώντας έτσι τη συγκεκριμένη μέθοδο εξαιρετικά σημαντική για τη μείωση χρόνου που απαιτείται για τους υπολογισμούς.

Χρησιμοποιώντας τη μέθοδο που αναπτύξαμε, αναλύσαμε εκ νέου τα δεδομένα της συσχέτισης του αλληλομόρφου 10-Repeat Allele ενός πολυμορφισμού VNTR στην 3' αμετάφραστη περιοχή του γονιδίου του μεταφορέα της ντοπαμίνης με τη ΔΕΠ-Y (Yang, Chan et al. 2007). Η συνδυασμένη ανάλυση επέτρεψε την πραγματοποίηση πολλών διαγνωστικών ελέγχων για την ανίχνευση συστηματικού σφάλματος δημοσίευσης και τη πραγματοποίηση αθροιστικής μετα-ανάλυσης κάτι που δεν ήταν εφικτό έως τώρα. Έτσι, έγινε κατορθωτός ο υπολογισμός μιας συνολικής συνδυασμένης εκτίμησης που καταδεικνύει μια σημαντική συσχέτιση του υπό μελέτη πολυμορφισμού την ΔΕΠ-Y. Επιπλέον, δείξαμε την παρουσία του φαινομένου του Πρωτέα μιας και η πρώτη μελέτη είχε μεγαλύτερο Odds Ratio σε σχέση με τις επόμενες. Τέλος, η παρατηρούμενη ασυμμετρία στο funnel plot παρέχει ενδείξεις για συστηματικό σφάλμα δημοσίευσης.

Αν και η παρούσα ανάλυση δεν είχε σκοπό να πραγματοποιήσει μια ολοκληρωμένη μετα-ανάλυση για το συγκεκριμένο ερευνητικό ερώτημα αλλά μόνο να δείξει τη σημαντικότητα της μεθοδολογίας, αξίζει να συζητηθεί η διαφορά των εκτιμήσεων των μελετών γενετικής συσχέτισης σε οικογένειες (TDT και HHRR) και σε πληθυσμούς. Η ΔΕΠ-Y αποτελεί μια σύνθετη πολυπαραγοντική νόσο όπου πολλά γονίδια αλληλεπιδρούν τόσο μεταξύ τους (Coghill and Banaschewski 2009) όσο και με άλλους περιβαλλοντικούς παράγοντες (Max, Lansing et al. 2004). Η ΔΕΠ-Y αποτελεί έναν φαινότυπο που δεν μπορεί να περιγραφεί εύκολα κλινικά (Furman 2008). Συνεπώς, φαίνεται ότι οι παρατηρούμενες διαφορές στις εκτιμήσεις των αποτελεσμάτων των μελετών σε οικογένειες (TDT και HHRR) και σε πληθυσμούς θα μπορούσαν να εξηγηθούν από τους διαφορετικούς ορισμούς της νόσου και των υποτύπων αυτής. Επιπλέον, οι διαφορές στα αποτελέσματα των μελετών σε

οικογένειες (TDT και HHRR) και σε πληθυσμούς μπορούν να αποδοθούν εν μέρει στις ελλείπουσες τιμές που παρατηρούνται συχνά στις αναλύσεις TDT/HHRR. Σε μια μελέτη περιγράφεται ότι παιδιά με ΔΕΠ-Υ των οποίων τα δείγματα του πατέρα έλειπαν διέφεραν φαινοτυπικά από άλλα όπου καταγράφηκε ο γονότυπος και για τους δυο γονείς (West, Langley et al. 2002). Επιπλέον, σε μια ευρυγονιδιωματική μελέτη σε οικογένειες για τη ΔΕΠ-Υ καταγράφηκε μια συστηματική μεροληψία στη μεταβίβαση των επικρατών αλληλομόρφων. Οι συγγραφείς υποθέτουν ότι αυτό οφείλεται στα διαφορετικά ποσοστά γονοτύπων που λείπουν (Neale, Lasky-Su et al. 2008). Σε μια ευρυγονιδιωματική μελέτη σε πληθυσμούς, δεν βρέθηκε κάποια σημαντική συσχέτιση πολυμορφισμών του γονιδίου DAT1 (Neale, Medland et al. 2010), αλλά σε μια αντίστοιχη μελέτη σε οικογένειες ένας πολυμορφισμός του ίδιου γονιδίου (rs2937639) έδειξε σημαντική συσχέτιση ($p=0.004$) με την εμφάνιση της συγκεκριμένης νόσου (Lasky-Su, Anney et al. 2008). Είναι αρκετά ενδιαφέρον να αναφερθεί ότι μια μεγάλη πολυκεντρική μελέτη έδωσε άλλη μια πιθανή εξήγηση στηριζόμενη στη διαφορετική συσχέτιση του γονιδίου DAT1 με τη ΔΕΠ-Υ σε παιδιά και σε ενήλικες (Franke, Vasquez et al. 2010).

Συμπερασματικά, παρουσιάστηκε μια μέθοδος για την ενοποίηση των αποτελεσμάτων μελετών γενετικής συσχέτισης σε οικογένειες και σε πληθυσμούς που χρησιμοποιούν τους ίδιους ασθενείς. Η θεωρητική καινοτομία έγκειται στον αναλυτικό υπολογισμό της συνδιακύμανσης των συσχετισμένων εκτιμήσεων. Η υλοποίηση της μεθόδου είναι εύκολη ενώ έχει κατασκευαστεί ένα αντίστοιχο πρόγραμμα στο στατιστικό πακέτο Stata.

Μέρος Β: Μετα-ανάλυση πολυμορφισμών του γονιδίου FZD3 και η συσχέτισή τους με την εμφάνιση Σχιζοφρένειας

3.5. Εισαγωγή

Η σχιζοφρένεια είναι μια ψυχική νόσος που χαρακτηρίζεται από τρεις κύριες κατηγορίες συμπτωμάτων: θετικά (παράλογες ιδέες, ψευδαισθήσεις, κινητικές δυσκολίες), αρνητικά (διαταραχές της βούλησης, μείωση της ικανότητας για αυθόρμητο λόγο, κοινωνική απόσυρση, νευρογνωσιακού τύπου διαταραχές όπως διαταραχές μνήμης, προσοχής και εκτελεστικών λειτουργιών) και συναισθηματική αποδιοργάνωση που οδηγεί σε καταθλιπτικά ή μανικά (διπολικού τύπου) συμπτώματα. Αν και η σχιζοφρένεια βρίσκεται στο επίκεντρο του επιστημονικού ενδιαφέροντος για πάνω από έναν αιώνα, εξακολουθούν να υπάρχουν αναπάντητα ερωτήματα όπως από τι προκαλείται και αν αποτελεί νόσημα ή πρόκειται για μια ή πολλαπλές καταστάσεις (Bentall 2013). Είναι πλέον αποδεκτό, ότι πρόκειται για μια κληρονομική διαταραχή όπου αλληλεπιδρούν γενετικοί και περιβαλλοντικοί παράγοντες (Sullivan, Kendler et al. 2003; Lichtenstein, Yip et al. 2009).

Το γονίδιο Frizzled Class Receptor 3 (FZD3) χαρτογραφείται στην περιοχή του χρωμοσώματος 8p21, και μελέτες σύνδεσης δείχνουν συσχέτιση της περιοχής αυτής με την εμφάνιση σχιζοφρένειας (Pulver, Lasseter et al. 1995; Liu, Hwu et al. 2005). Το γονίδιο FZD3 αποτελείται από οκτώ εξώνια, είναι μήκους περίπου 70 kb και ανήκει στους συζευγμένους με G πρωτεΐνες υποδοχείς (GPCRs). Μεταφέρει τα σήματα των μορίων Wnt και διαδραματίζει ένα σημαντικό ρόλο στην διαμόρφωση πολλών δομών του νευρικού συστήματος συμπεριλαμβανομένης της νευρικής ακρολοφίας (Malbon 2004; Panaccione, Napoletano et al. 2013; Dijksterhuis, Petersen et al. 2014).

Αρχικά, ο Yang και συνεργάτες διερεύνησαν την πιθανή συσχέτιση πέντε πολυμορφισμών του γονιδίου FZD3 (rs2874940, rs2241802, rs181277, rs2323019 και rs352203) με την εμφάνιση σχιζοφρένειας με στατιστικά σημαντικά ευρήματα (Yang, Si et al. 2003). Παρόμοια είναι τα αποτελέσματα των Katsu και συνεργατών για τους πολυμορφισμούς rs2241802, rs375788 και rs960914 πληθυσμούς (Katsu, Ujike et al. 2003). Οι δυο αυτές ερευνητικές ομάδες χρησιμοποίησαν διαφορετικούς σχεδιασμούς, μελέτη ασθενών μαρτύρων σε οικογένειες (Yang, Si et al. 2003) και σε πληθυσμούς (Katsu, Ujike et al. 2003) σε Ασιάτες. Έκτοτε, οι διάφορες μελέτες που

πραγματοποιήθηκαν στον ίδιο ή σε διαφορετικούς πληθυσμούς οδήγησαν σε αντικρουόμενα συμπεράσματα (Hashimoto, Suzuki et al. 2005).

Στις ενότητες που ακολουθούν παρουσιάζονται τα αποτελέσματα της μετα-ανάλυσης που πραγματοποιήθηκε έτσι ώστε να συνοψιστεί η υπάρχουσα πληροφορία σχετικά με τη συσχέτιση πολυμορφισμών του γονιδίου FZD3 με τη σχιζοφρένεια συμπεριλαμβάνοντας μελέτες γενετικής συσχέτισης σε οικογένειες και σε πληθυσμούς. Για το σκοπό αυτό αξιοποιήθηκε η μεθοδολογία που αναλύθηκε στην Παράγραφο 3.2.3 η οποία επιτρέπει την από κοινού ανάλυση μελετών που έχουν διεξαχθεί σε οικογένειες και σε πληθυσμούς (Bagos, Dimou et al. 2011).

3.6. Μεθοδολογία

3.6.1. Πηγές δεδομένων, κριτήρια επιλεξιμότητας και εξαγωγή δεδομένων

Οι βιβλιογραφικές βάσεις δεδομένων Pubmed και Scopus χρησιμοποιήθηκαν για να ανακτηθούν μελέτες πολυμορφισμών του γονιδίου FZD3 και τη συσχέτισή τους με την εμφάνιση σχιζοφρένειας οι οποίες είχαν δημοσιευτεί μέχρι τον Φεβρουάριο του 2016. Χρησιμοποιήθηκαν λέξεις κλειδιά όπως “FZD3” ή “Frizzled-3” συνδυασμένες με ‘Schizophrenia’. Αφού πραγματοποιήθηκε ένας αρχικός έλεγχος του τίτλου και της περίληψης για να απορρίψουμε τις μη σχετικές με το θέμα εργασίες, προχωρήσαμε σε προσεκτική ανάγνωση των άρθρων που απέμειναν για να αποφασιστεί η καταλληλότητά τους καθώς και ο έλεγχος των βιβλιογραφικών αναφορών αυτών για επιπλέον μελέτες.

Στην μετα-ανάλυση συμπεριλάβαμε μελέτες που είχαν επαρκή δεδομένα για τον υπολογισμό ενός μέτρου σχετικού κινδύνου (Odds Ratio) των πολυμορφισμών του γονιδίου FZD3 με τη σχιζοφρένεια. Αξιολογήθηκαν μελέτες ασθενών-μαρτύρων σε οικογένειες και σε πληθυσμούς ενώ δεν τέθηκαν περιορισμοί γλώσσας. Για να αποφύγουμε το σφάλμα επιλογής μελετών δεν χρησιμοποιήσαμε κλίμακα αξιολόγησης των μελετών ως προς την ποιότητά τους (Stroup, Berlin et al. 2000). Από κάθε μελέτη καταχωρήθηκαν τα ακόλουθα: α) το όνομα του πρώτου συγγραφέα, το έτος δημοσίευσης, η καταγωγή του πληθυσμού και η χώρα, β) ο σχεδιασμός της μελέτης γ) ο αριθμός των family trios και ο αριθμός των ασθενών και υγιών μαζί με τα χαρακτηριστικά αυτών.

3.6.2. Στατιστική ανάλυση

Δεδομένου ότι οι μελέτες που ανακτήθηκαν είχαν υλοποιηθεί κάτω από δυο διαφορετικούς σχεδιασμούς (σε οικογένειες και σε πληθυσμούς) χρησιμοποιήθηκε η μεθοδολογία μετα-ανάλυσης δυο σταδίων όπου μοντελοποιούνται ταυτόχρονα οι εκτιμήσεις από τις δυο αυτές πηγές των δεδομένων (Janssen, Bont et al. 2007; Bagos, Dimou et al. 2011). Όλες οι μελέτες που είχαν διεξαχθεί σε οικογένειες χρησιμοποιούσαν τον έλεγχο TDT για την αξιολόγηση της συσχέτισης των υπό μελέτη πολυμορφισμών με την σχιζοφρένεια. Μια μόνο μελέτη παρείχε δεδομένα και για τους δυο σχεδιασμούς (Ide, Muratake et al. 2004). Διεξήγαμε μια κλασσική μετα-ανάλυση θέτοντας ως βάρη τις αντίστροφες ποσότητες των διασπορών (Normand 1999) χρησιμοποιώντας την μη επαναληπτική μέθοδο των ροπών που προτάθηκε από τους DerSimonian και Laird (DerSimonian and Laird 1986). Όπως έχει ήδη αναφερθεί, για τις μελέτες σε πληθυσμούς το Odds Ratio εκτιμήθηκε υπό τη θεώρηση συνεπικρατούς μοντέλου κληρονομικότητας καθώς μόνο αυτό παρουσιάζει αναλογία με τον έλεγχο TDT που χρησιμοποιείται σε μελέτες σε οικογένειες.

Το συστηματικό σφάλμα δημοσίευσης εκτιμήθηκε με τον έλεγχο των Begg και Mazumdar (Begg and Mazumdar 1994), τη μέθοδο παλινδρόμησης του Egger (Egger, Smith et al. 1997) και το ανάλογο αυτού χρησιμοποιώντας τυχαίες επιδράσεις (Thompson and Sharp 1999). Η πιθανή ύπαρξη διαχρονικής τάσης διερευνήθηκε χρησιμοποιώντας την κλασσική μέθοδο αθροιστικής μετα-ανάλυσης (Lau, Antman et al. 1992; Lau, Schmid et al. 1995) με δυο επίσημους στατιστικούς ελέγχους: τη σύγκριση της πρώτης έναντι των υπολοίπων μελετών ('First vs. subsequent') (Ioannidis and Trikalinos 2005) και τη μέθοδο παλινδρόμησης με γενικευμένα ελάχιστα τετράγωνα (Bagos and Nikolopoulos 2009).

Τέλος, στην περίπτωση για την οποία οι συνολικές εκτιμήσεις δεν ήταν στατιστικά σημαντικές, προβλέφθηκε ο αριθμός των επιπλέον ατόμων που απαιτούνται για την υπολογισμό στατιστικά σημαντικών εκτιμήσεων (Barrowman, Fang et al. 2003). Η πρόβλεψη λαμβάνει υπόψη τα αποτελέσματα της πρωταρχικής μετα-ανάλυσης χρησιμοποιώντας την ακόλουθη σχέση:

$$n = N \left(\frac{z_c^2}{z^2} - 1 \right) \quad (3.30)$$

όπου το n δηλώνει τον αριθμό των επιπλέον ατόμων που απαιτούνται για στατιστικά σημαντικές εκτιμήσεις, το N είναι ο συνολικός αριθμός των συμμετεχόντων στη

μετα-ανάλυση, το z είναι η τιμή του στατιστικού που υπολογίζεται από τη μετα-ανάλυση και το z_c είναι η κρίσιμη τιμή του z (για παράδειγμα $z_c = 1.96$, αντιστοιχεί σε έναν αμφίπλευρο έλεγχο σε επίπεδο σημαντικότητας 5%).

Η στατιστική ανάλυση πραγματοποιήθηκε χρησιμοποιώντας το στατιστικό πακέτο Stata 13 (Stata Corporation, College Station, Texas, USA), ενώ η στατιστική σημαντικότητα ορίστηκε σε επίπεδο σημαντικότητας $\alpha=0.05$.

3.7. Αποτελέσματα

Από τη βιβλιογραφική αναζήτηση στην Pubmed ανακτήθηκαν 15 μελέτες και επιπλέον 4 μελέτες από την Scopus. Αποκλείοντας τις μελέτες που δεν πληρούσαν τα κριτήρια επιλεξιμότητας, αξιοποιήθηκαν τελικά 9 μελέτες στην μετα-ανάλυση. Οι μελέτες περιείχαν δεδομένα για 366 family trios, 1,917 ασθενείς με σχιζοφρένεια και 2,192 υγιείς μάρτυρες. Όλες οι μελέτες αξιολογούσαν περισσότερους από έναν γονιδιακούς πολυμορφισμούς. Συμπεριλάβαμε έξι πολυμορφισμούς του γονιδίου FZD3, τον rs2241802 στο εξώνιο 5, τους rs2323019 και rs352203 στο εσώνιο 5, τους rs3757884, rs880481 στο εσώνιο 7 και τον rs960914 στο εσώνιο 3. Μια μελέτη είχε πληροφορίες και για τους δυο σχεδιασμούς (σε οικογένειες και σε πληθυσμούς) (Ide, Muratake et al. 2004), δυο μελέτες περιελάμβαναν μόνο family trios (Yang, Si et al. 2003; Wei and Hemmings 2004), ενώ οι υπόλοιπες αφορούσαν μελέτες ασθενών-μαρτύρων σε πληθυσμούς. Όλες οι μελέτες ήταν γραμμένες στα Αγγλικά. Η πλειοψηφία των μελετών περιλαμβάνει Ασιάτες ενώ μόνο δυο αφορούν Καυκάσιους (Wei and Hemmings 2004; Reif, Melchers et al. 2007). Τέσσερις μελέτες παρουσίαζαν δεδομένα απλοτύπων αλλά δεν μπορούσαν να αξιοποιηθούν για μετα-ανάλυση (Bagos 2011). Ο Πίνακας 3.4 παρουσιάζει τα χαρακτηριστικά κάθε μελέτης που χρησιμοποιήθηκε στη μετα-ανάλυση.

Στις ενότητες που ακολουθούν περιγράφονται τα αποτελέσματα της παρούσας μετα-ανάλυσης για κάθε πολυμορφισμό ξεχωριστά τα οποία συνοψίζονται στις Εικόνες 3.7-8 και στον Πίνακα 3.5.

Πίνακας 3.4. Χαρακτηριστικά των μελετών που περιλαμβάνονται στην παρούσα μετα-ανάλυση.

Μελέτη	Έτος	Χαρακτηριστικά ασθενών	Χαρακτηριστικά υγιών	Family trios	Καταγωγή	Πολυμορφισμός
Yang (Yang, Si et al. 2003)	2003			246 family trios (ασθενείς: 138 (56%) άντρες και 108 (44%) γυναίκες). Οι ασθενείς πληρούσαν τα κριτήρια ICD-10. Μέση διάρκεια νόσου: 5 έτη. Χρησιμοποιήθηκε ο έλεγχος transmission disequilibrium test	Chinese, Han	rs181277, rs2241802, rs2323019, rs2874940, rs352203
Katsu (Katsu, Ujike et al. 2003)	2003	200 ασθενείς (108 άντρες και 92 γυναίκες; μέση ηλικία 45.1 ± 12.5 έτη) που πληρούσαν τα κριτήρια της ICD-10, 91 είχαν διαγνωστεί με παρανοϊκού τύπου, 107 με ηβιφρενικού τύπου και 2 με κατατονικού τύπου σχιζοφρένεια	218 υγιείς (112 άντρες και 106 γυναίκες; μέση ηλικία 46.5 ± 15.7 years), κυρίως προερχόμενοι από το ιατρικό προσωπικό, άτομα με ιστορικό μείζονων ψυχιατρικών συμπτωμάτων αποκλείστηκαν		Japanese	rs2241802, rs3757888, rs960914
Ide (Ide, Muratake et al. 2004)	2004	540 ασθενείς (270 άντρες, 270 γυναίκες; μέση ηλικία, 45.6 ± 11.0 years). Η διάγνωση πραγματοποιήθηκε με βάση τα κριτήρια DSM-IV.	540 υγιείς χωρίς ιστορικό ψυχικών νοσημάτων (270 άντρες, 270 γυναίκες; μέση ηλικία, 45.1 ± 11.7 έτη)	212 οικογένειες με σχιζοφρένεια	Japanese	rs2241802, rs2323019, rs352203, rs352210, rs352222, rs352226, rs3757884, rs960914
Zhang (Zhang, Yu et al. 2004)	2004	236 ασθενείς με σχιζοφρένεια (130 άντρες και 106 γυναίκες; μέση ηλικία: 27±7.64 έτη) διαγνώστηκαν με βάση τα κριτήρια της ICD-10	275 υγιείς (158 άντρες και 117 γυναίκες; μέση ηλικία: 28.34±8.95 έτη) matched με τους ασθενείς για ηλικία, φύλο, εθνικότητα, και χωρίς ιστορικό ψυχικών νοσημάτων και νευρολογικών ασθενειών		Chinese, Han	rs2241802, rs2323019, rs352203, rs880481

Wei (Wei and Hemmings 2004)	2004			120 family trios (82 άντρες και 38 γυναίκες), ηλικίας 29.03 ± 6.95 έτη, οι οποίοι είχαν διαγνωστεί από τους ψυχιάτρους κατά την περίοδο 1990 – 2002, χρησιμοποιήθηκε ο έλεγχος transmission disequilibrium test	English, Welsh, Irish and Scottish	rs2241802, rs2323019, rs352203
Hashimoto (Hashimoto, Suzuki et al. 2005)	2005	427 ασθενείς (221 άντρες και 206 γυναίκες με μέση ηλικία 44.2 ± 14.5 έτη), 91 με διπολική διαταραχή και 396 με μείζονα κατάθλιψη. Η διάγνωση πραγματοποιήθηκε με βάση τα κριτήρια DSM-IV	473 υγιείς (228 και 245; 36.1 ± 12.5 έτη) χωρίς να έχει κατά το παρελθόν διαγνωστεί με κάποιο ψυχικό νόσημα		Japanese	rs2241802, rs2323019, rs352203, rs960914
Jeong (Jeong, Joo et al. 2006)	2006	241 ασθενείς (147 άντρες και 94 γυναίκες, μέση ηλικία= 32.3 ± 8.38) διαγνώστηκαν με σχιζοφρένεια από ψυχιάτρους	192 υγιείς (125 άντρες και 67 γυναίκες, μέση ηλικία= 22.6 ± 6.14) από το προσωπικό του νοσοκομείου και εθελοντές μαθητές. Τα άτομα με πρώτου βαθμού συγγενείς με ψυχικά νοσήματα αποκλείστηκαν		Korean	rs2241802, rs960914
Reif (Reif, Melchers et al. 2007)	2007	192 ασθενείς διαγνώστηκαν με βάση τα κριτήρια της ICD-10 και το σύστημα Leonhard	284 υγιείς δότες που προέρχονται από την ίδια περιοχή με τους ασθενείς		German	rs2241802, rs352203, rs960914

ChuanYuan	2011	81 ασθενείς (52% άντρες και 48% γυναίκες, μέση ηλικία =33.8±10.9 έτη) με βάση τα κριτήρια DSM-IV	210 υγιείς (107 ασθενείς και 103 γυναίκες, μέση ηλικία 34.10 ±10.33 years) matched για ηλικία και φύλο με τους ασθενείς	Chinese, Va	rs2241802, rs2323019, rs352203, rs3757888, rs880481
-----------	------	--------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------	-------------	-----------------------------------------------------

3.7.1. Ο πολυμορφισμός rs2241802

Ο πολυμορφισμός του γονιδίου rs2241802 εξετάστηκε για το σύνολο των 9 μελετών που περιλαμβάνονται στη μετα-ανάλυση. Δυο μελέτες είχαν πραγματοποιηθεί σε Ευρωπαίους ενώ οι υπόλοιπες σε Ασιάτες. Η σύγκριση του A έναντι του C αλληλομόρφου έδωσε ένα μη στατιστικά σημαντικό Odds Ratio (0.90, 95% C.I.: 0.79, 1.03). Εκτιμήθηκε μια μέτρια μεταξύ των μελετών μεταβλητότητα για τη σύγκριση των αλληλομόρφων ($I^2=57.3\%$, $p\text{-value}=0.016$) με στατιστικά σημαντικές διαφορές των εκτιμήσεων των μελετών σε οικογένειες και σε πληθυσμούς ($p\text{-value}=0.001$). Από την ανάλυση υποομάδας, στους Ευρωπαίους εκτιμήθηκε ένα Odds Ratio ίσο με 0.93 (95% C.I.: 0.74, 1.17) και απουσία ετερογένειας ($I^2=0.0\%$, $p\text{-value}=0.860$), ενώ για τους Ασιάτες το Odds Ratio ήταν 0.89 (95% CI: 0.76, 1.05) με στατιστικά σημαντική ετερογένεια ($I^2=67.9\%$, $p\text{-value}=0.005$). Όταν εστίασαμε την ανάλυση μόνο σε Κινέζους, το Odds Ratio παρέμεινε μη στατιστικά σημαντικό (OR: 0.81, 95% C.I.: 0.56, 1.17) αλλά με μεγαλύτερο I^2 ($I^2=79.6\%$, $p\text{-value}=0.007$).

Οι έλεγχοι του Begg και του Egger δεν έδειξαν την ύπαρξη συστηματικού σφάλματος δημοσίευσης. Ο οπτικός έλεγχος της αθροιστικής μετα-ανάλυσης παρείχε ενδείξεις διαχρονικής τάσης κάτι που επιβεβαιώθηκε από τους επίσημους στατιστικούς ελέγχους ($p\text{-value}<0.001$). Παρόμοια ήταν τα αποτελέσματα όταν αξιολογήθηκαν μόνο οι Ασιάτες. Ο εκτιμώμενος αριθμός των επιπρόσθετων ατόμων που απαιτούνται για να προκύψουν στατιστικά σημαντικές εκτιμήσεις ήταν αρκετά μεγάλος (3,228 άτομα) γεγονός που δηλώνει ότι είναι αρκετά δύσκολο να αντιστραφεί το αποτέλεσμα.

3.7.2. Ο πολυμορφισμός rs2323019

Η μετα-ανάλυση συμπεριέλαβε 6 μελέτες που αξιολογούσαν την πιθανή επίδραση του πολυμορφισμού rs2323019 στην εμφάνιση σχιζοφρένειας με ένα σύνολο 366 family trios (2 μελέτες σε οικογένειες), 1,284 ασθενείς με σχιζοφρένεια και 1,498 υγιείς (4 μελέτες ασθενών-μαρτύρων σε πληθυσμούς). Μια μόνο μελέτη (Wei and Hemmings 2004) περιελάμβανε Ευρωπαίους. Η μετα-ανάλυση για τη σύγκριση του αλληλομόρφου A έναντι του G κατέδειξε μια οριακά μη σημαντική συσχέτιση με Odds Ratio ίσο με 1.21 (95% C.I.: 0.98, 1.50). Η μεταξύ των μελετών ετερογένεια ήταν υψηλή ($I^2=75\%$, $p\text{-value}=0.001$), με εκτιμώμενη τιμή για τις μελέτες σε

οικογένειες ίση με 52.6% ($p=0.146$) ενώ για τις μελέτες σε πληθυσμούς ίση με 74.4% ($p\text{-value}=0.008$). Περιορίζοντας την ανάλυση σε Ασιάτες το Odds Ratio ήταν 1.24 (95% C.I.: 0.98, 1.56) με αρκετά μεγάλη ετερογένεια ($I^2=80\%$, $p\text{-value}=0.001$). Ωστόσο, αξίζει να αναφερθεί ότι, για τους Κινέζους το Odds Ratio ήταν 1.52 (95% C.I.: 1.29, 1.79, με απουσία ετερογένειας).

Οι στατιστικοί έλεγχοι δεν έδειξαν την ύπαρξη συστηματικού σφάλματος δημοσίευσης. Η ύπαρξη διαχρονικής τάσης ανιχνεύτηκε και από τους δυο επίσημους στατιστικούς ελέγχους ($p\text{-value}<0.001$). Η εφαρμογή της μεθοδολογίας που παρουσιάστηκε από τον Barrowman και συνεργάτες (Barrowman, Fang et al. 2003) έδειξε ότι απαιτούνται 772 επιπλέον άτομα για να υπολογιστούν στατιστικά σημαντικές επιδράσεις.

3.7.3. Ο πολυμορφισμός rs352203

Επτά μελέτες αξιολογούσαν τη συσχέτιση του πολυμορφισμού rs352203 με την σχιζοφρένεια περιλαμβάνοντας 366 family trios (2 μελέτες), 1476 σχιζοφρενείς και 1782 υγιείς (5 μελέτες). Μεταξύ αυτών, μόνο μια περιελάμβανε Ευρωπαίους (Reif, Melchers et al. 2007). Από τη σύγκριση του αλληλομόρφου C έναντι του T εκτιμήθηκε ένα Odds Ratio ίσο με 0.82 (95% C.I.: 0.69, 0.98) υποδηλώνοντας ένα προστατευτικό ρόλο του υπό μελέτη πολυμορφισμού με τη νόσο. Παρατηρήθηκε μέτρια ετερογένεια ($I^2=67.7\%$, $p\text{-value}=0.005$). Η ανάλυση υποομάδας σε Ασιάτες οδήγησε στην εκτίμηση μιας οριακά στατιστικά σημαντικής συσχέτισης με Odds Ratio ίσο με 0.80 (95% C.I.: 0.64, 1.00), αλλά αρκετά αυξημένη ετερογένεια ($I^2=77.2\%$, $p\text{-value}=0.002$). Ωστόσο, οι Κινέζοι παρουσίασαν μια ξεκάθαρη συσχέτιση του αλληλομόρφου C με τη σχιζοφρένεια (OR: 0.69, 95% C.I.: 0.51, 0.92) με ετερογένεια ίση με 66.7% ($p\text{-value}=0.05$).

Οι έλεγχοι του Begg και του Egger δεν έδειξαν την ύπαρξη συστηματικού σφάλματος δημοσίευσης. Με βάση τα ευρήματα της αθροιστικής μετα-ανάλυσης, το Odds Ratio της πρώτης μελέτης (Yang, Si et al. 2003) διαφέρει σημαντικά από τις υπόλοιπες μελέτες.

3.7.4. Ο πολυμορφισμός rs3757888

Ο πολυμορφισμός rs3757888 εξετάστηκε σε δυο μόνο μελέτες (Katsu, Ujike et al. 2003; ChuanYuan, Li et al. 2011) με μελέτες μόνο ασθενών-μαρτύρων σε Ασιάτες. Συνολικά, η μετα-ανάλυση περιελάμβανε 709 άτομα, με 281 σχιζοφρενείς και 428 υγιείς. Το Odds Ratio της σύγκρισης του αλληλομόρφου G έναντι του A ήταν μη στατιστικά σημαντικό (OR: 1.08, 95% C.I.: 0.63, 1.84). Η μεταξύ των μελετών ετερογένεια ήταν 54.4% (p -value=0.138). Οι στατιστικοί έλεγχοι για το συστηματικό σφάλμα δημοσίευσης και για διαχρονική τάση δεν πραγματοποιήθηκαν δεδομένου του μικρού αριθμού μελετών. Τέλος, χρησιμοποιώντας τη μέθοδο του Barrowman και συνεργατών απαιτείται ένας εξαιρετικά μεγάλος αριθμός ατόμων για να επιτύχουμε στατιστικά σημαντικά αποτελέσματα (34,032 άτομα).

3.7.5. Ο πολυμορφισμός rs880481

Η επίδραση του πολυμορφισμού rs880481 στην σχιζοφρένεια περιελάμβανε 2 μελέτες ασθενών-μαρτύρων σε πληθυσμούς, μόνο ασιατικής καταγωγής με 317 σχιζοφρενείς και 485 υγιείς. Από τη μετα-ανάλυση εκτιμήθηκε ένα μη σημαντικό Odds Ratio (OR: 1.22, 95% C.I.: 0.78, 1.91) για τη σύγκριση του αλληλομόρφου A έναντι του G με I^2 ίσο με 71.6% (p -value=0.041). Παρόμοια με τον πολυμορφισμό rs3757888 και εδώ δεν πραγματοποιήσαμε στατιστικούς ελέγχους για συστηματικό σφάλμα δημοσίευσης και διαχρονική τάση. Ο αριθμός των επιπλέον ατόμων για να υπολογιστούν στατιστικά σημαντικές επιδράσεις είναι 3,364.

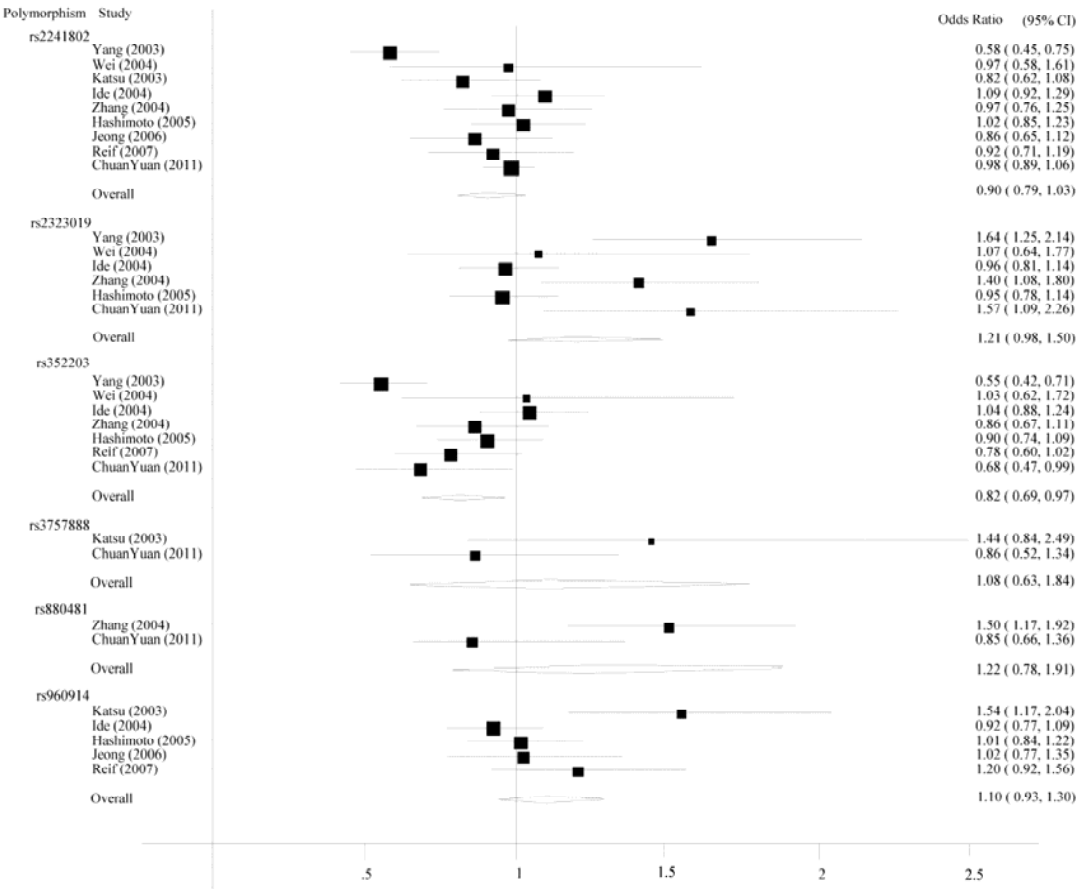
3.7.6. Ο πολυμορφισμός rs960914

Ένα σύνολο 5 μελετών εξέταζαν τη συσχέτιση του πολυμορφισμού rs960914 με τον κίνδυνο εμφάνισης σχιζοφρένειας, από τις οποίες οι τέσσερις αφορούσαν Ασιάτες και μια Ευρωπαίους (Reif, Melchers et al. 2007). Συνολικά, στρατολογήθηκαν 1600 σχιζοφρενείς και 1707 υγιείς. Δεν υπήρχε κάποια μελέτη που να περιλάμβανε σχεδιασμό σε οικογένειες. Η σύγκριση του αλληλομόρφου C έναντι του T οδήγησε σε ένα μη σημαντικό Odds Ratio (OR: 1.10, 95% C.I.: 0.93, 1.30) ενώ η μεταξύ των μελετών ετερογένεια ήταν αρκετά μεγάλη (I^2 =62.9%, p -value=0.029). Ένα μη σημαντικό Odds Ratio υπολογίστηκε επίσης στην ανάλυση που περιλάμβανε αποκλειστικά Ασιάτες (OR: 1.08, 95% C.I.: 0.88, 1.32) με σημαντική ετερογένεια (I^2 =69.2%, p -value=0.021).

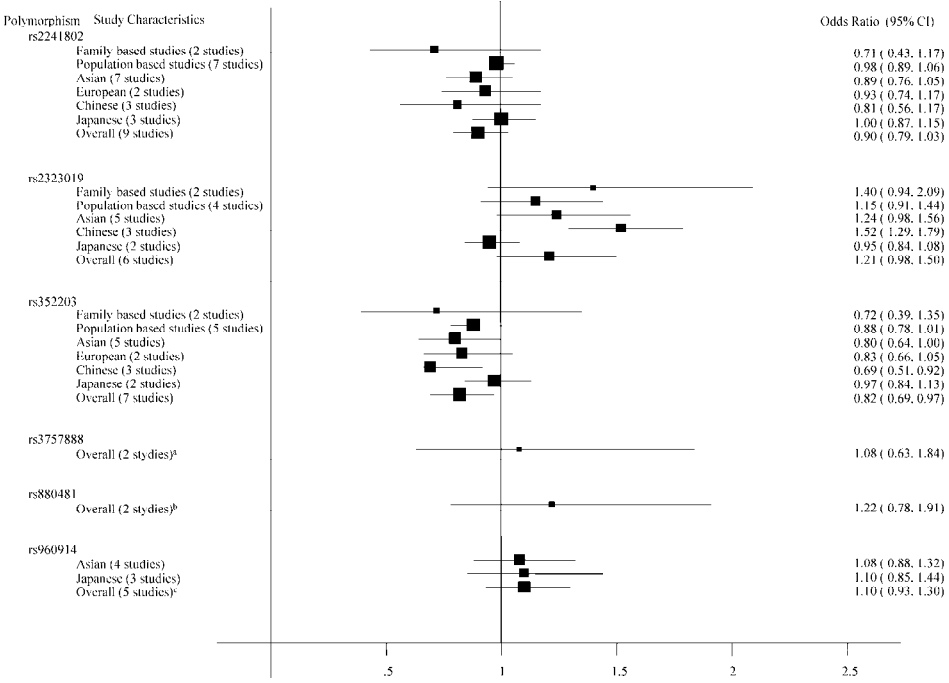
Οι έλεγχοι του Begg και του Egger δεν έδειξαν την ύπαρξη συστηματικού σφάλματος δημοσίευσης. Η αθροιστική μετα-ανάλυση έδειξε πιθανή ύπαρξη του φαινομένου του Πρωτέα και για τους δυο ελέγχους (σύγκριση της πρώτης έναντι των υπολοίπων μελετών και τη μέθοδο παλινδρόμησης με γενικευμένα ελάχιστα τετράγωνα) που χρησιμοποιήθηκαν (p -value=0.002 και p -value=0.001, αντίστοιχα). Η μέθοδος των Barrowman και συνεργατών (Barrowman, Fang et al. 2003) έδειξε ότι απαιτούνται 7585 άτομα για στατιστικά σημαντικές εκτιμήσεις.

3.7.7. Ανισορροπία σύνδεσης

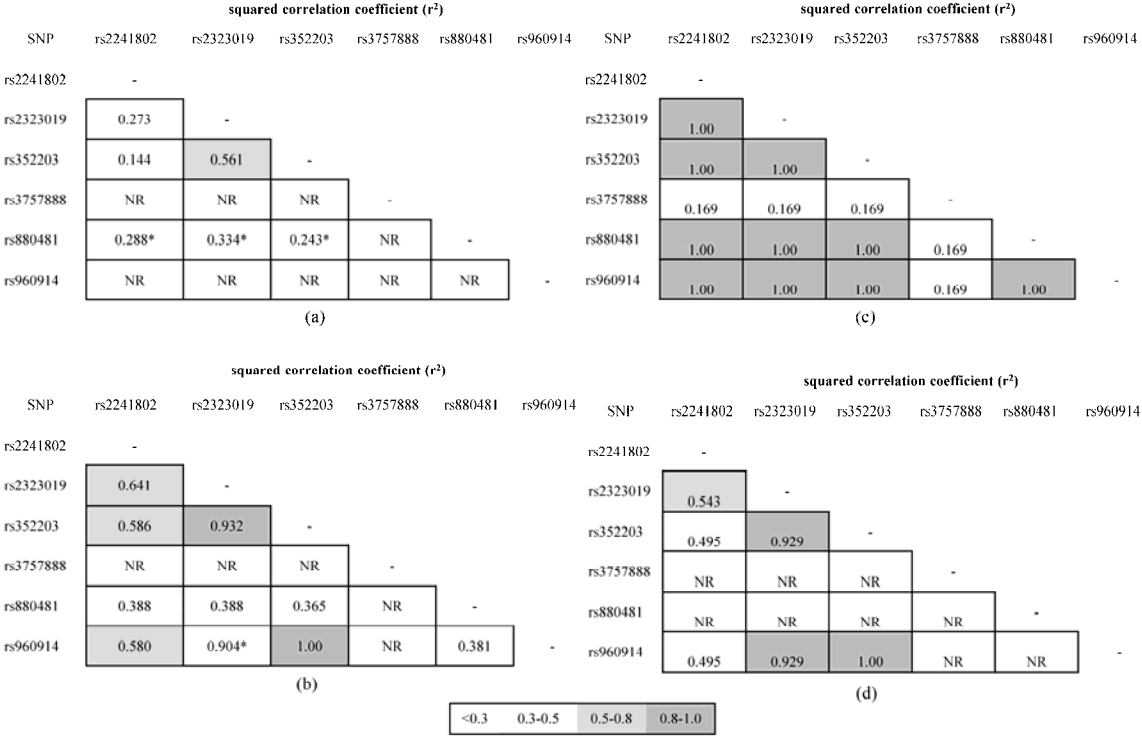
Δεδομένα για το βαθμό της ανισορροπίας σύνδεσης μεταξύ των υπό μελέτη πολυμορφισμών αντλήθηκαν από τη βιβλιογραφία για τους Κινέζους και Ιάπωνες ενώ για τους Ευρωπαίους και Ασιάτες, χρησιμοποιήθηκε η SNAP (<https://www.broadinstitute.org/mpg/snap/ldsearch.php>) (Εικόνα 3.9). Η ανισορροπία σύνδεσης ήταν τέλεια ($r^2=1.00$) μεταξύ των ζευγαριών rs2241802, rs2323019, rs352203, rs880481 και rs960914 σε Ευρωπαίους. Για τους Ασιάτες, εκτιμήθηκε ένας μέτριος βαθμός ανισορροπίας σύνδεσης (με μέσο $r^2 \approx 0.5$) μεταξύ των rs2241802 - rs2323019 και rs352203 - rs960914, ενώ μεγαλύτερος για τα ζεύγη rs2323019 - rs960914 και rs352203 - rs960914. Ο βαθμός ανισορροπίας σύνδεσης ήταν χαμηλότερος στους Κινέζους με το εκτιμώμενο r^2 μικρότερο από 0.3.



Εικόνα 3.7. Γραφική απεικόνιση των αποτελεσμάτων της μετα-ανάλυσης για τη συσχέτιση των πολυμορφισμών rs2241802, rs2323019, rs352203, rs3757888, rs880481 και rs960914 του γονιδίου FZD3 με την εμφάνιση σχιζοφρένειας.



Εικόνα 3.8. Γραφική απεικόνιση των αποτελεσμάτων των μετα-αναλύσεων σε υποομάδες για τη συσχέτιση των πολυμορφισμών rs2241802, rs2323019, rs352203, rs3757888, rs880481 και rs960914 του γονιδίου FZD3 με την εμφάνιση σχιζοφρένειας.^a και οι δυο μελέτες ήταν μελέτες ασθενών-μαρτύρων σε πληθυσμούς και περιελάμβαναν Ασιάτες; ^b και οι δυο μελέτες ήταν μελέτες ασθενών-μαρτύρων σε πληθυσμούς και περιελάμβαναν Κινέζους; ^c όλες οι μελέτες ήταν μελέτες ασθενών-μαρτύρων σε πληθυσμούς



Εικόνα 3.9. Εκτιμήσεις του βαθμού ανισορροπίας σύνδεσης μεταξύ των πολυμορφισμών του γονιδίου FZD3 για (a) Κινέζους (b) Ιάπωνες (c) Ευρωπαίους και (d) Ασιάτες. NR: Not Reported.

Πίνακας 3.5. Τα αποτελέσματα της μετα-ανάλυσης των μελετών που εξέταζαν τη συσχέτιση μεταξύ πολυμορφισμών του γονιδίου FZD3 με την εμφάνιση σχιζοφρένειας.

SNP	Αριθμός μελετών	Αριθμός των family trios / ασθενών/υγιών	Odds ratio (p-value)	95% Confidence Interval	Begg/Egger bias test /Random effect regression (p-values)	Meta-cumulative 'First vs. Subsequent' / GLS (p-values)
rs2241802	9	366/1917/2192	0.90 (0.125)	0.79, 1.03	0.175/0.350/0.667	<0.001/0.001
rs2323019	6	366/1284/1498	1.215 (0.073)	0.98, 1.50	0.260/0.207/0.394	0.032/<0.001
rs352203	7	366/1476/1782	0.819 (0.024)	0.69, 0.98	0.230/0.352/0.726	<0.001/0.001
rs3757888	2	0/281/428	1.08 (0.783)	0.63, 1.84	-	-
rs880481	2	0/317/485	1.217 (0.390)	0.78, 1.91	-	-
rs960914	5	0/1600/1707	1.10 (0.281)	0.93, 1.30	0.221/0.141/0.014	0.002/0.001

3.8. Συζήτηση

Η πρωτεΐνη που κωδικοποιείται από το γονίδιο FZD3 αποτελεί έναν υποδοχέα των εκκρινόμενων Wnt γλυκοπρωτεϊνών. Η τροποποίηση της μεταγωγής σήματος των πρωτεϊνών αυτών έχει συσχετιστεί με δομικές κυτταρικές αλλοιώσεις που παρατηρούνται στη σχιζοφρένεια. Το γονίδιο FZD3 χαρτογραφείται στο χρωμόσωμα 8p21, μια περιοχή που έχει συνδεθεί σε παλαιότερες εργασίες με την εμφάνιση σχιζοφρένειας (Pulver, Lasseter et al. 1995; Liu, Hwu et al. 2005).

Πολλοί ερευνητές έχουν εξετάσει την πιθανή συσχέτιση πολυμορφισμών του γονιδίου FZD3 με την εμφάνιση σχιζοφρένειας, με αντικρουόμενα όμως αποτελέσματα. Αρχικά, ο Yang και συνεργάτες (Yang, Si et al. 2003) ανέφεραν συσχέτιση των πολυμορφισμών του γονιδίου FZD3 σε Κινέζους πυροδοτώντας έτσι την επερχόμενη έρευνα. Μελέτες που αφορούσαν την ίδια φυλή έδειξαν συσχετίσεις των πολυμορφισμών rs2241802 (Yang, Si et al. 2003; ChuanYuan, Li et al. 2011), rs2323019 (Yang, Si et al. 2003; Zhang, Yu et al. 2004; ChuanYuan, Li et al. 2011), rs352203 (Yang, Si et al. 2003; ChuanYuan, Li et al. 2011) και rs880481 (Zhang, Yu et al. 2004) με την υπό μελέτη νόσο. Επιπλέον, βρέθηκε μια θετική συσχέτιση του πολυμορφισμού rs960914 με τη σχιζοφρένεια σε Ιάπωνες (Katsu, Ujike et al. 2003).

Η παρούσα μετα-ανάλυση, αν και απέτυχε να δείξει στατιστικά σημαντική συσχέτιση των πολυμορφισμών rs2241802, rs3757884, rs880481 και rs960914 με τη σχιζοφρένεια, υπογράμμισε τον ρόλο του πολυμορφισμού rs352203. Πιο συγκεκριμένα η παρουσία του αλληλομόρφου C έχει προστατευτικό ρόλο για την εμφάνιση της νόσου. Εκτιμήθηκε επίσης ένας αυξημένος κίνδυνος για τους φορείς του αλληλομόρφου A του πολυμορφισμού rs2323019 στους Κινέζους, κάτι που μπορεί να αποδοθεί στο διαφορετικό βαθμό ανισορροπίας σύνδεσης που εκτιμήθηκε για τη φυλή αυτή.

Οι περιορισμοί της παρούσας μετα-ανάλυσης εστιάζονται κυρίως στον μικρό αριθμό μελετών με την πλειοψηφία αυτών να περιλαμβάνουν Ασιάτες. Τα αποτελέσματά μας είναι προς την ίδια κατεύθυνση με εκείνα του Jeong και συνεργατών (Jeong, Joo et al. 2006) αν και μόνο δυο πολυμορφισμοί (rs2241802 και rs960914) είχαν μελετηθεί στην εργασία αυτή. Ωστόσο, εμείς επικαιροποιήσαμε την ανάλυση συμπεριλαμβάνοντας μεγαλύτερο αριθμό μελετών, υπολογίζοντας παράλληλα πιο ακριβείς εκτιμήσεις εφαρμόζοντας σύγχρονη μεθοδολογία που

συνυπολογίζει τα αποτελέσματα από μελέτες ασθενών μαρτύρων σε πληθυσμούς και σε οικογένειες. Τέλος, υπολογίσαμε τα μέτρα ανισορροπίας σύνδεσης αιτιολογώντας την παρατηρούμενη ετερογένεια μεταξύ των πληθυσμών καθώς και τα επιπρόσθετα άτομα που θα οδηγούσαν σε στατιστικά σημαντικές εκτιμήσεις.

Κεφάλαιο 4: Ανάπτυξη μεθοδολογίας μετα-ανάλυσης πολλαπλών εκβάσεων σε μελέτες γενετικής συσχέτισης και εφαρμογές

Μέρος Α: Πολυμεταβλητή μέθοδος μετα-ανάλυσης πολλαπλών εκβάσεων σε μελέτες γενετικής συσχέτισης

4.1. Εισαγωγή

Υπάρχουν αρκετές μελέτες γενετικής συσχέτισης που μελετούν ταυτόχρονα περισσότερες της μιας έκβασης οι οποίες μπορεί να είναι συσχετισμένες. Κάποιες άλλες μελέτες αναφέρουν αποτελέσματα για περισσότερους του ενός πολυμορφισμούς, του ίδιου ή γειτονικών γονιδίων που μπορεί να βρίσκονται σε ανισορροπία σύνδεσης. Σε αυτές τις περιπτώσεις, μπορεί να πραγματοποιηθεί μια πολυμεταβλητή μετα-ανάλυση τυχαίων επιδράσεων όπου λαμβάνονται υπόψη οι συσχετίσεις μεταξύ των εκβάσεων (van Houwelingen, Arends et al. 2002), πραγματοποιούνται επίσημοι στατιστικοί έλεγχοι και οι συνδυασμένες εκτιμήσεις έχουν καλύτερες στατιστικές ιδιότητες σε σχέση με τις κλασσικές μονομεταβλητές προσεγγίσεις (μικρότερο τυπικό σφάλμα και μέσο τετραγωνικό σφάλμα) υπό την υπόθεση ότι οι ελλείπουσες τιμές στις εκβάσεις οφείλονται στην τύχη (Riley, Abrams et al. 2007; Riley, Abrams et al. 2007). Ένα πρόβλημα που ανακύπτει στις πολυμεταβλητές μεθοδολογίες είναι ότι χρειαζόμαστε μια εκτίμηση της εντός των μελετών μεταβλητότητας των πολλαπλών εκβάσεων ή των πολλαπλών γενετικών τόπων η οποία σπανίως αναφέρεται στις δημοσιευμένες εργασίες, ενώ ο μεταξύ των μελετών πίνακας διακυμάνσεων-συνδιακυμάνσεων εκτιμάται από τα δεδομένα (Riley, Abrams et al. 2007).

Για διάφορες περιπτώσεις, ο υπολογισμός της εντός των μελετών μεταβλητότητας είναι εύκολος, όπως στην περίπτωση των πολλαπλών θεραπειών σε κλινικές μελέτες (Higgins and Whitehead 1996), ή στην περίπτωση των αμοιβαίως αποκλειόμενων εκβάσεων (Trikalinos and Olkin 2008). Για πιο πολύπλοκες περιπτώσεις μελετών με επικαλυπτόμενα άτομα (Lin and Sullivan 2009) ή μελετών γενετικής συσχέτισης με περισσότερους από έναν πολυμορφισμούς (Bagos and Liakopoulos 2010), έχουν προταθεί διάφορες επεκτάσεις. Στη γενική περίπτωση

βέβαια, η εντός των μελετών μεταβλητότητα μπορεί να υπολογιστεί εάν για όλες τις μελέτες υπάρχουν διαθέσιμα ατομικά δεδομένα (Riley, Lambert et al. 2008). Εναλλακτικά, η εντός των μελετών μεταβλητότητα μπορεί να εκτιμηθεί χρησιμοποιώντας κάποια βιολογικά λογική τιμή ή μια προσέγγιση με τον συντελεστή συσχέτισης του Pearson (Kirkham, Riley et al. 2012). Επιπλέον, έχει προταθεί ένα εναλλακτικό μοντέλο διμεταβλητής μετα-ανάλυσης τυχαίων επιδράσεων, όπου χρησιμοποιείται μια μόνο παράμετρος για την από κοινού μοντελοποίηση της εντός και μεταξύ των μελετών μεταβλητότητας (Riley, Thompson et al. 2008) το οποίο επεκτάθηκε για την περίπτωση πολλαπλών θεραπειών στη μετα-ανάλυση δικτύων (Efthimiou, Mavridis et al. 2014).

Πρόσφατα, προτάθηκε μια γενική μέθοδος υπολογισμού της συνδιακύμανσης δυο συσχετισμένων λογαρίθμων των Odds Ratios, η οποία περιλαμβάνει όλες τις μεθόδους που έχουν ήδη αναφερθεί σαν ειδικές περιπτώσεις, και χρησιμοποιεί μόνο τις παρατηρηθείσες ποσότητες από τους πίνακες συνάφειας (Bagos 2012). Μια διαφορετική αλλά αρκετά συναφής μέθοδος έχει παρουσιαστεί τόσο για διακριτές όσο και για συνεχείς εκβάσεις (Wei and Higgins 2013), όπου απαιτείται η γνώση του συντελεστή συσχέτισης μεταξύ των δυο εκβάσεων. Τέλος, παρουσιάστηκε μια σχέση για τον υπολογισμό της εντός των μελετών μεταβλητότητας χρησιμοποιώντας ένα σύνολο δεσμευμένων πιθανοτήτων που εκμαιεύονται από κλινικούς (Efthimiou, Mavridis et al. 2013) και επεκτάθηκε για διχοτομικές και συνεχείς εκβάσεις καθώς επίσης και για δεδομένα επιβίωσης αλλά οι δυο αυτές προσεγγίσεις προσανατολίζονται κυρίως στη μελέτη δικτύων παρεμβάσεων (Efthimiou, Mavridis et al. 2014).

Η προτεινόμενη μεθοδολογία που θα περιγραφεί στη συνέχεια είναι απλή, αλλά παρουσιάζει αυξημένη ισχύ για την πραγματοποίηση πολυμεταβλητής μετα-ανάλυσης σε μελέτες γενετικής συσχέτισης. Αποτελεί μια άμεση επέκταση της διμεταβλητής μεθόδου μετα-ανάλυσης μελετών γενετικής συσχέτισης (van Houwelingen, Arends et al. 2002) και λαμβάνει υπόψη τόσο συνεχείς όσο και διακριτές εκβάσεις. Παρέχονται σχέσεις για τον αναλυτικό υπολογισμό της εντός των μελετών μεταβλητότητας χρησιμοποιώντας την ταξινόμηση των ατόμων για κάθε έκβαση ανάλογα με το γονότυπο (Bagos 2012). Για την περίπτωση που αυτή δεν είναι γνωστή, προτείνουμε μια απλή μέθοδο εκτίμησης χρησιμοποιώντας ένα log-linear μοντέλο και τον iterative proportional fitting algorithm (IPF algorithm) υπό τη θεώρηση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης. Εφόσον το μοντέλο στηρίζεται

στο γενικό μοντέλο πολυμεταβλητής μετα-ανάλυσης (van Houwelingen, Arends et al. 2002; Mavridis and Salanti 2012), έχει κάποια σημαντικά χαρακτηριστικά: μπορεί εύκολα να συμπεριλάβει μελέτες που έχουν εκτιμήσεις για μόνο μια από τις εκβάσεις υπό τη θεώρηση ότι οι ελλείπουσες αυτές παρατηρήσεις οφείλονται στην τύχη, δανειζόμενη ισχύ από εξωτερικές μελέτες (Higgins and Whitehead 1996); μπορεί εύκολα να επεκταθεί για περισσότερες από δυο εκβάσεις; η εκτίμηση των παραμέτρων αυτού μπορεί να γίνει χρησιμοποιώντας ένα οποιοδήποτε στατιστικό πακέτο που χρησιμοποιείται για πολυμεταβλητή μετα-ανάλυση (Jackson, Riley et al. 2011; Mavridis and Salanti 2012); και τέλος επιτρέπει την απευθείας σύγκριση των εκτιμώμενων παραμέτρων με επίσημους στατιστικούς ελέγχους.

4.2. Μεθοδολογία

4.2.1. Το γενικό πολυμεταβλητό μοντέλο μετα-ανάλυσης πολλαπλών διακριτών εκβάσεων

Όπως έχει ήδη παρουσιαστεί στην εισαγωγή, η πλειοψηφία των μετα-αναλύσεων μελετών γενετικής συσχέτισης για έναν γενετικό τόπο με δυο αλληλόμορφα (A,B), πραγματοποιείται συνενώνοντας τους γονοτύπους σε δυο κατηγορίες, υποθέτοντας κάποιο μοντέλο κληρονομικότητας, και κάνοντας συγκρίσεις υπολογίζοντας το Odds Ratio. Έτσι, εάν θεωρήσουμε τη σύγκριση των γονοτύπων BB+AB έναντι του AA, υποθέτουμε επικρατές μοντέλο κληρονομικότητας, BB έναντι των AB+AA υπολειπόμενο και αν συγκρίνουμε τις συχνότητες των αλληλομόρφων μεταξύ ασθενών και υγιών υπονοούμε συνεπικρατές μοντέλο. Η Minelli και συνεργάτες (Minelli, Thompson et al. 2005) πρότειναν την από κοινού μοντελοποίηση του λογαρίθμου των Odds Ratios για τη σύγκριση του γονοτύπου AB έναντι του AA και BB έναντι του AA (Bagos 2008).

Συνήθως, τα δεδομένα μιας μελέτης γενετικής συσχέτισης με δυο εκβάσεις (Y_{1i} , Y_{2i}) απεικονίζονται με τη μορφή του Πίνακα 4.1, όπου με την ορολογία των πινάκων συνάφειας είναι ένας περιθώριος πίνακας αφού για κάθε έκβαση η κατανομή των γονοτύπων της δεύτερης έκβασης δεν λαμβάνεται υπόψη, ενώ αν αγνοήσουμε την κατανομή των γονοτύπων τα δεδομένα απεικονίζονται με τη μορφή του Πίνακα 4.2. Ο Πίνακας 4.3 περιγράφει την από κοινού ταξινόμηση των ατόμων για κάθε έκβαση ανάλογα με το γονότυπο σε κάθε i μελέτη ($i=1,2,...,k$). Ο Trikalinos και

Olkin έχουν προτείνει μια προσεγγιστική μέθοδο για τον υπολογισμό των συνδιακυμάνσεων δυο αμοιβαίως αποκλειόμενων διακριτών εκβάσεων (Trikalinos and Olkin 2008). Εδώ περιγράφουμε την γενική περίπτωση δυο εκβάσεων που δεν αποκλείουν η μια την άλλη.

Πίνακας 4.1. Οι δυο περιθώριοι 2X3 πίνακες συνάφειας που χρησιμοποιούνται για την ταξινόμηση των ατόμων για κάθε έκβαση (Y_{1i} , και Y_{2i}) ανάλογα με το γονότυπό τους για κάθε i μελέτη ($i=1,2,\dots,k$).

Genotype	Y_{1i}		Y_{2i}	
	$Y_{1i}=1$	$Y_{1i}=0$	$Y_{2i}=1$	$Y_{2i}=0$
AA	nAA_{1+i}	nAA_{0+i}	nAA_{+1i}	nAA_{+0i}
AB	nAB_{1+i}	nAB_{0+i}	nAB_{+1i}	nAB_{+0i}
BB	nBB_{1+i}	nBB_{0+i}	nBB_{+1i}	nBB_{+0i}

Πίνακας 4.2. Η ταξινόμηση των δυο εκβάσεων (Y_{1i} , Y_{2i}) για κάθε i μελέτη ($i=1,2,\dots,k$). Με την ορολογία πινάκων συνάφειας, πρόκειται για έναν περιθώριο πίνακα. Βλέπε επίσης (Agresti 2002; Bagos 2012).

		Y_{2i}	
		$Y_{2i}=1$	$Y_{2i}=0$
Y_{1i}	$Y_{1i}=1$	n_{+11i}	n_{+10i}
	$Y_{1i}=0$	n_{+01i}	n_{+00i}

Πίνακας 4.3. Η από κοινού ταξινόμηση των ατόμων για τις δυο εκβάσεις (Y_{1i} , Y_{2i}) ανάλογα με το γονότυπό τους για κάθε i μελέτη ($i=1,2,\dots,k$).

		$Y_{1i}=1$			$Y_{1i}=0$		
		AA	AB	BB	AA	AB	BB
Y_{2i}	$Y_{2i}=1$	nAA_{11i}	nAB_{11i}	nBB_{11i}	nAA_{01i}	nAB_{01i}	nBB_{01i}
	$Y_{2i}=0$	nAA_{10i}	nAB_{10i}	nBB_{10i}	nAA_{00i}	nAB_{00i}	nBB_{00i}

Επεκτείνοντας τη γενική προσέγγιση της διμεταβλητής μετα-ανάλυσης για την περίπτωση δυο εκβάσεων (van Houwelingen, Arends et al. 2002), προκύπτουν τέσσερις λογάριθμοι των Odds Ratios από τη σύγκριση των γονοτύπων που φέρουν το αλληλόμορφο κινδύνου (AB, BB) έναντι του γονοτύπου με το φυσιολογικό αλληλόμορφο (AA). Έτσι, για κάθε έκβαση θα έχουμε δυο λογαρίθμους των Odds Ratios (AB έναντι AA και BB έναντι AA) που μπορούν να μοντελοποιηθούν ταυτόχρονα κάτω από ένα πολυμεταβλητό πλαίσιο. Ξεκινάμε την περιγραφή με την προσέγγιση όπου δεν υποθέτουμε εκ των προτέρων κάποιο γενετικό μοντέλο (genetic model-free approach) και συνεχίζουμε εξετάζοντας τα διάφορα μοντέλα κληρονομικότητας (υπολειπόμενο, συνεπικρατές, επικρατές). Οι λογάριθμοι των Odds Ratios (AB έναντι AA και BB έναντι AA) για τις δυο εκβάσεις με τις αντίστοιχες διασπορές αυτών, για κάθε i μελέτη ($i=1,2,\dots,k$) δίνονται από τη γνωστή σχέση του Woolf:

$$y_{1i} = \log \left(\frac{(nAB_{11i} + nAB_{10i})(nAA_{00i} + nAA_{01i})}{(nAA_{11i} + nAA_{10i})(nAB_{01i} + nAB_{00i})} \right) \quad (4.1)$$

$$y_{2i} = \log \left(\frac{(nBB_{11i} + nBB_{10i})(nAA_{00i} + nAA_{01i})}{(nAA_{11i} + nAA_{10i})(nBB_{01i} + nBB_{00i})} \right) \quad (4.2)$$

$$y_{3i} = \log \left(\frac{(nAB_{01i} + nAB_{11i})(nAA_{00i} + nAA_{10i})}{(nAA_{11i} + nAA_{01i})(nAB_{10i} + nAB_{00i})} \right) \quad (4.3)$$

$$y_{4i} = \log \left(\frac{(nBB_{01i} + nBB_{11i})(nAA_{00i} + nAA_{10i})}{(nAA_{11i} + nAA_{01i})(nBB_{10i} + nBB_{00i})} \right) \quad (4.4)$$

με προσεγγιστικές διασπορές που υπολογίζονται από τις ακόλουθες σχέσεις:

$$s_{1i}^2 = \frac{1}{(nAB_{11i} + nAB_{10i})} + \frac{1}{(nAA_{00i} + nAA_{01i})} + \frac{1}{(nAA_{11i} + nAA_{10i})} + \frac{1}{(nAB_{01i} + nAB_{00i})} \quad (4.5)$$

$$s_{2i}^2 = \frac{1}{(nBB_{11i} + nBB_{10i})} + \frac{1}{(nAA_{00i} + nAA_{01i})} + \frac{1}{(nAA_{11i} + nAA_{10i})} + \frac{1}{(nBB_{01i} + nBB_{00i})} \quad (4.6)$$

$$s_{3i}^2 = \frac{1}{(nAB_{01i} + nAB_{11i})} + \frac{1}{(nAA_{00i} + nAA_{10i})} + \frac{1}{(nAA_{11i} + nAA_{01i})} + \frac{1}{(nAB_{10i} + nAB_{00i})} \quad (4.7)$$

$$s_{4i}^2 = \frac{1}{(nBB_{0li} + nBB_{1li})} + \frac{1}{(nAA_{00i} + nAA_{10i})} + \frac{1}{(nAA_{11i} + nAA_{01i})} + \frac{1}{(nBB_{10i} + nBB_{00i})} \quad (4.8)$$

Στις ανωτέρω σχέσεις, τα y_{1i} και y_{2i} είναι οι λογάριθμοι των Odds Ratios για τη σύγκριση του γονοτύπου AB έναντι του AA και BB έναντι του AA, αντίστοιχα για την πρώτη έκβαση, ενώ τα y_{3i} και y_{4i} αποτελούν τις εκτιμήσεις για τις ίδιες συγκρίσεις για τη δεύτερη έκβαση. Έτσι, ακολουθώντας την γενική πολυμεταβλητή προσέγγιση που έχει περιγραφεί στην Παράγραφο 1.7.4 (van Houwelingen, Arends et al. 2002), ορίζουμε με \mathbf{y}_i το διάνυσμα που περιλαμβάνει τις 4 διαφορετικές εκτιμήσεις και με $\boldsymbol{\beta}$ το διάνυσμα των μέσων. Σε μια πολυμεταβλητή μετα-ανάλυση τυχαίων επιδράσεων υποθέτουμε ότι το \mathbf{y}_i ακολουθεί μια πολυμεταβλητή κανονική κατανομή γύρω από τους πραγματικούς μέσους $\boldsymbol{\beta}$, σύμφωνα με το ακόλουθο μοντέλο:

$$\mathbf{y}_i \sim MVN(\boldsymbol{\beta}, \boldsymbol{\Sigma} + \mathbf{C}_i) \quad (4.9)$$

Στο μοντέλο αυτό, ορίζουμε ως \mathbf{C}_i τον πίνακα των εντός των μελετών συνδιακυμάνσεων:

$$\mathbf{C}_i = \begin{pmatrix} s_{1i}^2 & \rho_{w12i}s_{1i}s_{2i} & \rho_{w13i}s_{1i}s_{3i} & \rho_{w14i}s_{1i}s_{4i} \\ \rho_{w12i}s_{1i}s_{2i} & s_{2i}^2 & \rho_{w23i}s_{2i}s_{3i} & \rho_{w24i}s_{2i}s_{4i} \\ \rho_{w13i}s_{1i}s_{3i} & \rho_{w23i}s_{2i}s_{3i} & s_{3i}^2 & \rho_{w34i}s_{3i}s_{4i} \\ \rho_{w14i}s_{1i}s_{4i} & \rho_{w24i}s_{2i}s_{4i} & \rho_{w34i}s_{3i}s_{4i} & s_{4i}^2 \end{pmatrix} \quad (4.10)$$

και με $\boldsymbol{\Sigma}$ τον πίνακα των μεταξύ των μελετών συνδιακυμάνσεων:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \tau_1^2 & \rho_{B12}\tau_1\tau_2 & \rho_{B13}\tau_1\tau_3 & \rho_{B14}\tau_1\tau_4 \\ \rho_{B12}\tau_1\tau_2 & \tau_2^2 & \rho_{B23}\tau_2\tau_3 & \rho_{B24}\tau_2\tau_4 \\ \rho_{B13}\tau_1\tau_3 & \rho_{B23}\tau_2\tau_3 & \tau_3^2 & \rho_{B34}\tau_3\tau_4 \\ \rho_{B14}\tau_1\tau_4 & \rho_{B24}\tau_2\tau_4 & \rho_{B34}\tau_3\tau_4 & \tau_4^2 \end{pmatrix} \quad (4.11)$$

Τα διαγώνια στοιχεία του \mathbf{C}_i αποτελούν τις ειδικές ανά μελέτη εκτιμήσεις της διασποράς ενώ τα στοιχεία εκτός της διαγωνίου αντιστοιχούν στις εντός των μελετών συνδιακυμάνσεις, για παράδειγμα $\rho_{w12}s_{1i}s_{2i} = \text{cov}(y_{1i}, y_{2i})$. Σχέσεις για τον υπολογισμό των συνδιακυμάνσεων μεταξύ των λογαρίθμων των Odds Ratios για την ίδια έκβαση έχουν παρουσιαστεί σε παλαιότερη εργασία ($\text{cov}(y_{1i}, y_{2i}) = 1/(nAA_{00i} + nAA_{01i}) + 1/(nAA_{11i} + nAA_{10i})$ και $\text{cov}(y_{3i}, y_{4i}) = 1/(nAA_{00i} + nAA_{10i}) + 1/(nAA_{11i} + nAA_{01i})$) (Bagos 2008). Έτσι, οι υπόλοιπες συνδιακυμάνσεις πρέπει να υπολογιστούν (βλέπε Παράγραφο 4.4.2.). Τα

διαγώνια στοιχεία του Σ αποτελούν τις μεταξύ των μελετών εκτιμήσεις της διασποράς ενώ τα στοιχεία εκτός της διαγωνίου αντιστοιχούν στις μεταξύ των μελετών συνδιακυμάνσεις που εκτιμώνται από τα δεδομένα.

Η μέθοδος δεν υποθέτει κάποιο γενετικό μοντέλο εκ των προτέρων (genetic model-free approach) (Minelli, Thompson et al. 2005) αλλά μπορούμε να υποθέσουμε κάποιο μοντέλο κληρονομικότητας (υπολειπόμενο, συνεπικρατές, επικρατές) υπολογίζοντας μια τιμή για την παράμετρο λ για κάθε έκβαση διαιρώντας τις εκτιμήσεις που προκύπτουν από τις εξισώσεις 4.1-2 και 4.3-4 (Minelli, Thompson et al. 2005). Η τιμή του λ δεν λαμβάνει καθορισμένες τιμές, αλλά τιμές ίσες με 0, 0.5 και 1 αντιστοιχούν σε υπολειπόμενο, συνεπικρατές και επικρατές μοντέλο κληρονομικότητας, ενώ τιμές >1 ή <0 υποδηλώνουν θετική ή αρνητική ετέρωση. Συνεπώς, στην περίπτωση αυτή καταλήγουμε με ένα διμεταβλητό μοντέλο όπου απαιτείται ξανά ο υπολογισμός των εντός των μελετών συνδιακυμάνσεων.

4.2.2. Ο υπολογισμός των εντός των μελετών συνδιακυμάνσεων

Η μεθοδολογία που χρησιμοποιήθηκε αποτελεί μια ειδική περίπτωση της γενικής μεθοδολογίας που έχει προταθεί πρόσφατα (Bagos 2012). Έτσι, προχωρούμε με υπολογισμούς (Παράρτημα Β) όπου αποδεικνύονται οι ακόλουθες σχέσεις:

$$\text{cov}(y_{1i}, y_{3i}) = \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{nAB_{j+i}nAB_{+ki}} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right) \quad (4.12)$$

$$\text{cov}(y_{2i}, y_{4i}) = \sum_j \sum_k (-1)^{j-k} \left(\frac{nBB_{jki}}{nBB_{j+i}nBB_{+ki}} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right) \quad (4.13)$$

$$\text{cov}(y_{1i}, y_{4i}) = \text{cov}(y_{2i}, y_{3i}) = \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right) \quad (4.14)$$

Παρόμοιες σχέσεις διαμορφώνονται εάν για κάθε έκβαση υποθέσουμε κάποιο γενετικό μοντέλο και συνοψίζονται στον Πίνακα 4.4.

Πίνακας 4.4. Εκτιμητές της συνδιακύμανσης του λογαρίθμου των Odds Ratios κάτω από διαφορετικά μοντέλα κληρονομικότητας.

Μοντέλο κληρονομικότητας για την πρώτη έκβαση	Μοντέλο κληρονομικότητας για τη δεύτερη έκβαση	Συνδιακυμάνσεις	Εξίσωση
Επικρατές	Επικρατές	$\sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{(nAB_{jki} + nBB_{jki})}{(nAB_{j+i} + nBB_{j+i})(nAB_{+ki} + nBB_{+ki})} \right)$	(4.15)
Επικρατές	Υπολειπόμενο	$\sum_j \sum_k (-1)^{j-k} \left(\frac{nBB_{jki}}{(nAB_{j+i} + nBB_{j+i})(nBB_{+ki})} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{(nAB_{j+i} + nBB_{j+i})(nAA_{+ki} + nAB_{+ki})} \right)$ $+ \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{(nAA_{j+i})(nAA_{+ki} + nAB_{+ki})} \right)$	(4.16)
Επικρατές	Συνεπικρατές	$\sum_j \sum_k (-1)^{j-k} \left(\frac{2nBB_{jki} + nAB_{jki}}{(nAB_{j+i} + nBB_{j+i})(nAB_{+ki} + 2nBB_{+ki})} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{(nAB_{j+i} + nBB_{j+i})(2nAA_{+ki} + nAB_{+ki})} \right)$ $+ \sum_j \sum_k (-1)^{j-k} \left(\frac{2nAA_{jki}}{(nAA_{j+i})(2nAA_{+ki} + nAB_{+ki})} \right)$	(4.17)
Υπολειπόμενο	Υπολειπόμενο	$\sum_j \sum_k (-1)^{j-k} \left(\frac{nBB_{jki}}{nBB_{j+i}nBB_{+ki}} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{(nAA_{jki} + nAB_{jki})}{(nAA_{j+i} + nAB_{j+i})(nAA_{+ki} + nAB_{+ki})} \right)$	(4.18)

Μοντέλο κληρονομικότητας για την πρώτη έκβαση	Μοντέλο κληρονομικότητας για τη δεύτερη έκβαση	Συνδιακυμάνσεις	Εξίσωση
Υπολειπόμενο	Συνεπικρατές	$\sum_j \sum_k (-1)^{j-k} \left(\frac{2nBB_{jki}}{(nBB_{j+i})(2nBB_{+ki} + nAB_{+ki})} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{(nAA_{j+i} + nAB_{j+i})(2nBB_{+ki} + nAB_{+ki})} \right)$ $+ \sum_j \sum_k (-1)^{j-k} \left(\frac{(2nAA_{jki} + nAB_{jki})}{(nAA_{j+i} + nAB_{j+i})(2nAA_{+ki} + nAB_{+ki})} \right)$	(4.19)
Συνεπικρατές	Συνεπικρατές	$\sum_j \sum_k (-1)^{j-k} \left(\frac{4nAA_{jki} + nAB_{jki}}{(2nAA_{j+i} + nAB_{j+i})(2nAA_{+ki} + nAB_{+ki})} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{4nBB_{jki} + nAB_{jki}}{(2nBB_{j+i} + nAB_{j+i})(2nBB_{+ki} + nAB_{+ki})} \right)$ $+ \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{(2nAA_{j+i} + nAB_{j+i})(2nBB_{+ki} + nAB_{+ki})} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{(2nBB_{j+i} + nAB_{j+i})(2nAA_{+ki} + nAB_{+ki})} \right)$	(4.20)

4.2.3. Διαχείριση ελλειπουσών πληροφοριών

Όπως είναι εμφανές από τις εξισώσεις 4.12-14, για τον υπολογισμό των εντός των μελετών συνδιακυμάνσεων πρέπει να είναι γνωστές οι ακριβείς τιμές του Πίνακα 4.3. Στην απλούστερη περίπτωση, όπου για κάθε i μελέτη οι τιμές αυτές είναι γνωστές προχωρούμε απευθείας στον υπολογισμό των συνδιακυμάνσεων. Σε πρακτικές εφαρμογές ωστόσο, δεν αναμένουμε να είναι διαθέσιμη αυτή η πληροφορία για όλες τις μελέτες. Σε αυτή την περίπτωση (ή ακόμα και όταν καμία μελέτη δεν παρέχει δεδομένα), η ταξινόμηση των ατόμων για κάθε έκβαση ανάλογα με το γονότυπο υπολογίζεται χρησιμοποιώντας ένα log-linear μοντέλο και τον αλγόριθμο IPF (Fienberg 1970). Ο αλγόριθμος IPF χρησιμοποιείται για εκτίμηση των δεδομένων ενός πίνακα συνάφειας έτσι ώστε να αθροίζουν σε συγκεκριμένα σύνολα γραμμών και στηλών (στην απλούστερη περίπτωση των δισδιάστατων πινάκων). Θεωρούμε επίσης ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης στον $2 \times 2 \times 3$ πίνακα συνάφειας, το οποίο αντικατοπτρίζει το γεγονός ότι τα τρία Odds Ratios που υπολογίζουν τη συσχέτιση των δυο εκβάσεων για κάθε γονότυπο (AA, AB και BB) είναι ίσα, αλλά όχι απαραίτητα ίσα με το περιθώριο Odds Ratio για τις δυο εκβάσεις (με άλλα λόγια δεν υποθέτουμε ότι οι πίνακες είναι collapsible).

Τεχνικά, η εντολή `ipf` στο Stata, που χρησιμοποιείται για την υλοποίηση του αλγορίθμου IPF, χρησιμοποιεί τα δεδομένα σε long format ως ένα γενικευμένο γραμμικό μοντέλο, και έτσι απαιτεί κάποιες αρχικές τιμές του Πίνακα 4.3. Για το σκοπό αυτό, μια γενική και εύκολη μέθοδος περιγράφεται με λεπτομέρεια στο παράρτημα Γ). Εν συντομία, για κάθε μελέτη υπολογίζουμε τα κελιά του Πίνακα 4.2 που περιγράφει τη συσχέτιση των δυο εκβάσεων χρησιμοποιώντας τα περιθώρια αθροίσματα (που θεωρούνται γνωστά) και τα Odds Ratios που δείχνουν τη συσχέτιση των δυο εκβάσεων. Τα Odds Ratios αυτά εξάγονται από τη βιβλιογραφία ή από τις συγκεκριμένες μελέτες που χρησιμοποιούνται στη μετα-ανάλυση. Οι τιμές που υπολογίζονται από το προηγούμενο βήμα είναι τα περιθώρια αθροίσματα του Πίνακα 4.3 και χρησιμοποιούνται για τη διαμόρφωση εξισώσεων έτσι ώστε να υπολογιστούν οι ποσότητες των κελιών του πίνακα αυτού (Γ26-35). Τέλος, αξίζει να αναφερθεί ότι οι εξισώσεις Γ26-35 δεν δίνουν μοναδική λύση, αλλά η επιλογή αυτής που θα χρησιμοποιηθεί δεν τροποποιεί τις τελικές εκτιμήσεις που προκύπτουν από τον αλγόριθμο IPF.

4.2.4. Το γενικό πολυμεταβλητό μοντέλο μετα-ανάλυσης πολλαπλών συνεχών εκβάσεων

Για την περίπτωση των πολλαπλών συνεχών εκβάσεων, οι υπολογισμοί είναι παρόμοιοι αλλά πιο εύκολοι. Μια τυπική μορφή των δεδομένων δύο συνεχών εκβάσεων για έναν γενετικό τόπο με δυο αλληλόμορφα παρουσιάζεται στον Πίνακα 4.5. Έτσι, ορίζουμε για παράδειγμα ως y_{AA1i} (sd_{AA1i}) το μέσο όρο και την τυπική απόκλιση της πρώτης έκβασης των ατόμων με γονότυπο AA ενώ με y_{AA3i} (sd_{AA3i}) τις αντίστοιχες ποσότητες για τη δεύτερη έκβαση.

Πίνακας 4.5. Οι μέσες τιμές και τυπικές αποκλίσεις για τις δυο εκβάσεις (Y_{1i} , Y_{2i}) ανάλογα με το γονότυπο για κάθε i μελέτη ($i=1,2,\dots,k$).

Έκβαση	Γονότυποι		
	AA	AB	BB
	mean±sd (N)	mean±sd (N)	mean±sd (N)
Y_{1i}	$y_{AA1i} \pm sd_{AA1i} (nAA_{1i})$	$y_{AB1i} \pm sd_{AB1i} (nAB_{1i})$	$y_{BB1i} \pm sd_{BB1i} (nBB_{1i})$
Y_{2i}	$y_{AA2i} \pm sd_{AA2i} (nAA_{2i})$	$y_{AB2i} \pm sd_{AB2i} (nAB_{2i})$	$y_{BB2i} \pm sd_{BB2i} (nBB_{2i})$

Οι ανά δυο διαφορές των δυο εκβάσεων για τα άτομα που φέρουν το μεταλλαγμένο γονότυπο (AB, BB) από τα άτομα με φυσιολογικό γονότυπο (AA) μπορούν να μοντελοποιηθούν ως ακολούθως:

$$y_{1i} = y_{AB1i} - y_{AA1i} \quad (4.21)$$

$$y_{2i} = y_{BB1i} - y_{AA1i} \quad (4.22)$$

$$y_{3i} = y_{AB2i} - y_{AA2i} \quad (4.23)$$

$$y_{4i} = y_{BB2i} - y_{AA2i} \quad (4.24)$$

Αποδεικνύεται εύκολα ότι οι τέσσερις διαφορές κατανέμονται κανονικά με μέσο που ισούται με το μέσο όρο της διαφοράς και διασπορά ίση με το άθροισμα των επιμέρους διασπορών. Έτσι, το μοντέλο που παρουσιάστηκε στην εξίσωση 4.9 μπορεί να χρησιμοποιηθεί για την πολυμεταβλητή μοντελοποίηση των τεσσάρων διαφορών. Οι δυο μέσες διαφορές κάθε έκβασης είναι συσχετισμένες καθώς για τον υπολογισμό τους χρησιμοποιείται η ίδια κατηγορία αναφοράς. Μέσω απλών υπολογισμών

πιθανοτήτων και παρατηρώντας ότι οι τυχαίες μεταβλητές $y_{AA1i}, y_{AB1i}, y_{BB1i}$ για την πρώτη έκβαση είναι αμοιβαίως αποκλειόμενες (δηλαδή $\text{cov}(y_{AB1i}, y_{BB1i}) = 0$ και λοιπά), αποδεικνύεται ότι η συνδιακύμανση υπολογίζεται από την ακόλουθη σχέση (Bagos 2008):

$$\begin{aligned}\text{cov}(y_{1i}, y_{2i}) &= \text{cov}(y_{AB1i} - y_{AA1i}, y_{BB1i} - y_{AA1i}) \\ &= \text{cov}(y_{AB1i}, y_{BB1i}) - \text{cov}(y_{AB1i}, y_{AA1i}) \\ &\quad - \text{cov}(y_{AA1i}, y_{BB1i}) + \text{cov}(y_{AA1i}, y_{AA1i}) \\ &= \text{var}(y_{AA1i}) = \frac{sd_{AA1i}^2}{nAA_{1i}}\end{aligned}\quad (4.25)$$

Παρόμοια, για τη δεύτερη έκβαση ισχύει ότι $\text{cov}(y_{3i}, y_{4i}) = \text{var}(y_{AA2i}) = \frac{sd_{AA2i}^2}{nAA_{2i}}$.

Ακολούθως, πρέπει να υπολογιστούν οι υπόλοιπες συνδιακυμάνσεις. Έτσι, αν ορίσουμε ως $r_{y_1 y_2}$ το συντελεστή συσχέτισης των δυο εκβάσεων και ως $r_{y_{AA1} y_{AA2}}$, $r_{y_{AB1} y_{AB2}}$, $r_{y_{BB1} y_{BB2}}$ τους συντελεστές συσχέτισης των δυο εκβάσεων για κάθε γονότυπο ξεχωριστά αποδεικνύεται ότι:

$$\begin{aligned}\text{cov}(y_{1i}, y_{3i}) &= \text{cov}(y_{AB1i} - y_{AA1i}, y_{AB2i} - y_{AA2i}) \\ &= \text{cov}(y_{AB1i}, y_{AB2i}) - \text{cov}(y_{AA1i}, y_{AB2i}) - \text{cov}(y_{AB1i}, y_{AA2i}) + \text{cov}(y_{AA1i}, y_{AA2i}) \\ &= r_{y_{AB1} y_{AB2}} \sqrt{\text{var}(y_{AB1i}) \text{var}(y_{AB2i})} + r_{y_{AA1} y_{AA2}} \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})} \\ &\approx r_{y_1 y_2} \left(\sqrt{\text{var}(y_{AB1i}) \text{var}(y_{AB2i})} + \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})} \right)\end{aligned}\quad (4.26)$$

$$\begin{aligned}\text{cov}(y_{2i}, y_{4i}) &= \text{cov}(y_{BB1i} - y_{AA1i}, y_{BB2i} - y_{AA2i}) \\ &= \text{cov}(y_{BB1i}, y_{BB2i}) - \text{cov}(y_{AA1i}, y_{BB2i}) - \text{cov}(y_{BB1i}, y_{AA2i}) + \text{cov}(y_{AA1i}, y_{AA2i}) \\ &= r_{y_{BB1} y_{BB2}} \sqrt{\text{var}(y_{BB1i}) \text{var}(y_{BB2i})} + r_{y_{AA1} y_{AA2}} \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})} \\ &\approx r_{y_1 y_2} \left(\sqrt{\text{var}(y_{BB1i}) \text{var}(y_{BB2i})} + \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})} \right)\end{aligned}\quad (4.27)$$

$$\begin{aligned}
\text{cov}(y_{1i}, y_{4i}) &= \text{cov}(y_{AB1i} - y_{AA1i}, y_{BB2i} - y_{AA2i}) \\
&= \text{cov}(y_{AB1i}, y_{BB2i}) - \text{cov}(y_{AB1i}, y_{AA2i}) - \text{cov}(y_{AA1i}, y_{BB2i}) + \text{cov}(y_{AA1i}, y_{AA2i}) \\
&= r_{y_{AA1} y_{AA2}} \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})} \\
&\approx r_{y_1 y_2} \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})}
\end{aligned}
\tag{4.28}$$

$$\begin{aligned}
\text{cov}(y_{2i}, y_{3i}) &= \text{cov}(y_{BB1i} - y_{AA1i}, y_{AB2i} - y_{AA2i}) \\
&= \text{cov}(y_{BB1i}, y_{AB2i}) - \text{cov}(y_{BB1i}, y_{AA2i}) - \text{cov}(y_{AA1i}, y_{AB2i}) + \text{cov}(y_{AA1i}, y_{AA2i}) \\
&= r_{y_{AA1} y_{AA2}} \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})} \\
&\approx r_{y_1 y_2} \sqrt{\text{var}(y_{AA1i}) \text{var}(y_{AA2i})}
\end{aligned}
\tag{4.29}$$

Για τον υπολογισμό της τελευταίας ισότητας στις εξισώσεις 4.26-29 υποθέτουμε ότι ο συντελεστής συσχέτισης των δυο εκβάσεων είναι ο ίδιος, ανεξάρτητα από το γονότυπο (δηλαδή δεν υπάρχει αλληλεπίδραση). Ο συντελεστής αυτός μπορεί να ανακτηθεί εύκολα από τη βιβλιογραφία, τουλάχιστον για την περίπτωση κοινών εκβάσεων (όπως για αιματολογικές παραμέτρους, μεταβολίτες ή ανθρωπομετρικά χαρακτηριστικά). Φυσικά, όπως και στην περίπτωση των διακριτών εκβάσεων, μπορούμε να υποθέσουμε εξ'αρχής κάποιο γενετικό μοντέλο και να πραγματοποιήσουμε μια κλασσική διμεταβλητή μετα-ανάλυση με ένα μόνο μέτρο επίδρασης για κάθε έκβαση.

Συμπερασματικά, είναι κατανοητό ότι είτε για την περίπτωση των διακριτών είτε των συνεχών εκβάσεων που δεν είναι αμοιβαίως αποκλειόμενες μπορεί να χρησιμοποιηθεί το γενικό πολυμεταβλητό μοντέλο για την εκτίμηση των μέτρων επίδρασης. Το μοντέλο μπορεί να υλοποιηθεί εύκολα χρησιμοποιώντας ένα οποιοδήποτε στατιστικό πακέτο με σταθμισμένα μοντέλα παλινδρόμησης μεικτών επιδράσεων, όπως το SAS (με την εντολή PROC MIXED) ή το Stata (με την εντολή mvmeta). Κώδικας για την υλοποίηση του μοντέλου στο πακέτο Stata διατίθεται στην ιστοσελίδα www.compgen.org/tools/multiple-outcomes. Η εντολή mvmeta χρησιμοποιεί για την εκτίμηση των παραμέτρων είτε μέγιστη πιθανοφάνεια είτε δεσμευμένη μέγιστη πιθανοφάνεια με απευθείας μεγιστοποίηση των παραμέτρων μέσω του αλγορίθμου Newton-Raphson (White 2009).

4.3. Εφαρμογή της μεθόδου

4.3.1. Διακριτές εκβάσεις

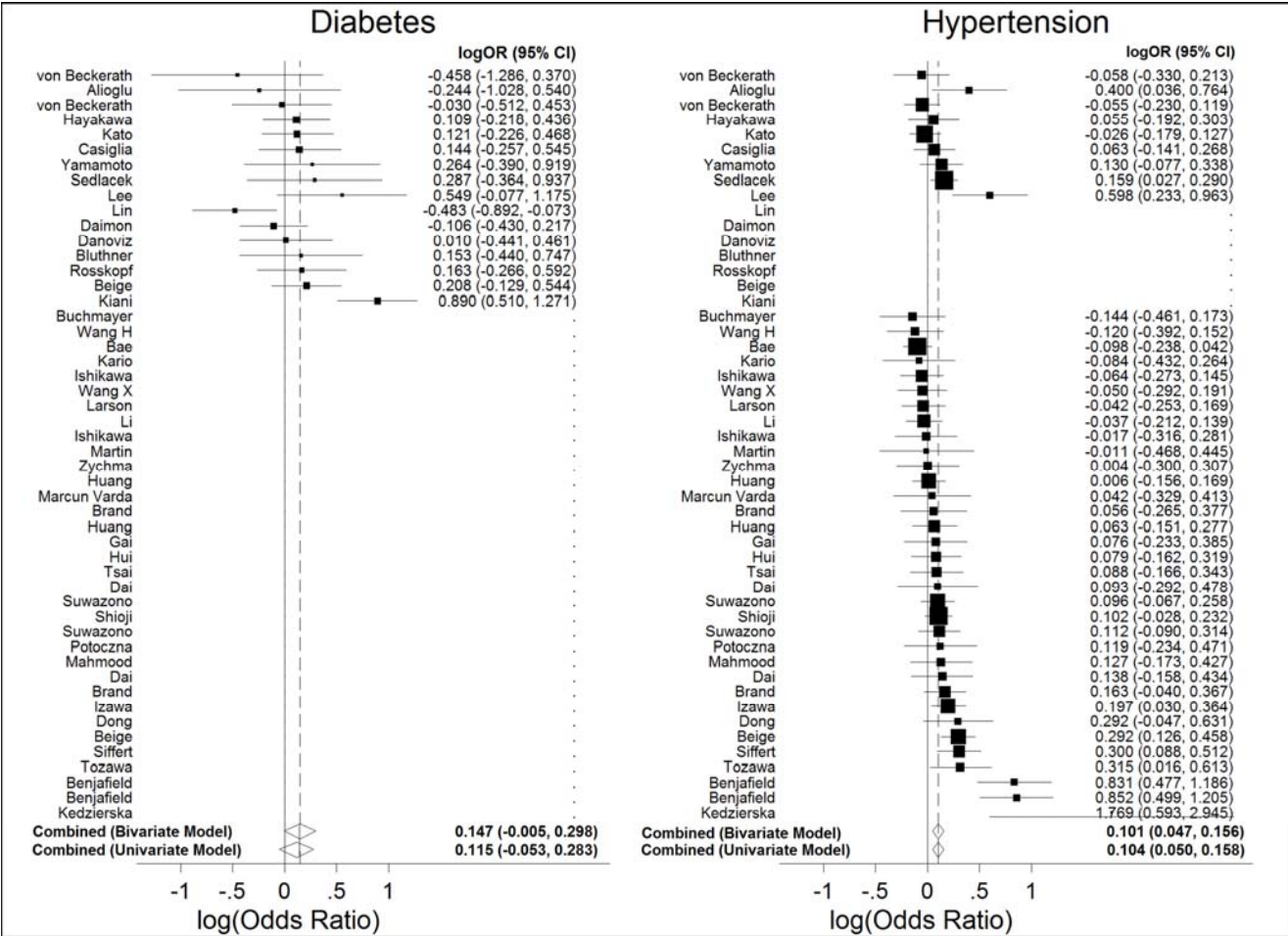
Η ιδιοπαθής υπέρταση και ο διαβήτης τύπου 2 αποτελούν δυο νοσήματα που συχνά συνυπάρχουν. Ο κίνδυνος εμφάνισης διαβήτη είναι περίπου 2.4 φορές μεγαλύτερος στους υπερτασικούς σε σχέση με αυτούς που δεν έχουν υπέρταση (Gress, Nieto et al. 2000). Έτσι, οι λογάριθμοι των Odds Ratios ενός γονιδίου που συσχετίζεται με τα χαρακτηριστικά αυτά είναι συσχετισμένοι εντός της ίδιας μελέτης καθώς υπολογίζονται χρησιμοποιώντας τα ίδια άτομα. Ως παράδειγμα, χρησιμοποιήσαμε τα δεδομένα μιας δημοσιευμένης μετα-ανάλυσης για τη διερεύνηση της συσχέτισης του πολυμορφισμού C825T της β3 υπομονάδας της G πρωτεΐνης (GNB3) με την ιδιοπαθή υπέρταση (Bagos, Elefsinioti et al. 2007) και πραγματοποιήσαμε μια επιπλέον βιβλιογραφική αναζήτηση για τη διερεύνηση της συσχέτισης του ίδιου γονιδίου με τον διαβήτη τύπου 2. Η ιδέα στηρίζεται στην από κοινού μοντελοποίηση των δεδομένων των δυο εκβάσεων (υπέρταση και διαβήτη) έτσι ώστε εκμεταλλευτούμε τα πλεονεκτήματα της πολυμεταβλητής προσέγγισης.

Πιο συγκεκριμένα, 9 μελέτες αξιολογούσαν και τις δυο εκβάσεις για τα ίδια άτομα. Ένα σύνολο 7 και 34 μελετών αντίστοιχα είχαν δεδομένα μόνο για το διαβήτη ή μόνο για την υπέρταση. Δυο από τις εννιά μελέτες, με εκτιμήσεις και για τις δυο εκβάσεις, παρείχαν τα δεδομένα για να υπολογιστούν τα κελιά του πίνακα της από κοινού κατανομής των ασθενειών (Y_{1i} , Y_{2i}) ανάλογα με το γονότυπο (Kato, Sugiyama et al. 1998; Hayakawa, Takamura et al. 2007). Για τις υπόλοιπες επτά μελέτες, οι τιμές αυτές υπολογίστηκαν χρησιμοποιώντας τον αλγόριθμο IPF. Τα κελιά της ταξινόμησης των συμμετεχόντων ανά κατηγορία έκβασης υπολογίστηκαν χρησιμοποιώντας ένα Odds Ratio ίσο με 2.70 (Gress, Nieto et al. 2000). Οι εκτιμήσεις υπό την προτεινόμενη πολυμεταβλητή ή την υπάρχουσα διμεταβλητή προσέγγιση ήταν πολύ κοντά δείχνοντας ότι ο συγκεκριμένος πολυμορφισμός συσχετίζεται με την υπέρταση αλλά όχι με το διαβήτη (Wald test p-value=0.004) (Πίνακας 4.6).

Πίνακας 4.6. Οι εκτιμήσεις των παραμέτρων του πολυμεταβλητού και διμεταβλητού μοντέλου του λογαρίθμου του Odds Ratio των ετερόζυγων ή ομόζυγων για το μεταλλαγμένο αλληλόμορφο έναντι των ομόζυγων για το φυσιολογικό αλληλόμορφο για τις δυο εκβάσεις (β_1 - β_2 για το διαβήτη και β_3 - β_4 για την υπέρταση) χρησιμοποιώντας την εντολή `pnmmeta` στο Stata. Δίνονται επίσης οι εκτιμήσεις της μεταξύ των μελετών μεταβλητότητας.

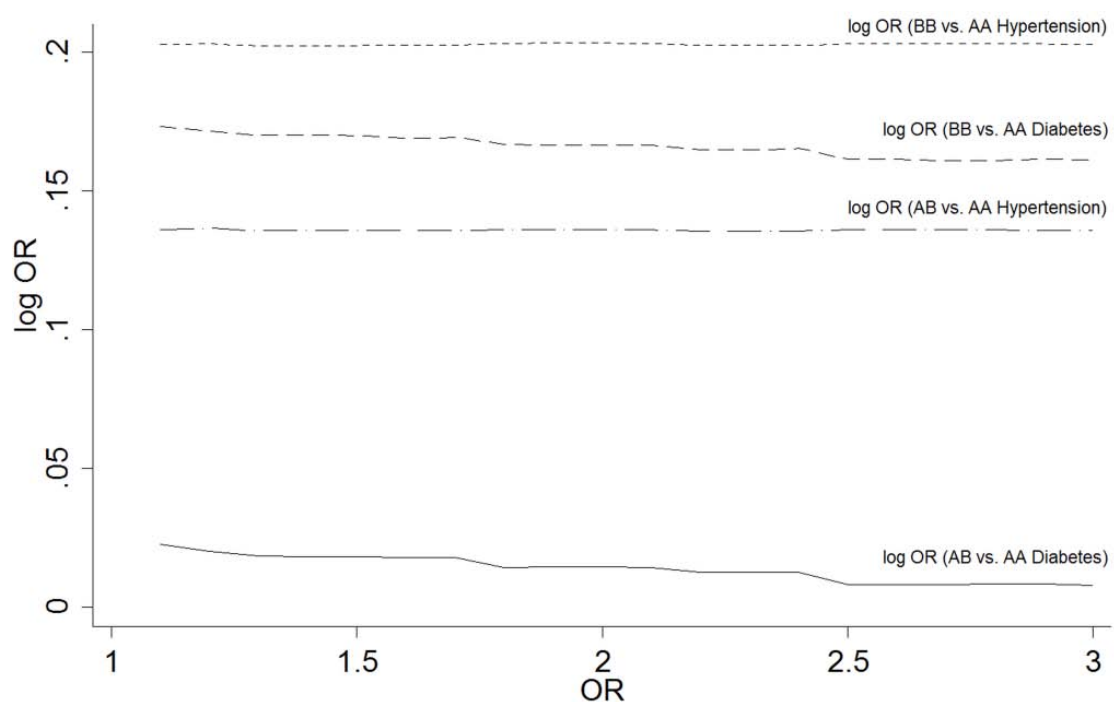
	Πολυμεταβλητό μοντέλο			Διμεταβλητό μοντέλο		
	logOR (95% CI)	z	τ^2	logOR (95% CI)	z	τ^2
β_1	0.008 (-0.167, 0.184)	0.09	0.094	-0.013 (-0.201, 0.174)	-0.14	0.099
β_2	0.161 (-0.057, 0.379)	1.45	0.180	0.118 (-0.142, 0.378)	0.89	0.198
β_3	0.136 (0.041, 0.232)	2.79	0.059	0.142 (0.046, 0.237)	2.91	0.052
β_4	0.203 (0.093, 0.313)	3.62	0.071	0.209 (0.100, 0.317)	3.76	0.059
$\rho_{B12}=0.913, \rho_{B13}=0.731, \rho_{B14}=0.750,$						
$\rho_{B23}=0.945, \rho_{B24}=0.954$						
$\rho_{B34}=1$				$\rho_{B12}=0.934, \rho_{B34}=1$		

Παρατηρήθηκε ωστόσο ότι η συνδυασμένη προσέγγιση έδινε οριακά μη στατιστικά σημαντικές εκτιμήσεις για το διαβήτη. Συνεπώς, προχωρήσαμε σε διμεταβλητή μοντελοποίηση εκτιμώντας το μοντέλο κληρονομικότητας για κάθε έκβαση. Οι τιμές της παραμέτρου λ ήταν 0.052 (95% CI: -0.985, 1.089) και 0.670 (95% CI: 0.392, 0.948), υποδεικνύοντας υπολειπόμενο και συνεπικρατές μοντέλο κληρονομικότητας για το διαβήτη και την υπέρταση αντίστοιχα. Ο λογάριθμος του Odds Ratio για το διαβήτη υπό την μονοπαραγοντική προσέγγιση ήταν 0.115 (95% CI: -0.053, 0.283), αλλά χρησιμοποιώντας τη διμεταβλητή προσέγγιση μέσω δεσμευμένης μέγιστης πιθανοφάνειας μεταβλήθηκε σε 0.147 (95% CI: -0.005, 0.298), καταδεικνύοντας μια οριακά μη στατιστικά σημαντική συσχέτιση. Το γεγονός αυτό δείχνει τη σημαντικότητα της προτεινόμενης μεθοδολογίας προσδίδοντας αυξημένη ακρίβεια στην εκτίμηση των παραμέτρων. Οι αντίστοιχες εκτιμήσεις για την υπέρταση ήταν 0.104 (95% CI: 0.050, 0.158) και 0.101 (95% CI: 0.047, 0.156) δείχνοντας σε κάθε περίπτωση μια στατιστικά σημαντική συσχέτιση του υπό μελέτη γονιδιακού πολυμορφισμού με την υπέρταση υπό τη θεώρηση συνεπικρατούς μοντέλου κληρονομικότητας (Εικόνα 4.1).

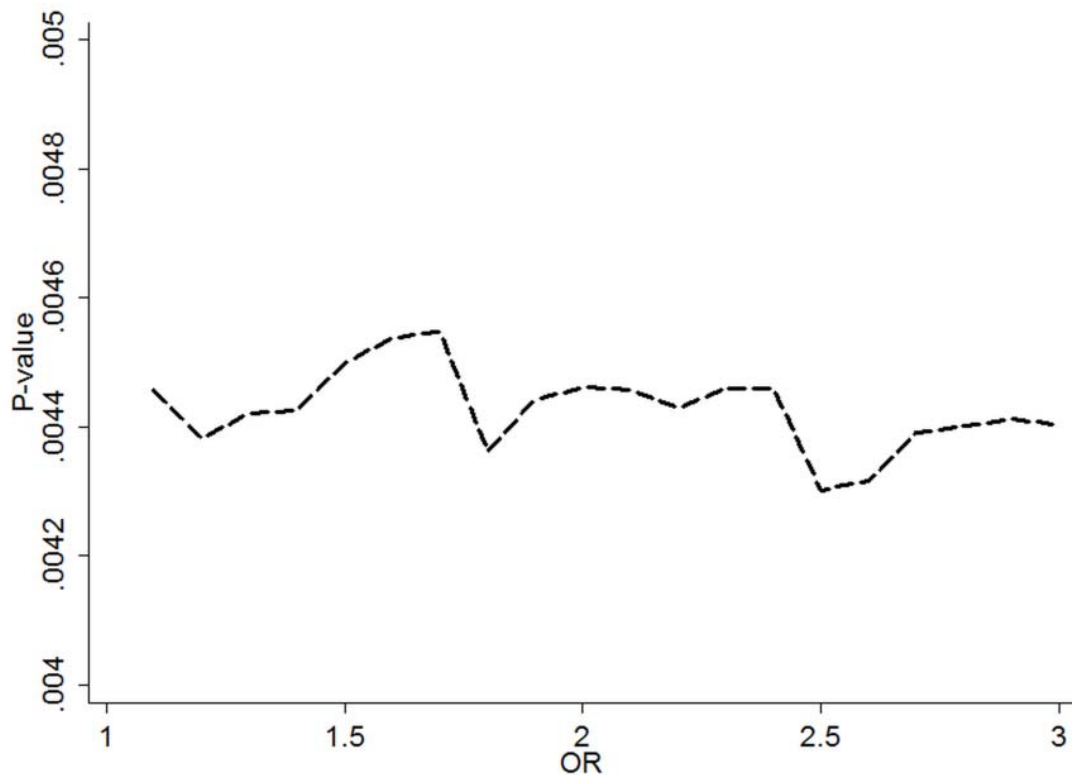


Εικόνα 4.1. Το forest plot και οι συνδυασμένες εκτιμήσεις από την διμεταβλητή και μονομεταβλητή προσέγγιση υιοθετώντας υπολειπόμενο και συνεπικρατές μοντέλο για το διαβήτη και για την υπέρταση αντίστοιχα. Οι οριζόντιες μπάρες αντιπροσωπεύουν τα 95% διαστήματα εμπιστοσύνης.

Πραγματοποιήθηκε ανάλυση ευαισθησίας για να διερευνηθεί αν η αρχική επιλογή των Odds Ratios, για τη συσχέτιση των δυο εκβάσεων, μπορεί να επηρεάσει τις εκτιμήσεις των παραμέτρων, τα 95% διαστήματα εμπιστοσύνης καθώς επίσης την συνολική στατιστική σημαντικότητα του μοντέλου (συνολικό p-value). Αξιολογήθηκαν σενάρια με Odds Ratios να ποικίλουν από 1 έως 3 δείχνοντας μια μικρή διακύμανση στην εκτίμηση του λογαρίθμου του Odds Ratio της σύγκρισης του γονότυπου AB έναντι του AA για το διαβήτη καθώς επίσης και για τα p-values που προκύπτουν από το συνολικό έλεγχο Wald, μολονότι όλα τα ευρήματα αυτά ήταν μη στατιστικά σημαντικά (Εικόνες 4.2-3).



Εικόνα 4.2. Ανάλυση ευαισθησίας για το διαβήτη και την υπέρταση. Απεικονίζονται οι εκτιμήσεις των παραμέτρων σε σχέση με τα Odds Ratios, που αντικατοπτρίζουν τη συσχέτιση των δυο εκβάσεων.



Εικόνα 4.3. Ανάλυση ευαισθησίας για το διαβήτη και την υπέρταση. Απεικονίζονται τα p-values, που υπολογίζονται μέσω ενός συνολικού ελέγχου Wald, σε σχέση με τα Odds Ratios, που αντικατοπτρίζουν τη συσχέτιση των δυο εκβάσεων.

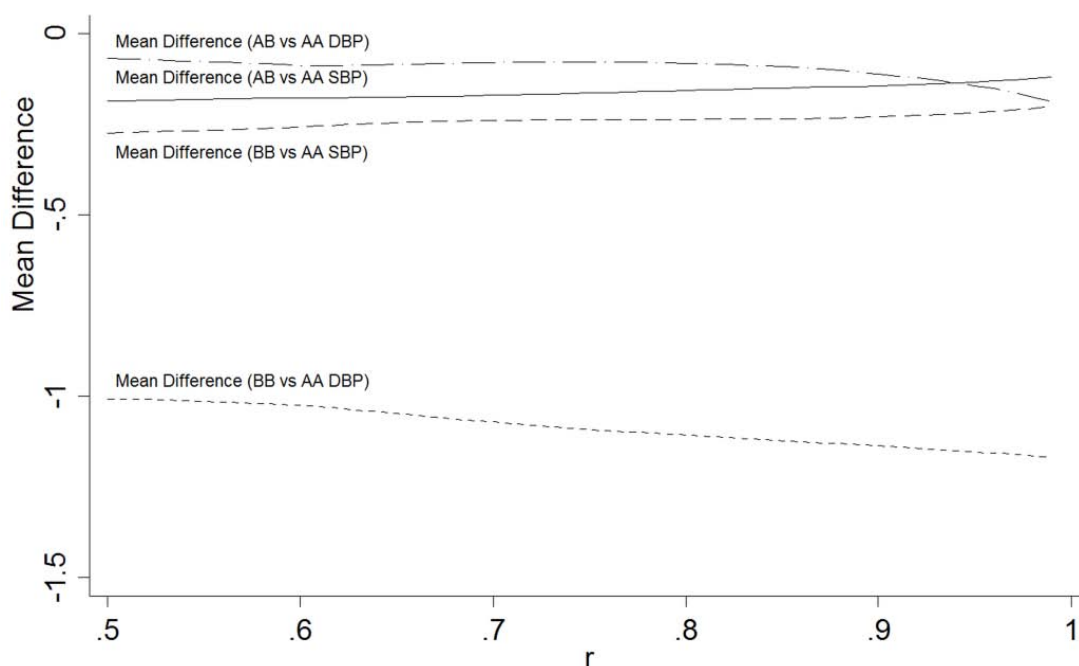
4.3.2. Συνεχείς εκβάσεις

Η εκτίμηση των επιδράσεων γενετικών παραγόντων σε πολλαπλά συσχετισμένα συνεχή χαρακτηριστικά όπως η συστολική (systolic blood pressure-SBP) και η διαστολική πίεση (diastolic blood pressure-DBP) είναι ιδιαίτερου ενδιαφέροντος. Ως παράδειγμα, χρησιμοποιήθηκαν τα δεδομένα μιας δημοσιευμένης μετα-ανάλυσης που στόχευε να διαλευκάνει εάν ο πολυμορφισμός Gly482Ser του γονιδίου Peroxisome Proliferator Activated Receptor-γ Coactivator 1α (PPARGC1A) μπορούσε να τροποποιήσει την πίεση (συστολική και διαστολική) (Vimalleswaran, Luan et al. 2008). Πιο συγκεκριμένα, ένα σύνολο 16 μελετών είχαν δεδομένα και για τις δυο εκβάσεις. Ο πίνακας διακυμάνσεων-συνδιακυμάνσεων υπολογίστηκε χρησιμοποιώντας τον συντελεστή συσχέτισης από ένα άρθρο ανασκόπησης (Collins and MacMahon 1994) ίσο με 0.70. Μη στατιστικά σημαντικά αποτελέσματα προέκυψαν υπό την διμεταβλητή ή την πολυμεταβλητή προσέγγιση (Πίνακας 4.7). Η ανάλυση ευαισθησίας, χρησιμοποιώντας διάφορους πιθανούς συντελεστές συσχέτισης που κυμαίνονταν από 0.5 έως 1.0, έδειξε μια μικρή διακύμανση για τη διαφορά της διαστολικής πίεσης των ατόμων που φέρουν το γονότυπο BB έναντι των

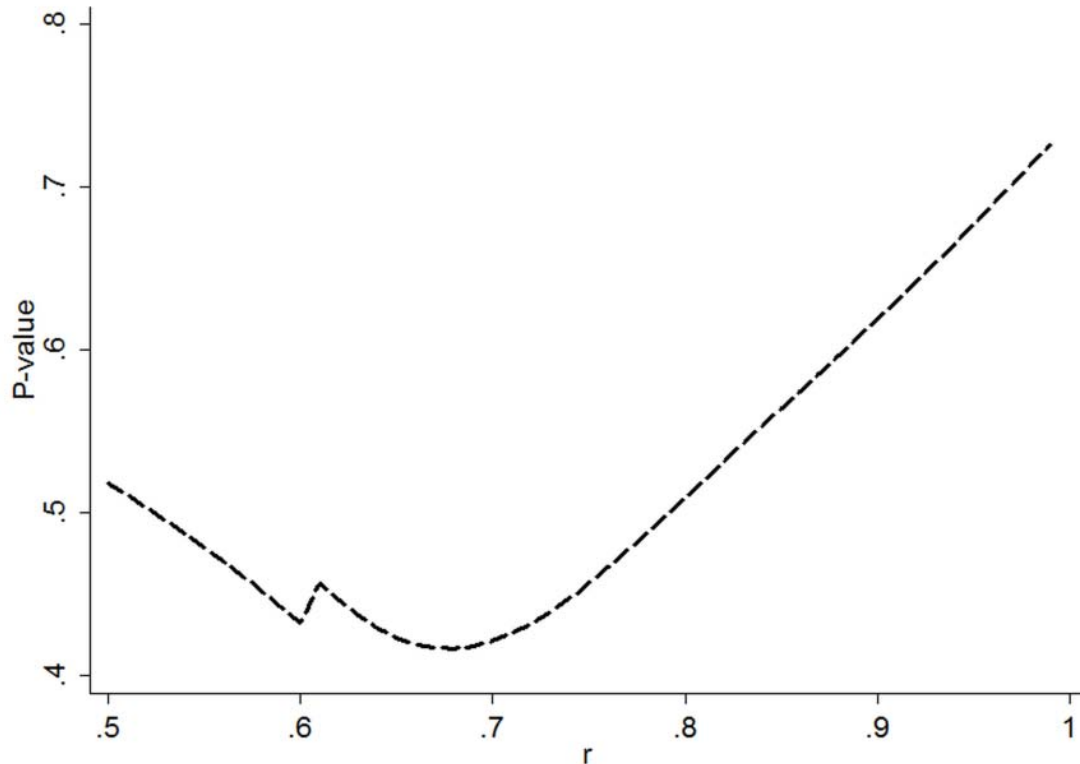
ατόμων με φυσιολογικό γονότυπο (AA) και μη στατιστικά σημαντικά p-values από το συνολικό έλεγχο Wald (Εικόνες 4.4-5).

Πίνακας 4.7. Οι εκτιμήσεις των παραμέτρων του πολυμεταβλητού και διμεταβλητού μοντέλου των μέσων διαφορών των ετερόζυγων ή ομόζυγων για το μεταλλαγμένο αλληλόμορφο έναντι των ομόζυγων για το φυσιολογικό αλληλόμορφο για τις δυο εκβάσεις (β_1 - β_2 για τη συστολική και β_3 - β_4 για τη διαστολική πίεση) χρησιμοποιώντας την εντολή mvmeta στο Stata.

	Πολυμεταβλητό μοντέλο			Διμεταβλητό μοντέλο		
	Mean (95% CI)	z	τ^2	Mean (95% CI)	z	τ^2
β_1	-0.170 (-0.693, 0.354)	-0.64	0.319	-0.155 (-0.653, 0.343)	-0.61	0.267
β_2	-0.241 (-1.397, 0.916)	-0.41	3.126	-0.296 (-1.396, 0.805)	-0.53	2.638
β_3	-0.080 (-1.054, 0.894)	-0.16	1.458	-0.079 (-0.973, 0.816)	-0.17	1.030
β_4	-1.072 (-2.693, 0.550)	-1.30	4.702	-1.132 (-2.767, 0.502)	-1.36	4.859
$\rho_{B12}=0.975, \rho_{B13}=0.488, \rho_{B14}=0.775,$						
$\rho_{B23}=0.372, \rho_{B24}=0.632$						
$\rho_{B34}=0.865$						
				$\rho_{B12}=1, \rho_{B34}=1$		



Εικόνα 4.4. Ανάλυση ευαισθησίας για τη συστολική και διαστολική πίεση. Απεικονίζονται οι εκτιμήσεις των παραμέτρων σε σχέση με τους συντελεστές συσχέτισης των δυο εκβάσεων.

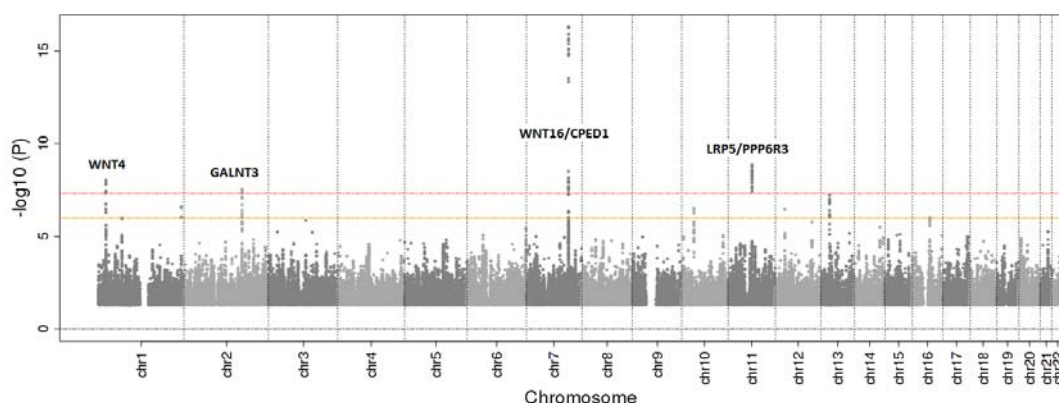


Εικόνα 4.5. Ανάλυση ευαισθησίας για τη συστολική και διαστολική πίεση. Απεικονίζονται τα p-values, που υπολογίζονται μέσω ενός συνολικού ελέγχου Wald, σε σχέση με τους συντελεστές συσχέτισης των δυο εκβάσεων.

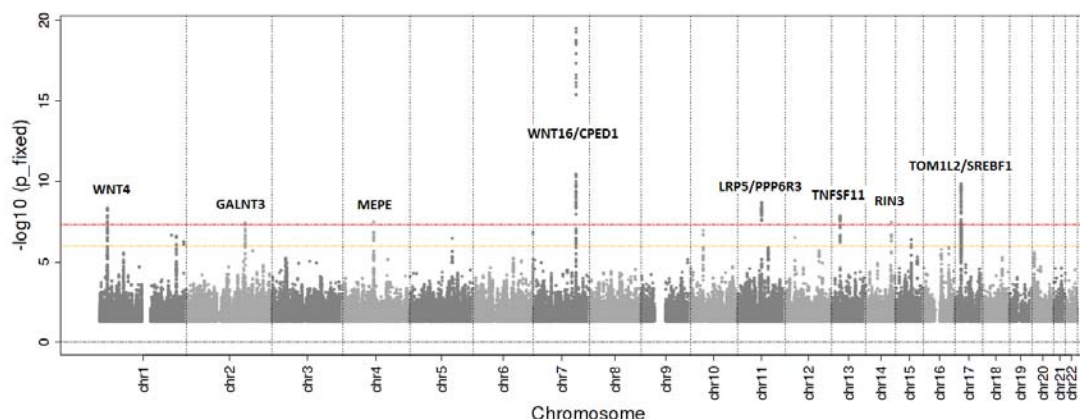
Επιπλέον, εφαρμόσαμε την πολυμεταβλητή μεθοδολογία μετα-ανάλυσης για τις συνεχείς εκβάσεις Total-body lean mass (TB-LM) και Bone mineral density (TB-BMD) έτσι ώστε να διερευνηθούν οι πιθανές πλειοτροπικές επιδράσεις και να διαλευκανθούν τα μονοπάτια που αυτά εμπλέκονται κατά τη διαμόρφωση του σκελετικού συστήματος. Συνολικά, αξιολογήθηκαν περισσότεροι από 2.5 εκατομμύρια γονιδιακοί πολυμορφισμοί τεσσάρων μελετών κοορτής παιδικού πληθυσμού (Generation R Study, Avon Longitudinal Study of Parents and their Children (ALSPAC), Bone Mineral Density in Childhood Study (BMDCS) και Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) cohort) με 10,414 συμμετέχοντες. Η συσχέτιση μεταξύ BMD και lean mass (σταθμίζοντας ως προς το φύλο, την ηλικία, το ύψος, την ποσοστιαία κατανομή του λίπους και genomic principal components) ήταν σημαντική και προς την ίδια κατεύθυνση για τις τέσσερις μελέτες κοορτής: Generation R ($\rho=0.437$), ALSPAC ($\rho=0.449$), BMD-CS ($\rho=0.486$) και COPSAC ($\rho=0.420$). Για λόγους σύγκρισης, πραγματοποιήθηκε μονομεταβλητή και διμεταβλητή μετα-ανάλυση ευρυγονιδιωματικών μελετών με

μοντέλα σταθερών επιδράσεων (υποθέτοντας συνεπικρατές μοντέλο κληρονομικότητας).

Τα αποτελέσματα της μονομεταβλητής μετα-ανάλυσης οδήγησαν στην εύρεση τεσσάρων γενετικών τόπων (με genome-wide significant p -values $< 5 \times 10^{-8}$) που συσχετίζονται με την έκβαση TB-BMD: *WNT16/CPED1* locus, leading SNP rs917727-T (beta=0.129SD, p -value= 1.28×10^{-16}); *LRP5/PPP6R3* locus, leading SNP rs12272917-C (beta=-0.097SD, p -value= 1.42×10^{-9}); *WNT4* locus, leading SNP rs3765350-G (beta=-0.094SD, p -value= 9.75×10^{-9}); και *GALNT3* locus, leading SNP rs6726821-G (beta=-0.077SD, p -value= 2.95×10^{-8}) (Εικόνα 4.6). Για την TB-LM δεν βρέθηκε κάποιος πολυμορφισμός με p -value $< 5 \times 10^{-8}$. Από την άλλη πλευρά, η διμεταβλητή μετα-ανάλυση αποκάλυψε 8 γενετικούς τόπους με p -values $< 5 \times 10^{-8}$ (Εικόνα 4.7): 1p36.12, 2q24.3, 4q22.1, 7q31.31, 11q13.2, 13q14.11, 14q.2.12, 17p11.2, οι περισσότεροι εκ των οποίων συσχετίζονται με την έκβαση TB-BMD, εκτός από τα σήματα στις θέσεις 2q24.3, 11q13.2 και 17p11.2 όπου υπάρχουν ενδείξεις για συσχέτιση και με τις δυο εκβάσεις (p -value < 0.05).



Εικόνα 4.6. Manhattan plot του μονομεταβλητού μοντέλου μετα-ανάλυσης σταθερών επιδράσεων ($-\log_{10}(P\text{-values})$) για την έκβαση TB-BMD. Οι διακεκομμένες κόκκινες και κίτρινες γραμμές ορίζουν ως κατώφλι στατιστικής σημαντικότητας τα p -values 5×10^{-8} και 1×10^{-6} αντίστοιχα.



Εικόνα 4.7. Manhattan Plot του διμεταβλητού μοντέλου μετα-ανάλυσης σταθερών επιδράσεων για τις εκβάσεις TB-BMD και TB-LM. Οι διακεκομμένες κόκκινες και κίτρινες γραμμές ορίζουν ως κατώφλι στατιστικής σημαντικότητας τα p -values 5×10^{-8} και 1×10^{-6} αντίστοιχα.

4.4. Συζήτηση

Η μέθοδος που παρουσιάστηκε αποτελεί μια απευθείας επέκταση της κλασσικής μεθόδου πολυμεταβλητής μετα-ανάλυσης έτσι ώστε να μοντελοποιούνται ταυτόχρονα οι πολλαπλές συσχετισμένες εκβάσεις. Είναι γενική και μπορεί να χρησιμοποιηθεί είτε υπό την αρχική υπόθεση κάποιου γενετικού μοντέλου είτε χωρίς να κάνουμε κάποια υπόθεση. Η μέθοδος είναι απλή και γρήγορη και μπορεί να επεκταθεί για πολλαπλές εκβάσεις. Δεν απαιτούνται ατομικά δεδομένα ή η ταυτόχρονη αξιολόγηση και των δυο εκβάσεων σε κάθε μελέτη, δανειζόμενη έτσι ισχύ από εξωτερικές μελέτες (Higgins and Whitehead 1996). Μπορεί να υλοποιηθεί χρησιμοποιώντας οποιοδήποτε στατιστικό πακέτο για πολυμεταβλητή μετα-ανάλυση (Jackson, Riley et al. 2011; Mavridis and Salanti 2012), και τέλος, επιτρέπει την απευθείας σύγκριση των εκτιμώμενων παραμέτρων με επίσημους στατιστικούς ελέγχους.

Η μέθοδος εφαρμόστηκε σε διακριτές και συνεχείς εκβάσεις. Πιο συγκεκριμένα, βρέθηκε συσχέτιση του πολυμορφισμού C825T του γονιδίου GNB3 με την υπέρταση, είτε χωρίς να υποθέσουμε κάποιο μοντέλο κληρονομικότητας εκ των προτέρων είτε υποθέτοντας συνεπικρατές μοντέλο, και μια οριακά μη στατιστικά σημαντική εκτίμηση για το διαβήτη (υπολειπόμενο μοντέλο κληρονομικότητας). Στο συγκεκριμένο παράδειγμα, είναι προφανής η αύξηση της ισχύος της από κοινού μοντελοποίησης των δυο εκβάσεων, καθώς η εκτίμηση για το διαβήτη και το 95% C.I. αλλάζουν σημαντικά προς την κατεύθυνση απόρριψης της μηδενικής υπόθεσης. Στην περίπτωση του πολυμορφισμού Gly482Ser του γονιδίου PPARGC1A, η ανάλυση δεν

έδειξε κάποια διαφοροποίηση των επιπέδων της συστολικής και διαστολικής πίεσης. Η ανάλυση ευαισθησίας έδειξε μια μικρή-ανεπαίσθητη επίδραση στις εκτιμήσεις των παραμέτρων και στη συνολική στατιστική σημαντικότητα του μοντέλου. Επιπλέον, στη διμεταβλητή μετα-ανάλυση ευρυγονιδιωματικών μελετών των εκβάσεων TB-BMD και TB-LM (συνεπικρατές μοντέλο κληρονομικότητας και για τις δυο εκβάσεις) σε μελέτες κοορτής παιδικού πληθυσμού, ανιχνεύτηκαν οκτώ γενετικοί τόποι (in/near *WNT4*, *GALNT3*, *MEPE*, *CPED1/WNT16*, *TNFSF11*, *RIN3*, *TOM1L2/SREBF1* και *PPP6R3/LRP5*), τέσσερις επιπλέον σε σχέση με τη κλασσική μονομεταβλητή προσέγγιση, αν και για τους περισσότερους από αυτούς η στατιστική σημαντικότητα οφείλεται κυρίως στην έκβαση TB-BMD. Μόνο ο γενετικός τόπος 17p11.2 που περιλαμβάνει τα γονίδια *TOM1L2* και *SREBF1*, έδειξε πιο ισχυρή συσχέτιση με την TB-LM σε σχέση με την TB-BMD.

Η προτεινόμενη μεθοδολογία διαφέρει από άλλες υπάρχουσες μεθοδολογίες που παρέχουν σχέσεις για τον υπολογισμό των συνδιακυμάνσεων. Αποτελεί μια άμεση εφαρμογή της γενικής μεθόδου που έχει προταθεί παλαιότερα (Bagos 2012). Η μέθοδος της Wei και Higgins (Wei and Higgins 2013) χρησιμοποιήθηκε για την επίλυση του ίδιου προβλήματος, αλλά οι συγγραφείς ακολούθησαν μια ελαφρώς διαφορετική εκδοχή της μεθόδου Δέλτα και μια εκτίμηση για το συντελεστή συσχέτισης των δυο εκβάσεων. Όπως αποδείχτηκε αργότερα από τον Efthimiou και συνεργάτες (Efthimiou, Mavridis et al. 2013) οι δυο προσεγγίσεις είναι μαθηματικά ισοδύναμες (βλέπε το παράρτημα (Efthimiou, Mavridis et al. 2013)). Εκτός από το ότι η μέθοδος που προτείνουμε εστιάζεται στις μελέτες γενετικής συσχέτισης, η σημαντική διαφορά συνίσταται στην εύκολη λύση του προβλήματος των ελλειπουσών παρατηρήσεων με τη χρήση του αλγόριθμου IPF. Με άλλα λόγια, εάν είναι γνωστή η από κοινού ταξινόμηση των ατόμων, οι δυο μέθοδοι ταυτίζονται. Διαφορετικά, η προτεινόμενη μεθοδολογία παρέχει μια εύκολη λύση, εκτιμώντας τις ποσότητες του $2 \times 2 \times 3$ πίνακα συνάφειας (Πίνακας 4.3), χρησιμοποιώντας ένα log-linear μοντέλο υπό την υπόθεση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης. Η υπόθεση αυτή είναι πολύ γενική και απαιτούνται ισχυρές ενδείξεις για να απορριφθεί (για παράδειγμα η αλληλεπίδραση τρίτης τάξης θα σήμαινε ότι η συσχέτιση του διαβήτη με την υπέρταση διαφέρει στους διαφόρους γονοτύπους του γονιδίου GNB3).

Ένα σημείο που πρέπει να αναλυθεί είναι ο λόγος για τον οποίο η μέθοδος των Wei και Higgins, δεν απαιτεί επαναληπτική διαδικασία (και έτσι είναι μια φαινομενικά ευκολότερη μέθοδος). Σε κάποιο σημείο στους υπολογισμούς, οι Wei

και Higgins υποθέτουν ότι η συσχέτιση μεταξύ των δυο εκβάσεων είναι ίδια μεταξύ των ομάδων που ελέγχονται. Η υπόθεση αυτή ταυτίζεται με την υπόθεση που κάνουμε ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης; χρησιμοποιούμε το Odds Ratio για τη συσχέτιση διαβήτη-υπέρτασης, ενώ οι Wei και Higgins τον συντελεστή συσχέτισης. Ωστόσο, ένα σημαντικό ζήτημα αποτελεί το γεγονός ότι η υπόθεση ίσων Odds Ratios (ή ισοδύναμα ίσων συντελεστών συσχέτισης) ανάμεσα στους γονότυπους, δε σημαίνει απαραίτητα ότι αυτά είναι ίσα με το περιθώριο Odds Ratio (ή τον συντελεστή συσχέτισης). Με άλλα λόγια, δεν μπορούμε πάντα να χρησιμοποιήσουμε μια εκτίμηση της συσχέτισης διαβήτη-υπέρτασης και να την εξισώσουμε με τις επιμέρους συσχετίσεις ανάλογα με το γονότυπο (AA, AB και BB). Με την ορολογία πινάκων συνάφειας, η υπόθεση αυτή αναφέρεται ως «collapsibility» (Agresti 2002) και οι συνθήκες κάτω από τις οποίες ισχύει δεν μπορούν να καθοριστούν εκ των προτέρων. Για απλούστευση, εάν κάποιος ήθελε να υλοποιήσει τη μέθοδο των Wei και Higgins στο συγκεκριμένο παράδειγμα, όπου δεν είναι γνωστή η ταξινόμηση των ατόμων, θα έπρεπε να καταφύγει στον αλγόριθμο IPF ή σε κάποια παρόμοια προσέγγιση. Η μέθοδος των Efthimiou και συνεργατών από την άλλη πλευρά (Efthimiou, Mavridis et al. 2013) (η οποία προσανατολίζεται σε δίκτυα παρεμβάσεων, παρά σε μελέτες γενετικής συσχέτισης), μελετά ένα παρόμοιο πρόβλημα κάτω από μια μπεϋζιανή προσέγγιση αντλώντας πληροφορίες από κλινικούς. Σε μια παρόμοια περίπτωση που περιλαμβάνει την από κοινού μοντελοποίηση πολυμορφισμών που βρίσκονται σε ανισορροπία σύνδεσης και συσχετίζονται με την ίδια ασθένεια (Bagos and Liakopoulos 2010), η προσέγγιση είναι ελαφρώς διαφορετική καθώς η συσχέτιση των πολυμορφισμών (r^2 για την ανισορροπία σύνδεσης) προέρχεται από υγιείς (δηλαδή από το Harmap project ή τη σχετική βιβλιογραφία), και έτσι, δεν αποτελεί μέτρο περιθώριας συσχέτισης, αλλά μερικής συσχέτισης (και φυσικά υπό τη θεώρηση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης μπορεί να εξισωθεί με την αντίστοιχη εκτίμηση στους ασθενείς).

Τέλος, πρέπει να τονιστεί ότι μολονότι η μέθοδος αναπτύχθηκε και υλοποιήθηκε για την επίλυση ενός συγκεκριμένου προβλήματος σε μελέτες γενετικής συσχέτισης, μπορεί να επεκταθεί και σε άλλες κατηγορίες μελετών. Δεδομένα πολλαπλών εκβάσεων μπορεί να υπάρξουν σε διάφορα ερευνητικά πεδία, όπως οι κλινικές δοκιμές και οι μελέτες παρατήρησης, όπου η προτεινόμενη μεθοδολογία θα μπορούσε να αποτελεί ένα πολύτιμο εργαλείο. Η μέθοδος (και πιο συγκεκριμένα ο αλγόριθμος IPF) μπορεί να χρησιμοποιηθεί σε κάθε περίπτωση συσχετισμένων

εκβάσεων που δεν είναι αμοιβαίως αποκλειόμενες (η περίπτωση των αμοιβαίως αποκλειόμενων εκβάσεων σε κλινικές δοκιμές έχει παρουσιαστεί από τους Trikalinos και Olkin (Trikalinos and Olkin 2008)) όπως στην περίπτωση όπου ενδιαφερόμαστε να μελετήσουμε τις παρενέργειες μια συγκεκριμένης θεραπείας (Ioannidis, Mulrow et al. 2006; Golder, Loke et al. 2011; Loke, Golder et al. 2011). Έχει δειχθεί ότι οι παρενέργειες των φαρμάκων παρουσιάζουν αρκετά κοινά πρότυπα (κατά αναλογία με την παρουσία συνοδών νοσημάτων (Luo, Zhang et al. 2013)) και μπορούν να χρησιμοποιηθούν για τη διερεύνηση κοινών φαρμακευτικών στόχων (Campillos, Kuhn et al. 2008). Συνεπώς, στις περιπτώσεις αυτές, η προτεινόμενη μεθοδολογία μπορεί να χρησιμοποιηθεί εύκολα χωρίς τροποποιήσεις, αντλώντας πληροφορίες από τη βιβλιογραφία για τη συσχέτιση δυο ή περισσότερων παρενεργειών. Φυσικά, η μέθοδος μπορεί να χρησιμοποιηθεί σε πολυάριθμες εφαρμογές μελετών παρατήρησης, όπου ένας παράγοντας κινδύνου αξιολογείται για την εμπλοκή του σε διάφορες ασθένειες ή χαρακτηριστικά που είναι συσχετισμένα.

Συμπερασματικά, η προτεινόμενη μέθοδος υλοποιείται εύκολα και αποτελεί επέκταση της κλασσικής μετα-ανάλυσης μελετών γενετικής συσχέτισης. Πρόκειται για μια γενική μεθοδολογία που εφαρμόζεται ακόμα και όταν υπάρχουν ελλείπουσες παρατηρήσεις με την προϋπόθεση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης. Επικεντρωθήκαμε σε μελέτες γενετικής συσχέτισης, αλλά η μεθοδολογία, μέσω κατάλληλων τροποποιήσεων, μπορεί να επεκταθεί και σε άλλα ερευνητικά πεδία όπως η μετα-ανάλυση δικτύων. Κώδικας για την υλοποίηση του μοντέλου στο πακέτο Stata διατίθεται στην ιστοσελίδα www.compgen.org/tools/multiple-outcomes.

Μέρος Β: Διερεύνηση της συσχέτισης πολυμορφισμών της απολιποπρωτεΐνης Ε με την ανεπάρκεια αριστερής κοιλίας σε ασθενείς με βήτα θαλασσαιμία-μια πολυμεταβλητή μετα-ανάλυση

4.5. Εισαγωγή

Η βήτα θαλασσαιμία είναι μια κοινή νόσος και θεωρείται ένα παγκόσμιο πρόβλημα υγείας (Kremastinos, Farmakis et al. 2010). Ο επιπολασμός της νόσου είναι υψηλότερος στις μεσογειακές χώρες, με το υψηλότερο ποσοστό να καταγράφεται στην Κύπρο (14%) (Flint, Harding et al. 1998). Η βήτα θαλασσαιμία είναι επίσης κοινή σε πληθυσμούς αφρικανικής καταγωγής, Μέσω και Άπω Ανατολής, Κεντρικής Ασίας και Ινδίας. Η επίπτωση της νόσου εκτιμάται σε 1/100,000 παγκοσμίως ανά έτος, ενώ για την Ευρώπη υπολογίζεται σε 1/10,000 (Galanello and Origa 2010).

Η βήτα θαλασσαιμία προκαλείται από μειωμένη ή ελλιπή σύνθεση των βήτα αλυσίδων του τετραμερούς της αιμοσφαιρίνης με συνεπακόλουθο τον μη ικανοποιητικό βαθμό ερυθροποίησης και την αιμόλυση (Galanello and Origa 2010). Τα άτομα με μείζονα βήτα θαλασσαιμία, η οποία είναι η πιο συχνή μορφή της νόσου, εμφανίζουν σοβαρή αιμολυτική αναιμία και χρειάζονται τακτικές μεταγγίσεις αίματος κάθε τρεις ή τέσσερις εβδομάδες για την επιβίωσή τους. Στις επιπλοκές της θεραπείας τους περιλαμβάνεται η συγκέντρωση πλεονάζουσας ποσότητας σιδήρου. Μολονότι υπάρχουν διάφορες θεραπείες που στηρίζονται στη χρήση χηλικών παραγόντων, οι επιπλοκές στην καρδιακή λειτουργία είναι οι πιο συχνές αιτίες θνησιμότητας σε ασθενείς με περίσσεια σιδήρου (Rachmilewitz and Giardina 2011; Delaporta, Kattamis et al. 2013). Η ανεπάρκεια αριστερής κοιλίας (Left Ventricular Failure-LVF) είναι η πιο κοινή αιτία θανάτου σε ασθενείς με μείζονα βήτα θαλασσαιμία (Hahalis, Alexopoulos et al. 2005). Οι ασθενείς ταξινομούνται σε τρεις ομάδες ανάλογα με την σοβαρότητα των συμπτωμάτων που εμφανίζουν, με βάση την κλινική εκτίμηση των ιατρών και των αποτελεσμάτων των υπερηχογραφήματων. Πιο συγκεκριμένα, οι ομάδες αυτές είναι: i) ασυμπτωματικοί βήτα θαλασσαιμικοί ασθενείς (Ομάδα I) ii) ασθενείς με ενδείξεις LVF (Ομάδα II) και iii) ασθενείς με κλινικά και υπερηχογραφικά ευρήματα LVF (Ομάδα III) (Economou-Petersen, Aessopos et al. 1998).

Περισσότερες από 200 μεταλλάξεις έχουν αναφερθεί να συσχετίζονται με τη βήτα θαλασσαιμία (Weatherall 2001; Rund and Rachmilewitz 2005; Cao and

Galanello 2010). Το γονίδιο της απολιποπρωτεΐνης E (Apolipoprotein E-ApoE) εδράζεται στο χρωμόσωμα 19 και στον άνθρωπο υπάρχουν τρία κύρια αλληλόμορφα (E2, E3 και E4) με μειωμένο βαθμό αντιοξειδωτικής δράσης για τα αλληλόμορφα E3 και E4 σε σχέση με το E2. Τα αλληλόμορφα αυτά διαφέρουν σε τρία αμινοξικά κατάλοιπα στις θέσεις 112 και 158 και προκαλούν δυσλειτουργίες τόσο σε κυτταρικό όσο και σε μοριακό επίπεδο. Το αλληλόμορφο E3 είναι το πιο κοινό (>60%) σε όλους τους πληθυσμούς που έχουν μελετηθεί ως τώρα (Eichner, Dunn et al. 2002), ενώ τα άλλα δυο αλληλόμορφα είναι δυσλειτουργικά (Mahley and Rall 2000). Οι πολυμορφισμοί του γονιδίου ApoE έχουν συσχετιστεί με πολλές ασθένειες όπως η καρδιαγγειακή νόσος (Eichner, Dunn et al. 2002), η νόσος Alzheimer (Rubinsztein and Easton 1999), η ιδιοπαθής υπέρταση (Stoumpos, Hamodrakas et al. 2012) και ο διαβήτης τύπου 2 (Anthopoulos, Hamodrakas et al. 2010). Η απολιποπρωτεΐνη E δρα ως χηλικός παράγοντας ελευθέρων ριζών και οι ασθενείς που φέρουν το αλληλόμορφο E4 έχουν αυξημένο κίνδυνο για την καταστροφή κυτταρικών και υποκυτταρικών τμημάτων λόγω της αυξημένης ποσότητας σιδήρου (Gibson, Haroutunian et al. 2000) με συνεπακόλουθες επιπλοκές που περιλαμβάνουν την ανεπάρκεια αριστερής κοιλίας (Bazrgar, Karimi et al. 2007; El-Tagui, Hamdy et al. 2013). Η παρουσία του αλληλομόρφου E4 αν και δεν είναι από μόνη της ικανή να οδηγήσει στην ανάπτυξη LVF, λόγω της μειωμένης αντιοξειδωτικής του δράσης πρέπει να λαμβάνεται υπόψη ως προβλεπτικός γενετικός παράγοντας για την εμφάνιση καρδιακών επιπλοκών σε ομόζυγους βήτα θαλασσαιμικούς ασθενείς (Ferrara, Matarese et al. 2001).

Ένας μικρός αριθμός μελετών είναι διαθέσιμος στην βιβλιογραφία για τη διερεύνηση της συσχέτισης των πολυμορφισμών του γονιδίου ApoE με την εμφάνιση LVF. Αν και όλα τα μέχρι τώρα ευρήματα δείχνουν ότι το αλληλόμορφο E4 είναι ένας γενετικός παράγοντας κινδύνου για την εμφάνιση LVF, αξίζει να υπολογιστεί το πραγματικό μέτρο επίδρασης αυτού. Στην κατεύθυνση αυτή, μια μετα-αναλυτική προσέγγιση (Glass 1976; Petiti 1994; Greenland 1998; Normand 1999) θα μπορούσε να διαλευκάνει τους πιθανούς παράγοντες κινδύνου συνδυάζοντας τα αποτελέσματα των επιμέρους, ανεξάρτητων μελετών.

Στο συγκεκριμένο σχεδιασμό, οι μελέτες περιλαμβάνουν περισσότερες από μια εκβάσεις, μιας και οι ασθενείς ταξινομούνται σε τρεις ομάδες ανάλογα με την σοβαρότητα των συμπτωμάτων, και επιπλέον το γονίδιο ApoE έχει πολλαπλά αλληλόμορφα εισάγοντας έτσι έναν αυξημένο βαθμό πολυπλοκότητας στην ανάλυση.

Σε αυτές τις περιπτώσεις, συνίσταται η χρήση ενός πολυμεταβλητού μοντέλου μετα-ανάλυσης τυχαίων επιδράσεων έτσι ώστε να συνυπολογιστεί η συσχέτιση μεταξύ του λογαρίθμου των Odds Ratios μιας και για την εκτίμησή τους χρησιμοποιείται η ίδια κατηγορία αναφοράς (Bagos 2008; Bagos 2012). Η παρουσία των πολλαπλών αλληλομόρφων του γονιδίου ApoE, παραπέμπει στη μεθοδολογία που χρησιμοποιείται στην ανάλυση των απλότυπων (Bagos 2011) ενώ η παρουσία των πολλαπλών εκβάσεων μπορεί να αναλυθεί χρησιμοποιώντας τη γενική μεθοδολογία που περιγράφηκε στην Παράγραφο 4.2. Συνεπώς, στην εργασία αυτή, υιοθετούμε ένα σύνθετο σχήμα, το οποίο στηρίζεται στη γενική μέθοδο εκτίμησης των συνδιακυμάνσεων δυο συσχετισμένων λογαρίθμων των Odds Ratios (Bagos 2012) και παρουσιάζουμε σχέσεις για τις ανά δυο συγκρίσεις όλων των ομάδων των ασθενών που φέρουν ένα συγκεκριμένο αλληλόμορφο έναντι αυτών που βρίσκονται στην κατηγορία αναφοράς. Αξίζει να σημειώσουμε ότι αυτή η προσέγγιση είναι νέα, καθώς σε όλες τις προϋπάρχουσες μεθοδολογίες, υπήρχαν είτε δυο ομάδες με πολλαπλά αλληλόμορφα (Bagos 2011) ή γονοτύπους (Bagos 2008), είτε δυο ομάδες έκθεσης με πολλαπλές εκβάσεις (Trikalinos and Olkin 2008) ή θεραπείες (Higgins and Whitehead 1996), αλλά ποτέ και τα δυο στην ίδια ανάλυση.

4.6. Μεθοδολογία

4.6.1. Ανάκτηση δημοσιευμένων μελετών

Πραγματοποιήθηκε εκτεταμένη βιβλιογραφική αναζήτηση στη βάση δεδομένων PubMed έτσι ώστε να ανιχνευτούν οι μελέτες ασθενών-μαρτύρων που εστιάζονται σε πολυμορφισμούς του γονιδίου ApoE και συσχετίζονται με την LVF σε ασθενείς με μείζονα βήτα θαλασσαιμία έως και το Μάιο του 2015. Συναξιολογήθηκαν οι βιβλιογραφικές αναφορές κάθε άρθρου για σχετικές με το θέμα εργασίες. Για την αναζήτηση της βιβλιογραφίας, χρησιμοποιήθηκαν διάφορες λέξεις κλειδιά, όπως "Apolipoprotein E", "APOE", "allele", "polymorphism", "thalassemia" και "left ventricular failure". Το πλήρες κείμενο των άρθρων που ανακτήθηκαν διαβάστηκε με λεπτομέρεια έτσι ώστε να διαπιστωθεί η καταλληλότητα της κάθε μελέτης. Δεν θέσαμε κάποιο περιορισμό ως προς τη γλώσσα που χρησιμοποιείται. Όλες οι μελέτες είχαν επαρκώς ποιοτικά δεδομένα αν και δεν χρησιμοποιήθηκε η κλίμακα διαβάθμισης της ποιότητας των μελετών που συμπεριλάβαμε στην ανάλυση (Greenland 1998; Stroup, Berlin et al. 2000).

4.6.2. Εξαγωγή δεδομένων

Από κάθε μελέτη καταγράφηκαν το όνομα του πρώτου συγγραφέα, το έτος δημοσίευσης, η χώρα διεξαγωγής της μελέτης, βασικά χαρακτηριστικά των ασθενών και των μαρτύρων (π.χ. αριθμός ανδρών-γυναικών, μέση ηλικία και λοιπά), ο αριθμός των ασθενών και των μαρτύρων σε κάθε ομάδα καθώς και η αντίστοιχη συχνότητα των αλληλομόρφων E2, E3 και E4 της ApoE. Οι γονότυποι καταχωρήθηκαν αλλά δεν χρησιμοποιήθηκαν στην ανάλυση καθώς υπήρχαν δεδομένα μόνο για δυο μελέτες. Επίσης, για κάθε ομάδα ασθενών και μαρτύρων καταγράφηκαν η μέση τιμή της Τελοδιαστολικής Διαμέτρου Αριστερής Κοιλίας (Left Ventricular End Diastolic Diameter, LVEDD), της Τελοσυστολικής Διαμέτρου Αριστερής Κοιλίας (Left Ventricular End Systolic Diameter, LVESD), του Μεσοκοιλιακού Διαφράγματος (Interventricular Septum, IVS), του Πάχους Οπίσθιου Τοιχώματος (Posterior Wall Thickness, PWT), της Αιμογλοβίνης (Hemoglobin, Hb) και του αριθμού των Μονάδων Μετάγγισης Αίματος (Blood Transfusion Units, BTU).

4.6.3. Στατιστική ανάλυση

Η σύγκριση των αλληλομόρφων μεταξύ των διαφορετικών ομάδων ασθενών και μαρτύρων πραγματοποιήθηκε με τον υπολογισμό του Odds Ratio. Τα μηδενικά κελιά αντικαταστάθηκαν προσθέτοντας το 0.5 σε όλα τα κελιά του πίνακα συνάφειας. Αρχικά, για να αποφύγουμε το σφάλμα πολλαπλών συγκρίσεων που προκύπτει από τη σύγκριση του κάθε αλληλομόρφου με ένα αλληλόμορφο αναφοράς, εφαρμόστηκε η πολυμεταβλητή μέθοδος μετα-ανάλυσης τυχαίων επιδράσεων όπου λαμβάνονται υπόψη οι συσχετίσεις των Odds Ratios (Bagos 2012). Το αλληλόμορφο E3 επιλέχτηκε ως κατηγορία αναφοράς καθώς ήταν το πιο συχνό ανάμεσα σε υγιείς και ασθενείς.

Στον Πίνακα 4.8 παρουσιάζεται η τυπική μορφή των δεδομένων μιας μετα-ανάλυσης ασθενών-μαρτύρων μελετών γενετικής συσχέτισης για έναν γενετικό τόπο με πολλαπλά αλληλόμορφα ($m=1, 2, \dots, p$) και πολλαπλές εκβάσεις ($j=1, 2, \dots, r$). Οι λογάριθμοι των $(r-1)(p-1)$ Odds Ratios για κάθε έκβαση ($j=2, \dots, r$) και κάθε αλληλόμορφο ($m=2, \dots, p$) έναντι των κατηγοριών αναφοράς (δηλαδή για την έκβαση 1 και αλληλόμορφο 1) με τις αντίστοιχες διασπορές, για κάθε i μελέτη ($i=1, 2, \dots, k$) δίνονται από τη γνωστή σχέση του Woolf :

$$y_{jmi} = \log \left(\frac{n_{jmi} n_{11i}}{n_{1mi} n_{j1i}} \right) \quad (4.30)$$

και:

$$s_{jmi}^2 = \frac{1}{n_{jmi}} + \frac{1}{n_{1mi}} + \frac{1}{n_{j1i}} + \frac{1}{n_{11i}} \quad (4.31)$$

Ακολουθώντας το γενικό πολυμεταβλητό μοντέλο μετα-ανάλυσης (van Houwelingen, Arends et al. 2002), ορίζουμε με \mathbf{y}_i το διάνυσμα που περιέχει τις $(r-1)(p-1)$ διαφορετικές εκτιμήσεις και με $\boldsymbol{\beta}$ το διάνυσμα των μέσων. Σε μια πολυμεταβλητή μετα-ανάλυση τυχαίων επιδράσεων υποθέτουμε ότι το \mathbf{y}_i ακολουθεί μια πολυμεταβλητή κανονική κατανομή με μέσους τους πραγματικούς μέσους $\boldsymbol{\beta}$, σύμφωνα με το ακόλουθο περιθώριο μοντέλο:

$$\mathbf{y}_i \sim MVN(\boldsymbol{\beta}, \boldsymbol{\Sigma} + \mathbf{C}_i) \quad (4.32)$$

Σχέση για τον υπολογισμό της συνδιακύμανσης των λογαρίθμων των Odds Ratios για την ίδια έκβαση, αν δεν υποθέσουμε κάποιο μοντέλο κληρονομικότητας (Minelli, Thompson et al. 2005), έχουν παρουσιαστεί κατά το παρελθόν (Bagos 2008). Με την ίδια λογική, υπολογίστηκαν σχέσεις για τον υπολογισμό των συνδιακυμάνσεων στην περίπτωση των πολλαπλών αλληλομόρφων και πολλαπλών εκβάσεων. Πιο συγκεκριμένα, η συνδιακύμανση του λογαρίθμου των Odds Ratios για κάθε έκβαση ($j=2, \dots, r$) έναντι της κατηγορίας αναφοράς για την ίδια σύγκριση αλληλομόρφων, $\forall j \neq j'$, υπολογίζεται ως:

$$\text{cov}(y_{jmi}, y_{j'mi}) = \frac{1}{n_{1mi}} + \frac{1}{n_{11i}} \quad (4.33)$$

Η συνδιακύμανση του λογαρίθμου των Odds Ratios για κάθε αλληλόμορφο ($m=2, \dots, p$) έναντι της κατηγορίας αναφοράς για την ίδια σύγκριση εκβάσεων, $\forall m \neq m'$, υπολογίζεται ως:

$$\text{cov}(y_{jmi}, y_{jm'i}) = \frac{1}{n_{j1i}} + \frac{1}{n_{11i}} \quad (4.34)$$

Η συνδιακύμανση του λογαρίθμου των Odds Ratios για κάθε έκβαση και αλληλόμορφο, έναντι των κατηγοριών αναφοράς, $\forall j \neq j'$ και $\forall m \neq m'$, υπολογίζεται ως:

$$\text{cov}(y_{jmi}, y_{j'm'i}) = \frac{1}{n_{11i}} \quad (4.35)$$

Εφόσον ολοκληρωθεί η εκτίμηση των παραμέτρων του μοντέλου, μπορεί να χρησιμοποιηθεί ο έλεγχος Wald για να υπολογιστεί η στατιστική σημαντικότητα του μοντέλου. Επιπλέον, μπορούν να διαμορφωθούν και άλλοι έλεγχοι αν εκμεταλλευτούμε την αναλογία του παρόντος προβλήματος με την γνωστή μέθοδο ANOVA διαμορφώνοντας τους λεγόμενους “planned or a priori contrasts” και “post hoc or a posteriori contrasts”. Για παράδειγμα, μπορούμε να διαμορφώσουμε μια “planned” ανάλυση τάσης των $(r-1)$ Odds Ratios για κάθε έκβαση έναντι της κατηγορίας αναφοράς εντός των $(p-1)$ συγκρίσεων των αλληλομόρφων. Προτού πραγματοποιήσουμε μια “planned” ανάλυση πολλαπλών συγκρίσεων, πρέπει να αξιολογήσουμε εάν οι συγκρίσεις αυτές είναι ορθογώνιες. Δυο ή περισσότερες συγκρίσεις είναι ορθογώνιες όταν οι συντελεστές των αντίστοιχων συγκρίσεων δεν είναι συσχετισμένοι και αυτού του είδους η ανεξαρτησία οδηγεί σε πιο ισχυρούς στατιστικούς ελέγχους. Έτσι, η ανάλυση τάσης μπορεί να διαμορφωθεί ως ένα σύνολο ορθογώνιων συγκρίσεων με τη χρήση καθορισμένων συντελεστών $(-1, 0, 1)$ για τη γραμμική μορφή τάσης). Οι έλεγχοι μπορούν να υλοποιηθούν εύκολα με τη μέθοδο Δέλτα. Χρησιμοποιώντας την προσέγγιση αυτή μπορούμε να εκτιμήσουμε εάν ο υπό μελέτη πολυμορφισμός συσχετίζεται με αυξημένο κίνδυνο εμφάνισης βαρύτερων συμπτωμάτων της νόσου (δηλαδή μεταβαίνοντας από την Ομάδα I στην Ομάδα III) και συγκεκριμένα ποιο αλληλόμορφο παρουσιάζει το μεγαλύτερο κίνδυνο. Τέλος, ο έλεγχος Wald θα μπορούσε να δείξει εάν η διαφορά του λογαρίθμου των Odds Ratios κάθε σύγκρισης αλληλομόρφων διαφέρει μεταξύ των εκβάσεων. Ο έλεγχος αυτός παρέχει ένδειξη για παράλληλη τάση της επίδρασης του γονιδιακού πολυμορφισμού μεταξύ των πολλαπλών εκβάσεων.

Τα αποτελέσματα της ανάλυσης τάσης και του ελέγχου του Wald μας κατευθύνουν στην ενοποίηση των αλληλομόρφων ή των ομάδων των ασθενών και στον υπολογισμό των αντίστοιχων συνδυασμένων Odds Ratios (δηλαδή τα αλληλόμορφα E2 συγκρίνονται έναντι όλων των υπολοίπων και λοιπά) ακολουθώντας ξανά μια πολυμεταβλητή προσέγγιση. Ο δείκτης ασυνέπειας I^2 (με τιμές από 0–100%) υπολογίστηκε για να εκτιμηθεί η ετερογένεια μεταξύ των μελετών (Higgins, Thompson et al. 2003). Επιπρόσθετα, η χρήση των μοντέλων τυχαίων επιδράσεων επιτρέπει την επέκταση της ανάλυσης λαμβάνοντας υπόψη

συμμεταβλητές, έτσι ώστε να εκτιμηθεί ο βαθμός που αυτές συνεισφέρουν στην ερμηνεία της παρατηρούμενης μεταβλητότητας. Έτσι, μπορούμε να διεξάγουμε μια μετα-παλινδρόμηση (Thompson and Sharp 1999) όπου αξιολογείται η σχέση μεταξύ της επίδρασης του αλληλομόρφου με κλινικές ή αιματολογικές παραμέτρους. Πιο συγκεκριμένα, ως συμμεταβλητές χρησιμοποιήθηκαν διάφορα συνοπτικά χαρακτηριστικά που ανακτήθηκαν από κάθε μελέτη, όπως η ποσοστιαία κατανομή του φύλου σε ασθενείς και υγιείς, η μέση ηλικία, LVEDD, LVESD, IVS, PWT, Hb, ο αριθμός των BTU, το γεωγραφικό πλάτος που διεξήχθη κάθε μελέτη και λοιπά. Πραγματοποιήθηκε αθροιστική μετα-ανάλυση (Lau, Antman et al. 1992; Lau, Schmid et al. 1995) για να διερευνηθεί εάν οι συνδυασμένες εκτιμήσεις διαφοροποιούνται όσο συσσωρεύεται ο όγκος της πληροφορίας (Ioannidis and Trikalinos 2005) καθώς και μια μεθοδολογία που στηρίζεται σε μεθόδους παλινδρόμησης (Bagos and Nikolopoulos 2009).

Για όλες τις αναλύσεις που πραγματοποιήθηκαν, χρησιμοποιήθηκε το στατιστικό πακέτο Stata v.13 (Stata Corporation, College Station, Texas, USA). Σε όλες τις αναλύσεις, ως στατιστικά σημαντικά αποτελέσματα δηλώθηκαν εκείνα με $p\text{-value} < 0.05$.

Πίνακας 4.8. Η τυπική μορφή των δεδομένων μιας μετα-ανάλυσης ασθενών-μαρτύρων μελετών γενετικής συσχέτισης για έναν γενετικό τόπο με πολλαπλά αλληλόμορφα και πολλαπλές εκβάσεις. Παρουσιάζεται η κατανομή των διαφόρων αλληλομόρφων ($m=1, 2, \dots, p$) ανάλογα με την έκβαση ($j=1,2, \dots, r$), ενώ σε κάθε γραμμή καταγράφεται η κάθε μελέτη ($i=1, 2, \dots, k$) που περιλαμβάνεται στην μετα-ανάλυση.

Μελέτη	Έκβαση 1				Έκβαση 2				...	Έκβαση r			
	Αλληλόμορφο 1	Αλληλόμορφο 2	...	Αλληλόμορφο p	Αλληλόμορφο 1	Αλληλόμορφο 2	...	Αλληλόμορφο p		Αλληλόμορφο 1	Αλληλόμορφο 2	...	Αλληλόμορφο p
1	n_{111}	n_{121}		n_{1p1}	n_{211}	n_{221}		n_{2p1}		n_{r11}	n_{r21}		n_{rp1}
2	n_{112}	n_{122}		n_{1p2}	n_{212}	n_{222}		n_{2p2}		n_{r12}	n_{r22}		n_{rp2}
...													
k	n_{11k}	n_{12k}		n_{1pk}	n_{21k}	n_{22k}		n_{2pk}		n_{r1k}	n_{r2k}		n_{rpk}

4.7. Αποτελέσματα

Από τη βιβλιογραφική αναζήτηση ανακτήθηκαν 10 μελέτες από τις οποίες οι 4 πληρούσαν τα απαραίτητα κριτήρια για να συμπεριληφθούν στη μετα-ανάλυση με συνολικά 613 βήτα θαλασσαιμικούς ασθενείς και 664 μάρτυρες. Όλες οι μελέτες ήταν στα αγγλικά. Στο Πίνακα 4.9 συνοψίζονται τα χαρακτηριστικά των μελετών που συμπεριλήφθηκαν στη μετα-ανάλυση (η χώρα, τα χαρακτηριστικά των ασθενών και των υγιών, το μέγεθος δείγματος και η συχνότητα των αλληλομόρφων). Τρεις μελέτες είχαν διεξαχθεί σε χώρες της Μεσογείου και μια στην Ασία. Δεν μπορούσε να πραγματοποιηθεί μετα-ανάλυση για τις συγκρίσεις των γονοτύπων καθώς μόνο δυο μελέτες παρείχαν την πληροφορία αυτή.

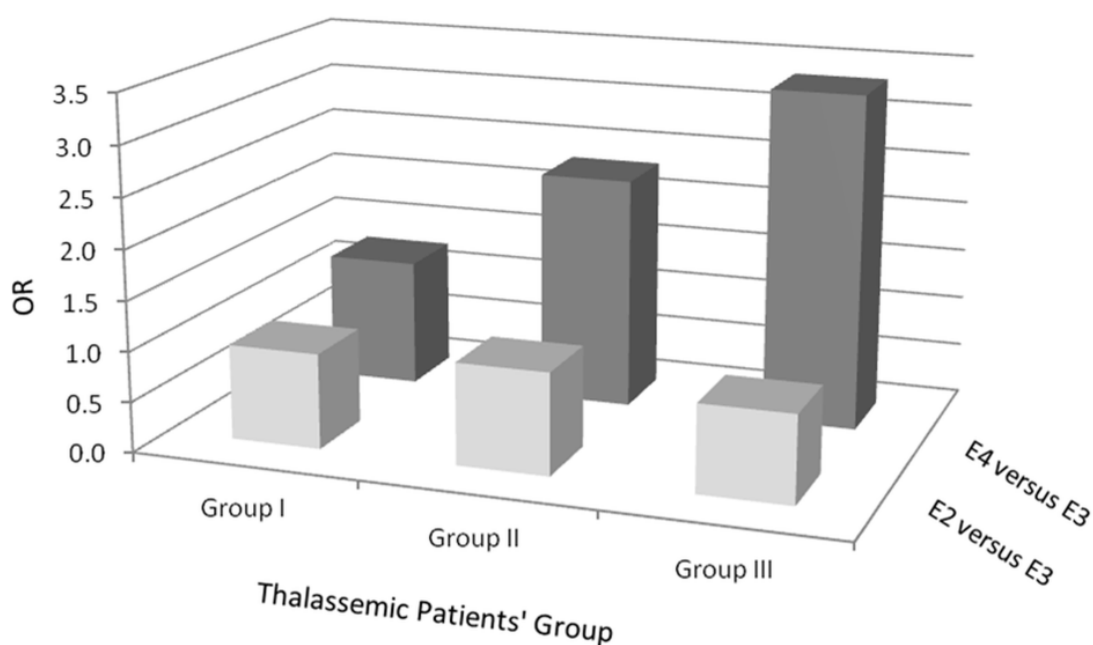
Η πολυμεταβλητή μετα-ανάλυση, θέτοντας το E3 αλληλόμορφο ως αλληλόμορφο αναφοράς, αποκάλυψε μια ισχυρά σημαντική συσχέτιση του υπό μελέτη γονιδίου με την εμφάνιση LVF (Wald test p -value=0.009). Πιο συγκεκριμένα, τα Odds Ratios του αλληλομόρφου E4 έναντι του E3 υπολογίστηκαν 2.32 (95% C.I.: 1.19, 4.53) και 3.34 (95% C.I.: 1.78, 6.26) για την Ομάδα II και III αντίστοιχα (Εικόνα 4.8 και Πίνακας 4.10). Η ανάλυση τάσης έδειξε οριακά μη στατιστικά σημαντικό αποτέλεσμα ($\chi^2(2)=5.55$, p -value=0.062) υποδηλώνοντας πιθανή γραμμική τάση των 3 Odds Ratios κάθε έκβασης (ένα για κάθε ομάδα) έναντι της κατηγορίας αναφοράς εντός των δύο επιπέδων σύγκρισης αλληλομόρφων (σύγκριση του αλληλομόρφου E2 και E4 έναντι του αλληλομόρφου αναφοράς E3). Επιπλέον, κατά τον έλεγχο της ισότητας των διαφορών του λογαρίθμου των Odds Ratios των αλληλομόρφων E2 και E4 έναντι του E3 μεταξύ των τριών ομάδων ασθενών δεν φάνηκε να αποτυπώνεται κάποιο πρότυπο στο προφίλ γονιδιακής επίδρασης.

Πίνακας 4.9. Τα χαρακτηριστικά των μελετών που συμπεριλήφθησαν στη μετα-ανάλυση.

Μελέτη	Έτος	Χαρακτηριστικά ασθενών	Χαρακτηριστικά υγιών	Χώρα	Σύνολο				Αλληλόμορφο E2				Αλληλόμορφο E3				Αλληλόμορφο E4			
					Ομάδες			Υγιείς	Ομάδες			Υγιείς	Ομάδες			Υγιείς	Ομάδες			Υγιείς
					I	II	III		I	II	III		I	II	III		I	II	III	
El-Tagui (El-Tagui, Hamdy et al. 2013)	2013	50 ασθενείς (29 άντρες και 21 γυναίκες, ηλικίας ≥11 ετών) που μεταγγίζονταν κάθε 2-4 εβδομάδες για να συντηρούν τα επίπεδα αιμοσφαιρίνης >9 g/dl και λάμβαναν θεραπεία αποσιδήρωσης	50 υγιείς εθελοντές (εξομοίωση με ασθενείς για φύλο και ηλικία)	Αίγυπτος	9	26	15	50	2	5	4	6	15	38	20	88	1	9	6	6
Bazrgar (Bazrgar, Karimi et al. 2007)	2007	202 ασθενείς (ηλικίας ≥10 ετών) μεταγγίζονταν για να συντηρούν τα επίπεδα αιμοσφαιρίνης > 9 g/dL και λάμβαναν υποδόρια θεραπεία αποσιδήρωσης με deferoxamine	198 υγιείς (95 άντρες και 103 γυναίκες, ηλικίας 23.7±5.2 ετών) από την ίδια γεωγραφική περιοχή	Ιράν	135	38	29	198	14	7	1	25	234	65	49	351	22	4	8	20
Ferrera (Ferrara, Matarese et al. 2001)	2001	110 ασθενείς και των δυο φύλων, ηλικίας μεταξύ 5 και 34 ετών, που μεταγγίζονταν σε τακτά χρονικά διαστήματα (επίπεδα αιμοσφαιρίνης 9.5 ±1.2 g/dl) και λάμβαναν θεραπεία αποσιδήρωσης	200 υγιή άτομα	Ιταλία	88	12	10	200	18	0	0	52	151	20	15	332	7	4	5	16
Economou-Petersen (Economou-Petersen, Aessopos et al. 1998)	1998	251 ασθενείς (117 άντρες και 134 γυναίκες), που μεταγγίζονταν κάθε 15-25 ημέρες για να συντηρούν τα επίπεδα αιμοσφαιρίνης > 9 g/dL και λάμβαναν υποδόρια θεραπεία αποσιδήρωσης με deferoxamine	216 υγιείς δότες αίματος (146 άντρες και 70 γυναίκες, ηλικίας 35.6 ετών με εύρος 19-64 έτη)	Ελλάδα	151	53	47	216	17	5	6	23	261	88	76	381	24	13	12	28

Πίνακας 4.10. Τα αποτελέσματα της πολυμεταβλητής μετα-ανάλυσης. Παρουσιάζονται τα Odds Ratios και τα 95% Confidence Intervals για κάθε ομάδα ασθενών σε σύγκριση με τους υγιείς (αλληλόμορφο αναφοράς το E3), ο δείκτης ασυνέπειας (I^2) και τα αποτελέσματα από το συνολικό έλεγχο του Wald.

Σύγκριση	Ομάδα	OR	95% CI		I^2 (%)	χ^2	p-value
E2 έναντι του E3 αλληλομόρφου	Ομάδα I	0.95	0.63	1.44	18	17.19	0.009
	Ομάδα II	1.01	0.41	2.49	33		
	Ομάδα III	0.87	0.26	2.95	47		
	Ομάδα I	1.27	0.80	2.02	19		
E4 έναντι του E3 αλληλομόρφου	Ομάδα II	2.32	1.19	4.53	44		
	Ομάδα III	3.34	1.78	6.26	40		



Εικόνα 4.8. Γραφική απεικόνιση των αποτελεσμάτων της πολυμεταβλητής μετα-ανάλυσης. Οι μπάρες απεικονίζουν τα Odds Ratios για κάθε ομάδα ασθενών σε σύγκριση με τους υγιείς (αλληλόμορφο αναφοράς το E3).

Βάσει των ανωτέρω αποτελεσμάτων, τα αλληλόμορφα ομαδοποιήθηκαν και ακολούθησε η σύγκριση του αλληλομόρφου E4 με τα E2+E3 για τις διάφορες ομάδες των ασθενών. Από την πολυμεταβλητή μετα-ανάλυση υπολογίστηκε Odds Ratio ίσο με 2.49 (95% C.I.: 1.28, 4.86) για την ομάδα ασθενών II, ενώ για την ομάδα ασθενών III το Odds Ratio ήταν 3.43 (95% C.I.: 1.84, 6.41) (Πίνακας 4.11). Επιπλέον, ο έλεγχος του Wald έδωσε στατιστικά σημαντικά αποτελέσματα ($\chi^2(3)=16.80$, $p\text{-value}<0.001$), γεγονός που ισχυροποιεί τη σημαντικότητα της συσχέτισης του αλληλομόρφου E4 με την LVF σε ασθενείς με βήτα θαλασσαιμία. Η ετερογένεια μεταξύ των μελετών ήταν μέτρια ($I^2<46.0\%$) ενώ διαπιστώθηκε διαχρονική τάση στην περίπτωση των Ομάδων I και III.

Από την ανάλυση μετα-παλινδρόμησης δεν προέκυψαν αποτελέσματα που να αποδεικνύουν ότι ο κίνδυνος που αποδίδεται στο αλληλόμορφο E4 του γονιδίου ApoE συσχετίζεται με κλινικές ή αιματολογικές παραμέτρους αν και αυτό μπορεί να οφείλεται στα λίγα διαθέσιμα δεδομένα.

Πίνακας 4.11. Τα αποτελέσματα της πολυμεταβλητής μετα-ανάλυσης με ομαδοποιημένα αλληλόμορφα. Παρουσιάζονται τα Odds Ratios και τα 95% Confidence Intervals για κάθε ομάδα ασθενών σε σύγκριση με τους υγιείς, ο δείκτης ασυνέπειας (I^2) και τα αποτελέσματα από τον έλεγχο του Wald (Το αλληλόμορφο E4 συγκρίνεται έναντι των αλληλομόρφων E2+E3).

Σύγκριση	Ομάδα	OR	95% CI		I^2 (%)	χ^2	p-value
E4 έναντι των E2+E3 αλληλομόρφων	Ομάδα I	1.27	0.80	2.01	16		
	Ομάδα II	2.49	1.28	4.86	46	16.80	<0.001
	Ομάδα III	3.43	1.84	6.41	40		

4.8. Συζήτηση

Οι ασθενείς με βήτα θαλασσαιμία χαρακτηρίζονται από μειωμένη ή μηδενική έκφραση του γονιδίου της βήτα σφαιρίνης. Παρά τις συνεχείς μεταγγίσεις αίματος και τη θεραπεία με χηλικούς παράγοντες, οι ασθενείς εξακολουθούν να εμφανίζουν επιπλοκές σε διάφορα όργανα. Η καρδιακή ανεπάρκεια αποτελεί την κύρια αιτία θανάτου σε ομόζυγους ασθενείς για βήτα θαλασσαιμία. Το αλληλόμορφο E4 του γονιδίου ApoE, συγκρινόμενο με τα αλληλόμορφα E2 και E3, παρουσιάζει τη χαμηλότερη αντιοξειδωτική δράση, και έτσι οι ασθενείς που φέρουν το αλληλόμορφο αυτό έχουν πιθανώς αυξημένο κίνδυνο για οργανικές βλάβες οφειλούμενες στην υπερφόρτωση σιδήρου. Δεδομένου ότι ένας μικρός αριθμός μελετών έχουν διερευνήσει την πιθανή επίδραση πολυμορφισμών του γονιδίου ApoE στην LVF κρίνεται αναγκαίος ο συνδυασμός των αποτελεσμάτων των επιμέρους μελετών μέσω διενέργειας μετα-ανάλυσης έτσι ώστε να εκτιμηθεί το συνολικό μέγεθος της επίδρασης.

Η παρούσα μελέτη αποτελεί την πρώτη μετα-ανάλυση για τη διερεύνηση της συσχέτισης πολυμορφισμών του γονιδίου ApoE στην LVF. Συνολικά, συμπεριλάβαμε 613 ασθενείς και 664 μάρτυρες από 4 μελέτες. Προηγούμενες μελέτες έχουν δείξει ότι δεν υπάρχει διαφορά στη συχνότητα του αλληλομόρφου E4 του γονιδίου ApoE μεταξύ ομόζυγων ασθενών με βήτα θαλασσαιμία που δεν εμφανίζουν καρδιακές επιπλοκές και υγιών, ενώ καταδεικνύουν υψηλότερη συχνότητα σε ασθενείς με LVF. Πιο συγκεκριμένα, εκτιμήθηκε ότι το αλληλόμορφο E4 διπλασιάζει τον κίνδυνο για LVF σε ασθενείς με βήτα θαλασσαιμία (OR: 2.11, 95% C.I.: 1.03, 4.32) (Economou-Petersen, Aessopos et al. 1998). Τα αποτελέσματα της παρούσας μελέτης, συνοψίζοντας όλη την υπάρχουσα πληροφορία, υποδεικνύουν ότι το αλληλόμορφο E4 του γονιδίου ApoE τριπλασιάζει τον κίνδυνο για την εμφάνιση LVF στους ασθενείς με βήτα θαλασσαιμία. Η ανάλυση μετα-παλινδρόμησης απέτυχε να δείξει κάποια συσχέτιση μεταξύ κλινικών ή αιματολογικών παραμέτρων με την επίδραση του γονιδίου ApoE.

Η συσχέτιση του αλληλομόρφου E4 του γονιδίου ApoE με την LVF σε ασθενείς με βήτα θαλασσαιμία ενισχύει την υπάρχουσα θεωρία των ελευθέρων ριζών για την πρόκληση οργανικών βλαβών. Η χαμηλή αντιοξειδωτική δράση που αποδίδεται στο αλληλόμορφο E4 έχει συσχετιστεί με αυξημένο κίνδυνο για τη νόσο Alzheimer, καρδιαγγειακά νοσήματα και μικρότερη διάρκεια ζωής (Miyata and Smith

1996). Μολονότι το αλληλόμορφο E4 φαίνεται να αποτελεί παράγοντα κινδύνου για την LVF, αξίζει να αναφερθεί ότι η έκβαση αυτή είναι πολυπαραγοντικής αιτιολογίας. Πιο συγκεκριμένα, ομόζυγοι θαλασσαιμικοί ασθενείς για τους πολυμορφισμούς H63D (Melis, Cau et al. 2002) ή C282Y (Piperno, Mariani et al. 2000) του γονιδίου HFE φαίνεται να παρουσιάζουν αυξημένη συγκέντρωση σιδήρου έχοντας έτσι αυξημένο κίνδυνο για την πρόκληση πολλαπλών επιπλοκών. Αντικρουόμενα ήταν τα αποτελέσματα μιας προγενέστερης μελέτης, που απέτυχε να δείξει συσχέτιση μεταξύ των πολυμορφισμών H63D ή C282Y με αυξημένα επίπεδα σιδήρου κυρίως λόγω της αποτελεσματικότητας της παρεχόμενης θεραπείας (Longo, Zecchina et al. 1999).

Το κυριότερο μειονέκτημα της παρούσας μελέτης ήταν η απουσία της ανάλυσης των γονοτύπων λόγω της απουσίας των πρωταρχικών δεδομένων. Παρόλα αυτά, πραγματοποιήσαμε ενδελεχή αναζήτηση της βιβλιογραφίας, ακόμα και άρθρων που δεν υπήρχαν σε ηλεκτρονικές βάσεις δεδομένων (Conn, Valentine et al. 2003), χωρίς να είναι κατ' ανάγκη γραμμένα στα Αγγλικά (Pan, Trikalinos et al. 2005). Επιπλέον, εφαρμόσαμε νέες προηγμένες μεθόδους μετα-ανάλυσης (Bagos and Nikolopoulos 2007; Bagos 2008) και στατιστικούς ελέγχους διερεύνησης του συστηματικού σφάλματος δημοσίευσης (Egger, Davey Smith et al. 1997; Sterne, Gavaghan et al. 2000). Το στατιστικό μοντέλο που παρουσιάστηκε, μπορεί να είναι χρήσιμο σε πολλές εφαρμογές, καθώς αποτελεί την πρώτη μέθοδο που αξιολογεί ταυτόχρονα πολλαπλές εκθέσεις (δηλαδή αλληλόμορφα) και πολλαπλές ομάδες σύγκρισης σε μελέτες γενετικής συσχέτισης. Μελλοντικές μελέτες απαιτούνται για να διαλευκανθεί ο μηχανισμός της συσχέτισης του γονιδίου ApoE με την LVF σε ομόζυγους βήτα θαλασσαιμικούς ασθενείς και να τονιστεί η πιθανή αλληλεπίδραση του γονιδίου αυτού με άλλα γονίδια.

Κεφάλαιο 5: Μετα-αναλύσεις μελετών γενετικής συσχέτισης, κλινικών μελετών και μελετών φαρμακογονιδιωματικής

Μέρος Α: Πολυμορφισμοί των γονιδίων των ιντερλευκινών και προδιάθεση για μόλυνση από τον HIV-1

5.1. Εισαγωγή

Παρά τη μείωση της νοσηρότητας και της θνησιμότητας των ατόμων που έχουν μολυνθεί από τον ιό της ανθρώπινης ανοσοανεπάρκειας (Human Immunodeficiency Virus-HIV) (Puhan, Van Natta et al. 2010), η ασθένεια που προκαλείται από τον HIV παραμένει ένα μείζον παγκόσμιο πρόβλημα. Η έλλειψη θεραπείας ή εμβολίων, και οι περίπου 35 εκατομμύρια ασθενείς με HIV παγκοσμίως, καθιστούν αναγκαία την πρόληψη της νόσου. Η διαλεύκανση των αιτιολογικών μηχανισμών που οδηγούν στη μετάδοση του HIV είναι απαραίτητη για την διαμόρφωση καινοτόμων προληπτικών και θεραπευτικών μέτρων.

Εξαιρετικά ενδιαφέρον είναι το γεγονός ότι κάποια άτομα παραμένουν υγιή παρά τις πολλαπλές εκθέσεις στον HIV, ενώ άλλα μολύνονται από την πρώτη κιάλας έκθεση (Jennes, Vuylsteke et al. 2004; Levy 2009). Το φαινόμενο αυτό υπονοεί την ύπαρξη φυσικής ή επίκτητης ανοσίας (Kaur and Mehra 2009; Miyazawa, Lopalco et al. 2009). Έχει προταθεί ότι γενετικοί παράγοντες μπορεί να συσχετίζονται με την προδιάθεση για HIV και την εκδήλωση του συνδρόμου επίκτητης ανοσοανεπάρκειας (Acquired Immune Deficiency Syndrome-AIDS). Έτσι, η έλλειψη 32 ζευγών βάσεων του γονιδίου CCR5 (CCR5Δ32), που κωδικοποιεί τον κύριο υποδοχέα του HIV-1, οδηγεί σε ένα μικρότερο πρωτεϊνικό προϊόν που δεν εκφράζεται στην επιφάνεια των κυττάρων, προσδίδοντας έτσι πλήρη προστασία έναντι της μόλυνσης από τον HIV-1 σε ομόζυγους (Dean, Carrington et al. 1996; Samson, Libert et al. 1996; Ioannidis, Rosenberg et al. 2001).

Οι κυτταροκίνες είναι ρυθμιστικές πρωτεΐνες ή γλυκοπρωτεΐνες χαμηλού μοριακού βάρους, που εκκρίνονται από λεμφοκύτταρα και διάφορους άλλους κυτταρικούς τύπους του σώματος, ως απόκριση σε ένα πλήθος ερεθισμάτων. Οι πρωτεΐνες αυτές ασκούν επικουρικό ρόλο στη ρύθμιση της ανάπτυξης των

ανοσοδραστικών κυττάρων των ξενιστών (Alfano, Crotti et al. 2008; Levy 2009). Έχει βρεθεί ένας μεγάλος αριθμός γονιδίων που κωδικοποιούν κυτταροκίνες. Πολλοί ερευνητές έχουν εξετάσει τη συσχέτιση πολυμορφισμών των γονιδίων των ιντερλευκινών (interleukin-IL) με τον κίνδυνο μόλυνσης από τον HIV, με αντικρουόμενα αποτελέσματα (Pontillo, Oshiro et al. 2010; Sobti, Berhane et al. 2012). Στην εποχή των μαζικών και χαμηλού κόστους τεχνικών αλληλούχισης, ο καθορισμός γενετικών παραγόντων που συνεισφέρουν στον κίνδυνο μόλυνσης από τον HIV θα μπορούσε να αποτελέσει ένα σημαντικό εργαλείο στρατηγικών για πρόληψη και θεραπεία. Έτσι στοχεύσαμε στη σύνθεση της υπάρχουσας γνώσης, ώστε να ανιχνευτούν πολυμορφισμοί των γονιδίων των ιντερλευκινών που συσχετίζονται με την προδιάθεση για HIV, και να διερευνηθούν οι πηγές της μεταξύ των μελετών μεταβλητότητας.

5.2. Μεθοδολογία

5.2.1. Κριτήρια επιλεξιμότητας, πηγές δεδομένων και στρατηγική αναζήτησης

Η μετα-ανάλυση διεξήχθη ακολουθώντας συγκεκριμένες κατευθυντήριες γραμμές (MOOSE guidelines (Stroup, Berlin et al. 2000) και PRISMA statement (Moher, Liberati et al. 2010)). Μελέτες ασθενών-μαρτύρων περιλαμβάνονταν στην ανάλυση εάν αξιολογούσαν τον κίνδυνο για HIV καθώς και πολυμορφισμούς γονιδίων των ιντερλευκινών, και παρείχαν επαρκή δεδομένα για την εκτίμηση του Odds Ratio.

Σχετικές μελέτες ανιχνεύτηκαν στην Medline, Scopus, και ISI Web of Science χρησιμοποιώντας συνδυασμούς των ακόλουθων λέξεων κλειδιών: “HIV”, “interleukin”, “IL”, “polymorphism”, και “genotype” (τελευταία αναζήτηση: Απρίλιος 2015). Οι τίτλοι και οι περιλήψεις ελέγχονταν προσεκτικά έτσι ώστε να αποκλειστούν οι μη σχετικές με το θέμα εργασίες. Τα άρθρα που απέμειναν αξιολογήθηκαν για να καθοριστεί εάν πληρούσαν τα κριτήρια επιλεξιμότητας. Οι βιβλιογραφικές αναφορές σχετικών δημοσιεύσεων και οι περιλήψεις συνεδρίων εξετάστηκαν επίσης για να εντοπιστούν περαιτέρω μελέτες που θα μπορούσαν να χρησιμοποιηθούν στη μετα-ανάλυση.

5.2.2. Εξαγωγή δεδομένων και αξιολόγηση ποιότητας

Η εξαγωγή των δεδομένων έγινε από δυο ανεξάρτητους ερευνητές και συζητήθηκαν οι πιθανές ασυμφωνίες. Από κάθε μελέτη αντλήθηκαν οι ακόλουθες πληροφορίες: (i) όνομα του πρώτου συγγραφέα, έτος δημοσίευσης, περιοχή, και σχεδιασμός μελέτης; (ii) τα χαρακτηριστικά των ατόμων όπως η ηλικία, το φύλο, και η διάγνωση του AIDS; (iii) μέθοδοι αλληλούχισης; (iv) πολυμορφισμοί που μελετήθηκαν; και (v) η συχνότητα των γονοτύπων και των αλληλομόρφων σε ασθενείς και υγιείς.

Η ποιότητα των πρωταρχικών μελετών αξιολογήθηκε μέσω μιας καλά καθορισμένης περιγραφής των κριτηρίων ένταξης, και σωστής χρήσης στατιστικών μεθόδων. Η απόρριψη των μελετών λόγω χαμηλής ποιότητας ή μεροληψίας, αποφεύχθηκε. Ωστόσο, η πιθανή επίδραση της κάθε μελέτης εξετάστηκε πραγματοποιώντας μια εκ νέου μετα-ανάλυση αφαιρώντας μια μελέτη κάθε φορά.

5.2.3. Στατιστική ανάλυση

Το Odds Ratio χρησιμοποιήθηκε για τις συγκρίσεις των αλληλομόρφων και γονοτύπων μεταξύ ασθενών και υγιών. Οι αποκλίσεις από την κανονική κατανομή των λογαρίθμων των Odds Ratios αξιολογήθηκαν χρησιμοποιώντας τον έλεγχο Shapiro-Wilk. Οι συνδυασμένες εκτιμήσεις υπολογίστηκαν για τις διάφορες φυλές (Καυκάσιοι, Ασιάτες, μεικτοί/άλλοι) και τύπους της ομάδας ελέγχου (HIV-, HIV- που έχουν εκτεθεί (exposed seronegatives) [HES]). Οι αποκλίσεις από την ισορροπία Weinberg Equilibrium εξετάστηκαν χρησιμοποιώντας τον χ^2 έλεγχο.

Η μεταξύ των μελετών ετερογένεια αξιολογήθηκε χρησιμοποιώντας το Cochran's Q statistic και ποσοτικοποιήθηκε με το δείκτη I^2 (Higgins, Thompson et al. 2003). Τα Odds Ratios από τις πρωταρχικές μελέτες συνδυάστηκαν χρησιμοποιώντας μοντέλα σταθερών και τυχαίων επιδράσεων (DerSimonian and Laird 1986). Εάν δεν δηλώνεται διαφορετικά, παρουσιάζονται μόνο οι εκτιμήσεις τυχαίων επιδράσεων. Η ανάλυση επίσης πραγματοποιήθηκε χωρίς να υποθέσουμε εκ των προτέρων κάποιο γενετικό μοντέλο ακολουθώντας τη διμεταβλητή προσέγγιση μετα-ανάλυσης (Minelli, Thompson et al. 2005). Η παρουσία του συστηματικού σφάλματος δημοσίευσης διερευνήθηκε χρησιμοποιώντας τους ελέγχους των Begg και Egger (Begg and Mazumdar 1994; Egger, Davey Smith et al. 1997). Η διαχρονική τάση ανιχνεύτηκε μέσω αθροιστικής μετα-ανάλυσης (Lau, Antman et al. 1992).

Σε όλες τις αναλύσεις χρησιμοποιήθηκε το στατιστικό πακέτο Stata 12 (Stata Corporation, College Station, Texas, USA). Εκτός από τους στατιστικούς ελέγχους της ετερογένειας (όπου η σημαντικότητα δηλωνόταν για $p\text{-values} < 0.10$), τα αποτελέσματα θεωρούνταν στατιστικά σημαντικά εάν τα αντίστοιχα $p\text{-values}$ ήταν μικρότερα από 0.05.

5.3. Αποτελέσματα

5.3.1. Περιγραφή των μελετών

Μια σύνοψη της βιβλιογραφικής αναζήτησης και της διαδικασίας επιλογής μελετών παρουσιάζεται στην Εικόνα 5.1. Ένα σύνολο 42 μελετών, με 8,953 HIV+ και 6,774 HIV- άτομα, συμπεριλήφθησαν στην ανάλυση. Υπήρχαν δεδομένα για 8 SNPs: IL-1A C(-889)T (rs1800587), IL-1B C(+3953/4)T (rs1143634), IL-4 C(-589/90)T (rs2243250), IL-6 G(-174)C (rs1800795), IL-10 C(-592)A (rs1800872), IL-10 A(-1082)G (rs1800896), IL-12B A(-1188)C (rs3212227), και IL-28B (rs12979860) (C/T) (Πίνακας 5.1). Είκοσι μια μελέτες περιελάμβαναν Καυκάσιους, εννέα είχαν Ασιάτες, ενώ οι υπόλοιπες ήταν διαφόρων φυλών. Η συχνότητα των αλληλομόρφων και των γονοτύπων των διαφόρων πολυμορφισμών των γονιδίων των ιντερλευκινών παρουσιάζονται στον Πίνακα 5.2, ενώ τα αποτελέσματα της μετα-ανάλυσης παρουσιάζονται στον Πίνακα 5.3.

Πίνακας 5.1. Τα χαρακτηριστικά των πρωταρχικών μελετών που συμπεριλήφθησαν στη μετα-ανάλυση.

Μελέτη	Χώρα	Φυλή	Αριθμός ατόμων		Γονίδιο	SNP
			HIV(+)	HIV(-)		
Asensi 2008 (Asensi, Rego et al. 2008)	Ισπανία	Καυκάσιοι	228	109	IL-1B	C(+3953/4)T
		Καυκάσιοι	228	109	IL-1A	C(-889)T
Chatterjee 2009 (Chatterjee, Rathore et al. 2009)	Ινδία	Ασιάτες	180	355	IL-4	C(-589/90)T
Goncalves 2009 (Goncalves, Ferreira et al. 2009)	Βραζιλία	Μεικτοί	59	46	IL-1B	C(+3953/4)T
Kwa 2003 (Kwa, van Rij et al. 2003)	Ολλανδία	Καυκάσιοι	342	73	IL-4	C(-589/90)T
Martin 2010 (Martin, Qi et al. 2010)	ΗΠΑ	Καυκάσιοι	1,221	291	IL-28B	rs 12979860 (C/T)

Naicker 2009 (Naicker, Werner et al. 2009)	Νότια Αφρική	Αφρικάνοι	64	195	IL-10	A(-1082)G
		Αφρικάνοι	64	195	IL-10	C(-592)A
Nakayama 2000 (Nakayama, Hoshino et al. 2000)	Ιαπωνία	Ασιάτες	339	52	IL-4	C(-589/90)T
Nakayama 2002 (Nakayama, Meyer et al. 2002)	Γαλλία	Κauκάσιοι	427	86	IL-4	C(-589/90)T
Natterman 2007 (Nattermann, Vogel et al. 2007)	Γερμανία	Κauκάσιοι	309	310	IL-6	G(-174)C
Rallon 2011 (Rallon, Restrepo et al. 2011)	Ισπανία	Κauκάσιοι	29	29	IL-28B	rs 12979860 (C/T)
Ramaseri Sunder 2012 (Ramaseri Sunder, Hanumanth et al. 2012)	Ινδία	Ασιάτες	227	206	IL-10	A(-1082)G
Saumoy 2008 (Saumoy, Lopez-Dupla et al. 2008)	Ισπανία	Κauκάσιοι	273	385	IL-6	G(-174)C
Erikstrup 2007 (Erikstrup, Kallestrup et al. 2007)	Ζιμπάμπουε	Αφρικάνοι	194	174	IL-10	C(-592)A
		Αφρικάνοι	195	175	IL-10	A(-1082)G
Jablonowska 2010 (Jablonowska, Kolacinska et al. 2010)	Πολωνία	Κauκάσιοι	39	71	IL-6	G(-174)C
Naggie 2012 (Naggie, Osinusi et al. 2012)	ΗΠΑ, Γερμανία	Μεικτοί	44	44	IL-28B	rs 12979860 (C/T)
Pemberton 2008 (Pemberton, Stone et al. 2008)	Αυστραλία	Κauκάσιοι	228	60	IL-1A	C(-889)T
		Κauκάσιοι	242	60	IL-1B	C(+3953/4)T
		Κauκάσιοι	262	96	IL-12B	3' UTR A(-1188)C
Pontillo 2012 (Pontillo, Oshiro et al. 2012)	Βραζιλία	Αφρικάνοι, Κauκάσιοι	150	158	IL-1B	C(+3953/4)T
Price 1999 (Price, Calder et al. 1999)	Αυστραλία	Κauκάσιοι	33	60	IL-1A	C(-889)T
		Κauκάσιοι	33	60	IL-1B	C(+3953/4)T
Smolnikova 2001 (Smolnikova, Freidin et al. 2001)	Ρωσία	Κauκάσιοι	71	58	IL-4	C(-589/90)T
Sobti 2010 (Sobti, Salih et al. 2010)	Ινδία	Ασιάτες	300	300	IL-12B	3' UTR A(-1188)C
Sobti 2010 (Sobti, Berhane et al. 2010)	Ινδία	Ασιάτες	300	300	IL-6	G(-174)C
		Ασιάτες	300	300	IL-10	C(-592)A

Wichukchinda 2006 (Wichukchinda, Nakayama et al. 2006)	Ταϊλάνδη	Ασιάτες	246	119	IL-4	C(-589/90)T
Chaterjee 2009 (Chatterjee, Rathore et al. 2009)	Ινδία	Ασιάτες	180	355	IL-10	A(-1082)G
		Ασιάτες	180	355	IL-10	C(-592)A
Konenkov 2002 (Konenkov and Smol'nikova 2002)	Ρωσία	Κauκάσιοι	120	52	IL-4	C(-589/90)T
Modi 2003 (Modi, O'Brien et al. 2003)	ΗΠΑ	Αφρο-αμερικάνοι, Kauκάσιοι	865	230	IL-4	C(-589/90)T
Sajadi 2011 (Sajadi, Shakeri et al. 2011)	ΗΠΑ	Αφρο-αμερικάνοι	172	173	IL-28B	rs 12979860 (C/T)
Jablonowska 2012 (Jablonowska, Piekarska et al. 2012)	Πολωνία	Κauκάσιοι	33	90	IL-28B	rs 12979860 (C/T)
Price 2002 (Price, Morahan et al. 2002)	Αυστραλία	Κauκάσιοι	70	60	IL-1A	C(-889)T
		Κauκάσιοι	70	60	IL-1B	C(+3953/4)T
		Κauκάσιοι	82	100	IL-12B	A(-1188)C
Shin 2000 (Shin, Winkler et al. 2000)	ΗΠΑ	Μεικτοί	377	72	IL-10	C(-592)A
Wang 2004 (Wang, Song et al. 2004)	ΗΠΑ	Αφρο-αμερικάνοι, Ισπανική	319	258	IL-4	C(-589/90)T
		Αφρο-αμερικάνοι, Ισπανική	321	258	IL-6	G(-174)C
		Αφρο-αμερικάνοι, Ισπανική	415	334	IL-10	A(-1082)G
		Αφρο-αμερικάνοι, Ισπανική	321	258	IL-10	C(-592)A
Avihingsanon 2014 (Avihingsanon, Jitmitraparp et al. 2014)	Ταϊλάνδη	Ασιάτες	94	136	IL-28B	rs 12979860 (C/T)
Burlone 2013 (Burlone, Cerutti et al. 2013)	Ιταλία	Κauκάσιοι	75	112	IL-28B	rs 12979860 (C/T)
Di Lello 2013 (Di Lello, Caruz et al. 2013)	Ισπανία	Κauκάσιοι	314	109	IL-28B	rs 12979860 (C/T)
Grebely 2012 (Grebely, Hellard et al. 2012)	Καναδάς, Αυστραλία	Κauκάσιοι	31	56	IL-28B	rs 12979860 (C/T)
Neukam 2013 (Neukam, Barreiro et al. 2013)	Ισπανία	Κauκάσιοι	160	62	IL-28B	rs 12979860 (C/T)
Yuan 2012 (Yuan, Adams-Huet et al. 2012)	ΗΠΑ	Μεικτοί	25	8	IL-28B	rs 12979860 (C/T)

Stenkvist 2013 (Stenkvist, Sonnerborg et al. 2013)	Σουηδία	Κauκάσιοι	13	100	IL-28B	rs 12979860 (C/T)
Ramezani 2015 (Ramezani, Kalantar et al. 2015)	Ιράν	Ασιάτες	70	31	IL-10	A(-1082)G
		Ασιάτες	70	31	IL-10	C(-592)A
Kallas 2015 (Kallas, Huik et al. 2015)	Εσθονία	Κauκάσιοι	172	669	IL-10	A(-1082)G
		Κauκάσιοι	172	669	IL-10	C(-592)A
Corchado 2013 (Corchado, Marquez et al. 2013)	Ισπανία	Κauκάσιοι	88	51	IL-10	C(-592)A
Freitas 2015 (Freitas, Lima et al. 2015)	Βραζιλία	Μεικτοί	216	294	IL-10	A(-1082)G
			216	294	IL-6	G(-174)C
Zeremski 2013 (Zeremski, Dimova et al. 2013)	ΗΠΑ	Αφρο-αμερικάνοι, Ισπανική, Κauκάσιοι	113	222	IL-28B	rs 12979860 (C/T)

Πίνακας 5.2. Συχνότητες γονοτύπων και αλληλομόρφων των πολυμορφισμών των γονιδίων των ιντερλευκινών.

Γονίδιο	SNP	Μελέτη	Συχνότητα γονοτύπων						Συχνότητα αλληλομόρφων			
			HIV+			HIV–			HIV+		HIV–	
IL-1A	C(-889)T		CC	CT	TT	CC	CT	TT	C	T	C	T
		Asensi 2008 (Asensi, Rego et al. 2008)	117	87	24	57	43	9	321	135	157	61
		Pemberton 2008 (Pemberton, Stone et al. 2008)	129	80	19	33	24	3	338	118	90	30
		Price 1999 (Price, Calder et al. 1999)	24	4	5				52	14	90	30
		Price 2002 (Price, Morahan et al. 2002)							71	69	66	54
IL-1B	C(+3953/4)T		CC	CT	TT	CC	CT	TT	C	T	C	T
		Asensi 2008 (Asensi, Rego et al. 2008)	134	75	19	72	35	2	343	113	179	39
		Goncalves 2009 (Goncalves, Ferreira et al. 2009)	43	8	8	35	10	1	94	24	80	12
		Pemberton 2008 (Pemberton, Stone et al. 2008)	151	86	5	36	23	1	388	96	95	25
		Price 1999 (Price, Calder et al. 1999)	22	8	3				52	14	95	25
		Price 2002 (Price, Morahan et al. 2002)							81	59	64	56

	Pontillo 2012 (Pontillo, Oshiro et al. 2012)	57	62	31	106	44	8	176	124	256	60
IL-4	C(-589/90)T	CC	CT	TT	CC	CT	TT	C	T	C	T
	Wichukchinda 2006 (Wichukchinda, Nakayama et al. 2006)	12	87	147				111	381	48	190
	Chatterjee 2009 (Chatterjee, Rathore et al. 2009)	122	53	5	229	111	15	297	63	569	141
	Kwa 2003 (Kwa, van Rij et al. 2003)	243	89	10	55	16	2	575	109	126	20
	Nakayama 2000 (Nakayama, Hoshino et al. 2000)	44	157	138	5	22	25	245	433	32	72
	Nakayama 2002 (Nakayama, Meyer et al. 2002)	315	98	14	63	21	2	728	126	147	25
	Smolnikova 2001 (Smolnikova, Freidin et al. 2001)	39	26	6	35	17	6	104	38	87	29
	Konenkov 2002 (Konenkov and Smol'nikova 2002)	65	47	8	31	20	1	177	63	82	22
	Wang 2004 (Wang, Song et al. 2004)							293	345	251	265
		CC	CT+TT		CC	CT+TT		C	T	C	T

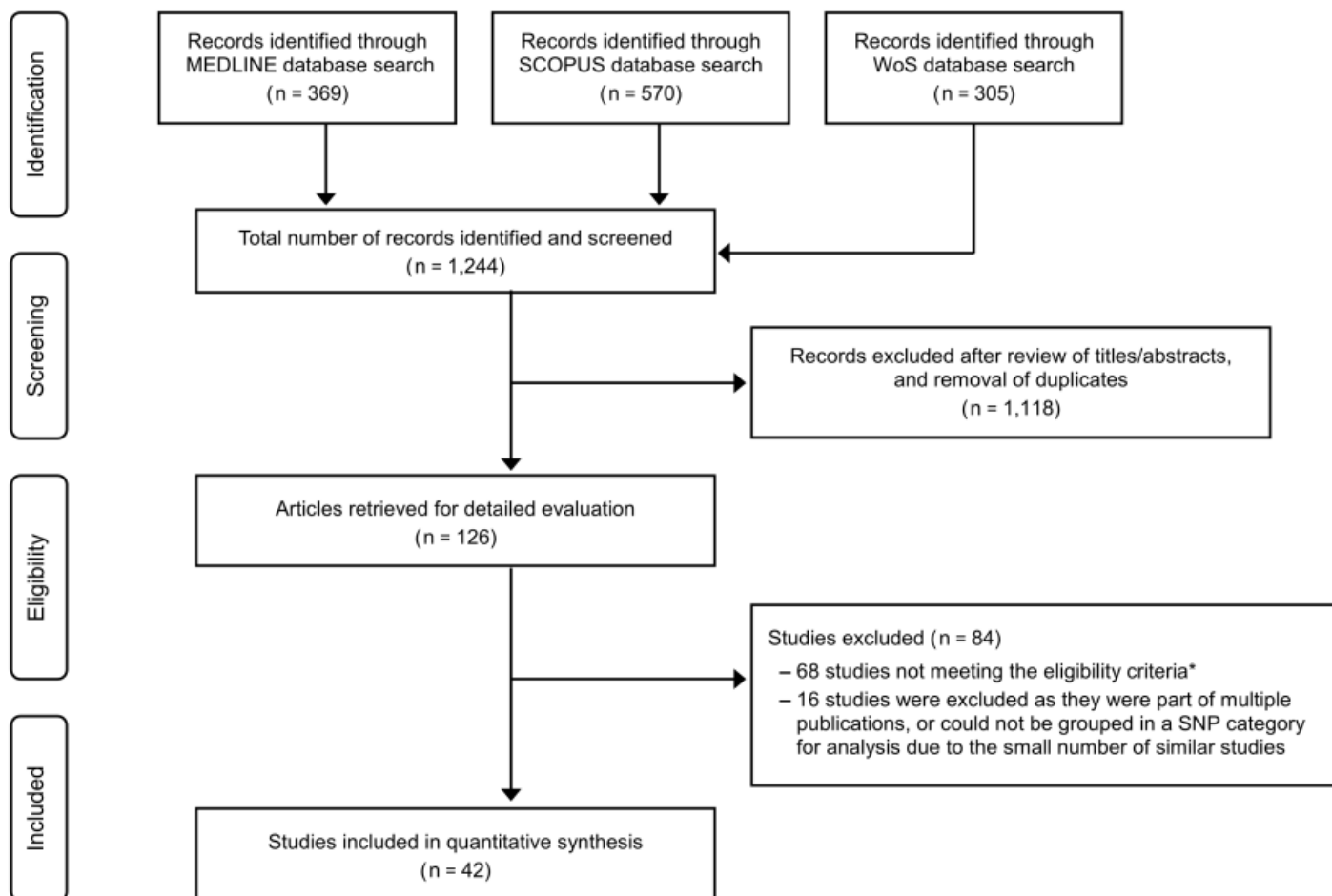
		Modi 2003. (Modi, O'Brien et al. 2003)	416	449	108	122						
IL-6	G(-174)C		GG	GC	CC	GG	GC	CC	G	C	G	C
		Natterman 2007 (Nattermann, Vogel et al. 2007)	114	135	60	99	150	61	363	255	348	272
		Saumoy 2008 (Saumoy, Lopez-Dupla et al. 2008)	108	114	51	182	154	49	330	216	518	252
		Jablonowska 2010 (Jablonowska, Kolacinska et al. 2010)	23	15	1	41	29	1	17	61	111	31
		Sobti 2010 (Sobti, Berhane et al. 2012)	161	94	45	152	106	42	416	184	410	190
		Freitas 2015 (Freitas, Lima et al. 2015)	142	70	4	184	96	14	354	78	464	124
		Wang 2004 (Wang, Song et al. 2004)							555	87	438	78
IL-10	C(-592)A		CC	CA	AA	CC	CA	AA	C	A	C	A
		Naicker 2009 (Naicker, Werner et al. 2009)	24	23	17	97	80	18	71	57	274	116
		Erikstrup 2007 (Erikstrup, Kallestrup et al. 2007)	80	71	43	68	81	25	231	157	217	131
		Sobti 2010 (Sobti, Berhane et al. 2012)	36	137	127	34	146	120	209	391	214	386

		Chatterjee 2009 (Chatterjee, Rathore et al. 2009)	67	74	39	163	141	51	208	152	467	243
		Ramezani 2015 (Ramezani, Kalantar et al. 2015)	31	35	4	16	11	4	97	43	43	19
		Kallas 2015 (Kallas, Huik et al. 2015)	113	49	10	405	232	32	275	69	1,042	296
		Corchado 2013 (Corchado, Marquez et al. 2013)	43	38	7	24	21	6	124	52	69	33
		Wang 2004 (Wang, Song et al. 2004)							420	222	321	195
			CC	CA+AA		CC	CA+AA		C	A	C	A
		Shin 2000 (Shin, Winkler et al. 2000)	207	170		50	22					
IL-10	A(-1082)G		AA	AG	GG	AA	AG	GG	A	G	A	G
		Naicker 2009 (Naicker, Werner et al. 2009)	37	22	5	88	80	27	96	32	256	134
		Ramaseri Sunder 2012 (Ramaseri Sunder, Hanumanth et al. 2012)	96	120	11	133	68	5	312	142	334	78
		Erikstrup 2007 (Erikstrup, Kallestrup et al. 2007)	100	73	22	76	82	17	273	117	234	116
		Chatterjee 2009 (Chatterjee, Rathore et al. 2009)	100	60	20	183	141	31	260	100	507	203

		Freitas 2015 (Freitas, Lima et al. 2015)	123	79	14	159	111	24	325	107	429	159
		Ramezani 2015 (Ramezani, Kalantar et al. 2015)	28	32	10	13	15	3	88	52	41	21
		Kallas 2015 (Kallas, Huik et al. 2015)	62	78	32	196	332	141	202	142	724	614
		Wang 2004 (Hispanic) (Wang, Song et al. 2004)	27	61	6	39	33	4	115	73	111	41
		Wang 2004 (mixed) (Wang, Song et al. 2004)							413	229	370	146
IL-12B	A(-1188)C		AA	AC	CC	AA	AC	CC	A	C	A	C
		Sobti 2010 (Sobti, Berhane et al. 2012)	135	159	6	141	150	9	429	171	432	168
		Pemberton 2008 (Pemberton, Stone et al. 2008)	155	91	16	68	24	4	401	123	160	32
		Price 2002 (Price, Morahan et al. 2002)							107	57	138	62
IL-28B	rs12979860(C/T)		CC	CT	TT	CC	CT	TT	C	T	C	T
		Martin 2010 (Martin, Qi et al. 2010)	446	586	189	105	136	50	1,478	964	346	236
		Rallon 2011 (Rallon, Restrepo et al. 2011)	13	11	5	18	10	1	37	21	46	12

Naggie 2012 (Naggie, Osinusi et al. 2012)	15	13	16	11	28	5	43	45	50	38
Sajadi 2011 (Sajadi, Shakeri et al. 2011)	27	87	58	20	95	58	141	203	135	211
Jablonowska 2012 (Jablonowska, Piekarska et al. 2012)	17	12	4	19	55	16	46	20	93	87
Avihingsanon 2014 (Avihingsanon, Jitmitraparp et al. 2014)	83	10	1	116	14	6	176	12	246	26
Burlone 2013 (Burlone, Cerutti et al. 2013)	34	29	12	39	60	13	97	53	138	86
Di Lello 2013 (Di Lello, Caruz et al. 2013)	180	121	13	62	41	6	481	147	165	53
Grebely 2012 (Grebely, Hellard et al. 2012)	18	9	4	24	27	5	45	17	75	37
Neukam 2013 (Neukam, Barreiro et al. 2013)	59	76	25	23	29	10	194	126	75	49
Stenkvis 2013 (Stenkvis, Sonnerborg et al. 2013)	6	6	1	44	52	4	18	8	140	60
Zeremski 2013 (Zeremski, Dimova et al. 2013)	34	55	24	66	119	37	123	103	251	193

	CC	CT+TT	CC	CT+TT	C	T	C	T
Yuan 2012 (Yuan, Adams-Huet et al. 2012)	13	12	0	8				



Εικόνα 5.1. Σύνοψη της διαδικασίας αναζήτησης και επιλογής των μελετών (διάγραμμα ροής).

5.3.2. Προδιάθεση στη μόλυνση από τον HIV

IL-1A C(-889)T (rs1800587): Η σύγκριση των συχνοτήτων των αλληλομόρφων των 4 πρωταρχικών μελετών απέτυχε να δείξει συσχέτιση του αλληλομόρφου T με τον κίνδυνο για μόλυνση από τον HIV (OR: 1.063, 95% C.I.: 0.843, 1.341). Η ανάλυση των γονοτύπων δεν πραγματοποιήθηκε καθώς μόνο δυο μελέτες είχαν τις απαραίτητες πληροφορίες.

IL-1B C(+3953/4)T (rs1143634): Από τη σύγκριση του T έναντι του C αλληλομόρφου 6 μελετών εκτιμήθηκε ένα Odds Ratio ίσο με 1.378 (95% C.I.: 0.875, 2.172). Ένα στατιστικά σημαντικό μέτρο επίδρασης (OR: 2.500, 95% C.I.: 1.484, 4.212) υπολογίστηκε στη μετα-ανάλυση δυο μελετών από την Βραζιλία (Εικόνα 5.2). Η σύνθεση των 4 μελετών, υπό τη θεώρηση υπολειπόμενου μοντέλου κληρονομικότητας, έδειξε ότι οι ασθενείς με HIV+ ήταν 4.5 φορές πιο πιθανό να φέρουν το γονότυπο TT σε σχέση με τους HIV- (OR: 4.474, 95% C.I.: 2.348, 8.524) (Εικόνα 5.3). Για τη σύγκριση αυτή, δεν υπήρχε ένδειξη για συστηματικό σφάλμα δημοσίευσης ή ετερογένεια (Πίνακας 5.3).

IL-4 C(-589/90)T (rs2243250): Η σύγκριση των αλληλομόρφων περιελάμβανε 8 μελέτες και έδειξε ότι δεν υπάρχει συσχέτιση του αλληλομόρφου T με τον κίνδυνο μόλυνσης από τον HIV (OR: 1.006, 95% C.I.: 0.879, 1.152). Η ανάλυση των γονοτύπων έδωσε παρόμοια αποτελέσματα θεωρώντας είτε υπολειπόμενο, είτε επικρατές μοντέλο κληρονομικότητας (OR: 0.843, 95% C.I.: 0.554, 1.284; και OR: 0.987, 95% C.I.: 0.824, 1.183; αντίστοιχα). Ο έλεγχος του Begg έδειξε την παρουσία συστηματικού σφάλματος δημοσίευσης κατά τη σύγκριση του TT έναντι των CT και CC (p -value=0.024).

IL-6 G(-174)C (rs1800795): Συνολικά, 6 μελέτες παρείχαν δεδομένα για τις συχνότητες των αλληλομόρφων. Σε μια μελέτη (Sobti, Berhane et al. 2012) παραβιαζόταν η υπόθεση της ισορροπίας Hardy-Weinberg. Το συνδυασμένο Odds Ratio του C έναντι του G αλληλομόρφου ήταν 1.348 (95% C.I.: 0.881, 2.063) και υπήρχε ένδειξη για πολύ υψηλή ετερογένεια μεταξύ των μελετών ($I^2=92.2\%$). Οι συγκρίσεις των γονοτύπων οδήγησαν σε μη σημαντικά αποτελέσματα.

IL-10 C(-592)A (rs1800872): Εννέα μελέτες αξιολογούσαν τη συσχέτιση του συγκεκριμένου πολυμορφισμού με τη μόλυνση από τον HIV. Υπήρχε απόκλιση από την ισορροπία Hardy-Weinberg σε μια μελέτη (Chatterjee, Rathore et al. 2009). Η εκτίμηση του Odds Ratio των αλληλομόρφων (A έναντι C) ήταν 1.096 (95% C.I.:

0.917, 1.310). Η σύγκριση των ατόμων με γονότυπο AA έναντι των AC+CC οδήγησε σε μια οριακά μη σημαντική εκτίμηση (OR: 1.393, 95% C.I.: 0.966, 2.008) (Πίνακας 5.3). Αποκλείοντας τις μελέτες των Ramezani και συνεργατών (Ramezani, Kalantar et al. 2015), ή Corchado και συνεργατών (Corchado, Marquez et al. 2013), το συνοπτικό Odds Ratio έγινε στατιστικά σημαντικό (OR: 1.487, 95% C.I.: 1.045, 2.116; και OR: 1.482, 95% C.I.: 1.017, 2.159, αντίστοιχα). Η ανάλυση υποομάδας για τους υγιείς HES έδειξε επίσης έναν αυξημένο κίνδυνο για μόλυνση από τον HIV για τα άτομα με γονότυπο AA (OR: 2.376, 95% C.I.: 1.241, 4.548). Η σύνθεση των μελετών σε Ασιάτες απέτυχε να δείξει κάποια συσχέτιση (OR: 1.189, 95% C.I.: 0.750, 1.885). Θεωρώντας επικρατές μοντέλο κληρονομικότητας, η σύγκριση των γονοτύπων (AA και AC έναντι του γονοτύπου CC), και η ανάλυση υποομάδας για τους υγιείς HES (Shin, Winkler et al. 2000; Chatterjee, Rathore et al. 2009; Naicker, Werner et al. 2009; Kallas, Huik et al. 2015), έδειξε μη στατιστικά σημαντικές εκτιμήσεις (OR: 1.147, 95% C.I.: 0.909, 1.447; και OR: 1.230, 95% C.I.: 0.685, 2.210; αντίστοιχα).

IL-10 A(-1082)G (rs1800896): Οκτώ μελέτες αξιολογούσαν το ρόλο του πολυμορφισμού αυτού στον κίνδυνο μόλυνσης από τον HIV. Όλες οι συγκρίσεις (G έναντι του αλληλομόρφου A, GG γονότυπος έναντι των AG και AA, GG και AG έναντι του AA γονότυπου) δεν έδωσε σημαντικά αποτελέσματα (OR: 1.032, 95% C.I.: 0.813, 1.308; OR: 1.005, 95% C.I.: 0.781, 1.294; OR: 1.054, 95% C.I.: 0.724, 1.535; αντίστοιχα).

IL-12B A(-1188)C (rs3212227): Η σύνθεση των 3 μελετών έδωσε μια μη σημαντική εκτίμηση για τα αλληλόμορφα (OR: 1.167, 95% C.I.: 0.927, 1.469). Τα δεδομένα για τις συχνότητες των γονοτύπων δεν ήταν διαθέσιμα.

IL-28B (rs12979860) C/T: Η επίδραση του αλληλομόρφου T του πολυμορφισμού rs12979860 του γονιδίου IL-28B αξιολογήθηκε σε 13 μελέτες. Αποκλίσεις από την ισορροπία Hardy-Weinberg βρέθηκαν σε δυο μελέτες (Stenkvist, Sonnerborg et al. 2013; Avihingsanon, Jitmitraparp et al. 2014). Το συνδυασμένο Odds Ratio του αλληλομόρφου T έναντι του C ήταν 0.949 (95% C.I.: 0.831, 1.084). Παρόμοια, οι συγκρίσεις των γονοτύπων δεν ήταν στατιστικά σημαντικές. Ωστόσο, σε μια ανάλυση 11 μελετών με ασθενείς μολυσμένους από τον ιό της ηπατίτιδας C (hepatitis C virus-HCV) (Sajadi, Shakeri et al. 2011; Grebely, Hellard et al. 2012; Jablonowska, Piekarska et al. 2012; Naggie, Osinusi et al. 2012; Yuan, Adams-Huet et al. 2012; Burlone, Cerutti et al. 2013; Di Lello, Caruz et al. 2013; Neukam, Barreiro et al. 2013; Stenkvist, Sonnerborg et al. 2013; Zeremski, Dimova et al. 2013; Avihingsanon,

Jitmitraparp et al. 2014) και υποθέτοντας επικρατές μοντέλο κληρονομικότητας, το συνδυασμένο Odds Ratio έδειξε προστατευτική δράση του αλληλομόρφου T έναντι της μόλυνσης από τον HIV (OR: 0.734, 95% C.I.: 0.569, 0.947) (Εικόνα 5.4). Ο προστατευτικός ρόλος του συγκεκριμένου γονιδιακού πολυμορφισμού παρατηρήθηκε επίσης όταν αποκλειόταν κάθε φορά, μια από τις δυο μελέτες (Martin, Qi et al. 2010; Rallon, Restrepo et al. 2011) με υγιείς HES που δεν περιλαμβάνονταν στην προηγούμενη ανάλυση υποομάδας (OR: 0.764, 95% C.I.: 0.585, 0.997; OR: 0.783, 95% C.I.: 0.627, 0.976; αντίστοιχα). Υπήρχε μέτρια ετερογένεια σε κάποιες αναλύσεις.

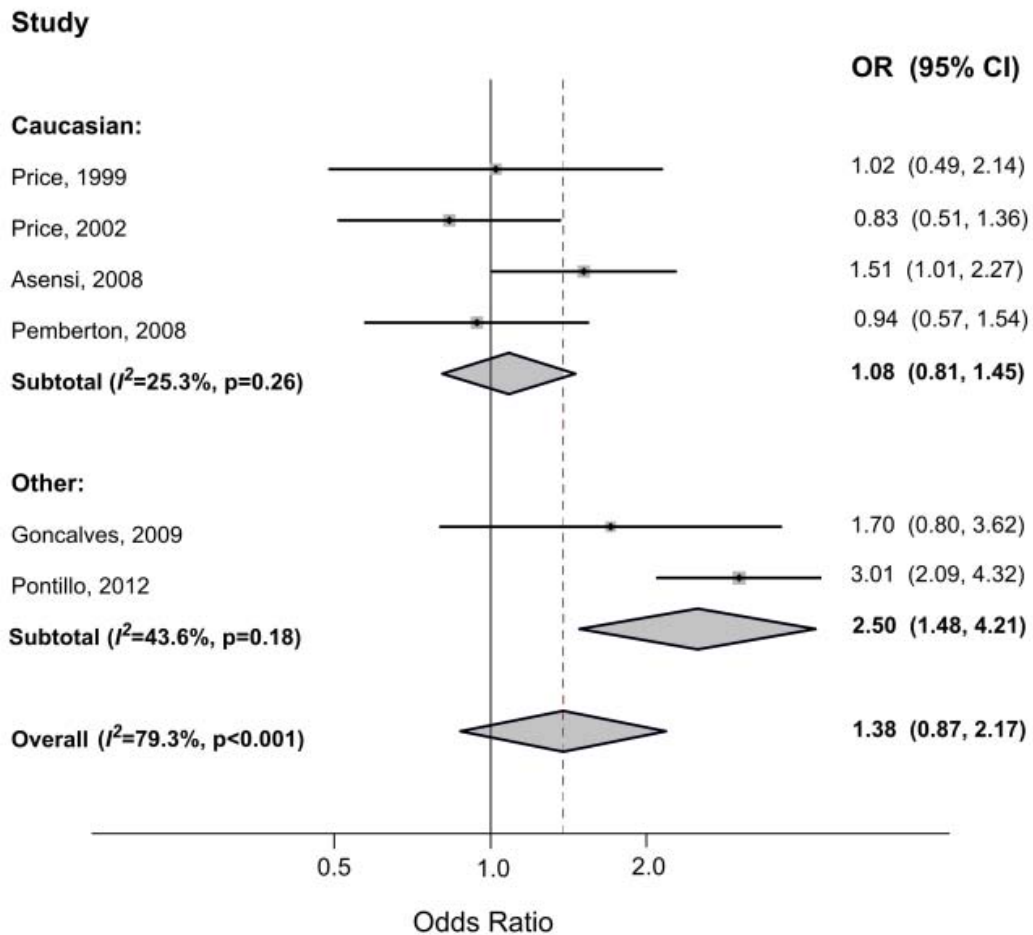
Διμεταβλητή μετα-ανάλυση: Συνδυάζοντας τις μελέτες που αξιολογούσαν τη συσχέτιση του πολυμορφισμού rs1143634 του γονιδίου της IL-1B C(+3953/4)T υπό τη διμεταβλητή προσέγγιση, χωρίς να υποθέσουμε εκ των προτέρων κάποιο μοντέλο κληρονομικότητας, υπολογίστηκε ένα σημαντικό Odds Ratio για τη σύγκριση του γονοτύπου TT έναντι του CC (OR: 4.484, 95% C.I.: 1.852, 10.857) και εκτίμησε ένα υπολειπόμενο μοντέλο κληρονομικότητας. Επιπλέον, επικυρώθηκε η προστατευτική επίδραση του πολυμορφισμού rs12979860 του γονιδίου IL-28B σε δεδομένα 11 μελετών που είχαν στρατολογήσει ασθενείς με μόλυνση από τον HCV. Πιο συγκεκριμένα, το συνοπτικό Odds Ratio του γονοτύπου CT έναντι του CC ήταν 0.692 (95% C.I.: 0.513, 0.934). Τέλος, η διμεταβλητή προσέγγιση κατέδειξε τον προστατευτικό ρόλο του πολυμορφισμού αυτού για το σύνολο των μελετών που συμμετείχαν στην ανάλυση (13 μελέτες), και υπέδειξε επικρατές μοντέλο κληρονομικότητας (CT έναντι του CC, OR: 0.764, 95% C.I.: 0.588, 0.993).

Πίνακας 5.3. Τα αποτελέσματα της μετα-ανάλυσης για τη συσχέτιση πολυμορφισμών των γονιδίων των ιντερλευκινών με τον κίνδυνο εμφάνισης HIV-1.

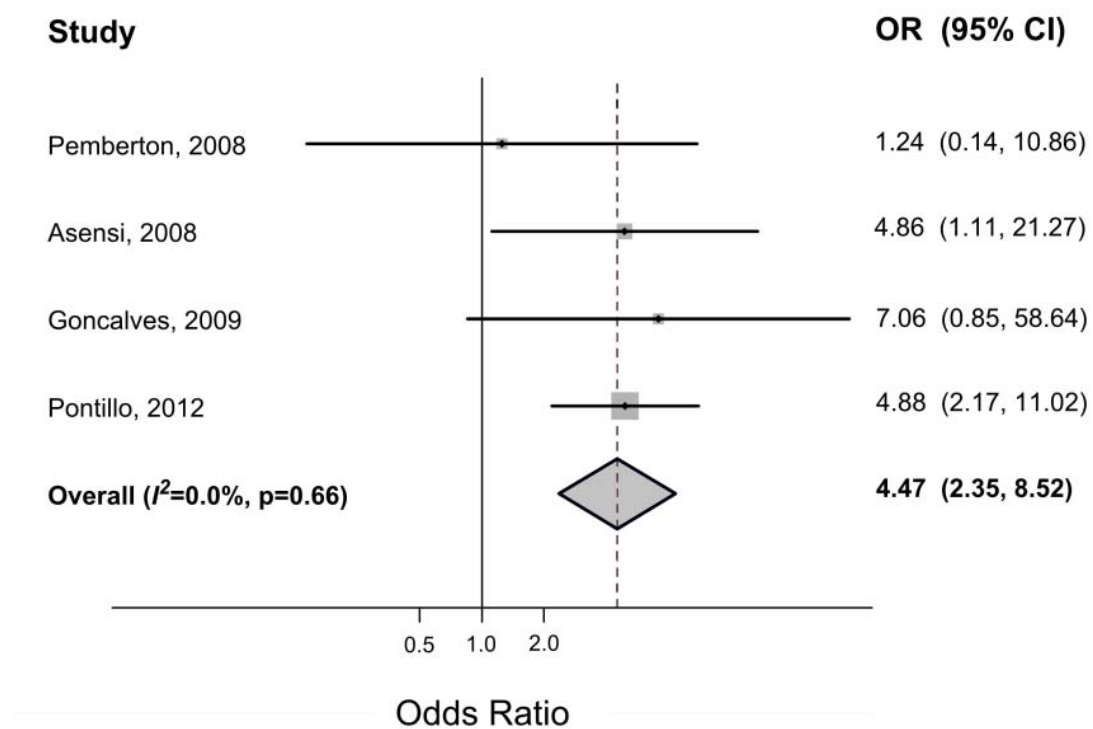
SNP	Σύγκριση	Φυλή	Μελέτες	HIV+ / HIV-	OR	(95% CI)	I ²
IL-1A C(-889)T	T έναντι του C αλληλομόρφου	Σύνολο	4	559 / 289	1.063	(0.843, 1.341)	0.0%
		Καυκάσιοι	4	559 / 289	1.063	(0.843, 1.341)	0.0%
IL-1B C(+3953/4)T	T έναντι του C αλληλομόρφου	Σύνολο	6	782 / 493	1.378	(0.875, 2.172)	79.3%
		Καυκάσιοι	4	573 / 289	1.083	(0.808, 1.453)	25.3%
		Άλλοι/μεικτοί	2	209 / 204	2.500	(1.484, 4.212)	43.6%
		TT έναντι των CT+CC (υπολειπόμενο μοντέλο)	4	679 / 373	4.474	(2.348, 8.524)	0.0%
		Καυκάσιοι	2	470 / 169	3.128	(0.896, 10.918)	3.7%
		Άλλοι/μεικτοί	2	209 / 204	5.121	(2.396, 10.945)	0.0%

IL-4 C(-589/90)T	TT+CT έναντι του CC (επικρατές μοντέλο)	Σύνολο	4	679 / 373	1.528 (0.824, 2.833)	78.5%
		Καυκάσιοι	2	470 / 169	1.149 (0.772, 1.712)	14.0%
		Άλλοι/μεικτοί	2	209 / 204	2.132 (0.782, 5.809)	75.4%
	T έναντι του C αλληλομόρφου	Σύνολο	8	2,044 / 1,053	1.006 (0.879, 1.152)	0.0%
		Καυκάσιοι	4	960 / 269	1.142 (0.881, 1.480)	0.0%
		Ασιάτες	3	765 / 526	0.842 (0.678, 1.046)	0.0%
		Άλλοι/μεικτοί	1	319 / 258	1.115 (0.884, 1.407)	–
	TT έναντι των CT+CC (υπολειπόμενο μοντέλο)	Σύνολο	6	1,479 / 676	0.843 (0.554, 1.284)	0.0%
		Καυκάσιοι	4	960 / 269	1.194 (0.567, 2.516)	0.0%
		Ασιάτες	2	519 / 407	0.717 (0.431, 1.193)	0.0%
IL-6 G(-174)C	TT+CT έναντι του CC (επικρατές μοντέλο)	Σύνολο	7	2,344 / 906	0.987 (0.824, 1.183)	0.0%
		Καυκάσιοι	5	1,571 / 421	1.017 (0.799, 1.294)	0.0%
		Ασιάτες	2	519 / 407	0.842 (0.591, 1.201)	0.0%
		Άλλοι/μεικτοί	1	254 / 78	0.724 (0.341, 1.537)	–
	C έναντι του G αλληλομόρφου	Σύνολο	6	1,458 / 1,618	1.348 (0.881, 2.063)	92.2%
		Καυκάσιοι	3	621 / 766	2.301 (0.919, 5.759)	96.4%
		Ασιάτες	1	300 / 300	0.954 (0.748, 1.219)	–
		Άλλοι/μεικτοί	2	537 / 552	0.851 (0.677, 1.068)	0.0%
	CC έναντι των GC+GG (υπολειπόμενο μοντέλο)	Σύνολο	5	1,137 / 1,360	1.089 (0.771, 1.537)	39.3%
		Καυκάσιοι	3	621 / 766	1.240 (0.863, 1.780)	22.6%
		Ασιάτες	1	300 / 300	1.084 (0.688, 1.708)	–
		Άλλοι/μεικτοί	1	216 / 294	0.377 (0.122, 0.163)	–
IL-10 C(-592)A	CC+GC έναντι του GG (επικρατές μοντέλο)	Σύνολο	5	1,137 / 1,360	0.966 (0.780, 1.197)	38.5%
		Καυκάσιοι	3	621 / 766	1.034 (0.694, 1.542)	62.4%
		Ασιάτες	1	300 / 300	0.887 (0.644, 1.222)	–
		Άλλοι/μεικτοί	1	216 / 294	0.872 (0.604, 1.258)	–
	A έναντι του C αλληλομόρφου	Σύνολο	8	1,389 / 2,033	1.096 (0.917, 1.310)	58.5%
		Καυκάσιοι	2	260 / 720	0.882 (0.682, 1.140)	0.0%
		Ασιάτες	3	550 / 686	1.175 (0.934, 1.477)	35.4%
		Άλλοι/μεικτοί	3	579 / 627	1.194 (0.795, 1.794)	80.7%
	AA έναντι των AC+CC (υπολειπόμενο μοντέλο)	Σύνολο	7	1,068 / 1,775	1.393 (0.966, 2.008)	56.4%
		Καυκάσιοι	2	260 / 720	1.022 (0.552, 1.894)	0.0%
		Ασιάτες	3	550 / 686	1.189 (0.750, 1.885)	52.1%
		Άλλοι/μεικτοί	2	258 / 369	2.352 (1.145, 4.831)	60.2%
IL-10 A(-1082)G	AA+AC έναντι του CC (επικρατές μοντέλο)	Σύνολο	8	1,445 / 1,847	1.147 (0.909, 1.447)	42.7%
		Καυκάσιοι	2	260 / 720	0.826 (0.604, 1.129)	0.0%
		Ασιάτες	3	550 / 686	1.245 (0.942, 1.646)	0.0%
		Άλλοι/μεικτοί	3	635 / 441	1.369 (0.856, 2.188)	60.8%
	G έναντι του A αλληλομόρφου	Σύνολο	8	1,445 / 2,183	1.032 (0.813, 1.308)	77.8%
		Καυκάσιοι	1	172 / 669	0.829 (0.652, 1.054)	–
		Asian	3	477 / 592	1.304 (0.784, 2.169)	81.6%
		Άλλοι/μεικτοί	4	796 / 922	0.938 (0.682, 1.290)	75.9%

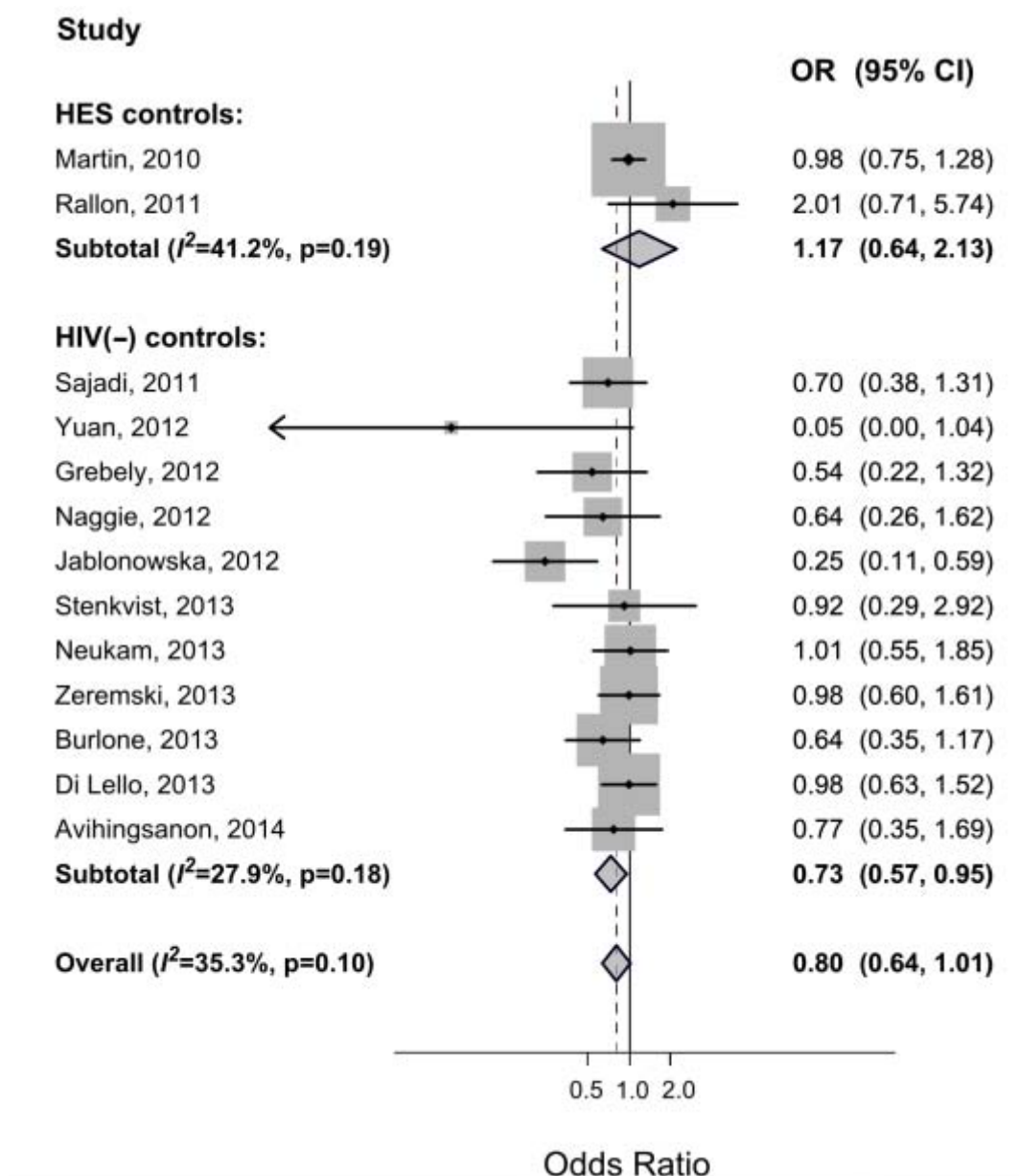
IL-12B A(-1188)C IL-28B(rs12979860)	GG έναντι των AG+AA (υπολειπόμενο μοντέλο)	Σύνολο	8	1,218 / 2,001	1.005 (0.781, 1.294)	0.0%
		Κauκάσιοι	1	172 / 669	0.856 (0.559, 1.311)	–
		Ασιάτες	3	477 / 592	1.464 (0.901, 2.378)	0.0%
		Άλλοι/μεικτοί	4	569 / 740	0.892 (0.593, 1.344)	0.0%
	GG+AG έναντι των AA (επικρατές μοντέλο)	Σύνολο	8	1,218 / 2,001	1.054 (0.724, 1.535)	82.2%
		Κauκάσιοι	1	172 / 669	0.735 (0.517, 1.046)	–
		Ασιάτες	3	477 / 592	1.340 (0.609, 2.947)	87.5%
		Άλλοι/μεικτοί	4	569 / 740	0.973 (0.584, 1.620)	78.3%
	C έναντι του A αλληλομόρφου	Σύνολο	3	644 / 496	1.167 (0.927, 1.469)	21.4%
		Κauκάσιοι	2	344 / 196	1.352 (0.994, 1.839)	0.0%
		Ασιάτες	1	300 / 300	1.025 (0.797, 1.318)	–
	T έναντι του C αλληλομόρφου	Σύνολο	12	2,299 / 1,424	0.949 (0.831, 1.084)	16.9%
		Κauκάσιοι	7	1,863 / 749	0.913 (0.744, 1.121)	38.3%
		Ασιάτες	1	94 / 136	0.645 (0.317, 1.313)	–
		Άλλοι/μεικτοί	4	342 / 539	1.037 (0.847, 1.269)	0.0%
	TT έναντι των CT+CC (υπολειπόμενο μοντέλο)	Σύνολο	12	2,299 / 1,424	1.115 (0.845, 1.473)	25.4%
		Κauκάσιοι	7	1,863 / 749	0.954 (0.730, 1.246)	0.0%
		Ασιάτες	1	94 / 136	0.233 (0.028, 1.968)	–
		Άλλοι/μεικτοί	4	342 / 539	1.513 (0.855, 2.679)	50.6%
	TT+CT έναντι του CC (επικρατές μοντέλο)	Σύνολο	13	2,324 / 1,432	0.805 (0.641, 1.010)	35.3%
		Κauκάσιοι	8	1,910 / 878	0.740 (0.516, 1.061)	61.8%
		Ασιάτες	1	94 / 136	0.769 (0.350, 1.690)	–
		Άλλοι/μεικτοί	5	328 / 500	0.807 (0.540, 1.207)	8.0%



Εικόνα 5.2. Επίδραση του αλληλομόρφου (+3953/4)T του γονιδίου της IL-1B (rs1143634) στην προδιάθεση για HIV-1.



Εικόνα 5.3. Επίδραση του πολυμορφισμού C(+3953/4)T του γονιδίου της IL-1B (rs1143634) στην προδιάθεση για HIV-1 υπό τη θεώρηση υπολειπόμενου μοντέλου κληρονομικότητας.



Εικόνα 5.4. Επίδραση του πολυμορφισμού rs12979860 του γονιδίου της IL-28B στην προδιάθεση για HIV-1 υπό τη θεώρηση επικρατούς μοντέλου κληρονομικότητας.

5.4. Συζήτηση

Τα δεδομένα 42 μελετών, με 8,953 HIV+ και 6,774 HIV- άτομα, συνδυάστηκαν σε μια προσπάθεια να διερευνηθεί η συσχέτιση μεταξύ της μόλυνσης από τον HIV με 8 πολυμορφισμούς των γονιδίων των ιντερλευκινών: IL-1A C(-889)T (rs1800587), IL-1B C(+3953/4)T (rs1143634), IL-4 C(-589/90)T (rs2243250), IL-6 G(-174)C (rs1800795), IL-10 C(-592)A (rs1800872), IL-10 A(-1082)G (rs1800896), IL-12B A(-1188)C (rs3212227), και IL-28B (rs12979860) (C/T). Υπό την υπόθεση του

υπολειπόμενου μοντέλου κληρονομικότητας, οι πολυμορφισμοί rs1143634 του γονιδίου της IL-1B και rs1800872 του γονιδίου της IL-10 συσχετίστηκαν με αυξημένο κίνδυνο οροθετικότητας. Από την άλλη πλευρά, τόσο η κλασσική μονομεταβλητή όσο και η διμεταβλητή προσέγγιση υποδεικνύουν ότι στους ασθενείς που έχουν μολυνθεί από τον HCV το αλληλόμορφο T του πολυμορφισμού rs12979860 του γονιδίου IL-28B, υποθέτοντας επικρατές μοντέλο κληρονομικότητας, μειώνει την πιθανότητα απόκτησης του HIV.

Η διαλεύκανση της μετάδοσης του HIV είναι σημαντική για να αναπτυχθούν νέες μέθοδοι προστασίας και να διαμορφωθούν αποτελεσματικές θεραπείες. Η προδιάθεση για τον HIV, ωστόσο, είναι το αποτέλεσμα μιας σύνθετης αλληλεπίδρασης μεταξύ του ιού, του ξενιστή, και του περιβάλλοντος. Οι γονιδιακοί πολυμορφισμοί μπορούν να επηρεάσουν κάποιες συνιστώσες του ξενιστή κατά τη διαδικασία της μόλυνσης περιλαμβάνοντας τμήματα της ανοσολογικής απόκρισης. Οι κυτταροκίνες, οι οποίες αποτελούν σημαντικούς ρυθμιστές της διαδικασίας της φλεγμονής, κατέχουν εξέχον ρόλο στην παθογένεση του HIV (Fellay 2009; Levy 2009).

Η συγκεκριμένη μετα-ανάλυση αξιολόγησε πολυμορφισμούς των γονιδίων των κυτταροκινών και οδήγησε σε ευρήματα με βιολογική ερμηνεία. Το αλληλόμορφο T του πολυμορφισμού IL-1B C(+3953/4)T το οποίο αυξάνει τον κίνδυνο για HIV έχει συσχετιστεί με αυξημένη έκφραση του υπό μελέτη μορίου (Pociot, Molvig et al. 1992). Η κυτταροκίνη αυτή έχει αναφερθεί ότι αυξάνει την αντιγραφή του HIV και έχει βρεθεί σε μεγαλύτερες συγκεντρώσεις σε μολυσμένους ασθενείς με τον ιό αυτό (Poli, Kinter et al. 1994). Η πλειοτροπική κυτταροκίνη που κωδικοποιείται από το γονίδιο IL-10 μπορεί να καταστείλει τις προφλεγμονώδεις αποκρίσεις (Mosser and Zhang 2008; Ouyang, Rutz et al. 2011). Μέσω των ανοσοκατασταλτικών της δράσεων, και πιο συγκεκριμένα, καταστέλλοντας μόρια όπως ο TNF-α, έχει φανεί ότι αναστέλλει την αντιγραφή του HIV σε μακροφάγα / μονοκύτταρα (Weissman, Poli et al. 1994; Chang, Naif et al. 1996). Το αλληλόμορφο A στη θέση 592 του υποκινητή του γονιδίου της IL-10 έχει συσχετιστεί με μειωμένη παραγωγή της συγκεκριμένης κυτταροκίνης (Crawley, Kay et al. 1999).

Το γονίδιο IL-28B που εδράζεται στο χρωμόσωμα 19 κωδικοποιεί την ιντερφερόνη λάμδα 3 (Interferon lambda 3-IFNλ3), που ανήκει στις ιντερφερόνες τύπου III (Sheppard, Kindsvogel et al. 2003; Ank, West et al. 2006; Li, Liu et al. 2009; Bellanti, Vendemiale et al. 2012). Η IFNλ3 παράγεται κυρίως από τα περιφερειακά

μονοπύρηννα κύτταρα (Peripheral blood mononuclear cells-PBMCs) και τα δενδριτικά κύτταρα (dendritic cells). Οι βιολογικές ιδιότητες της IFNλ3 περιλαμβάνουν ανοσορυθμιστικές δράσεις σε μια ποικιλία κυττάρων. Έχει βρεθεί επίσης ότι καταστέλλει τον HCV in vitro (Li, Liu et al. 2009; Bellanti, Vendemiale et al. 2012). Η IFNλ3 επάγει την έκφραση των Interferon-Stimulated Genes (ISG) και ενισχύει την επίκτητη ανοσία (Morrow, Pankhong et al. 2009). Επιπλέον, η IFNλ2, που είναι σχεδόν πανομοιότυπη με την IFNλ3, αναστέλλει την μόλυνση από τον HIV και καταστέλλει την αντιγραφή των μακροφάγων (Hou, Wang et al. 2009). Ωστόσο, σε μια μελέτη (Serra, Biolchini et al. 2008) παρουσιάστηκαν αντικρουόμενα αποτελέσματα, δηλαδή αυξημένη μόλυνση από τον HIV και αντιγραφή των ανθρώπινων PBMCs και C8166 T κυττάρων όταν χορηγήθηκε IFNλ2.

Ο πολυμορφισμός rs12979860, εδράζεται 3kB αναδρομικά του γονιδίου της IL-28B, και έχει μελετηθεί στην περίπτωση της μόλυνσης από τον HCV. Η σύνδεση, ωστόσο, μεταξύ του πολυμορφισμού αυτού και της σηματοδότησης της IFNλ3 δεν είναι γνωστή. Μια μελέτη έχει δείξει αυξημένα επίπεδα ορού της IFNλ3 σε φορείς για τα αλληλόμορφο C σε σχέση με τους ομόζυγους για τον TT γονότυπο (Langhans, Kupfer et al. 2011). Παρά την περιορισμένη γνώση του λειτουργικού ρόλου του πολυμορφισμού rs12979860, παλαιότερες μελέτες έχουν δείξει ότι τα άτομα με γονότυπους TT ή CT είναι λιγότερο πιθανό, σε σχέση με άτομα με γονότυπο CC, να εξαλείψουν τον HCV και να ανταποκριθούν σε θεραπευτικό σχήμα peginterferon/ribavirin (Thomas, Thio et al. 2009; Chen, Xu et al. 2012). Ωστόσο, οι συσχετίσεις αυτές δεν παρατηρήθηκαν σε μολυσμένους ή με επαναμόλυνση από τον HCV ομοφιλόφυλους άντρες (men who have sex with men-MSM) (Ingiliz, Krznaric et al. 2014; Seaberg, Witt et al. 2015). Επιπλέον, αν και ο γονότυπος CC έχει συσχετιστεί με αρχική μείωση του ιικού φορτίου (Hajarizadeh, Grady et al. 2015), οι χρόνιοι ασθενείς με HCV με γονότυπο CC είχαν υψηλότερα επίπεδα HCV RNA σε σχέση με εκείνους με γονότυπους CT/TT (Uccellini, Tseng et al. 2012; Grady, Prins et al. 2015).

Η παρούσα μετα-ανάλυση ανίχνευσε έναν πιθανό προστατευτικό ρόλο των ατόμων με HCV και γονότυπους CT/TT. Λαμβάνοντας υπόψη ότι έχει ήδη αναφερθεί, μπορούμε να υποθέσουμε ότι ο πολυμορφισμός rs12979860 C/T του γονιδίου της IL-28B μπορεί να έχει μια διαφορική επίδραση μεταξύ των διαφόρων πληθυσμιακών ομάδων στην μόλυνση από τον HIV. Για παράδειγμα, η μειωμένη επίκτητη ανοσία

των ατόμων με γονότυπους CT/TT καθιστά λιγότερο πιθανή την ίαση από τον HCV και δημιουργεί την ίδια στιγμή ένα δυσμενές περιβάλλον για την εδραίωση του HIV.

Η γενετική βάση των ξενιστών σε μολυσματικές νόσους έχει ιδιαίτερο ενδιαφέρον. Κάποιες ευρυγονιδιωματικές μελέτες έχουν επικεντρωθεί στο ιικό φορτίο ή την εξέλιξη της νόσου του HIV ανιχνεύοντας πολυμορφισμούς που εξηγούν μικρό κομμάτι της παρατηρούμενης ποικιλομορφίας μεταξύ των ασθενών με μόλυνση από τον HIV (Fellay, Shianna et al. 2007; Dalmasso, Carpentier et al. 2008; Le Clerc, Limou et al. 2009; Pereyra, Jia et al. 2010; van Manen, van 't Wout et al. 2012). Επιπλέον, μια ευρυγονιδιωματική μελέτη έδειξε ισχυρή σύνδεση της μόλυνσης του HIV με τον πολυμορφισμό rs6996198 T/C στο χρωμόσωμα 8 (το αλληλόμορφο T μειώνει τον κίνδυνο για HIV) (Limou, Delaneau et al. 2012). Αναλύσεις γονιδιακής έκφρασης παρέχουν ενδείξεις για την προστατευτική επίδραση του αλληλομόρφου G του πολυμορφισμού rs4878712 (Johnson, Hancock et al. 2015). Πρόσφατα, έχουν αναφερθεί νέες συσχετίσεις μεταξύ της νόσου από τον HIV και πολυμορφισμών των γονιδίων Zinc ribbon domain containing 1 (ZNRD1) και Bone marrow stromal cell antigen 2 (BST2) (Bigham, Mackelprang et al. 2014; Hancock, Gaddis et al. 2015).

Γενικώς, αν και αρκετές σημαντικές εργασίες έχουν παρουσιαστεί, μέχρι σήμερα μόνο πολυμορφισμοί του γονιδίου CCR5 έχει αποδειχθεί ότι επηρεάζουν σημαντικά την μόλυνση από τον HIV και οδήγησαν στη δημιουργία αναστολέων (Dean, Carrington et al. 1996; Samson, Libert et al. 1996; Ioannidis, Rosenberg et al. 2001). Φαίνεται ότι οι προσεγγίσεις στόχευσης γονιδιακών πολυμορφισμών (Bigham, Mackelprang et al. 2014) και νέες ευρυγονιδιωματικές μελέτες είναι απαραίτητες περιλαμβάνοντας την αλληλούχιση πολυμορφισμών των γονιδίων των κυτταροκινών. Επιπρόσθετα, μετα-αναλύσεις των δεδομένων των ευρυγονιδιωματικών μελετών θα μπορούσαν να οδηγήσουν στην ανίχνευση επιπλέον πολυμορφισμών αυξάνοντας έτσι την ισχύ.

Η παρούσα μετα-ανάλυση παρουσιάζει κάποιους περιορισμούς. Σε πολλές αναλύσεις, ο αριθμός των μελετών ήταν μικρός με συνεπακόλουθη μείωση της ισχύος. Επιπλέον, οι σημαντικές εκτιμήσεις μπορεί να έχουν προέλθει λόγω τύχης ή λόγω των μεροληπτικών εκτιμήσεων των πρωταρχικών μελετών. Για παράδειγμα, μεταξύ των τεσσάρων μελετών που αναλύθηκαν κάτω από το υπολειπόμενο μοντέλο κληρονομικότητας για τον πολυμορφισμό του γονιδίου IL-1B C(+3953/4)T, μόνο μια διερευνούσε αποκλειστικά το ρόλο αυτού του γενετικού παράγοντα στην προδιάθεση για HIV. Οι υπόλοιπες τρεις μελέτες αφορούσαν διαφορετικές εκβάσεις

περιλαμβάνοντας την περιοδοντική νόσο, τη σχετιζόμενη με τον HIV άνοια, και το σύνδρομο λιποδυστροφίας (Asensi, Rego et al. 2008; Pemberton, Stone et al. 2008; Goncalves, Ferreira et al. 2009). Παρόμοια, για τον πολυμορφισμό rs12979860 του γονιδίου της IL-28B, οι 11 από τις 13 μελέτες που αξιοποιήθηκαν στη μετα-ανάλυση περιελάμβαναν ασθενείς με συλλοίμωξη HIV/HCV. Συνεπώς, η μεροληπτική επιλογή του πληθυσμού ή άλλοι συγχυτικοί παράγοντες μπορεί να έχουν επηρεάσει τα αποτελέσματα. Φυσικά, από την άλλη πλευρά, οι αναλύσεις σε ασθενείς με συλλοίμωξη HIV/HCV πιθανώς επέτρεψαν τη χρήση συγκρίσιμων ομάδων υγιών ως προς τα επίπεδα έκθεσης στον HIV μιας και οι τρόποι μετάδοσης των HIV και HCV είναι κοινοί (για παράδειγμα με τη χρήση ενδοφλέβιων ναρκωτικών). Επιπλέον, σε μια προσπάθεια να ελαχιστοποιήσουμε τις πηγές σφαλμάτων, επεκτείναμε τη βιβλιογραφική αναζήτηση για μελέτες που δεν υπήρχαν στις ευρέως γνωστές βάσεις δεδομένων (Pubmed, Scopus), πραγματοποιήσαμε ελέγχους για το συστηματικό σφάλμα δημοσίευσης, και αναζητήσαμε, μέσω αθροιστικής μετα-ανάλυσης, ακραία ευρήματα των πρωταρχικών μελετών που θα μπορούσαν να υπό- ή υπερ-εκτιμήσουν τις πραγματικές επιδράσεις γονιδίων (Ioannidis and Trikalinos 2005).

Συμπερασματικά, η παρούσα μετα-ανάλυση έδειξε επιβαρυντική επίδραση του πολυμορφισμού rs1143634 του γονιδίου της IL-1B στην μόλυνση από τον HIV και προστατευτική επίδραση για το αλληλόμορφο T του πολυμορφισμού rs12979860 του γονιδίου IL-28B. Τα ευρήματα αυτά υποδηλώνουν τη σημαντικότητα των ιντερλευκινών στον κίνδυνο εμφάνισης HIV. Ωστόσο, απαιτείται περαιτέρω διερεύνηση. Στην κατεύθυνση αυτή θα βοηθούσε η διεξαγωγή μιας ευρυγονιδιωματικής μελέτης, με σαφώς καθορισμένους πληθυσμούς και ξεκάθαρα φαινοτυπικά χαρακτηριστικά. Τα άτομα που κάνουν χρήση ενδοφλέβιων ναρκωτικών θα μπορούσαν να αποτελέσουν μια ομάδα στόχο για τη διερεύνηση του πολυμορφισμού rs12979860 του γονιδίου της IL-28B.

Μέρος Β: Μελέτη της επίδρασης της συλλοίμωξης με τον ιό της Ηπατίτιδας C στην ανοσολογική και ιολογική ανταπόκριση HIV ασθενών μετά την έναρξη αντιρετροϊκής θεραπείας υψηλής δραστηριότητας

5.5. Εισαγωγή

Η αντιρετροϊκή θεραπεία υψηλής δραστηριότητας (Highly Active Antiretroviral Therapy-HAART) έχει οδηγήσει σε εντυπωσιακή μείωση της θνησιμότητας και της νοσηρότητας των ατόμων που έχουν μολυνθεί με τον ιό της ανθρώπινης ανοσοανεπάρκειας (Human Immunodeficiency Virus-HIV) (Puhan, Van Natta et al. 2010) αλλά η επιμήκυνση του χρόνου ζωής επέτρεψε την εμφάνιση συνοδών νοσημάτων, συμπεριλαμβάνοντας τη μόλυνση από τον ιό της ηπατίτιδας C (Hepatitis C Virus-HCV).

Η ταυτόχρονη μόλυνση από τους HIV και HCV είναι αρκετά συχνή μιας και μεταδίδονται με παρόμοιους τρόπους. Συνολικά, σχεδόν ένα 20% των ατόμων με HIV πάσχει από χρόνια ηπατίτιδα C (Soriano, Vispo et al. 2010). Μεταξύ των HIV+ που κάνουν χρήση ενδοφλέβιων ναρκωτικών (Injecting Drug Users-IDUs), ο επιπολασμός της μόλυνσης από τον HCV κυμαίνεται μεταξύ 82% και 93% (Mohsen, Murad et al. 2005; Solomon, Srikrishnan et al. 2008; Nelson, Mathers et al. 2011). Η επίπτωση του HCV έχει αυξηθεί επίσης στους HIV+ ομοφυλόφιλους άντρες την τελευταία δεκαετία (Danta and Rodger 2011).

Οι πρωταρχικές μελέτες για την επίδραση του HCV στην εξέλιξη του HIV έχουν οδηγήσει σε αντικρουόμενα αποτελέσματα (Macias, Pineda et al. 1998; Greub, Ledergerber et al. 2000; Sulkowski, Moore et al. 2002; Anderson, Guest et al. 2004). Μια πρόσφατη μετα-ανάλυση έδειξε ότι η συλλοίμωξη HIV/HCV δεν μετέβαλλε την επίπτωση του συνδρόμου επίκτητης ανοσοανεπάρκειας (Acquired Immune Deficiency Syndrome-AIDS) αλλά αύξησε τη συνολική θνησιμότητα στην εποχή της HAART (Chen, Ding et al. 2009). Μια ποσοτική ανασκόπηση που πραγματοποιήθηκε το 2005 έδειξε μειωμένη απόκριση του ανοσολογικού συστήματος σε ασθενείς με συλλοίμωξη που ξεκίνησαν αντιρετροϊκή θεραπεία (Miller, Haley et al. 2005).

Στο μεταξύ, έχουν διεξαχθεί πολλές σχετικές με το θέμα μελέτες. Ο μεγάλος όγκος της πληροφορίας που συσσωρεύεται και η απουσία συνοπτικών μέτρων εκτίμησης της ιολογικής ανταπόκρισης στη HAART σε ασθενείς με συλλοίμωξη

HIV/HCV καθιστά απαραίτητη τη διεξαγωγή μιας επικαιροποιημένης μετα-ανάλυσης. Ο σκοπός της παρούσας μετα-ανάλυσης είναι ο συνδυασμός και η αξιολόγηση της υπάρχουσας πληροφορίας σχετικά με την επίδραση του HCV στην ανοσολογική και ιολογική ανταπόκριση των ασθενών με HIV που λαμβάνουν HAART ή συνδυασμένη αντιρετροϊκή θεραπεία (combined Antiretroviral Treatment-c-ART) με τουλάχιστον 3 φάρμακα.

5.6. Μεθοδολογία

5.6.1. Κριτήρια επιλεξιμότητας, πηγές δεδομένων και διαδικασία αναζήτησης

Η μετα-ανάλυση διεξήχθη με συγκεκριμένες κατευθυντήριες γραμμές (MOOSE guidelines (Stroup, Berlin et al. 2000) και PRISMA statement (Moher, Liberati et al. 2010)). Μελέτες με HIV+ ασθενείς περιλαμβάνονταν στην ανάλυση εάν: (i) εξέταζαν την επίδραση της συλλοίμωξης HIV/HCV στην ανοσολογική και ιολογική ανταπόκριση των ασθενών που λαμβάνουν HAART ή c-ART, ii) περιελάμβαναν HAART-naïve ασθενείς κατά την έναρξη, και iii) παρείχαν κάποια μέτρηση για CD4 T κύτταρα ή τη διαφορά των επιπέδων HIV-RNA από την έναρξη έως 3-24 μήνες από τη λήψη ART, ή κάποια εκτίμηση του σχετικού κινδύνου (Relative Risk-RR) της ανοσολογικής και ιολογικής ανταπόκρισης ή επαρκή δεδομένα για τον υπολογισμό των μέτρων αυτών. Δεν τέθηκαν περιορισμοί γλώσσας.

Σχετικές μελέτες ανιχνεύτηκαν στην Medline, Scopus, και ISI Web of Science χρησιμοποιώντας συνδυασμούς των ακόλουθων όρων: “HIV”, “HCV”, “HAART”, “antiretroviral treatment”, “CD4 T”, “HIV-RNA” και “viral load” (τελευταία αναζήτηση: Φεβρουάριος 2012). Οι τίτλοι και οι περιλήψεις ελέγχονταν προσεκτικά έτσι ώστε να αποκλειστούν οι μη σχετικές με το θέμα εργασίες. Τα άρθρα που απέμειναν αξιολογήθηκαν για να καθοριστεί εάν πληρούσαν τα κριτήρια επιλεξιμότητας. Οι βιβλιογραφικές αναφορές σχετικών δημοσιεύσεων και οι περιλήψεις συνεδρίων εξετάστηκαν επίσης για να εντοπιστούν περαιτέρω μελέτες που θα μπορούσαν να χρησιμοποιηθούν στη μετα-ανάλυση.

5.6.2. Εξαγωγή δεδομένων

Η εξαγωγή των δεδομένων έγινε από δυο ανεξάρτητους ερευνητές και συζητήθηκαν οι πιθανές ασυμφωνίες. Από κάθε μελέτη, αντλήθηκαν οι ακόλουθες

πληροφορίες, εάν αυτές ήταν διαθέσιμες: i) όνομα του πρώτου συγγραφέα, έτος δημοσίευσης, περιοχή, και σχεδιασμός μελέτης, ii) διαγνωστική μέθοδος της μόλυνσης HCV, iii) αριθμός των ατόμων και χρόνος παρακολούθησης, iv) χαρακτηριστικά των ατόμων κατά την έναρξη και εργαστηριακές μετρήσεις, v) μέσες διαφορές (με τις τυπικές αποκλίσεις αυτών) του αριθμού των CD4 T κυττάρων και των επιπέδων HIV-RNA των μετρήσεων κατά την έναρξη της HAART ή c-ART και εκείνων 3-24 μηνών αργότερα, και vi) RRs για να επιτευχθεί αύξηση CD4 T κυττάρων και/ή μείωση HIV-RNA με τα αντίστοιχα 95% διαστήματα εμπιστοσύνης.

5.6.3. Στατιστική ανάλυση

Οι μη τυποποιημένες μέσες διαφορές με τα 95% C.I.s υπολογίστηκαν χρησιμοποιώντας μοντέλα σταθερών ή τυχαίων επιδράσεων. Για την εκτίμηση των συνδυασμένων RRs χρησιμοποιήθηκαν μοντέλα μονομεταβλητής μετα-ανάλυσης τυχαίων επιδράσεων (DerSimonian and Laird 1986). Για την περίπτωση των επιμέρους μελετών που δεν δίνονταν απευθείας τα Hazard Ratios (HRs), αυτά υπολογίστηκαν χρησιμοποιώντας τη μέθοδο του Pernerger και συνεργατών (Pernerger 2008). Οι έλεγχοι ετερογένειας περιελάμβαναν τη χρήση του Cochran Q statistic και του δείκτη I^2 (Higgins, Thompson et al. 2003). Η παρουσία του συστηματικού σφάλματος δημοσίευσης διερευνήθηκε χρησιμοποιώντας τους ελέγχους των Begg (Begg and Mazumdar 1994) και Egger (Egger, Davey Smith et al. 1997). Η διαχρονική τάση ανιχνεύτηκε μέσω αθροιστικής μετα-ανάλυσης (Lau, Antman et al. 1992). Αναλύσεις ευαισθησίας πραγματοποιήθηκαν μέσω αναλύσεων υποομάδων είτε αποκλείοντας από την μετα-ανάλυση μια μελέτη κάθε φορά και επαναλαμβάνοντας την ανάλυση. Επιπλέον, υλοποιήθηκε μια διμεταβλητή προσέγγιση για να ληφθεί υπόψη η πιθανή συσχέτιση μεταξύ ανοσολογικής και ιολογικής ανταπόκρισης (Riley, Thompson et al. 2008).

Σε όλες τις αναλύσεις χρησιμοποιήθηκε το στατιστικό πακέτο Stata 12 (Stata Corporation, College Station, Texas, USA). Εκτός από τους στατιστικούς ελέγχους της ετερογένειας (όπου η σημαντικότητα δηλωνόταν για p -values<0.10), τα αποτελέσματα ήταν στατιστικά σημαντικά εάν τα αντίστοιχα p -values ήταν μικρότερα από 0.05. Όλοι οι έλεγχοι ήταν αμφίπλευροι.

Εάν δεν δηλώνεται διαφορετικά, παρουσιάζονται μόνο οι εκτιμήσεις των μοντέλων τυχαίων επιδράσεων.

5.7. Αποτελέσματα

5.7.1. Περιγραφή των μελετών

Η αναζήτηση της βιβλιογραφίας έδωσε ένα σύνολο 31 σχετικών μελετών. Δέκα από αυτές αποκλείστηκαν επειδή δεν είχαν επαρκή δεδομένα ή είχαν δημοσιευτεί σε περισσότερες της μιας εργασίες (Εικόνα 5.5). Τελικά, στη μετα-ανάλυση συμπεριλήφθηκαν 21 μελέτες που περιελάμβαναν 5,278 ασθενείς με συλλοίμωξη HIV/HCV και 17,255 μόνο με HIV. Από αυτές, οι 10 παρείχαν πληροφορίες για την ανοσολογική απόκριση (Greub, Ledergerber et al. 2000; Sulkowski, Moore et al. 2002; Macias, Pineda et al. 2003; Antonucci, Girardi et al. 2005; Braitstein, Zala et al. 2006; Emokpae, Nwokedi et al. 2008; Santin, Mestre et al. 2008; Castagna, Galli et al. 2010; Isa, Gwamzhi et al. 2010; Seminari, Tinelli et al. 2010) και οι 11 και για τις δυο εκβάσεις (Filippini, Coppola et al. 2000; Chung, Evans et al. 2002; De Luca, Bugarini et al. 2002; Klein, Lalonde et al. 2003; Lincoln, Petoumenos et al. 2003; Rockstroh, Mocroft et al. 2005; Stapleton, Bennett et al. 2007; Zhou, Dore et al. 2007; Carmo, Guimaraes et al. 2008; Turner, Bansi et al. 2010; Zhao, Yu et al. 2011) (Πίνακας 5.4). Η ανοσολογική απόκριση μετρήθηκε είτε με την αλλαγή του αριθμού των CD4 T κυττάρων από την έναρξη της ART έως 3-24 μήνες μετά και/ή από την αύξηση των CD4 T κυττάρων κατά τουλάχιστον 50, 75, 100, ή 200 κύτταρα/mm³. Η ιολογική απόκριση καθορίστηκε κυρίως από το αν τα επίπεδα HIV-RNA ήταν κάτω από 50, 400, ή 500 αντίγραφα/ml ή κάτω από το όριο ανίχνευσης. Σε 7 μελέτες η διάγνωση του HCV πραγματοποιήθηκε με τη χρήση ενός θετικού anti-HCV test και με την ανίχνευση HCV-RNA στο πλάσμα (Filippini, Coppola et al. 2000; Chung, Evans et al. 2002; Macias, Pineda et al. 2003; Antonucci, Girardi et al. 2005; Santin, Mestre et al. 2008; Seminari, Tinelli et al. 2010; Zhao, Yu et al. 2011). Η πλειοψηφία των μελετών ήταν προοπτικές μελέτες κοορτής.

Πίνακας 5.4. Περιγραφή των πρωταρχικών μελετών.

Μελέτη	Είδος	N		ART naïve (%)		Αγωγή ART	Καταληκτικό σημείο	Έκβαση
		HIV	HIV/HCV	HIV	HIV/HCV			
Santin 2008 (Santin, Mestre et al. 2008)	Προοπτική μελέτη κοορτής	24	15	100	100	Συνδυασμένη ART, τουλάχιστον 3 φάρμακα	48 εβδομάδες μετά την έναρξη ART	Διαφορά CD4
Macias 2003 (Macias, Pineda et al. 2003)	Προοπτική μελέτη κοορτής	58	60	100	100	HAART	12 μήνες μετά την έναρξη HAART	Διαφορά CD4 ^a , HR ανάκτησης CD4 50 ^b και 200 κύτταρα/mm ³
Braitstein 2006 (Braitstein, Zala et al. 2006)	Nested μελέτη κοορτής	580	606	100	100	HAART	48 εβδομάδες μετά την έναρξη HAART, χρόνος μέχρι την ανοσολογική ανταπόκριση	Διαφορά CD4, HR ανάκτησης CD4 ≥75 κύτταρα/mm ³
Sulkowski 2002 (Sulkowski, Moore et al. 2002)	Προοπτική μελέτη κοορτής	122	65	NA	NA	HAART	≥ 1 έτους ≥ 2 έτη ≥ 3 έτη μετά την έναρξη HAART	Διαφορά CD4, HR ανάκτησης CD4 ≥50 κύτταρα/mm ^{3b} , HR ανάκτησης CD4 ≥100 κύτταρα/mm ^{3b}

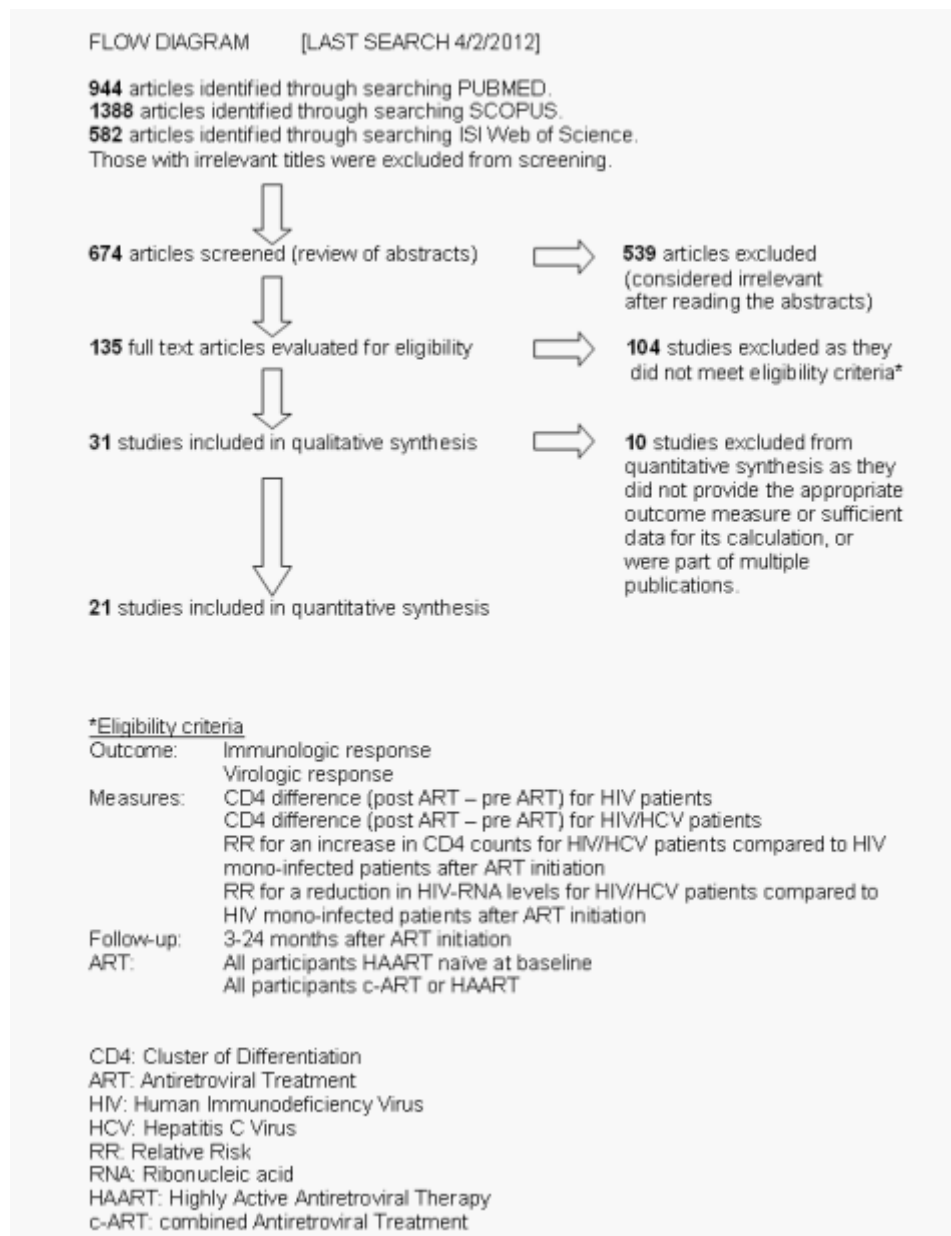
Filippini 2000 (Filippini, Coppola et al. 2000)	Προοπτική μελέτη κοορτής	8	8	NA	NA	HAART	1, 3, 6 μήνες μετά την έναρξη HAART	Διαφορά CD4 ^b , Ιικό φορτίο πριν και μετά, HR επίτευξης μη ανιχνεύσιμου ιικού φορτίου ^b
Antonucci 2005 (Antonucci, Girardi et al. 2005)	Προοπτική μελέτη κοορτής	1219	284	100	100	HAART	Χρόνος μέχρι την ανοσολογική ανταπόκριση	HR ανάκτησης CD4 ≥ 100 κύτταρα/mm ³
Castagna 2010 (Castagna, Galli et al. 2010)	Αναδρομική multicohort	1053	435	NA	NA	HAART	12, 24 εβδομάδες μετά την έναρξη HAART	Διαφορά CD4
Seminari 2010 (Seminari, Tinelli et al. 2010)	Αναδρομική μελέτη κοορτής	673	171	47	35	HAART	Χρόνος μέχρι να μην υπάρχει ανιχνεύσιμο ιικό φορτίο (<12 μήνες), 12 μήνες μετά από κατασταλακτική HAART(12<..<<24 μήνες)	Διαφορά CD4, OR ανάκτησης CD4 ≥ 100 κύτταρα /mm ³
De Luca 2002 (De Luca, Bugarini et al. 2002)	Προοπτική μελέτη κοορτής	674	556	100	100	Συνδυασμένη ART, τουλάχιστον 3 φάρμακα	Χρόνος μέχρι την ανοσολογική ανταπόκριση και το ιικό φορτίο	HR ανάκτησης CD4 ≥ 100 κύτταρα/mm ³ ή επίτευξης 500 κύτταρα/mm ³ , HR επίτευξης ιικού φορτίου<500 αντίγραφα/ml

Rockstroh 2005 (Rockstroh, Mocroft et al. 2005)	Προοπτική μελέτη κοορτής	1492	768	NA	NA	HAART	Χρόνος μέχρι την ανοσολογική ανταπόκριση και το ιικό φορτίο	HR ανάκτησης CD4 ≥ 50 κύτταρα/mm ³ , HR επίτευξης ιικού φορτίου <500 αντίγραφα/ml
Klein 2003 (Klein, Lalonde et al. 2003)	Προοπτική μελέτη κοορτής	263	52	NA	NA	HAART	24 μήνες μετά την έναρξη HAART, χρόνος μέχρι την ανοσολογική ανταπόκριση και το ιικό φορτίο	Διαφορά CD4, HR ανάκτησης CD4 ≥ 50 κύτταρα/mm ³ , HR επίτευξης ιικού φορτίου <500 αντίγραφα/ml
Chung 2002 (Chung, Evans et al. 2002)	Αναδρομική μελέτη κοορτής	129	40	NA	NA	HAART	16, 48 εβδομάδες μετά την έναρξη HAART	Διαφορά CD4, HR επίτευξης ιικού φορτίου <500 αντίγραφα/ml ^b
Zhou 2007 (Zhou, Dore et al. 2007)	Προοπτική μελέτη κοορτής	371	43	NA	NA	HAART	180 ημέρες μετά την έναρξη HAART	Διαφορά CD4, HR επίτευξης ιικού φορτίου <400 αντίγραφα/ml ^b
Carmo 2008 (Carmo, Guimaraes et al. 2008)	Αναδρομική μελέτη κοορτής	748	76	56	34	HAART	Χρόνος μέχρι την ανοσολογική ανταπόκριση και το ιικό φορτίο	HR ανάκτησης CD4 ≥ 50 κύτταρα/mm ³ , HR επίτευξης ιικού φορτίου <400 αντίγραφα/ml

Turner 2010 (Turner, Bansi et al. 2010)	Προοπτική μελέτη κοορτής	5500	396	100	100	HAART	Χρόνος μέχρι την ανοσολογική ανταπόκριση και το ιικό φορτίο	HR ανάκτησης CD4 >50 κύτταρα/mm ³ , HR επίτευξης ιικού φορτίου <50 αντίγραφα/ml
Stapleton 2007 (Stapleton, Bennett et al. 2007)	Longitudinal μελέτη κοορτής	1238	148	100	100	Συνδυασμένη ART, τουλάχιστον 3 φάρμακα και HAART	48 εβδομάδες μετά την έναρξη ART	Διαφορά CD4, HR επίτευξης ιικού φορτίου <50 αντίγραφα/ml ^b
Lincoln 2003 (Lincoln, Petoumenos et al. 2003)	Αναδρομική μελέτη κοορτής	772	112	NA	NA	HAART	12, 24 μήνες μετά την έναρξη HAART	Διαφορά CD4, HR επίτευξης μη ανιχνεύσιμου ιικού φορτίου ^b
Zhao 2011 (Zhao, Yu et al. 2011)	Αναδρομική μελέτη κοορτής	178	168	100	100	HAART	1 χρόνος μετά την έναρξη HAART	Διαφορά CD4 ^b , Ιικό φορτίο πριν και μετά
Isa 2010 (Isa, Gwamzhi et al. 2010)	Προοπτική μελέτη κοορτής	149	68	100	100	HAART	3, 6 μήνες μετά την έναρξη HAART	HR ανάκτησης CD4 ≥50 κύτταρα/mm ^{3b}
Emokpae 2008 (Emokpae, Nwokedi et al. 2008)	Αναδρομική μελέτη κοορτής	50	50	100	100	HAART	6 μήνες μετά την έναρξη HAART	Διαφορά CD4 ^b

Greub 2000 (Greub, Ledergerber et al. 2000)	Προοπτική μελέτη κοορτής	1954	1157	52	59	HAART	12 μήνες μετά την έναρξη HAART	Διαφορά CD4 ^a , HR ανάκτησης CD4 ≥ 50 κύτταρα/mm ³
---------------------------------------------------	-----------------------------	------	------	----	----	-------	-----------------------------------	-------------------------------------------------------------------------------------

^a τα δεδομένα προέρχονται από τη μετα-ανάλυση των Miller και συνεργατών (Miller, Haley et al. 2005), ^b υπολογίστηκαν από δεδομένα που παρέχονταν



Εικόνα 5.5. Διαδικασία επιλογής των μελετών.

5.7.2. Ανοσολογική και ιολογική ανταπόκριση

Ανοσολογική ανταπόκριση: Δεν υπήρχε διαφορά στον αριθμό των CD4 T κυττάρων κατά την έναρξη μεταξύ των δυο ομάδων ασθενών. Όπως φαίνεται στον Πίνακα 5.5 και στην Εικόνα 5.6, η συνολική διαφορά των CD4 T κυττάρων μεγαλώνει μεταξύ των HIV+ ασθενών και εκείνων με συλλοίμωξη HIV/HCV σε 12 μελέτες μετά από 3 έως 12 μήνες που λαμβάνουν HAART ή c-ART και εκτιμήθηκε 34.86 κύτταρα/mm³ (95% C.I.: 16.82, 52.89, I²=92.3%). Η διαφορά αυτή έγινε ακόμα μεγαλύτερη όταν ο αριθμός των CD4 T κυττάρων κατά την έναρξη ήταν κάτω από 350 κύτταρα/mm³ (38.97, 95% C.I.: 20.00, 57.93) (Πίνακας 5.5). Τα αποτελέσματα ήταν μη σημαντικά (27.26, 95% C.I.: -2.21, 56.72) (Πίνακας 5.5) για τις μελέτες με διπλή διάγνωση του HCV (δείγμα ορού και μοριακός έλεγχος). Τα αποτελέσματα όλων των αναλύσεων υποομάδας παρουσιάζονται στον Πίνακα 5.5.

Η ποσοτική σύνθεση των HRs για αύξηση του αριθμού των CD4 T κυττάρων (50 ή 75 ή 100 κύτταρα/mm³) σε ασθενείς με συλλοίμωξη έναντι των HIV+ μετά την έναρξη της ART έδωσε επίσης στατιστικά σημαντικές εκτιμήσεις (HR 8 μελετών: 0.82, 95% C.I.: 0.75, 0.91) (Πίνακας 5.5).

Ιολογική ανταπόκριση: Η διαφορά της ιολογικής ανταπόκρισης εξετάστηκε σε 10 μελέτες με 13,394 ασθενείς. Το συνοπτικό adjusted HR για να επιτευχθεί μη ανιχνεύσιμο ιικό φορτίο ή επίπεδα HIV-RNA χαμηλότερα από 50, 400 ή 500 αντίγραφα/ml μετά την έναρξη HAART ή c-ART σε ασθενείς με συλλοίμωξη έναντι των HIV+ ήταν 0.99 (95% C.I.: 0.91, 1.07, I²: 0.0%) (Πίνακας 5.5, Εικόνα 5.7).

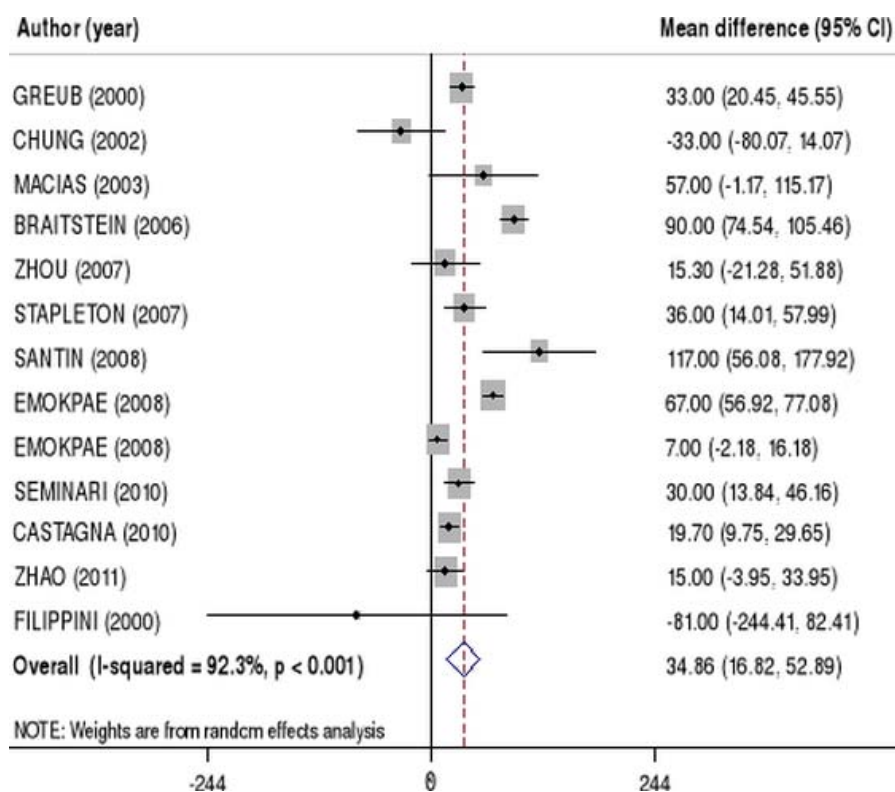
Διμεταβλητή μετα-ανάλυση: Η διμεταβλητή προσέγγιση έδωσε παρόμοιες εκτιμήσεις. Τα adjusted HRs της ανοσολογικής και ιολογικής απόκρισης σε ασθενείς με συλλοίμωξη έναντι των HIV+ ήταν 0.82 (95% C.I.: 0.74, 0.91) και 0.96 (95% C.I.: 0.89, 1.04), αντίστοιχα.

Συστηματικό σφάλμα δημοσίευσης: Σε όλες τις αναλύσεις, οι στατιστικοί έλεγχοι του Begg και Egger δεν έδειξαν την παρουσία σφάλματος δημοσίευσης.

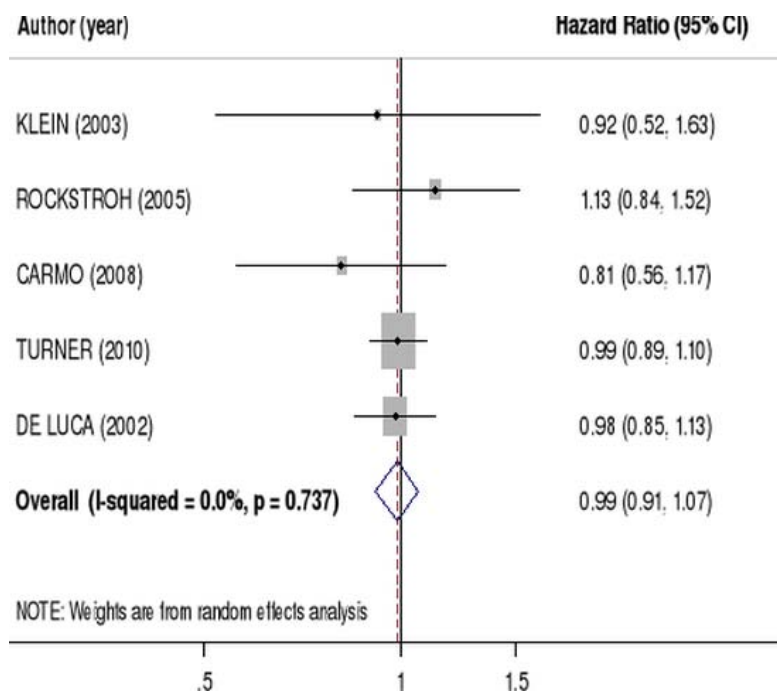
Πίνακας 5.5. Τα αποτελέσματα των μετα-αναλύσεων των μελετών που αξιολογούσαν τη διαφορά στην ανοσολογική και ιολογική ανταπόκριση ασθενών με HIV και ασθενών με συλλοίμωξη HIV/HCV που λάμβαναν αντιρετροϊκή θεραπεία

Αριθμός των							
Έκβαση	Μέτρο έκβασης	Καταληκτικό σημείο	HIV - HIV/HCV ασθενών	Αριθμός μελετών	Συνοπτική εκτίμηση	95% Διάστημα Εμπιστοσύνης	I ² (%)
Ανοσολογική ανταπόκριση	Διαφορά CD4 (κύτταρα/mm ³)	3 – 12 μήνες	5834, 2604	12	34.86	16.82 – 52.89	92.3
<u>Αναλύσεις υποομάδας</u>							
HCV διάγνωση με βάση θετικό δείγμα ορού μόνο			4764, 2142	6	38.91	15.51 – 62.30	95.5
HCV διάγνωση με βάση θετικό δείγμα ορού και ανίχνευση HCV RNA			1070, 462	6	27.26	-2.21 – 56.72	73.4
απουσία ART στο σημείο έναρξης			2127, 1047	6	51.87	21.66 – 82.08	95.4
αγωγή ART (απουσία HAART) στο σημείο έναρξης			3707, 1557	6	21.77	9.19 – 34.36	53.1
CD4<350 στο σημείο έναρξης			5288, 2497	9	38.97	20.00 – 57.93	91.7
CD4≥350 στο σημείο έναρξης			175, 64	3	1.61	-88.68 – 91.89	89.8

Ανοσολογική ανταπόκριση	Διαφορά CD4 (cells/mm ³)	13 – 24 μήνες	1830, 396	4	13.43	0.83 – 26.04	0.0
Ανοσολογική ανταπόκριση	Adjusted HR για HIV/HCV έναντι HIV	Χρόνος μέχρι ανοσολογική ανταπόκριση	12402, 3492	8	0.82	0.75 – 0.91	53.9
Ανοσολογική ανταπόκριση	Unadjusted HR για HIV/HCV έναντι HIV	Χρόνος μέχρι ανοσολογική ανταπόκριση	5523, 2550	8	0.72	0.64 – 0.81	57.6
Ιολογική ανταπόκριση	Adjusted HR για HIV/HCV έναντι HIV	Χρόνος μέχρι ιολογική ανταπόκριση	8677, 1848	5	0.99	0.91 – 1.07	0.0
Ιολογική ανταπόκριση	Unadjusted HR για HIV/HCV έναντι HIV	Χρόνος μέχρι ιολογική ανταπόκριση	4949, 1244	8	0.89	0.76 – 1.04	50.5



Εικόνα 5.6. Μονομεταβλητή μετα-ανάλυση τυχαίων επιδράσεων για τη διαφορά στην μεταβολή των CD4 T κυττάρων μεταξύ ασθενών HIV και ασθενών με συλλοίμωξη HIV/HCV μετά την έναρξη αντιρετροϊκής θεραπείας.



Εικόνα 5.7. Μονομεταβλητή μετα-ανάλυση τυχαίων επιδράσεων των adjusted Hazard Ratios για την ιολογική ανταπόκριση των ασθενών με συλλοίμωξη HIV/HCV σε σχέση με τους ασθενείς με HIV μετά την έναρξη αντιρετροϊκής θεραπείας.

5.8. Συζήτηση

Η παρούσα μετα-ανάλυση με ένα σύνολο 21 μελετών έδειξε ότι η συλλοίμωξη HIV/HCV μειώνει την ανοσολογική ανταπόκριση ασθενών που ξεκινούν τη HAART ή c-ART, κυρίως για τους ασθενείς εκείνους με αριθμό CD4 T κυττάρων χαμηλότερο από 350 κύτταρα/mm³ κατά την έναρξη. Η επίδραση ήταν μικρότερη δυο χρόνια μετά την έναρξη ART. Η ιολογική ανταπόκριση των ασθενών δεν διέφερε σημαντικά για τους ασθενείς με συλλοίμωξη σε σχέση με τους HIV+.

Η επίδραση του HCV στην εξέλιξη της νόσου του HIV δεν έχει διαλευκανθεί. Η μετα-ανάλυση που διεξήγαμε έδειξε επιβαρυντική επίδραση του HCV στην ανάκτηση του αριθμού των CD4 T κυττάρων σε ασθενείς HIV+ παρά τη λήψη θεραπείας. Το φαινόμενο αυτό έχει βιολογική βάση. Τα CD4 T κύτταρα μπορεί να μολυνθούν από τον HCV και οι αλληλεπιδράσεις των HIV και HCV σε κυτταρικό επίπεδο μπορούν να επηρεάσουν την ανοσολογική αποτελεσματικότητα της HAART (Laskus, Radkowski et al. 2004). Επιπλέον, δεδομένου ότι η χρόνια μόλυνση του HCV, ταυτόχρονα με τον HIV, αυξάνει την ενεργοποίηση του συστήματος ανοσίας των T κυττάρων, περιορίζεται ο αριθμός των CD4 T κυττάρων (Hunt, Martin et al. 2003; Gonzalez, Falconer et al. 2009; Sandberg, Falconer et al. 2010; Operskalski and Kovacs 2011). Τέλος, έχει αναφερθεί ότι ο HCV αυξάνει την απόπτωση των CD4 T κυττάρων (Korner, Kramer et al. 2009).

Η παρούσα μετα-ανάλυση έδειξε ότι το ποσοστό ιολογικής ανταπόκρισης στη HAART σε HIV+ άτομα δεν παρουσίαζε διαφοροποιήσεις ανάλογα με την παρουσία του HCV. Αντιθέτως, μια πρόσφατη μετα-ανάλυση τυχαιοποιημένων κλινικών δοκιμών έδειξε ότι η συλλοίμωξη HIV/HCV διαφοροποιούσε τα επίπεδα HIV-RNA (Pulido, Hill et al. 2012). Ωστόσο, η αναφορά αυτή, περιελάμβανε μελέτες που δεν είχαν σχεδιαστεί για να αξιολογήσουν την επίδραση της συλλοίμωξης σαν κύρια έκβαση, στρατολογούσε ασθενείς που είχαν ήδη λάβει HAART κατά την έναρξη και δεν είχε ξεκάθαρη περιγραφή των τεχνικών μετα-ανάλυσης. Δεδομένου ότι οι περισσότερες από αυτές τις δοκιμές υπήρχαν μόνο υπό τη μορφή της περίληψης, μια μετα-ανάλυση των επερχόμενων δημοσιεύσεων με τα πλήρη κείμενα ή τα ατομικά δεδομένα θα βοηθούσε στη διαλεύκανση των πιθανών διαφορών μεταξύ των κλινικών δοκιμών και των μελετών παρατήρησης που χρησιμοποιήθηκαν στη συγκεκριμένη μετα-ανάλυση.

Η μετα-ανάλυση της διαφοράς των CD4 T κυττάρων χαρακτηρίστηκε από μεγάλη μεταξύ των μελετών ετερογένεια η οποία έδωσε το έναυσμα για πολλαπλές αναλύσεις υποομάδων έτσι ώστε να διερευνηθεί από πού προέρχεται. Η λήψη αντιρετροϊκής θεραπείας κατά το παρελθόν θα μπορούσε να οδηγήσει σε ετερογενείς εκτιμήσεις. Στην παρούσα μετα-ανάλυση κάποιες μελέτες περιελάμβαναν ασθενείς που είχαν λάβει ART, κάτι που θα μπορούσε να μειώσει τη συμμόρφωση στη HAART, να ελέγξει το φορτίο του HIV και της ανοσολογικής απόκρισης (Horberg, Silverberg et al. 2008). Ωστόσο παρά κάποιες μικρές διαφορές, τα μέτρα επίδρασης ήταν στατιστικά σημαντικά και στις δυο περιπτώσεις δείχνοντας ότι τα μειωμένα ποσοστά ανταπόκρισης στα επίπεδα CD4 T κυττάρων των ασθενών με συλλοίμωξη ήταν προς την ίδια κατεύθυνση ανεξάρτητα από την πρότερη χορήγηση ART.

Είναι εξαιρετικά σημαντικό να διαγνώσουμε τη μόλυνση από τον HCV μέσω ανίχνευσης του HCV-RNA έτσι ώστε να αποφύγουμε τα σφάλματα που προκύπτουν από τη δισταξινόμηση με τη χρήση αντισωμάτων του HCV (Thomas, Astemborski et al. 2000). Έτσι, για να διερευνήσουμε την επίδραση της μεθόδου διάγνωσης, η μετα-ανάλυση της μεταβολής των CD4 T κυττάρων πραγματοποιήθηκε ξεχωριστά για κάθε μέθοδο διάγνωσης του HCV. Οι επιμέρους αναλύσεις έπαυαν να είναι στατιστικά σημαντικές σε μελέτες που υιοθετούσαν ταυτόχρονα μοριακό έλεγχο και έλεγχο του ορού. Αν και μπορεί να πρόκειται για τυχαίο εύρημα, σίγουρα καθίσταται αναγκαίος ο ακριβής καθορισμός της μόλυνσης από τον HCV.

Η επιβαρυντική επίδραση της συλλοίμωξης στην ανοσολογική ανταπόκριση στη HAART παρατηρήθηκε πιο έντονα σε ασθενείς με αριθμό CD4 T κυττάρων κατά την έναρξη μικρότερο από 350 κύτταρα/mm³. Τα ευρήματα αυτά συμφωνούν με τις υπάρχουσες ευρωπαϊκές κατευθυντήριες γραμμές χορήγησης αντιρετροϊκής θεραπείας, που συστήνουν άμεση έναρξη ART σε ασθενείς με συλλοίμωξη HIV/HCV όταν ο αριθμός των CD4 είναι μικρότερος από 350 κύτταρα/mm³.

Οι διαφορές στην ανάκτηση των CD4 T κυττάρων μπορεί να αποδοθεί σε συγχυτικούς παράγοντες. Για παράδειγμα, τα ποσοστά της μόλυνσης από τον HCV είναι εξαιρετικά υψηλά μεταξύ των IDUs και προηγούμενη μελέτη έχει δείξει ότι η ανάκαμψη των CD4 T κυττάρων είναι χαμηλότερη στους χρήστες ναρκωτικών σε σχέση με άλλες ομάδες (Collazos, Asensi et al. 2010) αν και το εύρημα αυτό δεν επιβεβαιώθηκε από άλλες παρόμοιες εργασίες (Mocroft, Madge et al. 1999; Wisaksana, Indrati et al. 2010). Φαίνεται, ωστόσο, ότι η ανοσολογική ανταπόκριση των IDUs στην ART μπορεί να επηρεάζεται από πολλούς παράγοντες όπως η χαμηλή

τους συμμόρφωση στη χορηγούμενη αγωγή, η συμμετοχή τους σε προγράμματα αποκατάστασης ή η παρουσία ψυχιατρικών καταστάσεων (Mehta, Lucas et al. 2007; Malta, Strathdee et al. 2008; Giordano, Hartman et al. 2009; Ortego, Huedo-Medina et al. 2011).

Η παρούσα μετα-ανάλυση επικαιροποίησε μια προηγούμενη μελέτη (Miller, Haley et al. 2005) χρησιμοποιώντας 13 επιπλέον μελέτες, πραγματοποιώντας πολλές αναλύσεις υποομάδων και υλοποιώντας νέες μεθόδους μετα-ανάλυσης. Καταδεικνύει μια καθυστερημένη ανοσολογική ανταπόκριση σε ασθενείς με συλλοίμωξη HIV/HCV οι οποίοι λαμβάνουν συνδυασμούς αντιρετροϊκών φαρμάκων, κυρίως μεταξύ αυτών με αριθμό CD4 T κυττάρων κάτω από 350 κύτταρα/mm³. Η καθυστέρηση αυτή στην ανάκαμψη κατά το πρώτο έτος χορήγησης της HAART θα μπορούσε να αυξήσει τον κίνδυνο τοξικότητας. Ωστόσο, η κλινική σημαντικότητα του γεγονότος αυτού δεν έχει διαλευκανθεί πλήρως και μελλοντικές μελέτες απαιτούνται για να διερευνηθούν οι βραχυπρόθεσμες και οι μακροπρόθεσμες συνέπειες.

Μέρος Γ: Διερεύνηση της συσχέτισης πολυμορφισμών του γονιδίου της αναγωγής του μεθυλενοτετραϋδροφυλλικού με την τοξικότητα από μεθοτρεξάτη

5.9. Εισαγωγή

Η μεθοτρεξάτη (Methotrexate-MTX) είναι ένα ευρέως συνταγογραφούμενο φάρμακο που χρησιμοποιείται για τη θεραπεία ποικίλων νοσημάτων όπως ο καρκίνος, η οξεία λεμφοβλαστική λευχαιμία ή λέμφωμα, η μεταμόσχευση αιμοποιητικών κυττάρων, η ρευματοειδής αρθρίτιδα, η ψωρίαση και άλλες (Kremer 1998; Kremer 1999). Ως ένα δομικό ανάλογο του φυλλικού οξέος, η MTX αναστέλλει το ένζυμο διυδροφυλλική αναγωγή (dihydrofolate reductase-DHFR), ένα ένζυμο που συμμετέχει στη σύνθεση του τετραϋδροφυλλικού (Cronstein 1996; Cronstein 1997; Chan and Cronstein 2002). Η MTX αναστέλλει επίσης άλλα ένζυμα στον κύκλο του φυλλικού οξέος, όπως η θυμιδική συνθετάση (thymidylate synthase-TS) και η αναγωγή του μεθυλενοτετραϋδροφυλλικού (methylene tetrahydrofolate reductase-MTHFR) (Baggott, Vaughn et al. 1986; Kremer 2004).

Ως αναστολέας της σύνθεσης του DNA και των πρωτεϊνών, η MTX αναστέλλει την ανάπτυξη των φυσιολογικών κυττάρων και ιστών εκτός από τη στόχευση μη φυσιολογικών κυττάρων. Συνεπώς, μπορεί να προκαλέσει ένα μεγάλο αριθμό παρενεργειών οι οποίες να οδηγήσουν τελικά σε διακοπή της θεραπείας. Οι κύριες παρενέργειες είναι η στοματίτιδα, η ναυτία, η διάρροια και η αλωπεκία. Επιπλέον, η τοξική δράση της MTX μπορεί να οδηγήσει σε βλάβη ζωτικών οργάνων, όπως το ήπαρ, οι πνεύμονες, το νευρικό σύστημα και ο μυελός των οστών (Braun and Rau 2009; Neves, Jorge et al. 2009). Μια ειδική μορφή τοξικότητας της MTX σε ασθενείς με μεταμόσχευση αιμοποιητικών κυττάρων είναι η acute graft –versus – host disease (GVHD).

Η MTHFR καταλύει την αντίδραση μετατροπής του 5,10-μεθυλενοτετραϋδροφυλλικού σε 5-μεθυλ-τετραϋδροφυλλικό στον κύκλο του φυλλικού οξέος. Από την άλλη πλευρά, το 5-μεθυλ-τετραϋδροφυλλικό μετέχει στην μετατροπή της ομοκυστεΐνης σε μεθειονίνη. Υπάρχουν δυο κοινοί πολυμορφισμοί της MTHFR: C677T και A1298C. Ο πολυμορφισμός C677T μετατρέπει την αλανίνη σε βαλίνη στη θέση 222 της αμινοξικής ακολουθίας. Η ενεργότητα του ενζύμου είναι μειωμένη για τους ομόζυγους με γονότυπο TT σε σχέση με τους ομόζυγους CC

(Goyette, Sumner et al. 1994; Rozen 1996). Ο πολυμορφισμός A1298C προκαλεί την αντικατάσταση του γλουταμικού από την αλανίνη στη θέση 429 της πρωτεΐνης (van der Put, Gabreels et al. 1998; Weisberg, Tran et al. 1998). Ο πολυμορφισμός αυτός έχει συσχετιστεί με μη μειωμένη ενεργότητα της MTHFR in vitro. Η μειωμένη ενεργότητα του ενζύμου οδηγεί σε υψηλά επίπεδα ομοκυστεΐνης στο αίμα με συνεπακόλουθη την πρόκληση στεφανιαίας νόσου (Frosst, Blom et al. 1995).

Διάφορες ομάδες έχουν εξετάσει τη συσχέτιση μεταξύ πολυμορφισμών της MTHFR και της τοξικότητας της MTX με αντικρουόμενα, ωστόσο, αποτελέσματα. Επιπλέον, οι πολλές ανεξάρτητες μελέτες έχουν μειωμένη ισχύ να ανιχνεύσουν ασθενείς επιδράσεις γονιδίων λόγω του μικρού μεγέθους δείγματος. Ωστόσο, διεξήγαμε μια μετα-ανάλυση μαζί με πολλές αναλύσεις υποομάδων για να συνοψίσουμε τα αποτελέσματα των ανεξάρτητων μελετών σχετικά με το συγκεκριμένο ερευνητικό ερώτημα και να μελετήσουμε τη μεταξύ των μελετών ετερογένεια.

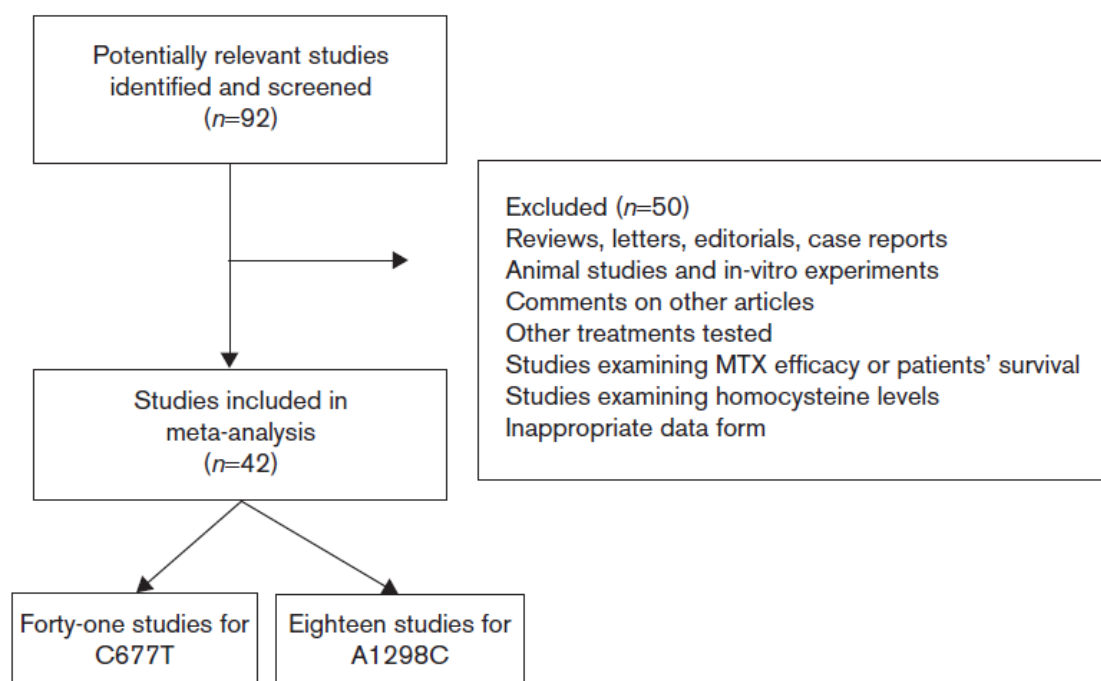
5.10. Μεθοδολογία

5.10.1. Επιλογή μελετών

Πραγματοποιήθηκε μια ηλεκτρονική αναζήτηση βιβλιογραφίας (χρησιμοποιώντας την PubMed, Scopus και Google Scholar) για δημοσιευμένες μελέτες πριν από το Δεκέμβριο του 2008. Επιπλέον, δυο ερευνητές εξέτασαν ανεξάρτητα τις μελέτες έτσι ώστε να προχωρήσουν σε εξαγωγή των δεδομένων συζητώντας τις όποιες ασυμφωνίες. Χρησιμοποιήθηκαν οι ακόλουθες λέξεις κλειδιά: “Methotrexate”, μαζί με “MTHFR” ή “Methylene Tetrahydrofolate Reductase” και “polymorphism” ή “variant”. Αποκλείστηκαν οι μη σχετικές με το θέμα μελέτες μέσω ενός αρχικού ελέγχου των τίτλων και των περιλήψεων. Τα πλήρη κείμενα των δημοσιεύσεων καθώς και οι βιβλιογραφικές αναφορές αυτών εξετάστηκαν προσεκτικά έτσι ώστε να διαπιστωθεί εάν αυτές ήταν σχετικές με το θέμα.. Επίσης, δεν αποκλείστηκαν μελέτες που ήταν γραμμένες σε άλλη γλώσσα πέρα της αγγλικής αν και στην τελική λίστα των εργασιών που επιλέχθηκαν δεν υπήρχε κάποια τέτοια μελέτη.

5.10.2. Κριτήρια επιλεξιμότητας και αποκλεισμού μελετών

Μελέτες ασθενών-μαρτύρων σε πληθυσμούς περιλαμβάνονταν στην μετα-ανάλυση εάν: (i) εξέταζαν τη συσχέτιση των πολυμορφισμών της MTHFR με την τοξικότητα σε MTX και (ii) παρείχαν την απαραίτητη πληροφορία (δηλαδή συχνότητες αλληλομόρφων και γονοτύπων) για την εκτίμηση του Odds Ratio. Για να περιορίσουμε το σφάλμα επιλογής, δεν θέσαμε περιορισμούς γλώσσας ή ποιότητας των μελετών (Stroup, Berlin et al. 2000; Pan, Trikalinos et al. 2005). Επιπλέον, για να μειωθεί η μεροληψία που οφείλεται στην «γκρίζα βιβλιογραφία», αξιολογήθηκαν επίσης μελέτες που είχαν δημοσιευτεί σε πρακτικά συνεδρίων (Conn, Valentine et al. 2003). Η Εικόνα 5.8 απεικονίζει τη διαδικασία επιλογής μελετών.



Εικόνα 5.8. Διάγραμμα ροής της διαδικασίας επιλογής των μελετών που συμπεριλάβαμε στην μετα-ανάλυση.

5.10.3. Εξαγωγή δεδομένων

Τα ακόλουθα δεδομένα αντλήθηκαν από κάθε μελέτη : (i) το όνομα του πρώτου συγγραφέα, το περιοδικό, το έτος δημοσίευσης, η εθνικότητα των συμμετεχόντων; (ii) ο συνολικός αριθμός ατόμων; (iii) ο υπό μελέτη πολυμορφισμός και η μορφή της νόσου; (iv) ο τύπος τοξικότητας της μεθοτρεξάτης και το συμπλήρωμα του φυλλικού οξέος; (v) η κατανομή των γονοτύπων και αλληλομόρφων των συμμετεχόντων.

5.10.4. Στατιστική ανάλυση

Το Odds Ratio χρησιμοποιήθηκε για τη σύγκριση της κατανομής των γονοτύπων και αλληλομόρφων μεταξύ των ατόμων που έχουν εμφανίσει ή όχι τοξικότητα στην MTX. Αρχικά, χρησιμοποιήσαμε την πολυμεταβλητή μέθοδο μετα-ανάλυσης τυχαίων επιδράσεων για τη σύγκριση της επίδρασης των γονοτύπων (TT και CT για C677T και CC και CA για A1298C) έναντι του φυσιολογικού γονοτύπου CC και AA, αντίστοιχα για τους πολυμορφισμούς C677T και A1298C (Bagos 2008). Επίσης, χρησιμοποιήθηκε η μεθοδολογία μετα-ανάλυσης μελετών γενετικής συσχέτισης που βασίζεται στη μέθοδο λογιστικής παλινδρόμησης τυχαίων επιδράσεων (Bagos and Nikolopoulos 2007). Οι δυο μεθοδολογίες ελέγχουν απευθείας το μοντέλο κληρονομικότητας αλλά απαιτούν την γνώση όλων των γονοτύπων σε κάθε μελέτη, κάτι που σε κάποιες περιπτώσεις δεν είναι γνωστό. Για το λόγο αυτό, χρησιμοποιήσαμε επίσης την κλασσική μονομεταβλητή προσέγγιση υπολογίζοντας τα Odds Ratios για τις τρεις διαφορετικές συγκρίσεις (T έναντι C αλληλομόρφου, TT έναντι CT+CC γονοτύπων και TT+CT έναντι του CC γονοτύπου για τον πολυμορφισμό C677T; C έναντι A αλληλομόρφου, CC έναντι AC+AA γονοτύπων, CC+AC έναντι AA γονοτύπων για τον πολυμορφισμό A1298C). Η μεταξύ των μελετών ετερογένεια εκτιμήθηκε χρησιμοποιώντας το Cochran Q statistic (Petiti 1994) και τον δείκτη ασυνέπειας I^2 (Higgins, Thompson et al. 2003). Τα συνδυασμένα Odds Ratios υπολογίστηκαν χρησιμοποιώντας τη μέθοδο τυχαίων επιδράσεων (DerSimonian and Laird 1986).

Σε δευτερεύουσες αναλύσεις, τα συνοπτικά Odds Ratios υπολογίστηκαν ανάλογα με τη φυλή και το αν τα άτομα χωρίς τοξικότητα σε MTX ήταν σε ισορροπία Hardy–Weinberg. Αποκλίσεις από την ισορροπία Hardy–Weinberg υπολογίστηκαν χρησιμοποιώντας τον έλεγχο χ^2 . Εκτός από τις αναλύσεις για τη συνολική τοξικότητα, πραγματοποιήσαμε μετα-αναλύσεις υποομάδων για έξι τύπους τοξικότητας και τρεις διαφορετικές μορφές νόσου. Η ανάλυση της συνολικής τοξικότητας (η οποία καθορίστηκε ως η εμφάνιση μιας οποιαδήποτε παρενέργειας κατά τη διάρκεια της θεραπείας με μεθοτρεξάτη, ήπια ή σοβαρή, αιματολογική ή όχι) περιλαμβάνει δυο κατηγορίες: (i) μελέτες με τουλάχιστον έναν τύπο τοξικότητας της MTX και (ii) μελέτες με ασθενείς με τουλάχιστον δυο διαφορετικές παρενέργειες από την MTX. Μια ξεχωριστή υπο-ανάλυση σχετικά με τη χρήση συμπληρώματος φυλλικού οξέος πραγματοποιήθηκε για να διερευνηθεί ο ρόλος και η πιθανή αλληλεπίδραση του

συμπλήρωματος αυτού με την τοξικότητα από τη χρήση της MTX. Έτσι, οι μελέτες ταξινομήθηκαν σε τρεις κατηγορίες: (i) αυτές που κανείς ασθενής δεν είχε λάβει συμπλήρωμα, (ii) αυτές που κάποιοι ασθενείς είχαν λάβει συμπλήρωμα και (iii) αυτές που όλοι οι ασθενείς είχαν λάβει συμπλήρωμα.

Η παρουσία του συστηματικού σφάλματος δημοσίευσης αξιολογήθηκε με τη χρήση των ελέγχων του Begg (Begg and Mazumdar 1994), του Egger (Egger, Davey Smith et al. 1997) και το ανάλογο αυτού με τυχαίες επιδράσεις (Thompson and Sharp 1999). Οι επιδραστικές μελέτες ανιχνεύτηκαν ελέγχοντας την επίδραση της αφαίρεσης μιας μελέτης κάθε φορά στο συνολικό αποτέλεσμα ή στην εκτιμώμενη ετερογένεια. Η αθροιστική μετα-ανάλυση (Lau, Antman et al. 1992; Lau, Schmid et al. 1995) χρησιμοποιήθηκε για να ανιχνεύσει την οποιαδήποτε αλλαγή στις συνολικές εκτιμήσεις με την πάροδο του χρόνου (Ioannidis and Trikalinos 2005). Για την ανίχνευση διαχρονικής τάσης χρησιμοποιήθηκαν οι κλασσικές μέθοδοι αθροιστικής μετα-ανάλυσης που περιλαμβάνουν τον οπτικό έλεγχο του γραφήματος και μια πρόσφατα προτεινόμενη μέθοδο παλινδρόμησης (Bagos and Nikolopoulos 2009).

Για όλες τις αναλύσεις, χρησιμοποιήθηκε το στατιστικό πακέτο STATA 10 (Stata Corporation, College Station, TX, USA). Στατιστικά σημαντικά αποτελέσματα ήταν εκείνα με p -value <0.05.

5.11. Αποτελέσματα

Συνολικά, συμπεριλάβαμε 44 μελέτες που αφορούσαν τη συσχέτιση των πολυμορφισμών του γονιδίου της MTHFR με την τοξικότητα σε μεθοτρεξάτη (Πίνακας Δ1). Πιο συγκεκριμένα, 26 μελέτες αξιολογούσαν μόνο τη συσχέτιση του πολυμορφισμού C677T, 3 εξέταζαν μόνο τον πολυμορφισμό A1298C και οι υπόλοιπες 15 είχαν δεδομένα και για τους δυο πολυμορφισμούς. Μια μελέτη (Ranganathan, Culverhouse et al. 2008) είχε πληροφορίες για τρεις ανεξάρτητους πληθυσμούς, και έτσι, υπολογίστηκε σαν τρεις μελέτες στην μετα-ανάλυση.

5.11.1. Ο πολυμορφισμός C677T του γονιδίου της MTHFR (rs1801133)

(i) Συνολική τοξικότητα – περισσότερες από δυο παρενέργειες

Συνολικά, δεκαοκτώ μελέτες (Chiusolo, Reddicono et al. 2002; Urano, Taniguchi et al. 2002; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Schmeling, Biber et al. 2005; Aggarwal, Naik et al.

2006; Kim, Jun et al. 2006; Weisman, Furst et al. 2006; Campalani, Arenas et al. 2007; Chiusolo, Reddicono et al. 2007; Taniguchi, Urano et al. 2007; Bohanec Grabar, Logar et al. 2008; Ghodke, Chopra et al. 2008; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Zeng, Wang et al. 2008; Warren, Smith et al. 2009) εξέτασαν τη συσχέτιση του πολυμορφισμού C677T με την εμφάνιση τουλάχιστον δυο παρενεργειών από τη χρήση μεθοτρεξάτης. Μεταξύ αυτών, οι έξι μελέτες είχαν διεξαχθεί σε Καυκάσιους, οι επτά σε Ασιάτες και οι πέντε μελέτες περιελάμβαναν άλλες εθνικότητες. Επτά μελέτες ήταν σε ισορροπία Hardy-Weinberg (Chiusolo, Reddicono et al. 2002; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Kim, Jun et al. 2006; Shimasaki, Mori et al. 2008; Warren, Smith et al. 2009). Το μοντέλο λογιστικής παλινδρόμησης εφαρμόστηκε σε 8 μελέτες (Chiusolo, Reddicono et al. 2002; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Kim, Jun et al. 2006; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Warren, Smith et al. 2009) και έδειχνε μια στατιστικά σημαντική επίδραση για τη σύγκριση του γονοτύπου TT έναντι του CC (OR: 1.672, 95% C.I.: 1.009, 2.769). Η πολυμεταβλητή μεθοδολογία (με 8 μελέτες (Chiusolo, Reddicono et al. 2002; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Kim, Jun et al. 2006; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Warren, Smith et al. 2009)) έδειξε επίσης μια παρόμοια εκτίμηση για τη σύγκριση του γονοτύπου TT έναντι του CC (OR: 1.670, 95% C.I.: 1.000, 2.770). Η τιμή του λ ήταν 0.39 υποδεικνύοντας πιθανά ένα συνεπικρατές μοντέλο κληρονομικότητας.

Η κλασική μετα-ανάλυση πραγματοποιήθηκε σε 18 μελέτες (Chiusolo, Reddicono et al. 2002; Urano, Taniguchi et al. 2002; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Schmeling, Biber et al. 2005; Aggarwal, Naik et al. 2006; Kim, Jun et al. 2006; Weisman, Furst et al. 2006; Campalani, Arenas et al. 2007; Chiusolo, Reddicono et al. 2007; Taniguchi, Urano et al. 2007; Bohanec Grabar, Logar et al. 2008; Ghodke, Chopra et al. 2008; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Zeng, Wang et al. 2008; Warren, Smith et al. 2009). Το Odds Ratio για τη σύγκριση του αλληλομόρφου T ήταν 1.217 (95% C.I.: 0.889, 1.665). Μια μεγαλύτερη αλλά οριακά μη σημαντική επίδραση υπολογίστηκε για τη σύγκριση του γονοτύπου TT έναντι των CT+CC (OR: 1.566, 95% C.I.: 0.949, 2.582). Ωστόσο, η ανάλυση αποκάλυψε ένα σημαντικό αποτέλεσμα για τη σύγκριση του γονοτύπου TT έναντι των CT+CC σε μελέτες που

περιλαμβάνουν Ασιάτες (OR: 1.655, 95% C.I.: 1.029, 2.660) και μελέτες που είναι σε ισορροπία Hardy–Weinberg (OR: 1.604, 95% C.I.: 1.063, 2.419). Δεν βρέθηκε ετερογένεια στις περισσότερες αναλύσεις που πραγματοποιήθηκαν ούτε ένδειξη για συστηματικό σφάλμα δημοσίευσης. Η ανάλυση ευαισθησίας αποκάλυψε μια επιδραστική μελέτη (Kim, Jun et al. 2006). Απομακρύνοντας τη μελέτη αυτή από την ανάλυση το Odds Ratio μειώθηκε σημαντικά (T vs. C OR: 1.043, 95% C.I.: 0.856, 1.269). Η αθροιστική μετα-ανάλυση έδειξε μια ελαττωμένη επίδραση με το πέρασμα του χρόνου (δεν παρουσιάζονται τα αποτελέσματα).

(ii) Συνολική τοξικότητα – περισσότερες από μια παρενέργειες

Σαράντα μια μελέτες (Fijnheer, Roest et al. 1998; van Ede, Laan et al. 2001; Chiusolo, Reddicono et al. 2002; Urano, Taniguchi et al. 2002; Bernbeck, Mauz-Korholz et al. 2003; Kalayoglu-Besisik, Caliskan et al. 2003; Kishi, Griener et al. 2003; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Pihusch, Lohse et al. 2004; Aplenc, Thompson et al. 2005; Jazbec, Kitanovski et al. 2005; Schmeling, Biber et al. 2005; Aggarwal, Naik et al. 2006; Kim, Jun et al. 2006; Murphy, Diviney et al. 2006; Robien, Bigler et al. 2006; Seidemann, Book et al. 2006; Shimasaki, Mori et al. 2006; Weisman, Furst et al. 2006; Azarpira, Ramzi et al. 2007; Campalani, Arenas et al. 2007; Chiusolo, Reddicono et al. 2007; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Ruiz-Arguelles, Coconi-Linares et al. 2007; Taniguchi, Urano et al. 2007; Bohanec Grabar, Logar et al. 2008; Ghodke, Chopra et al. 2008; Krull, Brouwers et al. 2008; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Sugimoto, Murata et al. 2008; Tsuchiya, Inoue et al. 2008; van Kooten Niekerk, Schmiegelow et al. 2008; Zeng, Wang et al. 2008; Warren, Smith et al. 2009) διερεύνησαν τη συσχέτιση του πολυμορφισμού C677T με την εμφάνιση τουλάχιστον μιας παρενέργειας από τη χρήση MTX. Δεκαπέντε μελέτες περιελάμβαναν Καυκάσιους, 12 μελέτες Ασιάτες και 14 άλλες εθνικότητες. Συνολικά, 18 μελέτες ήταν σε ισορροπία Hardy–Weinberg (Fijnheer, Roest et al. 1998; Chiusolo, Reddicono et al. 2002; Kalayoglu-Besisik, Caliskan et al. 2003; Kishi, Griener et al. 2003; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Jazbec, Kitanovski et al. 2005; Kim, Jun et al. 2006; Seidemann, Book et al. 2006; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Ruiz-Arguelles, Coconi-Linares et al. 2007; Ranganathan, Culverhouse et al. 2008;

Shimasaki, Mori et al. 2008; Warren, Smith et al. 2009). Η μέθοδος λογιστικής παλινδρόμησης (με 20 μελέτες (Fijnheer, Roest et al. 1998; Chiusolo, Reddicono et al. 2002; Kalayoglu-Besisik, Caliskan et al. 2003; Kishi, Griener et al. 2003; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Jazbec, Kitanovski et al. 2005; Kim, Jun et al. 2006; Robien, Bigler et al. 2006; Seidemann, Book et al. 2006; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Ruiz-Arguelles, Coconi-Linares et al. 2007; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Warren, Smith et al. 2009)) έδειξε μια οριακά μη σημαντική συσχέτιση (TT vs. CC, OR: 1.560, 95% C.I.: 0.950, 2.561). Το πολυμεταβλητό μοντέλο (με 20 μελέτες (Fijnheer, Roest et al. 1998; Chiusolo, Reddicono et al. 2002; Kalayoglu-Besisik, Caliskan et al. 2003; Kishi, Griener et al. 2003; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Jazbec, Kitanovski et al. 2005; Kim, Jun et al. 2006; Robien, Bigler et al. 2006; Seidemann, Book et al. 2006; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Ruiz-Arguelles, Coconi-Linares et al. 2007; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Warren, Smith et al. 2009)) οδήγησε σε παρόμοια αποτελέσματα. Η τιμή του λ ήταν 0.49 υποδεικνύοντας συνεπικρατές μοντέλο.

Η μονομεταβλητή μέθοδος εφαρμόστηκε σε 41 μελέτες (Fijnheer, Roest et al. 1998; van Ede, Laan et al. 2001; Chiusolo, Reddicono et al. 2002; Urano, Taniguchi et al. 2002; Bernbeck, Mauz-Korholz et al. 2003; Kalayoglu-Besisik, Caliskan et al. 2003; Kishi, Griener et al. 2003; Kumagai, Hiyama et al. 2003; Toffoli, Russo et al. 2003; Berkun, Levartovsky et al. 2004; Pihusch, Lohse et al. 2004; Aplenc, Thompson et al. 2005; Jazbec, Kitanovski et al. 2005; Schmeling, Biber et al. 2005; Aggarwal, Naik et al. 2006; Kim, Jun et al. 2006; Murphy, Diviney et al. 2006; Robien, Bigler et al. 2006; Seidemann, Book et al. 2006; Shimasaki, Mori et al. 2006; Weisman, Furst et al. 2006; Azarpira, Ramzi et al. 2007; Campalani, Arenas et al. 2007; Chiusolo, Reddicono et al. 2007; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Ruiz-Arguelles, Coconi-Linares et al. 2007; Taniguchi, Urano et al. 2007; Bohanec Grabar, Logar et al. 2008; Ghodke, Chopra et al. 2008; Krull, Brouwers et al. 2008; Ranganathan, Culverhouse et al. 2008; Shimasaki, Mori et al. 2008; Sugimoto, Murata et al. 2008; Tsuchiya, Inoue et al. 2008; van Kooten Niekerk, Schmiegelow et al. 2008; Zeng, Wang et al. 2008; Warren, Smith et al. 2009). Πιο συγκεκριμένα, η σύγκριση του αλληλομόρφου T έναντι του C

δεν ήταν στατιστικά σημαντική (OR: 1.158, 95% C.I.: 0.919, 1.459). Ωστόσο, η ανάλυση υποομάδας έδειξε μια σημαντική επίδραση του γονοτύπου TT έναντι των CT+CC σε μελέτες που ήταν σε ισορροπία Hardy–Weinberg (OR: 1.575, 95% C.I.: 1.100, 2.254). Δεν ανιχνεύτηκε συστηματικό σφάλμα δημοσίευσης, αλλά σε κάποιες περιπτώσεις παρατηρήθηκε μεταξύ των μελετών ετερογένεια. Τέλος, η αθροιστική μετα-ανάλυση αποκάλυψε μικρή μείωση στις εκτιμώμενες επιδράσεις με το πέρασμα του χρόνου οι οποίες σταθεροποιήθηκαν μετά το 2006 (δεν παρουσιάζονται τα δεδομένα).

(iii) Ηπατοτοξικότητα και Αιματολογική τοξικότητα

Η ανάλυση που αφορά την ηπατοτοξικότητα αποκάλυψε μια σημαντική συσχέτιση του πολυμορφισμού C677T με αυτό τον τύπο τοξικότητας. Πέντε μελέτες αξιοποιήθηκαν στη λογιστική παλινδρόμηση και στην πολυμεταβλητή μέθοδο (Chiusolo, Reddicono et al. 2002; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Shimasaki, Mori et al. 2008). Η εφαρμογή της λογιστικής παλινδρόμησης οδήγησε στην εκτίμηση ενός Odds Ratio ίσου με 4.191 (95% C.I.: 1.642, 10.698) για το γονότυπο TT έναντι του CC. Παρόμοια ήταν τα ευρήματα από την πολυμεταβλητή μετα-ανάλυση. Το λ ήταν 0.16 υποδεικνύοντας υπολειπόμενο μοντέλο κληρονομικότητας. Το κλασσικό μοντέλο μετα-ανάλυσης εφαρμόστηκε σε 13 μελέτες (van Ede, Laan et al. 2001; Chiusolo, Reddicono et al. 2002; Aplenc, Thompson et al. 2005; Schmeling, Biber et al. 2005; Aggarwal, Naik et al. 2006; Kim, Jun et al. 2006; Shimasaki, Mori et al. 2006; Campalani, Arenas et al. 2007; Chiusolo, Reddicono et al. 2007; Gemmati, Ongaro et al. 2007; Imanishi, Okamura et al. 2007; Kim, Lee et al. 2007; Shimasaki, Mori et al. 2008). Πιο συγκεκριμένα, το Odds Ratio του αλληλομόρφου T έναντι του C εκτιμήθηκε 1.588 (95% C.I.: 1.146, 2.202). Η ανάλυση υποομάδας για τους Καυκάσιους και η ανάλυση για μελέτες σε ισορροπία Hardy–Weinberg έδειξε μια στατιστικά σημαντική επίδραση (Καυκάσιοι: OR: 2.472, 95% C.I.: 1.341, 4.559; HWE: OR: 2.004, 95% C.I.: 1.074, 3.738). Στο σύνολο του δείγματος, για τα άτομα με γονότυπο TT έναντι των CT+CC υπολογίστηκαν στατιστικά σημαντικές εκτιμήσεις καθώς επίσης και για Καυκάσιους και πληθυσμούς σε ισορροπία Hardy–Weinberg. Η σύγκριση των γονοτύπων TT+CT έναντι του CC κατέδειξε σημαντική επίδραση στο σύνολο του δείγματος, καθώς επίσης και για τους Καυκάσιους και Ασιάτες. Δεν βρέθηκε ένδειξη για συστηματικό σφάλμα δημοσίευσης.

Μόνο τρεις μελέτες (Chiusolo, Reddicono et al. 2002; Gemmati, Ongaro et al. 2007; Shimasaki, Mori et al. 2008), οι οποίες αξιολογούσαν την επίδραση της θεραπείας στην τοξικότητα στο αίμα, μπορούσαν να χρησιμοποιηθούν σε ένα πολυμεταβλητό μοντέλο. Η ανάλυση έδειξε μια οριακή συσχέτιση του μεταλλαγμένου αλληλομόρφου με την τοξικότητα της MTX στο αίμα (TT έναντι του CC: OR: 2.590, 95% C.I.: 0.950, 7.080). Από το μονομεταβλητό μοντέλο (με 10 μελέτες (Chiusolo, Reddicono et al. 2002; Aplenc, Thompson et al. 2005; Aggarwal, Naik et al. 2006; Kim, Jun et al. 2006; Shimasaki, Mori et al. 2006; Chiusolo, Reddicono et al. 2007; Gemmati, Ongaro et al. 2007; Shimasaki, Mori et al. 2008; Tsuchiya, Inoue et al. 2008; van Kooten Niekerk, Schmiegelow et al. 2008)) εκτιμήθηκε ένα σημαντικό Odds Ratio μόνο για τις μελέτες σε ισορροπία Hardy–Weinberg για τα άτομα με γονότυπο TT έναντι των CT+CC (OR: 2.551, 95% C.I.: 1.055, 6.169). Δεν υπήρχε συστηματικό σφάλμα αλλά σε κάποιες συγκρίσεις βρέθηκε μεταξύ των μελετών ετερογένεια.

(iv) Τοξικότητα πεπτικού συστήματος, νευροτοξικότητα, GVHD και Mucositis

Η συσχέτιση μεταξύ του πολυμορφισμού C677T με την τοξικότητα του πεπτικού συστήματος, νευροτοξικότητα, GVHD και mucositis είναι καλά μελετημένη. Η υπο-ανάλυση σχετικά με την τοξικότητα της μεθοτρεξάτης στο πεπτικό σύστημα απέτυχε να δείξει κάποια συσχέτιση με τον πολυμορφισμό C677T. Σχετικά με τις παρενέργειες στο νευρικό σύστημα, η λογιστική παλινδρόμηση και τα πολυμεταβλητά μοντέλα δεν μπορούσαν να χρησιμοποιηθούν λόγω έλλειψης των πρωταρχικών δεδομένων. Τα μονομεταβλητά μοντέλα τυχαίων επιδράσεων ανίχνευαν σημαντική επίδραση στη νευροτοξικότητα κατά τη σύγκριση του γονοτύπου TT έναντι των CC+CT (OR: 3.398, 95% C.I.: 1.808, 6.389).

Η ανάλυση σχετικά με την GVHD καταδεικνύει έναν προστατευτικό ρόλο του πολυμορφισμού C677T στην τοξικότητα αυτή. Από τη λογιστική παλινδρόμηση εκτιμήθηκε ένα Odds Ratio ίσο με 0.446 (TT έναντι του CC; 95% C.I.: 0.242, 0.821). Η πολυμεταβλητή προσέγγιση οδήγησε σε παρόμοιες εκτιμήσεις. Χρησιμοποιώντας τη μονομεταβλητή μέθοδο (με 6 μελέτες (Pihusch, Lohse et al. 2004; Murphy, Diviney et al. 2006; Robien, Bigler et al. 2006; Azarpira, Ramzi et al. 2007; Kim, Lee et al. 2007; Sugimoto, Murata et al. 2008)) υπολογίστηκε ένα Odds Ratio ίσο με 0.441 (95% C.I.: 0.253, 0.771) κατά τη σύγκριση του γονοτύπου TT έναντι των CT+CC. Δεν υπήρχε ένδειξη για συστηματικό σφάλμα δημοσίευσης αλλά η ετερογένεια ήταν

παρούσα κατά τη σύγκριση των αλληλομόρφων (I^2 : 84.6%). Τέλος, η ανάλυση δεν έδειξε κάποια ένδειξη για τη συσχέτιση του πολυμορφισμού C677T με την mucositis. Για όλες τις υπο-αναλύσεις εκτιμήθηκε υπολειπόμενο μοντέλο κληρονομικότητας.

(v) Ρευματοειδής αρθρίτιδα, αιματολογικές διαταραχές και μεταμόσχευση αιμοποιητικών κυττάρων

Οι υποομάδες με ρευματοειδή αρθρίτιδα, αιματολογικές διαταραχές και μεταμόσχευση αιμοποιητικών κυττάρων περιλαμβάνουν άτομα με οποιαδήποτε μορφή τοξικότητας (ήπια ή σοβαρή, αιματολογική ή μη) κατά τη διάρκεια της θεραπείας με μεθοτρεξάτη για τη συγκεκριμένη πάθηση. Στην ανάλυση συμπεριλάβαμε άτομα με μια μόνο παρενέργεια.

Σχετικά με την ρευματοειδή αρθρίτιδα συγκεντρώθηκε ένα σύνολο 15 μελετών. Το Odds Ratio της σύγκρισης του γονοτύπου TT έναντι του CC από τη μέθοδο της λογιστικής παλινδρόμησης ήταν 1.991 (95% C.I.: 0.982, 4.036). Η πολυμεταβλητή μέθοδος έδωσε παρόμοιες εκτιμήσεις (με 6 μελέτες, (Kumagai, Hiyama et al. 2003; Berkun, Levartovsky et al. 2004; Kim, Jun et al. 2006; Ranganathan, Culverhouse et al. 2008)). Από την ανάλυση υποδεικνύεται συνεπικρατές μοντέλο κληρονομικότητας (λ : 0.7). Δεν ανιχνεύτηκε συσχέτιση κατά τη σύγκριση των γονοτύπων TT+CT έναντι του CC (OR: 1.546, 95% C.I.: 0.964, 2.479). Τέσσερις μελέτες (Kumagai, Hiyama et al. 2003; Berkun, Levartovsky et al. 2004; Kim, Jun et al. 2006; Ranganathan, Culverhouse et al. 2008) ήταν σε ισορροπία Hardy-Weinberg. Η ανάλυση της υποομάδας αυτής έδειξε σημαντική επίδραση της τοξικότητας σε ασθενείς με ρευματοειδή αρθρίτιδα (TT vs. CT+ CC; OR: 1.735, 95% C.I.: 1.129, 2.666). Δεν υπήρχε ένδειξη για συστηματικό σφάλμα δημοσίευσης, αλλά σε κάποιες περιπτώσεις υπήρχε ετερογένεια.

Από το πολυμεταβλητό μοντέλο για τα άτομα με αιματολογικές διαταραχές (με 7 μελέτες (Chiusolo, Reddicono et al. 2002; Kishi, Griener et al. 2003; Jazbec, Kitanovski et al. 2005; Seidemann, Book et al. 2006; Gemmati, Ongaro et al. 2007; Ruiz-Arguelles, Coconi-Linares et al. 2007; Shimasaki, Mori et al. 2008)) εκτιμήθηκε ένα Odds Ratio ίσο με 1.752 για τη σύγκριση των ατόμων με γονότυπο TT έναντι του CC (95% C.I.: 0.847, 3.626). Η κλασσική μεθοδολογία μετα-ανάλυσης (με 14 μελέτες) έδειξε μια οριακά σημαντική συσχέτιση του αλληλομόρφου T έναντι του C (OR: 1.237, 95% C.I.: 0.947, 1.616) και του μεταλλαγμένου γονοτύπου έναντι των CT+CC (OR: 1.957, 95% C.I.: 0.935, 4.099). Η ανάλυση των ασθενών με μεταμόσχευση

αιμοποιητικών κυττάρων υπογράμμισε τον προστατευτικό ρόλο του πολυμορφισμού. Πιο συγκεκριμένα, υπολογίστηκε ένα Odds Ratio ίσο με 0.413 (TT έναντι CC; 95% C.I.: 0.226, 0.753) χρησιμοποιώντας τη λογιστική παλινδρόμηση και το πολυμεταβλητό μοντέλο (με 3 μελέτες (Kalayoglu-Besisik, Caliskan et al. 2003; Robien, Bigler et al. 2006; Kim, Lee et al. 2007)). Ένα υπολειπόμενο μοντέλο κληρονομικότητας υποδεικνύεται από την ανάλυση (λ : 0.22). Επιπλέον, ο προστατευτικός ρόλος του πολυμορφισμού επιβεβαιώθηκε από την μονομεταβλητή μετα-ανάλυση σε πέντε μελέτες. Το Odds Ratio κατά τη σύγκριση του γονοτύπου TT έναντι των CC+CT ήταν 0.435 (95% C.I.: 0.252, 0.754). Δεν υπήρχε ένδειξη για συστηματικό σφάλμα δημοσίευσης, αλλά σε κάποιες περιπτώσεις υπήρχε ετερογένεια.

(vi) Ρόλος του συμπληρώματος του φυλλικού οξέος

Διερευνήθηκε η πιθανή επίδραση της εμφάνισης τοξικότητας από τη χρήση μεθοτρεξάτης λόγω της συνδυασμένης λήψης συμπληρώματος φυλλικού οξέος. Η πολυμεταβλητή μεθοδολογία περιελάμβανε 11 μελέτες χωρίς καθόλου συμπλήρωμα φυλλικού, τρεις μελέτες με μερική λήψη φυλλικού και έξι μελέτες με συνεχόμενη λήψη φυλλικού. Βρέθηκε ότι ο πολυμορφισμός C677T συσχετίζεται με αυξημένα επίπεδα τοξικότητας σε άτομα που λάμβαναν MTX και φυλλικό οξύ. Από τη λογιστική παλινδρόμηση για τη σύγκριση του γονοτύπου TT έναντι του CC το Odds Ratio ήταν 2.871 (95% C.I.: 1.107, 7.450) και 2.736 (95% C.I.: 1.347, 5.559) συγκρίνοντας τα άτομα με γονότυπο CT έναντι του CC όταν όλοι οι ασθενείς έκαναν λήψη φυλλικού οξέος. Παρόμοια ήταν τα αποτελέσματα κάτω από το πολυμεταβλητό μοντέλο. Αντίθετα, δεν βρέθηκε συσχέτιση μεταξύ των παρενεργειών από τη χρήση μεθοτρεξάτης και τον πολυμορφισμό C677T για τις υπόλοιπες υπο-αναλύσεις (καθόλου λήψη φυλλικού και μερική λήψη).

Η κλασσική μεθοδολογία μετα-ανάλυσης απέτυχε να δείξει συσχέτιση μεταξύ των πολυμορφισμών του γονιδίου της MTHFR και την τοξικότητα από MTX για όλες τις κατηγορίες λήψης φυλλικού οξέος. Στην ανάλυση στρατολογήθηκαν είκοσι τρεις μελέτες ατόμων που δεν έκαναν λήψη φυλλικού οξέος, δέκα μελέτες με μερική λήψη και οκτώ μελέτες με πλήρη λήψη φυλλικού. Δεν βρέθηκε συσχέτιση μεταξύ του μεταλλαγμένου αλληλομόρφου και την τοξικότητα από MTX για τα άτομα που έκαναν λήψη συμπληρώματος φυλλικού οξέος (OR: 1.511, 95% C.I.: 0.897, 2.546). Τα αποτελέσματα από τις αναλύσεις του πολυμορφισμού C677T παρουσιάζονται στους Πίνακες Δ2 και Δ3.

5.11.2. Ο πολυμορφισμός A1298C του γονιδίου της MTHFR (rs1801131)

(i) Συνολική τοξικότητα – περισσότερες από δυο παρενέργειες

Αρχικά, 12 μελέτες εξέταζαν τη συσχέτιση του πολυμορφισμού A1298C με την παρουσία τουλάχιστον δυο παρενεργειών από την θεραπεία με MTX. Τέσσερις μελέτες (Kumagai, Hiyama et al. 2003; Herrlinger, Cummings et al. 2005; Wessels, de Vries-Bouwstra et al. 2006; Warren, Smith et al. 2009) ήταν σε ισορροπία Hardy-Weinberg. Η μέθοδος λογιστικής παλινδρόμησης και η πολυμεταβλητή προσέγγιση δεν έδειξε στατιστικά σημαντικά αποτελέσματα. Η μονομεταβλητή μετα-ανάλυση είχε παρόμοια αποτελέσματα. Σε κάποιες συγκρίσεις εκτιμήθηκε ετερογένεια. Η αθροιστική μετα-ανάλυση δεν έδειξε την ύπαρξη διαχρονικής τάσης.

(ii) Συνολική τοξικότητα – περισσότερες από μια παρενέργειες

Συνολικά, 18 μελέτες εξέταζαν τη συσχέτιση του πολυμορφισμού A1298C με τουλάχιστον ένα είδος τοξικότητας από την MTX. Πιο συγκεκριμένα, 7 μελέτες περιελάμβαναν Καυκάσιους, 4 μελέτες Ασιάτες και τέσσερις άλλη εθνικότητα. Έξι μελέτες (Kumagai, Hiyama et al. 2003; Herrlinger, Cummings et al. 2005; Jazbec, Kitanovski et al. 2005; Wessels, de Vries-Bouwstra et al. 2006; Kim, Lee et al. 2007; Warren, Smith et al. 2009) βρίσκονταν σε ισορροπία Hardy-Weinberg. Το λογιστικό μοντέλο μετα-ανάλυσης (9 μελέτες, (Kumagai, Hiyama et al. 2003; Berkun, Levartovsky et al. 2004; Herrlinger, Cummings et al. 2005; Jazbec, Kitanovski et al. 2005; Robien, Bigler et al. 2006; Wessels, de Vries-Bouwstra et al. 2006; Gemmati, Ongaro et al. 2007; Kim, Lee et al. 2007; Warren, Smith et al. 2009)) έδειξε μια σημαντική επίδραση για τους φορείς του αλληλομόρφου C έναντι των ατόμων με γονότυπο AA (OR: 1.889, 95% C.I.: 1.232, 2.896) και για εκείνους με γονότυπο AC έναντι του AA (OR: 1.690, 95% C.I.: 1.011, 2.825). Η πολυμεταβλητή μετα-ανάλυση (9 μελέτες) έδειξε μη στατιστικά σημαντικά αποτελέσματα (CC vs. AA OR: 1.500, 95% C.I.: 0.720, 3.140; AC vs. AA OR: 1.520, 95% C.I.: 0.880, 2.610). Τα αποτελέσματα αυτά δείχνουν επικρατές μοντέλο κληρονομικότητας. Η μονομεταβλητή μετα-ανάλυση περιελάμβανε 18 μελέτες. Η σύγκριση του αλληλομόρφου C έναντι του A των μελετών σε ισορροπία Hardy-Weinberg έδωσε μια οριακά μη στατιστικά σημαντική συσχέτιση (OR: 1.346, 95% C.I.: 0.951, 1.905). Παρόμοια, η σύγκριση των φορέων του αλληλομόρφου C έναντι των ατόμων με γονότυπο AA οδήγησε στην εκτίμηση του Odds Ratio ίση με 1.415 (95% C.I.: 0.913,

2.193). Μεταξύ των μελετών ετερογένεια εκτιμήθηκε σε κάποιες συγκρίσεις. Δεν υπήρχε ένδειξη για συστηματικό σφάλμα δημοσίευσης.

(iii) Ρευματοειδής αρθρίτιδα, αιματολογικές διαταραχές και μεταμόσχευση αιμοποιητικών κυττάρων

Η ανάλυση για τα άτομα με ρευματοειδή αρθρίτιδα περιελάμβανε ένα σύνολο 8 μελετών. Το μοντέλο λογιστικής παλινδρόμησης και η πολυμεταβλητή ανάλυση απέτυχαν να δείξουν κάποια σημαντική συσχέτιση. Αντιθέτως, η μονομεταβλητή προσέγγιση εκτίμησε μια σημαντική επίδραση κατά τη σύγκριση του γονότυπου CC έναντι των AC+AA (OR: 0.506, 95% C.I.: 0.273, 0.936) υποδεικνύοντας προστατευτικό ρόλο σε ομόζυγους για ρευματοειδή αρθρίτιδα για την εμφάνιση παρενεργειών από τη χρήση MTX. Μεταξύ των μελετών ετερογένεια εκτιμήθηκε σε κάποιες συγκρίσεις. Καμία από τις υπόλοιπες συγκρίσεις δεν οδήγησε σε κάποιο σημαντικό εύρημα. Μελέτες με ασθενείς με αιματολογικές νόσους αξιοποιήθηκαν στη μετα-ανάλυση χωρίς κάποιο σημαντικό εύρημα.

(iv) Ρόλος του συμπληρώματος του φυλλικού οξέος

Εξετάστηκε η συσχέτιση μεταξύ της τοξικότητας σε μεθοτρεξάτη και του πολυμορφισμού A1298C σχετικά με την ταυτόχρονη λήψη συμπληρώματος φυλλικού οξέος. Το Odds Ratio που εκτιμήθηκε από τη λογιστική παλινδρόμηση για τη συσχέτιση των ατόμων με γονότυπο CC έναντι των AA ήταν 4.131 (95% C.I.: 1.864, 9.155) για τα άτομα που δεν είχαν λάβει συμπλήρωμα φυλλικού οξέος. Αυτό συνάδει με τα ευρήματα της πολυμεταβλητής μετα-ανάλυσης με το εκτιμώμενο Odds Ratio να είναι 3.83 (95% C.I.: 1.95, 7.49). Η μονομεταβλητή μέθοδος εφαρμόστηκε σε 8 μελέτες και έδειξε επίσης μια σημαντική συσχέτιση. Η σύγκριση των ατόμων με γονότυπο AC+CC έναντι του AA υποδεικνύει μια σημαντική συσχέτιση του πολυμορφισμού με τη τοξικότητα από MTX σε άτομα που δεν είχαν λάβει συμπλήρωμα (OR: 1.926, 95% C.I.: 1.098, 3.379). Ετερόγένεια υπήρχε σε κάποιες συγκρίσεις. Υποδεικνύεται συσχέτιση για τα άτομα που λαμβάνουν συμπλήρωμα φυλλικού οξέως σε μόνιμη βάση κατά τη σύγκριση των ατόμων με γονότυπους AC+CC έναντι των ατόμων με γονότυπο AA (OR: 2.425, 95% C.I.: 1.289 4.560) αν και μόνο μια μελέτη ήταν διαθέσιμη. Τα αποτελέσματα του πολυμορφισμού A1298C παρουσιάζονται στους Πίνακες Δ4 και Δ5.

5.12. Συζήτηση

Η παρούσα μετα-ανάλυση στοχεύει να διερευνήσει τη συσχέτιση των πολυμορφισμών C677T και A1298C του γονιδίου της MTHFR με τη τοξικότητα από τη χρήση μεθοτρεξάτης. Καταδεικνύεται σημαντική συσχέτιση του πολυμορφισμού C677T MTHFR με την τοξικότητα σε MTX. Πιο συγκεκριμένα, η λογιστική παλινδρόμηση και τα πολυμεταβλητά μοντέλα αποκαλύπτουν σημαντική επίδραση του γονοτύπου TT για ασθενείς με τουλάχιστον δυο παρενέργειες. Η επίδραση του πολυμορφισμού στην ηπατοτοξικότητα είναι ξεκάθαρη. Από την άλλη πλευρά, καταδεικνύεται ο προστατευτικός ρόλος του πολυμορφισμού C677T στην εμφάνιση GVHD. Η ανάλυση αποκάλυψε επίσης σημαντική συσχέτιση μεταξύ του πολυμορφισμού 677C>T με χαμηλότερο κίνδυνο για τα άτομα με GVHD που κάνουν μεταμόσχευση αιμοποιητικών κυττάρων. Τα ευρήματά μας συμφωνούν με την υπόθεση ότι η χαμηλότερη ενεργότητα της MTHFR λόγω του πολυμορφισμού 677C>T προκαλεί μεγαλύτερη ανοσοκαταστολή στα άτομα που λαμβάνουν MTX. Ωστόσο, υπάρχουν μελέτες που απέτυχαν να δείξουν την οποιαδήποτε συσχέτιση μεταξύ των πολυμορφισμών του γονιδίου της MTHFR και τη επίπτωση της GVHD. Υπήρχαν ενδείξεις από τη λογιστική παλινδρόμηση και τα πολυμεταβλητά μοντέλα ότι οι ομόζυγοι για τον πολυμορφισμό 677T που λάμβαναν συμπλήρωμα φυλλικού οξέος έχουν αυξημένο κίνδυνο να εμφανίσουν τοξικότητα σε MTX. Τα αποτελέσματα των αναλύσεων σχετικά με τον πολυμορφισμό C677T και τη τοξικότητα σε ασθενείς με ρευματοειδή αρθρίτιδα συμφωνούν με εκείνα που έχουν παρουσιαστεί σε παλαιότερη μετα-ανάλυση (Fisher and Cronstein 2009) υποδεικνύοντας μια μέτρια συσχέτιση χρησιμοποιώντας μοντέλα τυχαίων επιδράσεων.

Η λογιστική παλινδρόμηση του πολυμορφισμού A1298C υπογράμμισε μια σημαντική συσχέτιση μεταξύ των φορέων του αλληλομόρφου C και της τοξικότητας σε μεθοτρεξάτη σε ασθενείς με τουλάχιστον μια παρενέργεια. Επιπλέον, υπήρχε ένδειξη για την εμφάνιση τοξικότητας για τους φορείς του αλληλομόρφου C που δεν λάμβαναν συμπλήρωμα φυλλικού οξέος. Η ανάλυση έδειξε μειωμένο κίνδυνο των ομόζυγων για το αλληλόμορφο C σε ασθενείς με ρευματοειδή αρθρίτιδα στην εμφάνιση τοξικότητας κάτι που δεν συνάδει με ευρήματα προηγούμενης μετα-ανάλυσης (Fisher and Cronstein 2009).

Η μεθοτρεξάτη αποτελεί ένα φάρμακο με ευρεία χρήση σε πολλές διαταραχές. Οι ιστοί με αυξημένο ρυθμό κυτταρικής διαίρεσης είναι πιο επιρρεπείς σε

αντενδείξεις που προκύπτουν από τη χρήση της. Ο μηχανισμός αυτός εξηγεί επίσης τις σοβαρές παρενέργειες που προκαλούνται από τη μεθοτρεξάτη. Η τοξικότητα σε μεθοτρεξάτη περιλαμβάνει πολλές και διαφορετικές μορφές και αφορά όλα τα ανθρώπινα όργανα. Συνεπώς, αποτελεί ένα πολύπλοκο βιολογικό ερώτημα όπου μετέχουν πιθανά πολλοί παράγοντες. Πολλοί ερευνητές έχουν προσπαθήσει να διαλευκάνουν τα μεταβολικά μονοπάτια και το γενετικό υπόβαθρο της τοξικότητας σε μεθοτρεξάτη (van Ede, Laan et al. 1998; Gervasini 2009; Schmiegelow 2009). Οι παρενέργειες της MTX μπορούν να εξηγηθούν μόνο μερικώς από την ανταγωνιστική της δράση με το φυλλικό οξύ και μπορεί να συσχετίζονται με τη δράση της σε άλλα σχετικά μεταβολικά μονοπάτια. Πρόσφατα δεδομένα περιλαμβάνουν το μονοπάτι ομοκυστεΐνη-μεθιονίνη-πολυμαμίνη και το μεταβολισμό των πουρινών (van Ede, Laan et al. 1998). Πολλές μελέτες έχουν διερευνήσει το ρόλο του μονοπατιού της αδενοσίνης σε τοξικότητες οφειλούμενες στη χρήση μεθοτρεξάτης (Merrill, Shen et al. 1997; Tian and Cronstein 2007). Περαιτέρω διερεύνηση απαιτείται για να διαλευκανθεί το συγκεκριμένο ζήτημα.

Από την άλλη πλευρά, πολλοί ερευνητές έχουν προσπαθήσει να συσχετίσουν τον κίνδυνο της τοξικότητας της μεθοτρεξάτης με γονιδιακούς πολυμορφισμούς. Υπάρχει ένας μεγάλος αριθμός μελετών ασθενών-μαρτύρων που επικεντρώθηκαν στους δυο πιο κοινούς πολυμορφισμούς του γονιδίου της MTHFR. Πολλές πρωταρχικές μελέτες έχουν εκτιμήσει συσχέτιση των πολυμορφισμών C677T και A1298C με τη τοξικότητα της MTX ενώ άλλες κατέληξαν σε απουσία συσχέτισης. Έχουν αναλυθεί επίσης επιπλέον γενετικοί παράγοντες όπως ο ABCB1 3435C>T (ATP-binding cassette) και TLR4 +896A>G (toll-like receptor 4) (Kooloos, Wessels et al.). SNPs στην SLCO1B1 (solute carrier organic anion transporter family) και ADORA2a (adenosine receptor 2a) έδειξαν σημαντική συσχέτιση με τη τοξικότητα του πεπτικού συστήματος (Hider, Thomson et al. 2008; Trevino, Shimasaki et al. 2009). Συνεπώς, η τοξικότητα της MTX μπορεί να τροποποιείται από επιπρόσθετους πολυμορφισμούς οι οποίοι πρέπει να διερευνηθούν περαιτέρω από μεταγενέστερες μελέτες (Herrlinger and Jewell 2006; Pierik, Rutgeerts et al. 2006).

Οι πολυμορφισμοί του γονιδίου της MTHFR διαδραματίζουν ένα σημαντικό ρόλο στην επιστημονική έρευνα σχετικά με τον κύκλο του φυλλικού οξέος. Οι δυο πιο κοινοί πολυμορφισμοί οδηγούν σε μειωμένη ενεργότητα του ενζύμου ακολουθούμενη από υπερομοκυστεϊναιμία. Προηγούμενες μελέτες έχουν προσπαθήσει να διερευνήσουν πολυμορφισμούς του γονιδίου MTHFR με διάφορες

ασθένειες και διαταραχές. Ο πολυμορφισμός C677T θεωρείται παράγοντας κινδύνου για τη θρομβοφιλία και της επιπλοκές της κύησης (Kurzawinska, Seremak-Mrozikiewicz et al. 2009; Vucic, Frleta et al. 2009). Επιπλέον, υπάρχουν ενδείξεις ότι οι πολυμορφισμοί του γονιδίου MTHFR αποτελούν προδιαθεσικούς παράγοντες για ανωμαλίες του νευρικού σωλήνα και σχιζοφρένεια (Frankenburg 2007; Greene, Stanier et al. 2009). Ο πολυμορφισμός C677T φαίνεται επίσης να συσχετίζεται με χαμηλότερο κίνδυνο για τον καρκίνο του παχέος εντέρου, κάτι που δεν ισχύει για όλους τους πληθυσμούς (Taioli, Garza et al. 2009). Επιπλέον, υπάρχουν ενδείξεις ότι τα αυξημένα επίπεδα ομοκυστεΐνης στο πλάσμα συσχετίζονται με αυξημένο κίνδυνο αθηροσκλήρωσης και καρδιαγγειακής νόσου. Πολλές μελέτες έχουν προσπαθήσει να αποδείξουν την επίδραση του πολυμορφισμού C677T στο κίνδυνο καρδιαγγειακών επεισοδίων αλλά τα αποτελέσματα είναι αντικρουόμενα (Trabetti 2008).

Η ανάλυσή μας περιελάμβανε επίσης τη διερεύνηση του ρόλου των συμπληρωμάτων φυλλικού οξέως στην εμφάνιση τοξικότητας. Δείξαμε ότι ο πολυμορφισμός C677T συσχετίστηκε με αυξημένο κίνδυνο τοξικότητας σε ασθενείς που λάμβαναν συμπληρώματα διατροφής. Το εύρημα αυτό είναι παράδοξο μιας και κάποιος θα περίμενε η λήψη συμπληρώματος φυλλικού να έχει προστατευτικό ρόλο. Μια πιθανή εξήγηση αποτελεί το γεγονός ότι οι ασθενείς που λάμβαναν επιπλέον συμπλήρωμα φυλλικού είχαν βαρύτερα συμπτώματα. Το συγκεκριμένο φαινόμενο είναι κοινό στις μελέτες παρατήρησης, όπου υπάρχουν οι λεγόμενοι συγχυτικοί παράγοντες (Salas, Hofman et al. 1999). Μια μετα-ανάλυση δημοσιευμένων μελετών μπορεί να παρέχει μόνο ενδείξεις και απαιτούνται περαιτέρω μελέτες για να διερευνήσουν το εν λόγω ζήτημα. Από την άλλη πλευρά, ο πολυμορφισμός A1298C συσχετίστηκε με αυξημένη τοξικότητα της MTX σε ασθενείς που δεν λάμβαναν συμπλήρωμα φυλλικού οξέος. Ωστόσο, πολλοί συγγραφείς προτείνουν ότι η χρήση συμπληρωμάτων φυλλικού οξέος μειώνει τον κίνδυνο παρενεργειών από τη χρήση της MTX και ότι θα πρέπει να συνταγογραφείται σε κάθε περίπτωση (Harten 2005).

Συμπερασματικά, οι αναλύσεις που διεξήγαμε υποδεικνύουν ότι οι πολυμορφισμοί της MTHFR συνδέονται με τη συνολική τοξικότητα σε μεθοτρεξάτη. Τα αποτελέσματά μας έδειξαν επίσης μια σημαντική συσχέτιση με διαφόρους τύπους παρενεργειών. Μελλοντικές μελέτες απαιτούνται για να επιβεβαιώσουν τα ευρήματα αυτά έτσι ώστε να τεθούν οι βάσεις για την εξατομικευμένη θεραπεία. Τέλος, περαιτέρω διερεύνηση απαιτεί η ανίχνευση των ακριβών βιολογικών μονοπατιών και των μηχανισμών που εξηγούν την τοξικότητα σε MTX.

Κεφάλαιο 6: Πολυμεταβλητή μεθοδολογία μετα-ανάλυσης και σύγκρισης διαγνωστικών δοκιμασιών

6.1. Εισαγωγή

Ένα διαγνωστικό τεστ είναι ένα οποιοδήποτε ιατρικό τεστ το οποίο αποσκοπεί στη διάγνωση ή την ανίχνευση μιας ασθένειας. Συνήθως τα ευρήματα μιας διαγνωστικής δοκιμασίας ταξινομούνται σε έναν 2×2 πίνακα ανάλογα με το είδος τους (θετικά ή αρνητικά) και την προέλευσή τους από άτομα που έχουν ή όχι τη νόσο που μελετάται. Η ακρίβεια μιας διαγνωστικής δοκιμασίας, η οποία αναφέρεται στην ικανότητα ενός τεστ να διαχωρίζει τα άτομα που έχουν ή όχι τη νόσο, συνήθως μετράται χρησιμοποιώντας το ποσοστό των ορθώς θετικών (True Positive Rate-*TPR*) και ψευδώς θετικών ευρημάτων (False Positive Rate-*FPR*). Το *TPR* είναι ταυτόσημη ποσότητα με την ευαισθησία (Sensitivity-*Se*) μιας διαγνωστικής δοκιμασίας ενώ το *FPR* ισούται με την συμπληρωματική ποσότητα της ειδικότητας (Specificity-*Sp*). Τα δυο αυτά μέτρα συσχετίζονται θετικά και ποικίλουν ανάλογα με το κατώφλι της διαγνωστικής δοκιμασίας. Η καμπύλη που συνδέει τα σημεία με συντεταγμένες την ευαισθησία και τη συμπληρωματική της ειδικότητας τιμή, για δεδομένο κατώφλι της υπό εξέταση εργαστηριακής μεταβλητής, ονομάζεται καμπύλη ROC (Receiver Operating Characteristic Curve). Όσο μεγαλύτερη είναι η υπό την καμπύλη επιφάνεια (Area Under the Curve-AUC), τόσο μεγαλύτερη είναι η διαγνωστική αξία της διαγνωστικής δοκιμασίας (Mandrekar 2010).

Όπως έχει ήδη αναφερθεί στις προηγούμενες παραγράφους, στη μετα-ανάλυση ένα σύνολο πρωταρχικών μελετών συνδυάζεται και ανιχνεύονται οι πιθανές πηγές μεταβλητότητας (Glass 1976; Petiti 1994; Greenland 1998; Normand 1999). Διάφορες τεχνικές μετα-ανάλυσης έχουν προταθεί για να συνδυάσουν τα αποτελέσματα μελετών που αξιολογούν την ακρίβεια μιας διαγνωστικής δοκιμασίας (Moses, Shapiro et al. 1993; Rutter and Gatsonis 2001; Macaskill 2004; Harbord, Deeks et al. 2007; Arends, Hamza et al. 2008). Η κύρια διαφορά των μελετών διαγνωστικών δοκιμασιών είναι ότι συνήθως παραθέτουν ένα ζεύγος εκτιμήσεων (*TPR* και *FPR*) και συνεπώς απαιτείται η εφαρμογή ειδικής (δηλαδή διμεταβλητής) μεθοδολογίας.

Η απλούστερη (και σίγουρα μη προτεινόμενη) μεθοδολογία μετα-ανάλυσης διαγνωστικών δοκιμασιών, περιλαμβάνει έναν μη στρωματοποιημένο σχεδιασμό όπου προστίθεται ο αριθμός των αληθώς θετικών (True Positive-*TP*), ψευδώς θετικών (False Positive-*FP*), αληθώς αρνητικών (True Negatives-*TN*) και ψευδώς αρνητικών (False Negatives-*FN*) ευρημάτων από κάθε μελέτη. Έχει προταθεί ξεχωριστή μετα-ανάλυση μετασχηματισμένων ποσοτήτων σε κλίμακα των logits της ευαισθησίας και της ειδικότητας. Η προσέγγιση αυτή επιτρέπει την ύπαρξη ετερογένειας στην ευαισθησία και ειδικότητα αλλά αγνοεί τη μεταξύ των μελετών μεταβλητότητα (δηλαδή τη συσχέτιση των τυχαίων επιδράσεων). Επιπλέον, έχει προταθεί ξεχωριστή μετα-ανάλυση θετικών και αρνητικών πηλίκων πιθανοφάνειας (Deeks 2001) όπου επίσης αγνοείται η συσχέτιση μεταξύ των δυο αυτών παραμέτρων. Άλλοι συγγραφείς έχουν προτείνει τον υπολογισμό του διαγνωστικού Odds Ratio (Diagnostic Odds Ratio-*DOR*) από κάθε μελέτη και το συνδυασμό τους σε μια κλασσική μονομεταβλητή μετα-ανάλυση (Glas, Lijmer et al. 2003), με μοντέλα σταθερών ή τυχαίων επιδράσεων. Σε αυτή την περίπτωση, λαμβάνεται υπόψη η μεταξύ των μελετών μεταβλητότητα αλλά όχι και το γεγονός ότι σε κάθε μελέτη μπορεί να έχει χρησιμοποιηθεί ένα διαφορετικό κατώφλι; επιπλέον, δεν είναι εφικτός ο υπολογισμός μιας συνολικής εκτίμησης της ευαισθησίας και της ειδικότητας. Η ευρέως χρησιμοποιούμενη προσέγγιση είναι η μέθοδος summary ROC (SROC) (Littenberg and Moses 1993; Moses, Shapiro et al. 1993), που χρησιμοποιεί μετασχηματισμούς σε κλίμακα των logits των *TPR* και *FPR* και βασίζεται σε μια απλή γραμμική παλινδρόμηση της διαφοράς (*D* που ισούται με το λογάριθμο του *DOR*) ως προς το άθροισμα αυτών (*S*). Μια συνοπτική καμπύλη ROC μπορεί ακολούθως να κατασκευαστεί από την εκτίμηση των παραμέτρων του μοντέλου. Η μεθοδολογία αυτή συνυπολογίζει τη χρήση διαφορετικού κατωφλίου σε κάθε μελέτη αλλά ακόμα και στην περίπτωση που διεξάγεται σταθμισμένη ανάλυση για το άθροισμα των παραμέτρων, δε λαμβάνεται υπόψη το σφάλμα μέτρησης στο *S*.

Στη μέθοδο hierarchical SROC (HSROC) (Rutter and Gatsonis 2001) οι μετασχηματισμένες ποσότητες σε κλίμακα των logits των *TPR* and *FPR* εκφράζονται ως δυο παράμετροι (ακρίβεια και κατώφλι). Η μέθοδος αυτή επιτρέπει τη μεταξύ των μελετών μεταβλητότητα μοντελοποιώντας την ακρίβεια ως όρο τυχαίων επιδράσεων. Το μοντέλο αυτό απλοποιήθηκε χρησιμοποιώντας εκτιμητές Empirical Bayes με αποτελέσματα σχεδόν ταυτόσημα με εκείνα που προκύπτουν από τη μπεϋζιανή προσέγγιση (Macaskill 2004), και επεκτάθηκε για την περίπτωση όπου δεν υπάρχει

μέθοδος αναφοράς (gold standard) (Dendukuri, Schiller et al. 2012). Επιπρόσθετα, έχει προταθεί ένα διμεταβλητό μοντέλο μετα-ανάλυσης τυχαίων επιδράσεων μετασχηματισμένων ποσοτήτων σε κλίμακα των logits των *TPR* και *FPR* όπου επιτρέπεται η μεταξύ των μελετών ετερογένεια και η συσχέτιση μεταξύ των *TPR* και *FPR* (Reitsma, Glas et al. 2005; Arends, Hamza et al. 2008). Οι μεθοδολογίες όπου μοντελοποιείται απευθείας η διωνυμική μορφή των δεδομένων είναι προτιμότερες (Chu and Cole 2006; Arends, Hamza et al. 2008). Ο Harbord και συνεργάτες, έδειξαν ότι οι προαναφερθείσες πολυμεταβλητές μεθοδολογίες μετα-ανάλυσης αντιστοιχούν σε ισοδύναμα μοντέλα όπου απλώς χρησιμοποιείται διαφορετική παραμετροποίηση (Harbord, Deeks et al. 2007). Πρόσφατα, έχει προταθεί μια εναλλακτική μέθοδος composite likelihood για τη διμεταβλητή μετα-ανάλυση, όπου δεν παρουσιάζει προβλήματα σύγκλισης όταν ο αριθμός των μελετών είναι μικρός και είναι πιο robust όταν παραβιάζονται οι προϋποθέσεις των από κοινού κατανομών (Chen, Liu et al. 2014). Όταν ο επιπολασμός της νόσου είναι γνωστός, όπως συμβαίνει στις μελέτες κοορτής, έχει προταθεί η τριμεταβλητή μοντελοποίηση του επιπολασμού της νόσου, της ευαισθησίας και ειδικότητας (Chu, Nie et al. 2009). Η προσέγγιση αυτή λαμβάνει υπόψη την εξάρτηση της ευαισθησίας και ειδικότητας από τον επιπολασμό της νόσου κάτι που είναι προφανές όταν χρησιμοποιείται ένα συνεχές χαρακτηριστικό για την ταξινόμηση. Η συσχέτιση αυτή εκτιμήθηκε χρησιμοποιώντας το συντελεστή συσχέτισης του Pearson (Li and Fine 2011). Μειονέκτημα της τριμεταβλητής αυτής προσέγγισης είναι ότι μπορεί να περιλαμβάνει μόνο μελέτες κοορτής που παρέχουν πληροφορίες για την εκτίμηση του επιπολασμού της νόσου. Συνεπώς, έχουν προταθεί δυο κύριες εναλλακτικές προσεγγίσεις, που περιλαμβάνουν ένα μπεϋζιανό ιεραρχικό μοντέλο που ενσωματώνει δεδομένα μελετών κοορτής και μελετών ασθενών-μαρτύρων (Ma, Chen et al. 2014) και ένα υβριδικό μοντέλο, όπου χρησιμοποιείται μια εναλλακτική διαδικασία εκτίμησης παραμέτρων (composite likelihood) (Chen, Liu et al. 2015). Ο Ma και οι συνεργάτες παρουσίασαν μια πρόσφατη σύνοψη των πολυμεταβλητών μεθοδολογιών μετα-ανάλυσης διαγνωστικών μελετών (Ma, Nie et al. 2013).

Ο βαθμός της πολυπλοκότητας μεγαλώνει στην περίπτωση της σύγκρισης δυο ή περισσότερων διαγνωστικών δοκιμασιών. Η παραδοσιακή μεθοδολογία μετα-ανάλυσης για τη σύγκριση διαγνωστικών δοκιμασιών εξάγει τις ποσότητες *D* (δηλαδή τη διαφορά μετασχηματισμένων ποσοτήτων σε κλίμακα των logits των *TPR* και *FPR*) από κάθε μελέτη (ή τις εκτιμώμενες τιμές από ένα SROC μοντέλο) και

κατόπιν τις συνοψίζει. Συνεπώς, οι διαγνωστικές δοκιμασίες μπορούν να συγκριθούν χρησιμοποιώντας τη διαφορά των συνοπτικών Ds . Μια εναλλακτική προσέγγιση βασίζεται στο σχετικό Odds Ratio (relative odds ratio- ROR) που υπολογίζεται από τη σχετική ακρίβεια μιας διαγνωστικής δοκιμασίας σε σχέση με κάποια άλλη και στηρίζεται στην υπόθεση ότι οι δυο δοκιμασίες πραγματοποιούνται σε ζεύγη ατόμων εντός κάθε μελέτης. Από την άλλη πλευρά, εάν υπάρχουν επαρκή δεδομένα (το οποίο ωστόσο σπανίως συμβαίνει), μπορεί να χρησιμοποιηθεί ο υπολογισμός του δεσμευμένου σχετικού Odds Ratio (conditional relative $OR-CROR$) που βασίζεται στα μη συμφωνούντα αποτελέσματα των διαγνωστικών δοκιμασιών (Suzuki, Morooka et al. 2004). Επιπλέον, έχει υιοθετηθεί μεθοδολογία επαναλαμβανόμενων μετρήσεων για το συνδυασμό μελετών διαγνωστικών δοκιμασιών, όπου σε κάθε μελέτη παρέχονται εκτιμήσεις για περισσότερες της μιας δοκιμασίες (Siadaty, Philbrick et al. 2004). Στις περιπτώσεις όπου δεν υπάρχει μέθοδος αναφοράς, έχει προταθεί ένα πολυμεταβλητό μοντέλο τυχαίων επιδράσεων για την από κοινού μοντελοποίηση της ευαισθησίας, της ειδικότητας και του επιπολασμού της νόσου (Chu, Chen et al. 2009). Μοντέλα μεικτών επιδράσεων ή μπεϋζιανά ιεραρχικά μοντέλα μπορούν να χρησιμοποιηθούν για την εκτίμηση των παραμέτρων. Όταν η conditional independence assumption δεν ισχύει (Gardner, Stryhn et al. 2000), μπορεί να υποθεθεί μια συσχέτιση μεταξύ των διαγνωστικών δοκιμασιών. Η προαναφερθείσα προσέγγιση και η μέθοδος HSROC που παρουσιάστηκε από τον Dendukuri και συνεργάτες (Dendukuri, Schiller et al. 2012) ταυτίζονται σε αρκετά σημεία και πολλά από τα μοντέλα που παρουσιάζονται είναι ισοδύναμα (Liu, Chen et al. 2015). Τέλος, ο Trikalinos και οι συνεργάτες παρουσίασαν ένα μπεϋζιανό μοντέλο που λαμβάνει υπόψη τις συσχετίσεις των TPR και FPR μεταξύ δυο ή περισσότερων διαγνωστικών δοκιμασιών όταν αυτές πραγματοποιούνται σε διαφορετικά άτομα (Trikalinos, Hoaglin et al. 2014).

Παρουσιάζεται μια απλή μεθοδολογία αλλά με μεγάλη ισχύ για την πραγματοποίηση πολυμεταβλητής μετα-ανάλυσης σε μελέτες διαγνωστικών δοκιμασιών. Το μοντέλο που προτείνουμε αποτελεί μια απευθείας επέκταση της μεθόδου πολυμεταβλητής μετα-ανάλυσης μελετών διαγνωστικών δοκιμασιών (Reitsma, Glas et al. 2005; Arends, Hamza et al. 2008). Το σημείο κλειδί της προσέγγισής μας είναι ότι παρέχονται σε κλειστή μορφή σχέσεις για τον υπολογισμό των εντός των μελετών συνδιακυμάνσεων (Bagos 2012). Δεδομένου ότι το μοντέλο στηρίζεται στη γενική μεθοδολογία πολυμεταβλητής μετα-ανάλυσης, έχει κάποια

σημαντικά χαρακτηριστικά: μπορεί να συμπεριλάβει μελέτες που αναφέρουν τις εκτιμήσεις κάποιων, αλλά όχι απαραίτητα όλων των διαγνωστικών δοκιμασιών, κερδίζοντας έτσι ισχύ από εξωτερικές μελέτες (Higgins and Whitehead 1996); μπορεί να υλοποιηθεί εύκολα σε οποιοδήποτε στατιστικό πακέτο για πολυμεταβλητή μετα-ανάλυση (Jackson, Riley et al. 2011; Mavridis and Salanti 2012); επιτρέπει μέσω κατάλληλων μετασχηματισμών τον υπολογισμό σημαντικών μέτρων όπως το DOR, η καμπύλη SROC και η AUC; και τελικά, επιτρέπει την απευθείας σύγκριση των εκτιμώμενων παραμέτρων χρησιμοποιώντας επίσημους στατιστικούς ελέγχους.

6.2. Μεθοδολογία

6.2.1. Το πολυμεταβλητό μοντέλο

Έστω οι Y_i , X_{1i} και X_{2i} είναι τρεις κατηγορικές τυχαίες μεταβλητές με δυο επίπεδα (δηλαδή 0, 1) που χρησιμοποιούνται να ταξινομήσουν n_i άτομα για κάθε i μελέτη ($i=1,2,\dots,k$). Συνήθως η τυχαία μεταβλητή Y_i υποδηλώνει την παρουσία ή απουσία της νόσου ενώ οι X_{1i} , X_{2i} τα αποτελέσματα για την περίπτωση των δυο διαγνωστικών δοκιμασιών (θετικά ή αρνητικά) για κάθε i μελέτη ($i=1,2,\dots,k$). Στη γενική περίπτωση, τα δεδομένα μπορούν να παρουσιαστούν με τη μορφή ενός πίνακα συνάφειας ($2 \times 2 \times 2$) (Πίνακας 6.1) όπου τα άτομα υποδηλώνονται ως n_{clpi} με $c, l, p \in \{0,1\}$ και ο πίνακας αυτός αναφέρεται σαν μερικός πίνακας συσχέτισης. Συνήθως, τα αποτελέσματα δυο διαγνωστικών δοκιμασιών παρουσιάζονται με τη μορφή του Πίνακα 6.2, όπου, με την ορολογία πινάκων συνάφειας, είναι ένας περιθώριος πίνακας δεδομένου ότι για κάθε διαγνωστική δοκιμασία δεν λαμβάνονται υπόψη τα αποτελέσματα της δεύτερης. Οι μετασχηματισμοί σε κλίμακα των logits των \widehat{TPR}_{ji} (\widehat{Se}_{ji}) και \widehat{FPR}_{ji} ($1 - \widehat{Sp}_{ji}$) για τη j δοκιμασία ($j=1,2$) και για την i μελέτη ($i=1,2,\dots,k$) δίνονται ως ακολούθως:

$$\hat{y}_{1i} = \text{logit}(\widehat{TPR}_{1i}) = \text{logit}(\widehat{Se}_{1i}) = \log\left(\frac{TP_{1i}}{FN_{1i}}\right) = \log\left(\frac{n_{11+i}}{n_{10+i}}\right) \quad (6.1)$$

$$\hat{y}_{2i} = \text{logit}(\widehat{FPR}_{1i}) = \text{logit}(1 - \widehat{Sp}_{1i}) = \log\left(\frac{FP_{1i}}{TN_{1i}}\right) = \log\left(\frac{n_{01+i}}{n_{00+i}}\right) \quad (6.2)$$

$$\hat{y}_{3i} = \text{logit}(\widehat{TPR}_{2i}) = \text{logit}(\widehat{Se}_{2i}) = \log\left(\frac{TP_{2i}}{FN_{2i}}\right) = \log\left(\frac{n_{1+1i}}{n_{1+0i}}\right) \quad (6.3)$$

$$\hat{y}_{4i} = \text{logit}(\widehat{FPR}_{2i}) = \text{logit}(1 - \widehat{Sp}_{2i}) = \log\left(\frac{FP_{2i}}{TN_{2i}}\right) = \log\left(\frac{n_{0+1i}}{n_{0+0i}}\right) \quad (6.4)$$

με τις ακόλουθες εκτιμήσεις των διασπορών:

$$s_{1i}^2 = \frac{1}{TP_{1i}} + \frac{1}{FN_{1i}} = \frac{1}{n_{11+i}} + \frac{1}{n_{10+i}} \quad (6.5)$$

$$s_{2i}^2 = \frac{1}{FP_{1i}} + \frac{1}{TN_{1i}} = \frac{1}{n_{01+i}} + \frac{1}{n_{00+i}} \quad (6.6)$$

$$s_{3i}^2 = \frac{1}{TP_{2i}} + \frac{1}{FN_{2i}} = \frac{1}{n_{1+1i}} + \frac{1}{n_{1+0i}} \quad (6.7)$$

$$s_{4i}^2 = \frac{1}{FP_{2i}} + \frac{1}{TN_{2i}} = \frac{1}{n_{0+1i}} + \frac{1}{n_{0+0i}} \quad (6.8)$$

Πρέπει να τονιστεί ότι στην κλασσική προσέγγιση μετα-ανάλυσης για μια διαγνωστική δοκιμασία (Harbord, Deeks et al. 2007; Arends, Hamza et al. 2008), πραγματοποιείται διμεταβλητή μετα-ανάλυση για τα ζεύγη των εκβάσεων (δηλαδή $\hat{y}_{1i}, \hat{y}_{2i}$ ή $\hat{y}_{3i}, \hat{y}_{4i}$). Έτσι, αν θέλουμε να πραγματοποιήσουμε μια πολυμεταβλητή μετα-ανάλυση για την περίπτωση των δυο διαγνωστικών δοκιμασιών (Berkey, Hoaglin et al. 1998; van Houwelingen, Arends et al. 2002), ορίζουμε ως \mathbf{y}_i το διάνυσμα των 4 διαφορετικών εκτιμήσεων και με $\boldsymbol{\beta}$, το διάνυσμα των μέσων όρων:

$$\mathbf{y}_i = \begin{pmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \hat{y}_{3i} \\ \hat{y}_{4i} \end{pmatrix}, \text{ και } \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \quad (6.9)$$

Υποθέτουμε ότι τα τέσσερα \hat{y}_i 's ακολουθούν μια πολυμεταβλητή κανονική κατανομή:

$$\begin{bmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \hat{y}_{3i} \\ \hat{y}_{4i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{bmatrix}, [C_i] \right\} \quad (6.10)$$

όπου ο \mathbf{C}_i είναι ο εντός των μελετών πίνακας συνδιακυμάνσεων:

$$\mathbf{C}_i = \begin{pmatrix} s_{1i}^2 & \rho_{w12i}s_{1i}s_{2i} & \rho_{w13i}s_{1i}s_{3i} & \rho_{w14i}s_{1i}s_{4i} \\ \rho_{w12i}s_{1i}s_{2i} & s_{2i}^2 & \rho_{w23i}s_{2i}s_{3i} & \rho_{w24i}s_{2i}s_{4i} \\ \rho_{w13i}s_{1i}s_{3i} & \rho_{w23i}s_{2i}s_{3i} & s_{3i}^2 & \rho_{w34i}s_{3i}s_{4i} \\ \rho_{w14i}s_{1i}s_{4i} & \rho_{w24i}s_{2i}s_{4i} & \rho_{w34i}s_{3i}s_{4i} & s_{4i}^2 \end{pmatrix} \quad (6.11)$$

Οι μέσοι ($\beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i}$) θεωρούνται τυχαίοι όροι και κατανέμονται:

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, [\Sigma] \right\} \quad (6.12)$$

όπου ο Σ είναι ο μεταξύ των μελετών πίνακας συνδιακυμάνσεων:

$$\Sigma = \begin{pmatrix} \tau_1^2 & \rho_{B12}\tau_1\tau_2 & \rho_{B13}\tau_1\tau_3 & \rho_{B14}\tau_1\tau_4 \\ \rho_{B12}\tau_1\tau_2 & \tau_2^2 & \rho_{B23}\tau_2\tau_3 & \rho_{B24}\tau_2\tau_4 \\ \rho_{B13}\tau_1\tau_3 & \rho_{B23}\tau_2\tau_3 & \tau_3^2 & \rho_{B34}\tau_3\tau_4 \\ \rho_{B14}\tau_1\tau_4 & \rho_{B24}\tau_2\tau_4 & \rho_{B34}\tau_3\tau_4 & \tau_4^2 \end{pmatrix} \quad (6.13)$$

Συνεπώς, το τελικό περιθώριο μοντέλο είναι το ακόλουθο:

$$\mathbf{y}_i \sim MVN(\boldsymbol{\beta}, \Sigma + \mathbf{C}_i) \quad (6.14)$$

το οποίο μπορεί να παρουσιαστεί και ως ακολούθως:

$$\begin{bmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \hat{y}_{3i} \\ \hat{y}_{4i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \tau_1^2 & \rho_{w12i}s_{1i}s_{2i} + \rho_{B12}\tau_1\tau_2 & \rho_{w13i}s_{1i}s_{3i} + \rho_{B13}\tau_1\tau_3 & \rho_{w14i}s_{1i}s_{4i} + \rho_{B14}\tau_1\tau_4 \\ \rho_{w12i}s_{1i}s_{2i} + \rho_{B12}\tau_1\tau_2 & s_{2i}^2 + \tau_2^2 & \rho_{w23i}s_{2i}s_{3i} + \rho_{B23}\tau_2\tau_3 & \rho_{w24i}s_{2i}s_{4i} + \rho_{B24}\tau_2\tau_4 \\ \rho_{w13i}s_{1i}s_{3i} + \rho_{B13}\tau_1\tau_3 & \rho_{w23i}s_{2i}s_{3i} + \rho_{B23}\tau_2\tau_3 & s_{3i}^2 + \tau_3^2 & \rho_{w34i}s_{3i}s_{4i} + \rho_{B34}\tau_3\tau_4 \\ \rho_{w14i}s_{1i}s_{4i} + \rho_{B14}\tau_1\tau_4 & \rho_{w24i}s_{2i}s_{4i} + \rho_{B24}\tau_2\tau_4 & \rho_{w34i}s_{3i}s_{4i} + \rho_{B34}\tau_3\tau_4 & s_{4i}^2 + \tau_4^2 \end{bmatrix} \right\} \quad (6.15)$$

Όπως έχει ήδη αναφερθεί τα στοιχεία εκτός της διαγωνίου του πίνακα \mathbf{C}_i αντιστοιχούν στις εντός των μελετών συνδιακυμάνσεις, για παράδειγμα $\rho_{w12i}s_{1i}s_{2i} = \text{cov}(\hat{y}_{1i}, \hat{y}_{2i})$ (όπου ρ_{w12i} είναι ο εντός των μελετών συντελεστής συσχέτισης των TPR και FPR της πρώτης διαγνωστικής δοκιμασίας για την i μελέτη). Δεδομένου ότι $\text{cov}(\hat{y}_{1i}, \hat{y}_{2i}) = 0$ και $\text{cov}(\hat{y}_{3i}, \hat{y}_{4i}) = 0$ οι υπόλοιπες συνδιακυμάνσεις πρέπει να υπολογιστούν (βλέπε την επόμενη ενότητα).

Οι παράμετροι που εκτιμώνται από το γενικό πολυμεταβλητό μοντέλο της εξίσωσης 6.14 μπορούν να συγκριθούν απευθείας χρησιμοποιώντας για παράδειγμα τον έλεγχο του Wald. Άλλα μέτρα όπως το DOR, η AUC και παράμετροι της καμπύλης SROC μπορούν να υπολογιστούν εύκολα (Arends, Hamza et al. 2008) και να συγκριθούν. Η επέκταση της μεθόδου για περισσότερες από μια δοκιμασίες είναι άμεση.

Πίνακας 6.1. Ο πίνακας που καθορίζει την από κοινού κατανομή της συσχέτισης των δυο διαγνωστικών δοκιμασιών ανάλογα με την παρουσία ή απουσία της νόσου για κάθε i μελέτη ($i=1,2,\dots,k$). Ο πίνακας αυτός αναφέρεται συνήθως ως partial contingency table. Βλέπε επίσης (Agresti 2002; Bagos 2012).

		Y_i $Y_i=1$		$Y_i=0$	
		X_{2i} $X_{2i}=1$		$X_{2i}=0$	
X_{1i}	$X_{1i}=1$	n_{111i}	n_{110i}	n_{011i}	n_{010i}
	$X_{1i}=0$	n_{101i}	n_{100i}	n_{001i}	n_{000i}

Πίνακας 6.2. Ταξινόμηση των ευρημάτων των δυο διαγνωστικών δοκιμασιών ανάλογα με το αποτέλεσμα της δοκιμασίας (θετικό ή αρνητικό) και την παρουσία ή απουσία της νόσου για κάθε i μελέτη ($i=1,2,\dots,k$). Οι πίνακες αυτοί αναφέρονται συνήθως ως περιθώριοι πίνακες. Βλέπε επίσης (Agresti 2002; Bagos 2012).

		X_{1i} $X_{1i}=1$		X_{2i} $X_{2i}=1$	
		$X_{1i}=0$		$X_{2i}=0$	
Y_i	$Y_i=1$	n_{11+i} (TP_{1i})	n_{10+i} (FN_{1i})	n_{11+i} (TP_{2i})	n_{10+i} (FN_{2i})
	$Y_i=0$	n_{01+i} (FP_{1i})	n_{00+i} (TN_{1i})	n_{01+i} (FP_{2i})	n_{00+i} (TN_{2i})

6.2.2. Υπολογισμός των εντός των μελετών συνδιακυμάνσεων

Η μεθοδολογία που χρησιμοποιήθηκε αποτελεί μια ειδική περίπτωση της γενικής μεθοδολογίας που έχει προταθεί πρόσφατα (Bagos 2012). Συνεπώς, μπορούμε να προχωρήσουμε με κάποιους υπολογισμούς πιθανοτήτων (Παράρτημα Ε) χρησιμοποιώντας τις ιδιότητες των συνδιακυμάνσεων:

$$\begin{aligned} \text{cov}(\hat{y}_{1i}, \hat{y}_{3i}) &= \frac{n_{111i}}{(n_{111i} + n_{110i})(n_{111i} + n_{101i})} - \frac{n_{110i}}{(n_{111i} + n_{110i})(n_{110i} + n_{100i})} \\ &\quad - \frac{n_{101i}}{(n_{101i} + n_{100i})(n_{111i} + n_{101i})} + \frac{n_{100i}}{(n_{101i} + n_{100i})(n_{110i} + n_{100i})} \quad (6.16) \\ &= \sum_l \sum_p (-1)^{l-p} \left(\frac{n_{lpi}}{n_{l+i} n_{1+pi}} \right) \end{aligned}$$

$$\begin{aligned} \text{cov}(\hat{y}_{2i}, \hat{y}_{4i}) &= \frac{n_{011i}}{(n_{011i} + n_{010i})(n_{011i} + n_{001i})} - \frac{n_{010i}}{(n_{011i} + n_{010i})(n_{010i} + n_{000i})} \\ &\quad - \frac{n_{001i}}{(n_{001i} + n_{000i})(n_{011i} + n_{001i})} + \frac{n_{000i}}{(n_{001i} + n_{000i})(n_{010i} + n_{000i})} \quad (6.17) \\ &= \sum_l \sum_p (-1)^{l-p} \left(\frac{n_{0lpi}}{n_{0l+i} n_{0+p i}} \right) \end{aligned}$$

$$\text{cov}(\hat{y}_{1i}, \hat{y}_{4i}) = \text{cov}(\hat{y}_{2i}, \hat{y}_{3i}) = 0 \quad (6.18)$$

Για την περίπτωση όπου ο αριθμός των ατόμων δεν είναι ίσος και για τις δυο διαγνωστικές δοκιμασίες ανά μελέτη, μπορούμε να αυξήσουμε τα άτομα της δοκιμασίας με τα λιγότερα άτομα και να τα εξισώσουμε με εκείνα της δεύτερης που περιλαμβάνει τα περισσότερα άτομα.

6.2.3. Διαχείριση ελλειπουσών πληροφοριών

Όπως είναι προφανές από τις εξισώσεις 6.16 και 6.17, για τον υπολογισμό των εντός των μελετών συνδιακυμάνσεων πρέπει να είναι γνωστές οι ποσότητες των κελιών του Πίνακα 6.1. Στην απλούστερη περίπτωση, όπου για κάθε i μελέτη οι ποσότητες των κελιών αυτών είναι γνωστές ο υπολογισμός των συνδιακυμάνσεων είναι άμεσος. Στην πράξη όμως, δεν περιμένουμε όλες οι μελέτες να έχουν τις πληροφορίες αυτές.

Όταν οι πληροφορίες αυτές είναι διαθέσιμες για κάποιες μελέτες, μπορούμε να εργαστούμε ως ακολούθως. Αρχικά, για αυτές τις «πλήρεις» μελέτες, μπορούμε να υπολογίσουμε, τα δεσμευμένα Odds Ratios (conditional ORs) που περιγράφουν τη συσχέτιση των δυο διαγνωστικών δοκιμασιών δοθείσας της απουσίας ή παρουσίας της νόσου:

$$\widehat{OR}_{Y_i=1} = \frac{n_{111i} n_{100i}}{n_{101i} n_{110i}} \quad (6.19)$$

$$\widehat{OR}_{Y_i=0} = \frac{n_{011i} n_{00i}}{n_{010i} n_{001i}} \quad (6.20)$$

Ακολούθως, χρησιμοποιώντας την κλασσική μεθοδολογία μετα-ανάλυσης τυχαίων επιδράσεων, για αυτές τις «πλήρεις» μελέτες μπορούμε να υπολογίσουμε τις συνδυασμένες εκτιμήσεις των conditional Odds Ratios για τα άτομα που έχουν ή δεν έχουν τη νόσο. Οι εκτιμήσεις αυτές μπορούν να χρησιμοποιηθούν για τις υπόλοιπες μελέτες που δεν έχουν πλήρη δεδομένα μαζί με τα γνωστά περιθώρια αθροίσματα και

έτσι προκύπτει ένα σύστημα εξισώσεων από τις οποίες παίρνουμε μοναδική λύση για τα κελιά του Πίνακα 6.1 (Παράρτημα ΣΤ).

Εάν τα conditional Odds Ratios δεν διαφέρουν ανάλογα με την παρουσία ή απουσία της νόσου, ή έχουμε κάποια πρότερη γνώση ότι αυτό ισχύει, μπορεί να χρησιμοποιηθεί ο αλγόριθμος Iterative Proportional Fitting (IPF) (Deming 1940) έτσι ώστε να υπολογιστούν τα κελιά του Πίνακα 6.1, ακόμα και όταν καμιά μελέτη δεν αναφέρει τα δεδομένα του μερικού πίνακα συσχέτισης. Πρέπει να τονιστεί ότι ο αλγόριθμος IPF υποθέτει ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης στον $2 \times 2 \times 2$ πίνακα συνάφειας. Με άλλα λόγια, υποθέτει ότι τα δυο Odds Ratios των εξισώσεων 6.19 και 6.20 είναι ίσα. Συνεπώς, η προσέγγιση αυτή (αν και δεν είναι τέλεια) είναι πιο γενική από την υπόθεση της conditional independence που χρησιμοποιείται συνήθως στην ανάλυση (Hui and Zhou 1998; Enoe, Georgiadis et al. 2000) και μετα-ανάλυση μελετών διαγνωστικών δοκιμασιών όπου δεν υπάρχει μέθοδος αναφοράς (Chu, Chen et al. 2009; Dendukuri, Schiller et al. 2012)

6.2.4. Η μέθοδος SROC και η AUC

Στην προσέγγιση SROC (Littenberg and Moses 1993; Moses, Shapiro et al. 1993), χρησιμοποιούνται οι μετασχηματισμένες ποσότητες σε κλίμακα των logits των TPR_{ji} και FPR_{ji} και πραγματοποιείται γραμμική παλινδρόμηση της διαφοράς αυτών (D_{ji}) ως προς το άθροισμά τους (S_{ji}). Έτσι, οι παράμετροι της παλινδρόμησης για την πρώτη διαγνωστική δοκιμασία μπορούν να εκφραστούν χρησιμοποιώντας τις εκτιμώμενες μεταξύ των μελετών διασπορές (τ_1^2 και τ_2^2) και τη συνδιακύμανση ($\rho_{B12}\tau_1\tau_2$). Πιο συγκεκριμένα, από την εξίσωση 6.12 προκύπτει ότι η συνδιακύμανση των D_{1i} και S_{1i} ισούται με $\tau_1^2 - \tau_2^2$ και η διασπορά του S_{1i} είναι ίση με $\tau_1^2 + \tau_2^2 + 2\rho_{B12}\tau_1\tau_2$. Συνεπώς, η κλίση είναι $b_1 = (\tau_1^2 - \tau_2^2) / (\tau_1^2 + \tau_2^2 + 2\rho_{B12}\tau_1\tau_2)$, ο σταθερός όρος $a_1 = \beta_1 - \beta_2 - b(\beta_1 + \beta_2)$ και η διασπορά των καταλοίπων της παλινδρόμησης δίνεται

από $\sigma_{D_i|S_i}^2 = (\tau_1^2 + \tau_2^2 - 2\rho_{B12}\tau_1\tau_2) - \frac{(\tau_1^2 - \tau_2^2)^2}{\tau_1^2 + \tau_2^2 + 2\rho_{B12}\tau_1\tau_2}$ (Arends, Hamza et al. 2008). Οι

ποσότητες αυτές μπορούν να υπολογιστούν εύκολα από τις εκτιμήσεις των παραμέτρων του μοντέλου; τα τυπικά σφάλματα υπολογίζονται χρησιμοποιώντας τη μέθοδο Δέλτα (Agresti 2002). Παρόμοιες εκτιμήσεις υπολογίζονται και για τη δεύτερη διαγνωστική δοκιμασία χρησιμοποιώντας τις κατάλληλες εκτιμήσεις του

μεταξύ των μελετών πίνακα διακυμάνσεων-συνδιακυμάνσεων (δηλαδή $\tau_3^2, \tau_4^2, \rho_{B34}\tau_3\tau_4$). Αξίζει να αναφερθεί ότι ο συντελεστής b_j δείχνει την εξάρτηση της ακρίβειας της διαγνωστικής δοκιμασίας από το κατώφλι που επιλέγεται. Εάν το $b_j \approx 0$, τότε οι μελέτες είναι ομοιογενείς και μπορούν να συνοψιστούν από ένα συνολικό DOR_j σημειώνοντας ότι $a_j = \ln(OR_j)$.

Εφόσον εκτιμηθούν οι παράμετροι a_j και b_j , μια σχέση μεταξύ των TPR_{ji} και FPR_{ji} μπορεί να εξαχθεί:

$$TPR_{ji} = \frac{\exp\left(\frac{a_j}{1-b_j}\right)\left(\frac{FPR_{ji}}{1-FPR_{ji}}\right)^{(1+b_j)/(1-b_j)}}{1 + \exp\left(\frac{a_j}{1-b_j}\right)\left(\frac{FPR_{ji}}{1-FPR_{ji}}\right)^{(1+b_j)/(1-b_j)}} \quad (6.21)$$

Η εξίσωση 6.21 δίνει τα TPR_{ji} για κάθε τιμή των FPR_{ji} και συνεπώς ολόκληρη την καμπύλη SROC. Αν και μπορεί να υπάρχει ενδιαφέρον για την ανίχνευση συγκεκριμένων σημείων στην καμπύλη, συχνά είναι εξαιρετικά σημαντικό να έχουμε μια συνολική εκτίμηση από την καμπύλη. Ίσως το πιο κατάλληλο μέτρο είναι η AUC, που μπορεί να υπολογιστεί ως (Walter 2002):

$$AUC_j = \int_0^1 \frac{\exp\left(\frac{a_j}{1-b_j}\right)\left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)}}{1 + \exp\left(\frac{a_j}{1-b_j}\right)\left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)}} dx \quad (6.22)$$

Χρησιμοποιώντας τη μέθοδο-Δέλτα, μια προσεγγιστική εκτίμηση της διασποράς της \hat{AUC}_j είναι η ακόλουθη:

$$\text{var}(\hat{AUC}_j) = \left(\frac{\partial AUC_j}{\partial a_j}\right)^2 \text{var}(\hat{a}_j) + \left(\frac{\partial AUC_j}{\partial b_j}\right)^2 \text{var}(\hat{b}_j) + 2\left(\frac{\partial AUC_j}{\partial a_j}\right)\left(\frac{\partial AUC_j}{\partial b_j}\right) \text{cov}(\hat{a}_j, \hat{b}_j) \quad (6.23)$$

όπου:

$$\frac{\partial(AUC_j)}{\partial a_j} = \left(\frac{1}{1-b_j}\right) \exp\left(\frac{a_j}{1-b_j}\right) \int_0^1 \frac{\left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)}}{\left(1 + \left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)} \exp\left(\frac{a_j}{1-b_j}\right)\right)^2} dx \quad (6.24)$$

$$\frac{\partial(AUC_j)}{\partial b_j} = \left(\frac{1}{1-b_j}\right)^2 \exp\left(\frac{a_j}{1-b_j}\right) \int_0^1 \frac{\left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)} \left(a_j + 2\ln\left(\frac{x}{1-x}\right)\right)}{\left(1 + \left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)} \exp\left(\frac{a}{1-b_j}\right)\right)^2} dx \quad (6.25)$$

Στην ομογενή περίπτωση ισχύει ότι $b_j=0$, και η γενική έκφραση της εξίσωσης 6.22 γίνεται (Walter 2002):

$$AUC_{\text{hom}_j} = \frac{DOR_j}{(DOR_j - 1)^2} \left[(DOR_j - 1) - \ln(DOR_j) \right] \quad (6.26)$$

όπου η AUC_{hom_j} δηλώνει την AUC_j για τις ομοιογενείς μελέτες, και ισχύει ότι

$$DOR_j = \exp(a_j) \text{ . Εάν } a_j = 0, \text{ τότε } AUC_{\text{hom}_j} = \frac{1}{2} \text{ . Αν και η σχέση που περιγράφηκε}$$

ισχύει μόνο για τις ομοιογενείς μελέτες, αποτελεί το άνω όριο και μια καλή προσέγγιση της AUC_j σε μελέτες με ετερογένεια. Χρησιμοποιώντας τη μέθοδο Δέλτα μπορεί επίσης να υπολογιστεί ότι:

$$SE(\hat{AUC}_{\text{hom}_j}) = \frac{1}{(DOR_j - 1)^3} \left[(DOR_j + 1) \ln DOR_j - 2(DOR_j - 1) \right] SE(\hat{DOR}_j) \quad (6.27)$$

Τέλος, μπορεί να κατασκευαστεί ένας επίσημος έλεγχος της ισότητας των AUC των δυο διαγνωστικών δοκιμασιών (δηλαδή $AUC_1 = AUC_2$) χρησιμοποιώντας τις ιδιότητες της AUC (Walter 2002) και τη μέθοδο Δέλτα. Έτσι, θέτουμε την ακόλουθη μηδενική υπόθεση $H_0 : d = AUC_1 - AUC_2 = 0$, $H_a : d \neq 0$. Η διασπορά του \hat{d} είναι ίση με \mathbf{GVG}' όπου \mathbf{V} είναι ο εκτιμώμενος πίνακας διακυμάνσεων-συνδιακυμάνσεων και \mathbf{G} ο πίνακας των παραγώγων του d ως προς το διάνυσμα των εκτιμώμενων συντελεστών a_j, b_j $\left[\frac{\partial(d)}{\partial a_1} \quad \frac{\partial(d)}{\partial b_1} \quad \frac{\partial(d)}{\partial a_2} \quad \frac{\partial(d)}{\partial b_2} \right]$. Συνεπώς το \mathbf{GVG}' ισούται με:

$$\begin{aligned}
\mathbf{GVG}' = & \frac{\partial(d)}{\partial a_1} \left\{ \frac{\partial(d)}{\partial a_1} \text{var}(\hat{a}_1) + \frac{\partial(d)}{\partial \beta_1} \text{cov}(\hat{a}_1, \hat{b}_1) + \frac{\partial(d)}{\partial a_2} \text{cov}(\hat{a}_1, \hat{a}_2) + \frac{\partial(d)}{\partial \beta_2} \text{cov}(\hat{a}_1, \hat{b}_2) \right\} \\
& + \frac{\partial(d)}{\partial \beta_1} \left\{ \frac{\partial(d)}{\partial a_1} \text{cov}(\hat{a}_1, \hat{b}_1) + \frac{\partial(d)}{\partial \beta_1} \text{var}(\hat{b}_1) + \frac{\partial(d)}{\partial a_2} \text{cov}(\hat{a}_2, \hat{b}_1) + \frac{\partial(d)}{\partial \beta_2} \text{cov}(\hat{b}_1, \hat{b}_2) \right\} \\
& + \frac{\partial(d)}{\partial a_2} \left\{ \frac{\partial(d)}{\partial a_1} \text{cov}(\hat{a}_1, \hat{a}_2) + \frac{\partial(d)}{\partial \beta_1} \text{cov}(\hat{a}_2, \hat{b}_1) + \frac{\partial(d)}{\partial a_2} \text{var}(\hat{a}_2) + \frac{\partial(d)}{\partial \beta_2} \text{cov}(\hat{a}_2, \hat{b}_2) \right\} \\
& + \frac{\partial(d)}{\partial \beta_2} \left\{ \frac{\partial(d)}{\partial a_1} \text{cov}(\hat{a}_1, \hat{b}_2) + \frac{\partial(d)}{\partial \beta_1} \text{cov}(\hat{b}_1, \hat{b}_2) + \frac{\partial(d)}{\partial a_2} \text{cov}(\hat{a}_2, \hat{b}_2) + \frac{\partial(d)}{\partial \beta_2} \text{var}(\hat{b}_2) \right\}
\end{aligned} \tag{6.28}$$

με μερικές παραγώγους για κάθε διαγνωστική j δοκιμασία ($j=1,2$) να δίνονται από:

$$\frac{\partial(d)}{\partial a_j} = (-1)^{j+1} \left(\frac{1}{1-b_j} \right) \exp \left(\frac{\alpha_j}{1-b_j} \right) \int_0^1 \frac{\left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)}}{\left(1 + \left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)} \exp \left(\frac{a_j}{1-b_j} \right) \right)^2} dx
\tag{6.29}$$

$$\frac{\partial(d)}{\partial b_j} = (-1)^{j+1} \left(\frac{1}{1-b_j} \right)^2 \exp \left(\frac{a_j}{1-b_j} \right) \int_0^1 \frac{\left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)} \left(a_j + 2 \ln \left(\frac{x}{1-x} \right) \right)}{\left(1 + \left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)} \exp \left(\frac{a}{1-b_j} \right) \right)^2} dx
\tag{6.30}$$

Στην ομογενή περίπτωση (δηλαδή $b_j=0$) η γενική έκφραση της εξίσωσης 6.28 για την εκτίμηση της διασποράς του \hat{d} απλοποιείται αφού ακυρώνονται οι όροι που περιλαμβάνουν τα b_j . Σε αυτή την περίπτωση, η διασπορά του \hat{d}_{hom} υπολογίζεται ως ακολούθως:

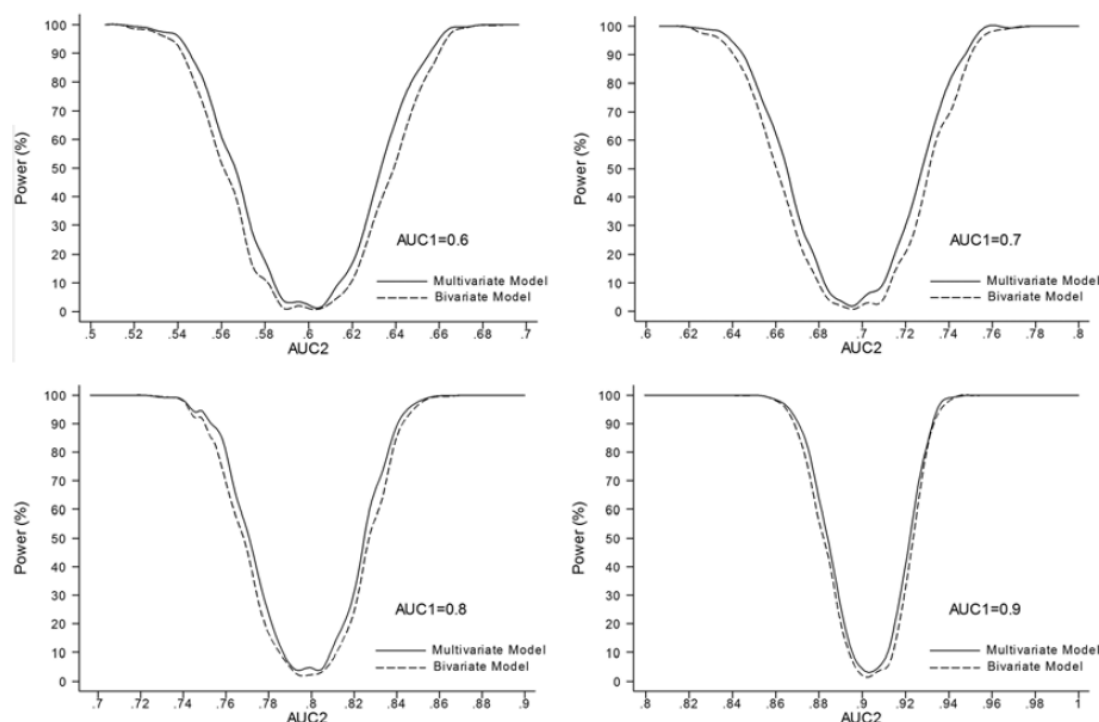
$$\text{var}(\hat{d}_{\text{hom}}) = \left(\frac{\partial d_{\text{hom}}}{\partial a_1} \right)^2 \text{var}(\hat{a}_1) + \left(\frac{\partial d_{\text{hom}}}{\partial a_2} \right)^2 \text{var}(\hat{a}_2) + 2 \left(\frac{\partial d_{\text{hom}}}{\partial a_1} \right) \left(\frac{\partial d_{\text{hom}}}{\partial a_2} \right) \text{cov}(\hat{a}_1, \hat{a}_2)
\tag{6.31}$$

και οι μερικές παράγωγοι για τη j διαγνωστική δοκιμασία ($j=1,2$) υπολογίζονται ως ακολούθως:

$$\frac{\partial(d_{\text{hom}})}{\partial a_j} = (-1)^{j+1} \frac{\exp(\alpha_j) [\exp(\alpha_j)(\alpha_j - 2) + \alpha_j + 2]}{[\exp(\alpha_j) - 1]^3}
\tag{6.32}$$

6.3. Μελέτη προσομοίωσης

Για να αξιολογήσουμε την προτεινόμενη μεθοδολογία πραγματοποιήσαμε μελέτες προσομοίωσης Στο παράρτημα Z περιγράφεται η διαδικασία που ακολουθήσαμε για την προσομοίωση και τα αποτελέσματα. Συγκρίνουμε την προτεινόμενη πολυμεταβλητή μέθοδο με την υπάρχουσα διμεταβλητή προσέγγιση. Συνοπτικά, οι προσομοιώσεις υποδεικνύουν ότι σε σενάρια με 20 μελέτες και 200 άτομα σε κάθε μελέτη, η πολυμεταβλητή μέθοδος σε σχέση με τη διμεταβλητή προσέγγιση οδηγεί σε αμερόληπτες εκτιμήσεις και έχει μεγαλύτερη ισχύ να ανιχνεύσει μικρές διαφορές στην ευαισθησία και ειδικότητα διατηρώντας σταθερό το σφάλμα τύπου I στο 5%. Επιπλέον, η πολυμεταβλητή μέθοδος έχει πάντα μεγαλύτερη ισχύ (έως και 12%) να ανιχνεύσει διαφορές στην \widehat{AUC} (Εικόνα 6.1). Τα πλεονεκτήματα της μεθοδολογίας ήταν λιγότερο εμφανή σε σχεδιασμούς με μεγάλο αριθμό συμμετεχόντων (>500) σε κάθε μελέτη με αποτελέσματα και για τις δυο διαγνωστικές δοκιμασίες (χωρίς ελλείπουσες παρατηρήσεις) και όταν υπήρχαν μεγάλες διαφορές είτε στην ευαισθησία είτε στην ειδικότητα.



Εικόνα 6.1. Υπολογισμοί της ισχύος για να ανιχνευτούν μεταβολές στη διαφορά των $\widehat{AUC}_1 - \widehat{AUC}_2$ για την πολυμεταβλητή και διμεταβλητή μεθοδολογία μετα-ανάλυσης διαγνωστικών δοκιμασιών. Πραγματοποιήσαμε 500 επαναλήψεις με 20 μελέτες σε κάθε μετα-ανάλυση. Ο επιπολασμός της νόσου ήταν 50% με 200 άτομα σε κάθε μελέτη. Για λόγους ευκολίας, χρησιμοποιήσαμε την αναλυτική έκφραση της εξίσωσης 6.26 έτσι ώστε να υπολογίσουμε την AUC υποθέτοντας ότι οι μελέτες μας ήταν ομοιογενείς.

6.4. Εφαρμογή της μεθοδολογίας

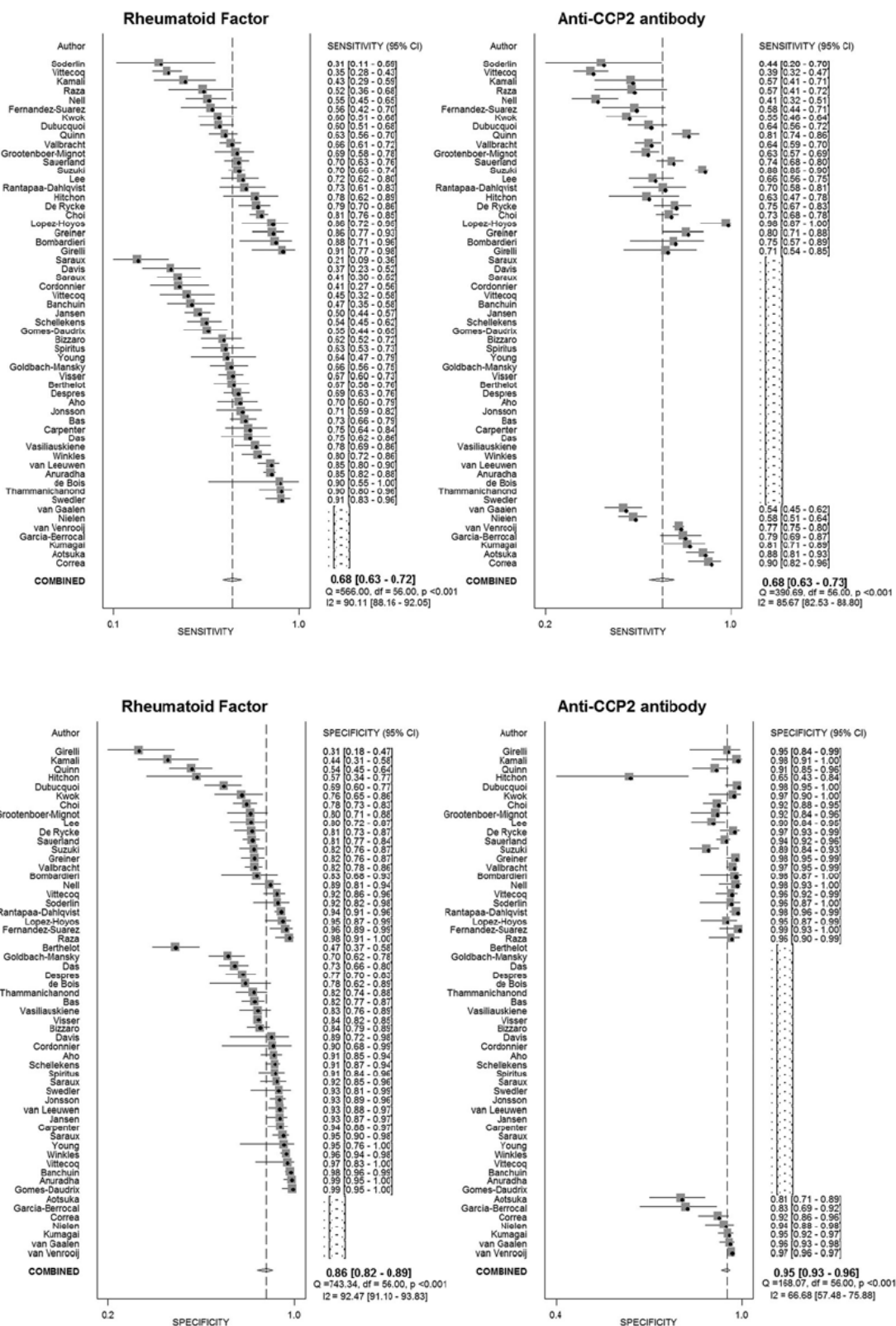
Η μέθοδος εφαρμόστηκε σε δεδομένα μιας μετα-ανάλυσης που είχε ως σκοπό να δείξει εάν το αντίσωμα ‘Anti–Cyclic Citrullinated Peptide Antibody’ (anti-CCP antibody) προσδιορίζει με μεγαλύτερη ακρίβεια ασθενείς με ρευματοειδή αρθρίτιδα (Rheumatoid Arthritis-RA) από ότι ο ρευματοειδής παράγοντας (Rheumatoid Factor-RF) (Nishimura, Sugiyama et al. 2007) (Παράρτημα Η). Πιο συγκεκριμένα, χρησιμοποιήθηκαν τα anti-CCP2 αντισώματα. Ένα σύνολο 50 και 29 μελετών είχαν δεδομένα για τον RF ή το αντίσωμα anti-CCP2 αντίστοιχα. Είκοσι δύο μελέτες αξιολογούσαν ταυτόχρονα και τις δυο διαγνωστικές δοκιμασίες.

Από τις μελέτες που αξιολογούσαν ταυτόχρονα και τις δυο διαγνωστικές δοκιμασίες, δεδομένα για τον υπολογισμό των κελιών του Πίνακα 6.1 ήταν διαθέσιμα για 2 μελέτες. Ένα σύνολο 6 μελετών είχαν εκτιμήσεις για την ευαισθησία και ειδικότητα όταν και οι δυο διαγνωστικές δοκιμασίες ήταν θετικές, ή όταν τουλάχιστον μια ήταν θετική. Σε αυτή την περίπτωση, δείχνουμε ότι τα κελιά του Πίνακα 6.1 μπορούν να υπολογιστούν αναλυτικά (Παράρτημα Θ). Μια μελέτη είχε πληροφορίες για την ευαισθησία, την ειδικότητα, τη θετική και αρνητική προγνωστική αξία για άτομα με αρνητικό αποτέλεσμα σε μια από τις δοκιμασίες (η δοκιμασία όπου χρησιμοποιείται ο RF στη συγκεκριμένη περίπτωση) και κατασκευάστηκαν αντίστοιχες σχέσεις για τον υπολογισμό των κελιών του Πίνακα 6.1 (Παράρτημα Θ). Ο συνολικός αριθμός των συμμετεχόντων και για τις δυο δοκιμασίες δεν ήταν ίσος σε μια μελέτη και συνεπώς επαυξήσαμε τα λιγότερα άτομα της μιας δοκιμασίας ώστε να εξισωθούν με τα περισσότερα άτομα της άλλης.

Τέλος, τα ακριβή δεδομένα για τα κελιά του Πίνακα 6.1 ήταν διαθέσιμα για έξι μελέτες μόνο όμως για ασθενείς. Δεν υπήρχαν δεδομένα για 7 μελέτες που αξιολογούσαν ταυτόχρονα και τις δυο διαγνωστικές δοκιμασίες και τα οποία έπρεπε να εκτιμηθούν. Αρχικά, υπολογίστηκε μια συνδυασμένη εκτίμηση του conditional Odds Ratio για τη συσχέτιση των δυο διαγνωστικών δοκιμασιών, για τους υγιείς και ασθενείς αντίστοιχα ($\widehat{OR}_{Y_i=0} = 6.537$ με $I^2=52.7\%$ και $\tau^2=0.7817$, $\widehat{OR}_{Y_i=1} = 12.112$ με $I^2=52.8\%$ και $\tau^2=0.2298$ αντίστοιχα) πραγματοποιώντας μονομεταβλητή μετα-ανάλυση τυχαίων επιδράσεων (DerSimonian and Laird 1986). Οι εκτιμήσεις αυτές διαφέρουν σημαντικά, υποδεικνύοντας ότι υπάρχει αλληλεπίδραση τρίτης-τάξης στον πίνακα συνάφειας. Συνεπώς, μπορούμε να

υποθέσουμε ότι οι εκτιμήσεις αυτές είναι κοντά στις πραγματικές εκτιμήσεις των αντίστοιχων Odds Ratios των μελετών που δεν έδιναν αυτή την πληροφορία. Ακολούθως, χρησιμοποιώντας τις εξισώσεις που περιγράφονται στο Παράρτημα ΣΤ μπορούμε να υπολογίσουμε τα κελιά του Πίνακα 6.1. Για την εκτίμηση των παραμέτρων του μοντέλου, χρησιμοποιήθηκε η εντολή `mnmeta` (White 2009) στο Stata (<http://www.compgen.org/tools/meta-roc>) χωρίς να υποθέσουμε κάποια δομή για τα στοιχεία του πίνακα Σ . Για λόγους σύγκρισης, εκτιμήθηκαν επίσης οι παράμετροι του διμεταβλητού μοντέλου.

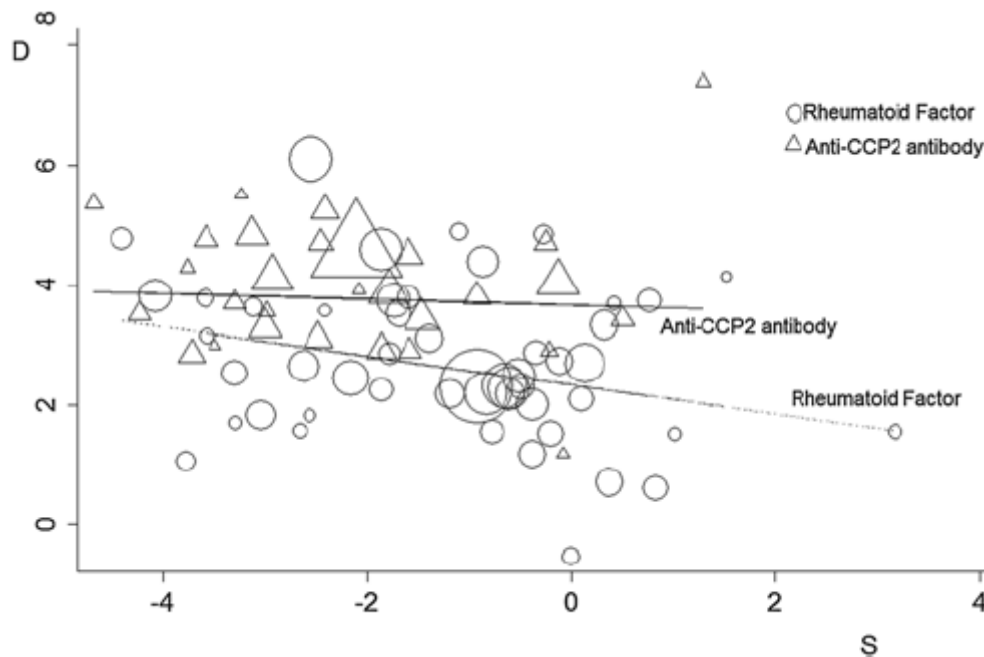
Δεν παρατηρήθηκαν στατιστικά σημαντικές διαφορές μεταξύ των \widehat{TPR} των δυο διαγνωστικών δοκιμασιών (Wald test: $z=-0.20$; $p\text{-value}=0.834$, Πίνακας 6.3). Το αντίσωμα anti-CCP2 ωστόσο, είχε χαμηλότερο \widehat{FPR} και συνεπώς μεγαλύτερη ευαισθησία σε σχέση με τον RF (Wald test: $z=5.57$; $p\text{-value}<0.001$) (Εικόνα 6.2). Από τις 22 μελέτες που αξιολογούσαν ταυτόχρονα και τις δυο διαγνωστικές δοκιμασίες, οι τιμές εντός των μελετών συντελεστών συσχέτισης των \widehat{TPR} των δυο διαγνωστικών δοκιμασιών (δηλαδή $\hat{\rho}_{w13i}$) κυμαίνονταν από 0.148-0.684 με μέση τιμή 0.463 ενώ οι αντίστοιχες των \widehat{FPR} (δηλαδή $\hat{\rho}_{w24i}$) κυμαίνονταν από 0-0.669 με μέση τιμή ίση με 0.197. Οι μεταξύ των μελετών συσχετίσεις των \widehat{TPR} και \widehat{FPR} των δυο διαγνωστικών δοκιμασιών ήταν $\hat{\rho}_{B13} = 0.779$ και $\hat{\rho}_{B24} = 0.410$. Επιπλέον, χρησιμοποιώντας τις εκτιμήσεις των μεταξύ των μελετών συνδιακυμάνσεων, υπολογίστηκαν οι παράμετροι της παλινδρόμησης των D_{ji} ως προς S_{ji} καθώς επίσης και η καμπύλη SROC (Arends, Hamza et al. 2008) με $\hat{b}_1 = -0.241$ (95% CI: -0.505, 0.022), $\hat{b}_2 = -0.049$ (95% CI: -0.344, 0.246), $\hat{a}_1 = 2.331$ (95% CI: 1.901, 2.761) και $\hat{a}_2 = 3.653$ (95% CI: 2.984, 4.323) (Εικόνα 6.3). Η \widehat{AUC} για τον RF και το αντίσωμα Anti-CCP2 ήταν 0.825 και 0.927 αντίστοιχα με αυτές τις τιμές να διαφέρουν στατιστικά σημαντικά, υποδεικνύοντας ότι συνολικά το αντίσωμα anti-CCP2 antibody έχει καλύτερη διαγνωστική ακρίβεια σε σχέση με τον RF ($z=-3.54$; $p\text{-value}<0.001$) (Εικόνα 6.4). Πρέπει επίσης να αναφέρουμε, ότι η μη ορθή εκτίμηση του μοντέλου (δηλαδή αγνοώντας τις διαφορές στα conditional Odds Ratios σε ασθενείς και υγιείς), οδηγεί σε μη σημαντικές διαφορές στις συνολικές εκτιμήσεις ή στα συμπεράσματα σχετικά με την υπεροχή του αντισώματος anti-CCP2.



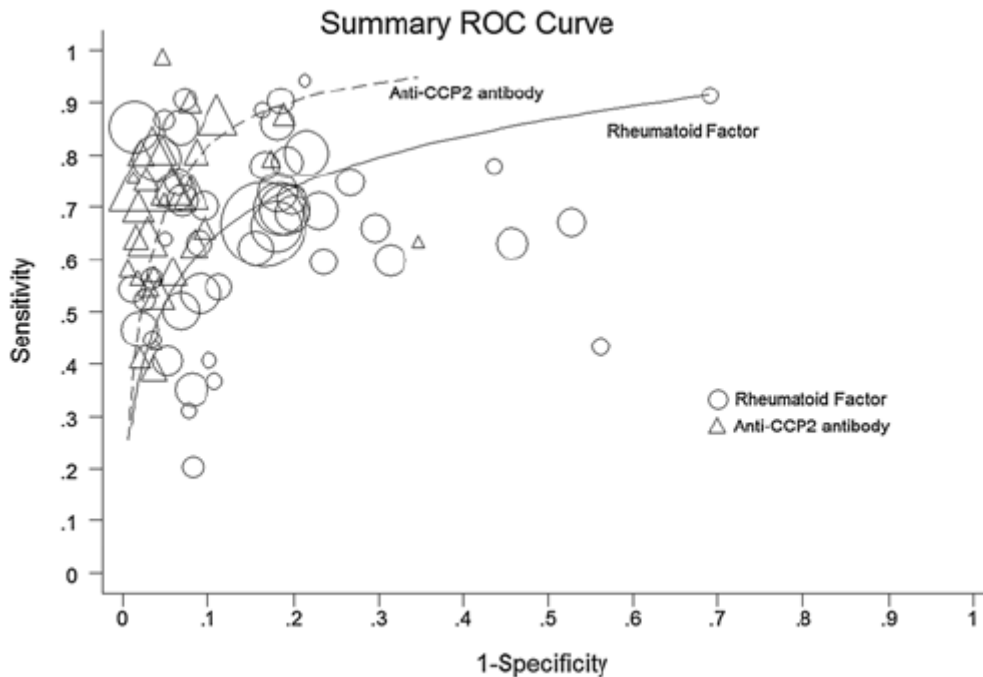
Εικόνα 6.2. Forest plots της ευαισθησίας και ειδικότητας των δυο διαγνωστικών δοκιμασιών όπως εκτιμώνται από την πολυμεταβλητή μετα-ανάλυση.

Πίνακας 6.3. Εκτιμήσεις του πολυμεταβλητού και διμεταβλητού μοντέλου των \widehat{TPR}_{ji} και \widehat{FPR}_{ji} (logits) των δυο διαγνωστικών δοκιμασιών ($\hat{\beta}_1-\hat{\beta}_2$ για τον RF και $\hat{\beta}_3-\hat{\beta}_4$ για το αντίσωμα anti-CCP2).

	Πολυμεταβλητό μοντέλο			Διμεταβλητό μοντέλο		
	Mean (95% CI)	z	$\hat{\tau}^2$	Mean (95% CI)	z	$\hat{\tau}^2$
$\hat{\beta}_1$	0.750 (0.538, 0.962)	6.94	0.537	0.708 (0.491, 0.925)	6.39	0.534
$\hat{\beta}_2$	-1.845 (-2.141, -1.550)	-12.23	0.977	-1.876 (-2.173, -1.579)	-12.37	0.990
$\hat{\beta}_3$	0.774 (0.534, 1.015)	6.31	0.449	0.823 (0.571, 1.075)	6.41	0.406
$\hat{\beta}_4$	-2.986 (-3.310, -2.663)	-18.08	0.519	-2.909 (-3.227, -2.590)	-17.90	0.527
$\hat{\rho}_{B12}=0.212, \hat{\rho}_{B13}=0.779, \hat{\rho}_{B14}=0.386,$						
$\hat{\rho}_{B23}=0.194, \hat{\rho}_{B24}=0.410$						
$\hat{\rho}_{B34}=0.487$						
				$\hat{\rho}_{B12}=0.220, \hat{\rho}_{B34}=0.477$		



Εικόνα 6.3. Γράφημα της διαφοράς των logits των \widehat{TPR}_{ji} και \widehat{FPR}_{ji} (\hat{D}_{ji}) ως προς το άθροισμά τους (\hat{S}_{ji}) για τις δυο διαγνωστικές δοκιμασίες με την SROC μέθοδο.



Εικόνα 6.4. Η καμπύλη SROC όπως προκύπτει από το πολυμεταβλητό μοντέλο μετα-ανάλυσης για τις δυο διαγνωστικές δοκιμασίες ($\widehat{AUC}=0.825$ και 0.927 για τον RF και το αντίσωμα anti-CCP2 αντίστοιχα).

6.5. Συζήτηση

Περιγράψαμε μια απλή αλλά με μεγάλη ισχύ μέθοδο μετα-ανάλυσης και σύγκρισης των διαγνωστικών δοκιμασιών επεκτείνοντας την υπάρχουσα διμεταβλητή προσέγγιση (Reitsma, Glas et al. 2005; Arends, Hamza et al. 2008). Το σημείο κλειδί της μεθόδου είναι ότι μπορούν να υπολογιστούν οι εντός των μελετών συνδιακυμάνσεις σε κλειστή μορφή χρησιμοποιώντας μια υπάρχουσα μεθοδολογία (Bagos 2012). Η προτεινόμενη μεθοδολογία έχει όλα τα πλεονεκτήματα του διμεταβλητού μοντέλου; δηλαδή, λαμβάνονται υπόψη οι μεταξύ των μελετών συσχετίσεις της ευαισθησίας και ειδικότητας, υπολογίζονται οι σημειακές εκτιμήσεις μαζί με τα διαστήματα εμπιστοσύνης και μπορεί να κατασκευαστεί η καμπύλη SROC. Παράμετροι όπως το *DOR*, η *AUC* και τα πηλίκα πιθανοφάνειας μπορούν να υπολογιστούν και να συγκριθούν για τις δυο ή περισσότερες διαγνωστικές δοκιμασίες. Κυριότερα, η μέθοδος μπορεί να συνυπολογίσει μελέτες που έχουν δεδομένα μόνο για μια κάποιες και όχι απαραίτητα για όλες τις διαγνωστικές δοκιμασίες, δανειζόμενη ισχύ από εξωτερικές μελέτες (Higgins and Whitehead 1996), μπορεί να υλοποιηθεί εύκολα χρησιμοποιώντας ένα οποιοδήποτε στατιστικό πακέτο πολυμεταβλητής μετα-ανάλυσης (Jackson, Riley et al. 2011; Mavridis and Salanti

2012), και επιτρέπει την απευθείας σύγκριση των εκτιμώμενων παραμέτρων των διαγνωστικών δοκιμασιών χρησιμοποιώντας επίσημους ελέγχους (Wald test). Ως ένα κλασσικό πολυμεταβλητό μοντέλο, μπορεί να συμπεριλάβει συμμεταβλητές που πιθανώς να συσχετίζονται με τη νόσο, τη διαδικασία της δοκιμής ή να λαμβάνουν υπόψη ομάδες με διαφορετικό επιπολασμό της νόσου (Thompson and Sharp 1999) έτσι ώστε να ερμηνευτεί η παρατηρούμενη ετερογένεια.

Οι προσομοιώσεις δείχνουν ότι αξιολογώντας ρεαλιστικά σενάρια, η προτεινόμενη μεθοδολογία οδηγεί σε αμερόληπτες εκτιμήσεις και διατηρεί το σφάλμα τύπου I στο 5%, ενώ έχει μεγαλύτερη ισχύ να ανιχνεύσει μικρές διαφορές στην ευαισθησία και ειδικότητα. Όταν η μετα-ανάλυση περιλαμβάνει μεγάλο αριθμό μελετών, με την κάθε μια μελέτη να έχει μεγάλο αριθμό συμμετεχόντων, όταν αξιολογούνται σε κάθε μελέτη και οι δυο διαγνωστικές δοκιμασίες και όταν υπάρχουν μεγάλες διαφορές στην ευαισθησία και ειδικότητα, περιμένουμε όχι τόσο σημαντικές διαφορές από τη χρήση της προτεινόμενης μεθοδολογίας. Ωστόσο, δεν περιμένουμε σε καμία περίπτωση να είναι χειρότερη σε σχέση με την διμεταβλητή προσέγγιση. Επιπλέον, οι προσομοιώσεις στις περιπτώσεις όπου δεν εκτιμάται ορθά το μοντέλο (δηλαδή υποθέτοντας ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης χρησιμοποιώντας τον αλγόριθμο IPF), έδειξαν ότι η προτεινόμενη μεθοδολογία είναι καλύτερη σε σχέση με την κλασσική διμεταβλητή προσέγγιση. Οι υπολογισμοί αυτοί συμφωνούν με παλαιότερες προσομοιώσεις σχετικά με την υπεροχή της διμεταβλητής μετα-ανάλυσης σε σχέση με την μονομεταβλητή προσέγγιση αλλά έχουν και θεωρητική αιτιολόγηση (Riley, Abrams et al. 2007; Riley, Abrams et al. 2007; Riley 2009). Αλγεβρικοί υπολογισμοί, αποκαλύπτουν ότι για \widehat{TPR} και $\widehat{FPR} > 0.5$, η συνδιακύμανση είναι πάντα θετική, και έτσι περιμένουμε ότι η διασπορά των διαφορών $(\hat{\beta}_1 - \hat{\beta}_3, \hat{\beta}_2 - \hat{\beta}_4)$ να είναι πάντα μικρότερη όταν λαμβάνεται υπόψη η συνδιακύμανση. Ωστόσο, η συνδιακύμανση είναι μικρή σε σχέση με τη διασπορά και έτσι η συνεισφορά της μικραίνει όσο αυξάνεται το μέγεθος του δείγματος ή μεγαλώνει η μεταξύ των μελετών ετερογένεια. Όλα αυτά μαζί, εξηγούν το γεγονός ότι δεν παρατηρούμε υπερβολικά μεγάλες διαφορές (τουλάχιστον για το εύρος των τιμών που ενδιαφερόμαστε), αλλά επίσης παρέχουν την ασφάλεια ότι το πολυμεταβλητό μοντέλο είναι τουλάχιστον τόσο ακριβές όσο και το διμεταβλητό κάτω από οποιεσδήποτε συνθήκες. Η ίδια λογική ισχύει και για την ισχύ να ανιχνεύσουμε διαφορές στην \widehat{AUC} . Όταν οι πραγματικές διαφορές μεταξύ των διαγνωστικών

δοκιμασιών είναι μεγάλες και οι δυο μέθοδοι έχουν 100% ισχύ να τις ανιχνεύσουν, αλλά η πολυμεταβλητή προσέγγιση έχει πάντα μεγαλύτερη ισχύ (έως και 12%) να ανιχνεύσει μικτές διαφορές που μπορεί να έχουν κλινική σημασία.

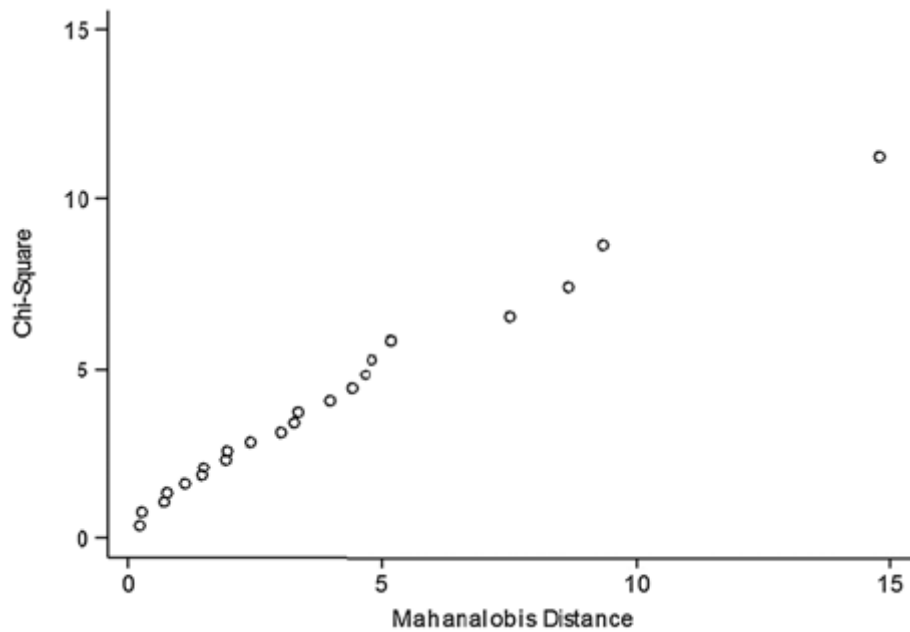
Διάφορες μεθοδολογίες έχουν προταθεί για τη σύγκριση διαγνωστικών δοκιμασιών σε μια μετα-ανάλυση. Αναφορικά με τα *DORs* (Glas, Lijmer et al. 2003), η απλή σύγκριση των μετασχηματισμών αυτών σε κλίμακα των logits ή αυτών που προκύπτουν από τη μέθοδο SROC, αγνοεί τη συσχέτιση μεταξύ των διαγνωστικών δοκιμασιών στην περίπτωση όπου εξετάζονται τα ίδια άτομα. Όταν οι δοκιμασίες εφαρμόζονται σε διαφορετικούς πληθυσμούς, η μέθοδος είναι ορθή αλλά παρέχει την εκτίμηση του *DOR* και όχι της ευαισθησίας και ειδικότητας. Η μέθοδος CROR (Suzuki, Moro-oka et al. 2004) είναι απλή αλλά απαιτεί ατομικά δεδομένα που συχνά δεν είναι συχνά διαθέσιμα. Συνδυαστικά με τη μεθοδολογία που προτάθηκε, αποτελεί το γεγονός ότι ο τρόπος που παρουσιάστηκε για την ανακατασκευή των πινάκων (Παράρτημα Θ), μπορεί να χρησιμοποιηθεί ευθέως στη μέθοδο CROR. Συνεπώς, μια μέθοδος που χρησιμοποιήθηκε αρχικά για ατομικά δεδομένα μπορεί να χρησιμοποιηθεί για συνοπτικά δεδομένα που συλλέγονται από τη βιβλιογραφία. Η συγκεκριμένη προσέγγιση, έχει το πλεονέκτημα ότι υλοποιείται εύκολα καθώς δεν απαιτεί παρά μόνο ένα συνηθισμένο στατιστικό πακέτο για μονομεταβλητή μετα-ανάλυση. Ως απόδειξη των όσων αναφέρθηκαν, από τα πρωταρχικά δεδομένα, 9 μελέτες παρείχαν τις απαραίτητες πληροφορίες για τον υπολογισμό του CROR. Ο συνδυασμός των μελετών αυτών, οδηγεί στην εκτίμηση του CROR να ισούται με 0.144 (95% C.I.: 0.066, 0.317) για τον RF σε σχέση με το αντίσωμα anti-CCP2. Ωστόσο, χρησιμοποιώντας τις εξισώσεις του παραρτήματος Θ, καταφέραμε να ανακατασκευάσουμε τον πίνακα για επιπλέον 13 μελέτες, κάτι που μας επιτρέπει να υπολογίσουμε μια πιο ακριβή εκτίμηση του CROR ίσης με 0.152 (95% C.I.: 0.090, 0.256), υποδεικνύοντας υπεροχή του αντισώματος anti-CCP2 σε σχέση με τον RF. Ωστόσο, η μέθοδος αυτή έχει ένα επιπλέον μειονέκτημα, ότι δηλαδή μπορεί να λάβει υπόψη μελέτες που αναφέρουν αποτελέσματα για μια διαγνωστική δοκιμασία και έτσι δεν επιτρέπεται η απευθείας σύγκριση της ευαισθησίας, της ειδικότητας ή της AUC. Αξίζει επίσης να σχολιάσουμε τη μπεύζιανή προσέγγιση που παρουσιάστηκε από τον Trikalinos και συνεργάτες (Trikalinos, Hoaglin et al. 2014), μια προσέγγιση που μοιράζεται πολλά κοινά χαρακτηριστικά με την προτεινόμενη μεθοδολογία. Η κύρια διαφορά εκτός από τη μπεύζιανή θεώρηση εντοπίζεται στο ότι ο Trikalinos και συνεργάτες μοντελοποιούν τη συσχέτιση των δυο διαγνωστικών δοκιμασιών ως

τυχαία παράμετρο, ενώ στην προσέγγιση που ακολουθήσαμε χρησιμοποιούμε απευθείας τις πραγματικές συσχετίσεις της κάθε μελέτης. Όταν όλες οι μελέτες αναφέρουν τα δεδομένα για την ταξινόμηση των ατόμων οι δυο προσεγγίσεις θα οδηγούν σε παρόμοιες εκτιμήσεις. Συνεπώς, το κύριο πλεονέκτημα της προσέγγισής μας είναι ότι είναι πιο απλό και ότι μπορεί να υλοποιηθεί με οποιοδήποτε στατιστικό πακέτο χωρίς να χάνει την ακρίβειά του. Ένα επιπλέον πλεονέκτημα αποτελεί το γεγονός ότι, όπως έχει ήδη αναφερθεί, έχει όλα τα πλεονεκτήματα του κλασσικού διμεταβλητού μοντέλου και επιπλέον επιτρέπει ελέγχους για τις παραμέτρους της καμπύλης SROC και της AUC.

Πρέπει να τονίσουμε, ότι η προτεινόμενη μεθοδολογία χρησιμοποιείται για τη σύγκριση διαγνωστικών δοκιμασιών σε περιπτώσεις όπου υπάρχει μέθοδος αναφοράς. Από την άλλη πλευρά, έχουν προταθεί κάποιες άλλες μέθοδοι για την από κοινού μοντελοποίηση της ευαισθησίας και ειδικότητας δυο διαγνωστικών δοκιμασιών μαζί με τον επιπολασμό της νόσου, στις περιπτώσεις όπου δεν υπάρχει μέθοδος αναφοράς (Chu, Chen et al. 2009; Dendukuri, Schiller et al. 2012). Οι μέθοδοι αυτές, βασίζονται στην υπόθεση της conditional independence (Hui and Zhou 1998; Enoe, Georgiadis et al. 2000; Gardner, Stryhn et al. 2000), που είναι αρκετά πιο ισχυρή σε σχέση με τη δική μας προσέγγιση. Όταν η υπόθεση αυτή δεν ικανοποιείται, μπορεί να χρησιμοποιηθεί μια συσχέτιση μεταξύ των διαγνωστικών δοκιμασιών. Ωστόσο η προτεινόμενη μεθοδολογία είναι πιο άμεση καθώς υπολογίζει απευθείας τη συσχέτιση από την από κοινού κατανομή των δυο διαγνωστικών δοκιμασιών ανάλογα με την παρουσία ή απουσία της νόσου. Οι υποθέσεις που θέτουμε για το μηχανισμό των δεδομένων που λείπουν είναι καθοριστικές καθώς όσο αυξάνεται ο αριθμός των διαγνωστικών δοκιμασιών που εφαρμόζονται, η υπόθεση ότι τα δεδομένα αυτά λείπουν λόγω τύχης μπορεί να ισχύει μόνο σε συγκεκριμένες περιπτώσεις, αποκλείοντας προφανώς τις περιπτώσεις όπου η σειρά εφαρμογής των δοκιμασιών εξαρτάται από την κατάσταση της υγείας. Δεδομένου ότι χρησιμοποιήσαμε το κλασσικό πολυμεταβλητό μοντέλο μετα-ανάλυσης, λαμβάνονται υπόψη περιπτώσεις όπου τα δεδομένα λείπουν εντελώς λόγω τύχης (missing completely at random) ή απλώς λόγω τύχης (completely at random) (Riley, Abrams et al. 2007; Jackson, Riley et al. 2011). Γενικώς, μπορεί να κατασκευαστούν απλοί διαγνωστικοί έλεγχοι για τη περίπτωση όπου τα δεδομένα λείπουν για κάποιο συγκεκριμένο λόγο, για παράδειγμα πραγματοποιώντας αναλύσεις υποομάδων για τις

μελέτες που περιγράφουν και τις δυο διαγνωστικές δοκιμασίες έναντι αυτών που περιγράφουν μόνο τη μια.

Η προτεινόμενη μεθοδολογία χρησιμοποιεί μόνο συνοπτικά δεδομένα τα οποία είναι διαθέσιμα στις δημοσιευμένες μελέτες. Η θεωρία στην οποία βασίζεται η μέθοδος είναι πολύ απλή και στηρίζεται σε προσεγγίσεις μεγάλου μεγέθους δείγματος και κανονικότητας (Clayton and Hills 1993; Agresti 2002) που χρησιμοποιούνται καθημερινά από τους ερευνητές που πραγματοποιούν μετα-αναλύσεις. Στο παράδειγμα των δυο διαγνωστικών δοκιμασιών που χρησιμοποιούνται στον εντοπισμό των ατόμων με ρευματοειδή αρθρίτιδα, ο έλεγχος Henze-Zirkler (Henze and Zirkler 1990) χρησιμοποιώντας την εντολή `multnorm` στο Stata (Stevens 1986; Thompson 1990), δεν παρείχε ενδείξεις για παραβίαση της υπόθεσης της πολυμεταβλητής κανονικής κατανομής (Εικόνα 6.5). Σε κάποιες ακραίες περιπτώσεις μπορεί να ανακύπτουν πρακτικά ζητήματα από τη χρήση της προσεγγιστικής πιθανοφάνειας (Hamza, Reitsma et al. 2008; Hamza, van Houwelingen et al. 2008). Θεωρητικά, η μέθοδος αναμένεται να αποτυγχάνει όταν το μέγεθος δείγματος είναι πολύ μικρό ή όταν υπάρχουν πολλές μελέτες με μικρά ή μηδενικά κελιά. Υπό αυτές τις καταστάσεις, βασιζόμαστε σε διορθώσεις κανονικότητας που περιλαμβάνουν την πρόσθεση $\frac{1}{2}$ στις συχνότητες των κελιών (Agresti 1999). Μια άλλη επιλογή είναι η χρήση `multinomial` κατανομών αλλά με όσα γνωρίζουμε η `multinomial likelihood` για το συγκεκριμένο πρόβλημα δεν μπορεί να υπολογιστεί χρησιμοποιώντας την ρουτίνα `xtmelogit` στο Stata ή `lmer` στην R. Η βελτιστοποίηση της πιθανοφάνειας για την πραγματοποίηση πολυμεταβλητής μετα-ανάλυσης χρησιμοποιώντας την `multinomial likelihood` περιλαμβάνει τον υπολογισμό πολύπλοκων ολοκληρωμάτων που ξεφεύγει από τους σκοπούς της παρούσας μελέτης.



Εικόνα 6.5. Γραφικός έλεγχος της υπόθεσης της πολυμεταβλητής κανονικής κατανομής των $\text{logit}(\widehat{TPR}_{ji})$ και $\text{logit}(\widehat{FPR}_{ji})$ για τις 22 μελέτες, που αξιολογούσαν ταυτόχρονα τις δυο διαγνωστικές δοκιμασίες για την ανίχνευση των ατόμων με ρευματοειδή αρθρίτιδα.

Συνολικά, παρουσιάστηκε μια απλή μεθοδολογία με μεγάλη ισχύ για τη πραγματοποίηση μετα-ανάλυσης και σύγκρισης των αποτελεσμάτων διαγνωστικών δοκιμασιών. Η μέθοδος μπορεί να υλοποιηθεί σε οποιοδήποτε στατιστικό πακέτο ενώ κώδικας σε Stata δίνεται στην ακόλουθη διεύθυνση <http://www.compgen.org/tools/meta-roc>.

Παραρτήματα

Παράρτημα Α

Για τον υπολογισμό των συνδιακυμάνσεων των εκτιμήσεων των μελετών ασθενών-μαρτύρων και του ελέγχου HHRR που εκτιμάται από τις μελέτες σε οικογένειες ισχύει ότι:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HHRR}) &= \text{cov}\left[\log\left(\frac{n_{11}n_{00}}{n_{01}n_{10}}\right), \log\left(\frac{wz}{xy}\right)\right] \\ &= \text{cov}\left[\log(n_{11}n_{00}) - \log(n_{01}n_{10}), \log(wz) - \log(xy)\right]\end{aligned}\quad (\text{A1})$$

όπου κατόπιν υπολογισμών με χρήση ιδιοτήτων των συνδιακυμάνσεων διαμορφώνεται η ακόλουθη σχέση:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HHRR}) &= \text{cov}[\log(n_{11}), \log(w)] + \text{cov}[\log(n_{10}), \log(x)] \\ &= \text{var}[\log(n_{11})] + \text{var}[\log(n_{10})] \\ &= \frac{1}{n_{11}} + \frac{1}{n_{10}}\end{aligned}\quad (\text{A2})$$

Η πρώτη ισότητα προκύπτει αν παρατηρήσουμε στον Πίνακα 3.3 ότι $n_{11}=w$ και $n_{10}=x$ (εφόσον χρησιμοποιούνται τα ίδια άτομα), ενώ οι άλλες δυο αποτελούν ιδιότητες της υπόθεσης μεγάλου μεγέθους δείγματος για τα κελιά του πίνακα συνάφειας.

Στην περίπτωση όπου μια μελέτη παρέχει εκτιμήσεις ασθενών-μαρτύρων και του ελέγχου TDT, η συνδιακύμανση γίνεται:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{TDT}) &= \text{cov}\left[\log\left(\frac{n_{11}n_{00}}{n_{01}n_{10}}\right), \log\left(\frac{b}{c}\right)\right] \\ &= \text{cov}[\log(n_{11}n_{00}) - \log(n_{01}n_{10}), \log(b) - \log(c)]\end{aligned}\quad (\text{A3})$$

όπου κατόπιν υπολογισμών ισχύει ότι:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \text{cov}[\log(n_{11}), \log(b)] + \text{cov}[\log(n_{10}), \log(c)] \quad (\text{A4})$$

Εφόσον $n_{11}=w=a+b$ και $n_{10}=x=c+d$, παρατηρούμε ότι το δεξί μέρος της εξίσωσης (A4) περιλαμβάνει όρους της μορφής $\text{cov}[\log(S+U), \log(S)]$. Παρατηρούμε επίσης ότι οι S και U είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με $E[S] = \text{var}[S] = s$ και $E[U] = \text{var}[U] = u$ και έτσι, η $Z = S + U$ είναι επίσης μια τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson με $E[Z] = \text{var}[Z] = z = s + u$. Συνεπώς, οι (Z, S) θα ακολουθούν μια διμεταβλητή κατανομή Poisson (Marshall and Olkin 1985) και η συνδιακύμανση μεταξύ των Z και S θα είναι:

$$\text{cov}[S+U, S] = \text{cov}[S, S] = \text{var}[S] = s \quad (\text{A5})$$

Η συνδιακύμανση μεταξύ των συναρτήσεων των τυχαίων μεταβλητών Z και S , απαιτεί τη γνώση των από κοινού και περιθώριων αθροιστικών συναρτήσεων κατανομής (Cuadras 2002). Ωστόσο, μπορεί εύκολα να προσεγγιστεί χρησιμοποιώντας τη μέθοδο Δέλτα (Agresti 2002):

$$\text{cov}[f(Z), g(S)] \approx \frac{\partial f(E[Z])}{\partial Z} \frac{\partial g(E[S])}{\partial S} \text{cov}(Z, S) \quad (\text{A6})$$

Συνεπώς προκύπτει ότι:

$$\begin{aligned} \text{cov}[\log(Z), \log(S)] &= \text{cov}(Z, S) \frac{\partial \log Z}{\partial Z} \frac{\partial \log S}{\partial S} \\ &= \frac{s}{zs} = \frac{1}{s+u} \end{aligned} \quad (\text{A7})$$

Χρησιμοποιώντας την εξίσωση (A7), ισχύει τελικά ότι:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \quad (\text{A8})$$

Στην πιο γενική περίπτωση όπου υπάρχουν ασθενείς που επικαλύπτονται μόνο μερικώς, οι εξισώσεις πρέπει να τροποποιηθούν. Για παράδειγμα, αν θεωρήσουμε ότι οι δύο σχεδιασμοί μοιράζονται $m/2$ ασθενείς που συνεισφέρουν m αλληλόμορφα (με $m < n_1$) οι αντίστοιχες ποσότητες της εξίσωσης A2 πρέπει να διαμεριστούν σε δυο ανεξάρτητα σύνολα τα οποία δηλώνονται με τους δείκτες s και d (s :shared και d :distinct). Έτσι, η εξίσωση A2 γίνεται:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{HHR}) &= \text{cov}[\log(n_{11s} + n_{11d}), \log(w_s + w_d)] \\ &+ \text{cov}[\log(n_{10s} + n_{10d}), \log(x_s + x_d)] \end{aligned} \quad (\text{A9})$$

Παρατηρώντας ότι $n_{11s}=w_s$, $n_{10s}=x_s$ και $m=n_{11s}+n_{10s}=w_s+x_s$ και ακολούθως χρησιμοποιώντας την εξίσωση A7 έτσι ώστε να υπολογιστεί η σχέση $\text{cov}[\log(S+U), \log(S+W)]$, έχουμε ότι:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{HHR}) &= \frac{n_{11s}}{(n_{11s} + n_{11d})(n_{11s} + w_d)} + \frac{n_{10s}}{(n_{10s} + n_{10d})(n_{10s} + x_d)} \\ &= \frac{n_{11s}}{n_{11}w} + \frac{n_{10s}}{n_{10}x} \end{aligned} \quad (\text{A10})$$

Παρομοίως, παρατηρώντας ότι $n_{11s}=w_s=a_s+b_s$ και $n_{10s}=x_s=c_s+d_s$, η γενικευμένη μορφή της εξίσωσης A8 γίνεται:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{TDT}) &= \text{cov}[\log(n_{11s} + n_{11d}), \log(b_s + b_d)] \\ &+ \text{cov}[\log(n_{10s} + n_{10d}), \log(b_s + b_d)] \end{aligned} \quad (\text{A11})$$

η οποία με τη σειρά της απλοποιείται σε:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{TDT}) &= \frac{b_s}{(n_{11s} + n_{11d})(b_s + b_d)} + \frac{b_s}{(n_{10s} + n_{10d})(b_s + b_d)} \\ &= \frac{b_s}{b} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right)\end{aligned}\tag{A12}$$

Παράρτημα Β

Ο υπολογισμός των δειγματικών εκτιμήσεων της διασποράς (Woolf's estimator (Woolf 1955)) μπορεί να βασιστεί στην προσέγγιση της διωνυμικής από την κανονική κατανομή για μεγάλο μέγεθος δείγματος. Ωστόσο, οι ίδιες ποσότητες μπορούν εύκολα να υπολογιστούν ακολουθώντας μια υπάρχουσα μεθοδολογία (Bagos 2012) με την υπόθεση ότι οι δειγματικές παρατηρήσεις είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με $E[n]=\text{var}[n]=m$ (Berrington and Cox 2003). Κατόπιν, χρησιμοποιώντας τη μέθοδο Δέλτα (Agresti 2002) υπολογίζεται η διασπορά των log-counts:

$$\text{var}[\log(n)] \approx \text{var}[n] \left(\frac{\partial \log(E[n])}{\partial (E[n])} \right)^2 = m \left(\frac{1}{m} \frac{\partial N}{\partial N} \right)^2 = \frac{1}{m}$$

Ακολουθώς, αντικαθιστώντας τις αναμενόμενες τιμές (m) με τις παρατηρηθείσες (n), εξάγεται η ακόλουθη σχέση:

$$\text{var}[\log(n)] = \frac{1}{n}$$

Προχωρούμε σε υπολογισμό των συνδιακυμάνσεων στηριζόμενοι σε θεμελιώδεις ιδιότητες:

$$\begin{aligned} \text{cov}(aX + bY, cW + dV) &= ac \text{cov}(X, W) + ad \text{cov}(X, V) \\ &\quad + bc \text{cov}(Y, W) + bd \text{cov}(Y, V) \end{aligned}$$

Συνεπώς:

$$\begin{aligned}
\text{cov}(y_{1i}, y_{3i}) &= \text{cov} \left(\log \left(\frac{(nAB_{11i} + nAB_{10i})(nAA_{00i} + nAA_{01i})}{(nAA_{11i} + nAA_{10i})(nAB_{01i} + nAB_{00i})} \right), \right. \\
&\quad \left. \log \left(\frac{(nAB_{01i} + nAB_{11i})(nAA_{00i} + nAA_{10i})}{(nAA_{11i} + nAA_{01i})(nAB_{10i} + nAB_{00i})} \right) \right) \\
&= \text{cov} \left(\log((nAB_{11i} + nAB_{10i})(nAA_{00i} + nAA_{01i})) - \log((nAA_{11i} + nAA_{10i})(nAB_{01i} + nAB_{00i})), \right. \\
&\quad \left. \log((nAB_{01i} + nAB_{11i})(nAA_{00i} + nAA_{10i})) - \log((nAA_{11i} + nAA_{01i})(nAB_{10i} + nAB_{00i})) \right) \\
&= \text{cov}(\log((nAB_{11i} + nAB_{10i})(nAA_{00i} + nAA_{01i})), \log((nAB_{01i} + nAB_{11i})(nAA_{00i} + nAA_{10i}))) \\
&\quad - \text{cov}(\log((nAB_{11i} + nAB_{10i})(nAA_{00i} + nAA_{01i})), \log((nAA_{11i} + nAA_{01i})(nAB_{10i} + nAB_{00i}))) \\
&\quad - \text{cov}(\log((nAA_{11i} + nAA_{10i})(nAB_{01i} + nAB_{00i})), \log((nAB_{01i} + nAB_{11i})(nAA_{00i} + nAA_{10i}))) \\
&\quad + \text{cov}(\log((nAA_{11i} + nAA_{10i})(nAB_{01i} + nAB_{00i})), \log((nAA_{11i} + nAA_{01i})(nAB_{10i} + nAB_{00i}))) \\
&= \text{cov}(\log(nAB_{11i} + nAB_{10i}) + \log(nAA_{00i} + nAA_{01i}), \log(nAB_{01i} + nAB_{11i}) + \log(nAA_{00i} + nAA_{10i})) \\
&\quad - \text{cov}(\log(nAB_{11i} + nAB_{10i}) + \log(nAA_{00i} + nAA_{01i}), \log(nAA_{11i} + nAA_{01i}) + \log(nAB_{10i} + nAB_{00i})) \\
&\quad - \text{cov}(\log(nAA_{11i} + nAA_{10i}) + \log(nAB_{01i} + nAB_{00i}), \log(nAB_{01i} + nAB_{11i}) + \log(nAA_{00i} + nAA_{10i})) \\
&\quad + \text{cov}(\log(nAA_{11i} + nAA_{10i}) + \log(nAB_{01i} + nAB_{00i}), \log(nAA_{11i} + nAA_{01i}) + \log(nAB_{10i} + nAB_{00i})) \\
&= \text{cov}(\log(nAB_{11i} + nAB_{10i}), \log(nAB_{01i} + nAB_{11i})) + \text{cov}(\log(nAA_{00i} + nAA_{01i}), \log(nAA_{00i} + nAA_{10i})) \\
&\quad - \text{cov}(\log(nAB_{11i} + nAB_{10i}), \log(nAB_{10i} + nAB_{00i})) - \text{cov}(\log(nAA_{00i} + nAA_{01i}), \log(nAA_{11i} + nAA_{01i})) \\
&\quad - \text{cov}(\log(nAA_{11i} + nAA_{10i}), \log(nAA_{00i} + nAA_{10i})) - \text{cov}(\log(nAB_{01i} + nAB_{00i}), \log(nAB_{01i} + nAB_{11i})) \\
&\quad + \text{cov}(\log(nAA_{11i} + nAA_{10i}), \log(nAA_{11i} + nAA_{01i})) + \text{cov}(\log(nAB_{01i} + nAB_{00i}), \log(nAB_{10i} + nAB_{00i})) \\
&\quad \quad \quad (B1)
\end{aligned}$$

Παρατηρούμε ότι το δεξί μέρος της εξίσωσης (B1) περιλαμβάνει όρους της μορφής $\text{cov}[\log(S+U), \log(S+W)]$, που πρέπει να υπολογιστούν. Αν υποθέσουμε ότι οι S , U , W είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με $E[S] = \text{var}[S] = s$, $E[U] = \text{var}[U] = u$ και $E[W] = \text{var}[W] = w$, τότε οι γραμμικοί συνδυασμοί $Z = S + U$ και $V = S + W$ είναι επίσης τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με $E[Z] = \text{var}[Z] = z = s + u$ και $E[V] = \text{var}[V] = v = s + w$, αντίστοιχα. Έτσι, οι (V, Z) θα ακολουθούν μια διμεταβλητή κατανομή Poisson (Marshall and Olkin 1985) και η συνδιακύμανση μεταξύ των Z και V θα είναι:

$$\text{cov}[S + U, S + W] = \text{cov}[S, S] = \text{var}[S] = s \quad (B2)$$

Η συνδιακύμανση μεταξύ των τυχαίων μεταβλητών Z και V , απαιτεί τον αναλυτικό υπολογισμό ολοκληρωμάτων των από κοινού και περιθώριων αθροιστικών συναρτήσεων κατανομής (Cuadras 2002). Ωστόσο, μπορεί κάποιος να υπολογίσει προσεγγιστικές εκτιμήσεις χρησιμοποιώντας τη μέθοδο Δέλτα (Agresti 2002):

$$\begin{aligned}\text{cov}[f(Z), g(V)] &\approx \text{cov}\left[\frac{\partial f(E[Z])}{\partial Z}(Z - E[Z]), \frac{\partial g(E[V])}{\partial V}(V - E[V])\right] \\ &= \frac{\partial f(E[Z])}{\partial Z} \frac{\partial g(E[V])}{\partial V} \text{cov}(Z, V)\end{aligned}$$

Συνεπώς ισχύει ότι:

$$\begin{aligned}\text{cov}[\log(S+U), \log(S+W)] &\approx \text{cov}[\log(Z), \log(V)] \\ &= \text{cov}(Z, V) \frac{\partial \log Z}{\partial Z} \frac{\partial \log V}{\partial V} \quad (\text{B3}) \\ &= \frac{s}{zv} = \frac{s}{(s+u)(s+w)}\end{aligned}$$

Η εξίσωση (B3), χρησιμοποιείται για την αναδιατύπωση της εξίσωσης (B1) ως εξής:

$$\begin{aligned}\text{cov}(y_{1i}, y_{3i}) &= \frac{nAB_{00i}}{(nAB_{01i} + nAB_{00i})(nAB_{10i} + nAB_{00i})} + \frac{nAB_{11i}}{(nAB_{11i} + nAB_{10i})(nAB_{01i} + nAB_{11i})} \\ &\quad - \frac{nAB_{10i}}{(nAB_{11i} + nAB_{10i})(nAB_{10i} + nAB_{00i})} - \frac{nAB_{01i}}{(nAB_{01i} + nAB_{00i})(nAB_{01i} + nAB_{11i})} \\ &\quad - \frac{nAA_{10i}}{(nAA_{11i} + nAA_{10i})(nAA_{00i} + nAA_{10i})} + \frac{nAA_{00i}}{(nAA_{00i} + nAA_{01i})(nAA_{00i} + nAA_{10i})} \\ &\quad + \frac{nAA_{11i}}{(nAA_{11i} + nAA_{10i})(nAA_{11i} + nAA_{01i})} - \frac{nAA_{01i}}{(nAA_{00i} + nAA_{01i})(nAA_{11i} + nAA_{01i})} \\ &= \sum_j \sum_k (-1)^{j-k} \left(\frac{nAB_{jki}}{nAB_{j+i} nAB_{+ki}} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i} nAA_{+ki}} \right) \quad (\text{B4})\end{aligned}$$

Παρόμοια, για τον υπολογισμό της $\text{cov}(y_{2i}, y_{4i})$ ισχύει ότι:

$$\begin{aligned}
 \text{cov}(y_{2i}, y_{4i}) &= \frac{nBB_{00i}}{(nBB_{01i} + nBB_{00i})(nBB_{10i} + nBB_{00i})} + \frac{nBB_{11i}}{(nBB_{11i} + nBB_{10i})(nBB_{01i} + nBB_{11i})} \\
 &\quad - \frac{nBB_{10i}}{(nBB_{11i} + nBB_{10i})(nBB_{10i} + nBB_{00i})} - \frac{nBB_{01i}}{(nBB_{01i} + nBB_{00i})(nBB_{01i} + nBB_{11i})} \\
 &\quad - \frac{nAA_{10i}}{(nAA_{11i} + nAA_{10i})(nAA_{00i} + nAA_{10i})} + \frac{nAA_{00i}}{(nAA_{00i} + nAA_{01i})(nAA_{00i} + nAA_{10i})} \\
 &\quad + \frac{nAA_{11i}}{(nAA_{11i} + nAA_{10i})(nAA_{11i} + nAA_{01i})} - \frac{nAA_{01i}}{(nAA_{00i} + nAA_{01i})(nAA_{11i} + nAA_{01i})} \\
 &= \sum_j \sum_k (-1)^{j-k} \left(\frac{nBB_{jki}}{nBB_{j+i}nBB_{+ki}} \right) + \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right)
 \end{aligned} \tag{B5}$$

Για τον υπολογισμό της $\text{cov}(y_{1i}, y_{4i})$ ισχύει ότι:

$$\begin{aligned}
 \text{cov}(y_{1i}, y_{4i}) &= \frac{nAA_{00i}}{(nAA_{00i} + nAA_{01i})(nAA_{00i} + nAA_{10i})} - \frac{nAA_{10i}}{(nAA_{11i} + nAA_{10i})(nAA_{00i} + nAA_{10i})} \\
 &\quad + \frac{nAA_{11i}}{(nAA_{11i} + nAA_{10i})(nAA_{11i} + nAA_{01i})} - \frac{nAA_{01i}}{(nAA_{00i} + nAA_{01i})(nAA_{11i} + nAA_{01i})} \\
 &= \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right)
 \end{aligned} \tag{B6}$$

Τέλος, για τον υπολογισμό της $\text{cov}(y_{2i}, y_{3i})$ ισχύει ότι:

$$\begin{aligned}
 \text{cov}(y_{2i}, y_{3i}) &= \frac{nAA_{00i}}{(nAA_{00i} + nAA_{01i})(nAA_{00i} + nAA_{10i})} - \frac{nAA_{01i}}{(nAA_{00i} + nAA_{01i})(nAA_{11i} + nAA_{01i})} \\
 &\quad - \frac{nAA_{10i}}{(nAA_{11i} + nAA_{10i})(nAA_{00i} + nAA_{10i})} + \frac{nAA_{11i}}{(nAA_{11i} + nAA_{10i})(nAA_{11i} + nAA_{01i})} \\
 &= \sum_j \sum_k (-1)^{j-k} \left(\frac{nAA_{jki}}{nAA_{j+i}nAA_{+ki}} \right)
 \end{aligned} \tag{B7}$$

Παράρτημα Γ

Για τον υπολογισμό των ποσοτήτων των κελιών του Πίνακα 4.2 θεωρώντας γνωστά τα περιθώρια αθροίσματα και το Odds Ratio, εργαζόμαστε ως ακολούθως. Στις εξισώσεις (Γ1-5), απεικονίζονται τα περιθώρια αθροίσματα καθώς επίσης και το Odds Ratio που περιγράφει τη συσχέτιση των δυο εκβάσεων. Για απλοποίηση των σχέσεων χρησιμοποιούμε το συμβολισμό n_{+jki} για να λαμβάνουμε υπόψη το $nAA_{jki}+nAB_{jki}+nBB_{jki}$.

$$(n_{+11i}) + (n_{+10i}) = n_{+1+i} \quad (\Gamma 1)$$

$$(n_{+01i}) + (n_{+00i}) = n_{+0+i} \quad (\Gamma 2)$$

$$(n_{+11i}) + (n_{+01i}) = n_{++1i} \quad (\Gamma 3)$$

$$(n_{+10i}) + (n_{+00i}) = n_{++0i} \quad (\Gamma 4)$$

$$OR = \frac{n_{+11i} n_{+00i}}{n_{+01i} n_{+10i}} \quad (\Gamma 5)$$

Συνεπώς, συνδυάζοντας τις εξισώσεις (Γ1-5) και λύνοντας ένα σύστημα 5 εξισώσεων με 4 αγνώστους, υπολογίζονται οι ποσότητες των κελιών της από κοινού κατανομής των δυο εκβάσεων ως ακολούθως:

$$n_{+11i} = \frac{-\sqrt{(-n_{+1+i} OR - n_{+0+i} - n_{++1i} OR + n_{++1i})^2 - 4n_{+1+i} n_{++1i} (OR - 1) OR}}{2(OR - 1)} + \frac{(n_{+1+i} OR + n_{+0+i} + n_{++1i} OR - n_{++1i})}{2(OR - 1)} \quad (\Gamma 6)$$

$$n_{+10i} = \frac{\sqrt{(-n_{+1+i} OR - n_{+0+i} - n_{++1i} OR + n_{++1i})^2 - 4n_{+1+i} n_{++1i} (OR - 1) OR}}{2(OR - 1)} + \frac{(n_{+1+i} OR - 2n_{+1+i} - n_{+0+i} - n_{++1i} OR + n_{++1i})}{2(OR - 1)} \quad (\Gamma 7)$$

$$n_{+01i} = \frac{\sqrt{(-n_{+1+i} OR - n_{+0+i} - n_{+1i} OR + n_{+1i})^2 - 4n_{+1+i} n_{+1i} (OR-1) OR}}{2(OR-1)} + \frac{(-n_{+1+i} OR - n_{+0+i} + n_{+1i} OR - n_{+1i})}{2(OR-1)} \quad (\Gamma 8)$$

$$n_{+00i} = \frac{-\sqrt{(-n_{+1+i} OR - n_{+0+i} - n_{+1i} OR + n_{+1i})^2 - 4n_{+1+i} n_{+1i} (OR-1) OR}}{2(OR-1)} + \frac{(n_{+1+i} OR + 2n_{+0+i} OR - n_{+0+i} - n_{+1i} OR + n_{+1i})}{2(OR-1)} \quad (\Gamma 9)$$

Ακολουθώς, υπολογίζονται τα κελιά του Πίνακα 4.3 που περιγράφει την από κοινού κατανομή της συσχέτισης των δυο εκβάσεων (Y_{1i} , Y_{2i}) ανάλογα με το γονότυπο συνδυάζοντας τις εξισώσεις (Γ10-25). Να σημειωθεί ότι τα περιθώρια αθροίσματα των εξισώσεων (Γ16, 17, 21, 22) προκύπτουν από τις εξισώσεις (Γ6-9).

$$nAA_{11i} + nAA_{01i} = nAA_{+1i} \quad (\Gamma 10)$$

$$nAB_{11i} + nAB_{01i} = nAB_{+1i} \quad (\Gamma 11)$$

$$nBB_{11i} + nBB_{01i} = nBB_{+1i} \quad (\Gamma 12)$$

$$nAA_{10i} + nAA_{00i} = nAA_{+0i} \quad (\Gamma 13)$$

$$nAB_{10i} + nAB_{00i} = nAB_{+0i} \quad (\Gamma 14)$$

$$nBB_{10i} + nBB_{00i} = nBB_{+0i} \quad (\Gamma 15)$$

$$nAA_{11i} + nAB_{11i} + nBB_{11i} = n_{+11i} \quad (\Gamma 16)$$

$$nAA_{10i} + nAB_{10i} + nBB_{10i} = n_{+10i} \quad (\Gamma 17)$$

$$nAA_{11i} + nAA_{10i} = nAA_{1+i} \quad (\Gamma 18)$$

$$nAB_{11i} + nAB_{10i} = nAB_{1+i} \quad (\Gamma 19)$$

$$nBB_{11i} + nBB_{10i} = nBB_{1+i} \quad (\Gamma 20)$$

$$nAA_{01i} + nAB_{01i} + nBB_{01i} = n_{+01i} \quad (\Gamma 21)$$

$$nAA_{00i} + nAB_{00i} + nBB_{00i} = n_{+00i} \quad (\Gamma 22)$$

$$nAA_{01i} + nAA_{00i} = nAA_{0+i} \quad (\Gamma 23)$$

$$nAB_{01i} + nAB_{00i} = nAB_{0+i} \quad (\Gamma 24)$$

$$nBB_{01i} + nBB_{00i} = nBB_{0+i} \quad (\Gamma 25)$$

Έτσι συνδυάζοντας τις εξισώσεις (Γ10-25) προκύπτουν οι ακόλουθες σχέσεις (Γ26-35):

$$nAA_{00i} = (n +_{11i}) + nAA_{0+i} + nAB_{0+i} + nBB_{+0i} - (nAA_{+1i} + nAB_{+1i} + nBB_{1+i}) - nAB_{00i} - nBB_{00i} \quad (\Gamma 26)$$

$$nBB_{01i} = nBB_{+0i} + nBB_{+1i} - nBB_{1+i} - nBB_{00i} \quad (\Gamma 27)$$

$$nAB_{01i} = nAB_{0+i} - nAB_{00i} \quad (\Gamma 28)$$

$$nAA_{01i} = nAA_{+1i} + nAB_{+1i} + nBB_{1+i} - ((n +_{11i}) + nAB_{0+i} + nBB_{+0i}) + nAB_{00i} + nBB_{00i} \quad (\Gamma 29)$$

$$nBB_{10i} = nBB_{+0i} - nBB_{00i} \quad (\Gamma 30)$$

$$nAB_{10i} = nAB_{+0i} - nAB_{00i} \quad (\Gamma 31)$$

$$nAA_{10i} = nAA_{+0i} + nAA_{+1i} + nAB_{+1i} + nBB_{1+i} - ((n +_{11i}) + nAA_{0+i} + nAB_{0+i} + nBB_{+0i}) + nAB_{00i} + nBB_{00i} \quad (\Gamma 32)$$

$$nBB_{11i} = nBB_{1+i} - nBB_{+0i} + nBB_{00i} \quad (\Gamma 33)$$

$$nAB_{11i} = nAB_{+1i} - nAB_{0+i} + nAB_{00i} \quad (\Gamma 34)$$

$$nAA_{11i} = (n +_{11i}) + nAB_{0+i} + nBB_{+0i} - (nAB_{+1i} + nBB_{1+i}) - nAB_{00i} - nBB_{00i} \quad (\Gamma 35)$$

Να σημειωθεί ότι για τον υπολογισμό των κελιών του 4.3, απαιτείται η γνώση τουλάχιστον δυο ποσοτήτων. Τέλος, πρέπει να αναφερθεί ότι οι εξισώσεις (Γ26-35) εφαρμόζονται μόνο εάν ισχύουν οι συνθήκες (Γ36-46):

$$(n +_{01i}) + (n +_{11i}) = nAA_{+1i} + nAB_{+1i} + nBB_{+1i} \quad (\Gamma 36)$$

$$(n +_{00i}) + (n +_{10i}) = nAA_{+0i} + nAB_{+0i} + nBB_{+0i} \quad (\Gamma 37)$$

$$nBB_{0+i} + nBB_{1+i} = nBB_{+0i} + nBB_{+1i} \quad (\Gamma 38)$$

$$nAA_{+0i} + nAA_{+1i} = nAA_{1+i} + nAA_{0+i} \quad (\Gamma 39)$$

$$(n +_{11i}) + nAA_{0+i} + nAB_{0+i} + (n +_{10i}) = nAA_{+1i} + nAB_{+1i} + nBB_{1+i} + nAA_{+0i} + nAB_{+0i} \quad (\Gamma 40)$$

$$nAB_{0+i} - nAB_{+1i} < nAB_{00i} < nAB_{0+i} \quad (\Gamma 41)$$

$$nBB_{+0i} - nBB_{+1i} < nBB_{00i} < nBB_{+0i} \quad (\Gamma 42)$$

$$nAB_{0+i} - nAB_{+1i} < nAB_{00i} < nAB_{+0i} \quad (\Gamma 43)$$

$$nBB_{+0i} - nBB_{+1i} < nBB_{00i} < nBB_{+0i} - nBB_{+1i} + nBB_{+1i} \quad (\Gamma 44)$$

$$\begin{aligned} & (n_{+11i}) + nAB_{0+i} + nBB_{+0i} - (nAA_{+1i} + nAB_{+1i} + nBB_{+1i}) < nAB_{00i} + nBB_{00i} \\ & < (n_{+11i}) + nAB_{0+i} + nBB_{+0i} + nAA_{0+i} - (nAA_{+1i} + nAB_{+1i} + nBB_{+1i}) \end{aligned} \quad (\Gamma 45)$$

$$\begin{aligned} & (n_{+11i}) + nAB_{0+i} + nBB_{+0i} + nAA_{0+i} - (nAA_{+1i} + nAB_{+1i} + nBB_{+1i} + nAA_{+0i}) < nAB_{00i} + nBB_{00i} \\ & < (n_{+11i}) + nAB_{0+i} + nBB_{+0i} - (nAB_{+1i} + nBB_{+1i}) \end{aligned} \quad (\Gamma 46)$$

Παράρτημα Δ

Πίνακας Δ1. Τα χαρακτηριστικά των μελετών που συμπεριλήφθησαν στην μετα-ανάλυση.

Author, Year	Descent	No. cases/controls	Toxicity genotypes	Non toxicity genotypes	Polymorphism	Disease	Patient's characteristics	Treatment	Toxicities *	Folate status**
van Ede, 2001 (van Ede, Laan et al. 2001)	Caucasian	30/206	CC: 10 CT+TT: 30	CC: 112 CT+TT: 94	C677T	RA	Of the 236 patients evaluated, 171 were females, 190 were positive for rheumatoid factor (RF) and the mean age was 55 years.	Patients were randomly assigned to receive either MTX alone, MTX plus folic acid (1.0 mg/day), or MTX plus folinic acid (2.5 mg/week). The initial MTX dosage was 7.5 mg/week and could be increased to a maximum of 25 mg/week if necessary. The dosages of folic and folinic acid were doubled when the MTX dosage reached \$15 mg/week.	hepatic	Not all
Bohanec, 2008 (Bohanec Grabar, Logar et al. 2008)	Caucasian	113/37	CC: 60 CT+TT: 53 AA+ CA: 102 CC: 11	CC: 14 CT+TT: 23 AA+ CA: 30 CC: 7	C677T, A1298C	RA	The study population consisted of 213 unrelated adult RA patients, 25 male (11.7%) and 188 female (88.3%), median age 61 (range 51–69) years.	157 out of 213 patients were treated with MTX with a median dose of 10 mg/week (range 10.0–12.5). The remaining 56 of 213 patients were treated with a median dose of 10 mg of MTX per week (range 10.0–14.4) for a median time of 24 months (range 8–48 months) but discontinued MTX prior to entry into the study.	overall	Not all
Aggarwal, 2006 (Aggarwal, Naik et al. 2006)	Other	30/120	CC: 19 CT+TT: 11	CC: 68 CT+TT: 52	C677T	RA	Of the 150 patients evaluated, 130 were females, 84.3 per cent were seropositive for rheumatoid factor (RF) and the mean age was 42.9 ± 11.1 yr with mean disease duration of 7.65 ± 5.2 yr.	Mean duration of MTX treatment was 26.1 ± 20.6 months and the mean MTX dose was 11.1 ± 3.1 mg/wk.	hepatic, haematological, GI, pulmonary	All

Zeng, 2008 (Zeng, Wang et al. 2008)	Asian	71/122	CC: 39 CT+TT: 32	CC: 60 CT+TT: 62	C677T	RA	193 Chinese Han consecutive outpatients with inflammatory arthritis who received	Low-dose MTX treatment (5–15 mg/week, mean (SD) 9.4 (1.9) mg/week)	GI, nervous system	None
Urano, 2002 (Urano, Taniguchi et al. 2002)	Asian	21/81	CC: 3 CT+TT: 18	CC: 32 CT+TT: 49	C677T, A1298C	RA	One hundred and six rheumatoid arthritis patients (all Japanese nationals; 93 females, 13 males) who had been administered MTX were selected.	MTX treatment had been started at initial doses of 2.5 mg to 7.5 mg per week.	overall	Not all
Tanigushi, 2007 (Taniguchi, Urano et al. 2007)	Asian	43/113	AA: 15 AC+CC: 6 CC: 10 CT+TT: 33	AA: 56 AC+CC: 25 CC: 56 CT+TT: 57	C677T, A1298C	RA	A total of 384 patients with RA were randomly selected from the outpatient clinic at the Institute of Rheumatology, Tokyo Women's Medical University.	MTX treatment was started at initial doses of 4–6 mg/ week in almost all patients.	overall	Not all
Krull, 2008 (Krull, Brouwers et al. 2008)	Other	11/37	CC: 6 CT+TT: 5 AA: 3 AC+CC: 8	CC: 18 CT+TT: 19 AA: 22 AC+CC: 15	C677T, A1298C	ALL	Mean age at diagnosis for the genotyped sample was 4.1 years (range, 0.7 to 9.1 years), and median time off therapy was 3.8 years (range, 1.5 to 9.3 years) for the study population. Thirty-two of the patients (67%) were male.	Patients were treated with various Pediatric Oncology Group protocols (ie, POG 8617, 8698, 8699, 9005, 9006, 9297, 9405, 9406, 9411, 9605).	nervous system	None
Schmeling, 2005 (Schmeling, Biber et al. 2005)	Caucasian	26/32	CC: 9 CT+TT: 17	CC: 20 CT+TT: 12	C677T	arthritis	58 patients (43 girls, 15 boys). 21 with seronegative polyarticular onset, 14 oligoarticular onset, 7 with enthesitis related arthritis, 4 with psoriatic arthritis, 2 systemic onset and 2 unclassified arthritis.	The treatment included MTX orally for at least 3 months at a mean dosage of $12.8 \pm 2.6 \text{ mg/m}^2$ (range 8.4–20.5 mg/m^2).	hepatic, GI	None
Bernbeck, 2003 (Bernbeck, Mauz-Korholz et al. 2003)	Caucasian	14/73	CC: 4 CT+TT: 10	CC: 32 CT+TT: 41	C677T	ALL	14 of 209 patients with ALL BME and/or AON where diagnosed by MRI while they suffered from increasing pain of the affected region, with an incidence of 6.7%.	The therapy included oral MTX ($20 \text{ mg/m}^2/\text{week}$)	bones	None

Azarpira, 2007 (Azarpira, Ramzi et al. 2007)	Other	9/21	CC: 6 CT+TT: 3	CC: 15 CT+TT: 6	C677T	hem. cell transplant ation	Subjects in this retrospective cohort study were 30 patients (20 males, 10 females; median age, 29 years; range, 14-48 years) receiving an allogeneic hematopoietic cell transplant at the Bone Marrow Transplant Center of the Shiraz University of Medical Sciences in Shiraz, Iran, between 2006 and 2007.	Prophylactic MTX	GVHD	None
Sugimoto, 2008 (Sugimoto, Murata et al. 2008)	Asian	33/123	CC+ CT: 32 TT: 1	CC +CT: 104 TT: 19	C677T	hem. cell transplant ation	The study population included adult patients who had received bone marrow or peripheral blood stem cell transplantation from an HLA-identical sibling donor. Median age 38 (range 15–62) years. 56% were males.	Prophylactic MTX	GVHD	None
Imanishi, 2007 (Imanishi, Okamura et al. 2007)	Asian	5/13	CC+ CT: 2 TT: 3	CC+ CT: 10 TT: 3	C677T	ALL, lymphoma	Blood samples were obtained from 26 patients with ALL (n = 20) or ML (n = 6). The age range was 10 months to 15 years, average 6.7 ± 4.7.	The patients were administered a high dose of methotrexate (3 g/m ²) followed by leucovorin rescue treatment in the event of complete remission.	hepatic	None
Campalani, 2007 (Campalani, Arenas et al. 2007)	Other	103/73	C allele: 131 T allele: 75 A allele: 158 C allele: 48	C allele: 92 T allele: 54 A allele: 103 C allele: 43	C677T, A1298C	psoriasis	203 patients with psoriasis (192 with chronic plaque, 8 with palmoplantar pustular (PPP), 2 with erythrodermic, and 1 with guttate psoriasis), who had previously or currently been treated with oral methotrexate were recruited in this study.	Patients had previously or currently been treated with oral methotrexate.	hepatic	Not all
Ghodke, 2008 (Ghodke, Chopra et al. 2008)	Asian	13/21	C allele: 22 T allele: 4 A allele: 158 C allele: 48	C allele: 34 T allele: 8 A allele: 103 C allele: 43	C677T, A1298C	RA	Thirty-four naïve RA patients (27 females) and American College of Rheumatology (ACR) classified.	6 months of supervised MTX therapy (7.5–17.5 mg weekly)	overall	None

Chiusolo, 2007 (Chiusolo, Reddicono et al. 2007)	Caucasian	31/24	CC+ CT: 19 TT: 12 AA+ CA: 31 CC: 0	CC+ CT: 23 TT: 1 AA+ CA: 20 CC: 4	C677T, A1298C	ALL	Patients with ALL receiving chemotherapy.	Homogeneous maintenance therapy with MTX administered weekly intramuscularly at a variable dosage 15–30 mg/m ² for almost 2 years.	hepatic, haematological	None
Pihusch, 2004 (Pihusch, Lohse et al. 2004)	Caucasian	28/4	CC+ CT: 26 TT: 2	CC+ CT: 3 TT: 1	C677T	hem. cell transplantation	Median age 43 (range 14–62) years. 62% were males.	MTX, intravenous methotrexate (15 mg/m ² on day 1, 10 mg/m ² on day 3 and 6)	GVHD, mucositis	All
Seidemann, 2006 (Seidemann, Book et al. 2006)	Caucasian	***	-	-	C677T	NHL	791 eligible patients up to 18 years of age with newly diagnosed NHL were enrolled in multicenter therapy trial NHL-BFM 95.	High-dose MTX infusion regimens were randomized (4- vs 24-h infusion) in patients with Bcell lymphoma; MTX was applied as 24-h infusion in all patients with lymphoblastic lymphoma and anaplastic large cell lymphoma	GI, mucositis, infection, secondary cancer	All
Toffoli, 2003 (Toffoli, Russo et al. 2003)	Caucasian	33/10	CC: 9 CT: 11 TT: 13	CC: 4 CT: 6 TT: 0	C677T	ovarian cancer	Forty-three patients with ovarian cancer, age 18–75 years	The first 15 patients were treated with oral MTX 1.25 mg every 12hr for 21 days. After a 2-week rest, this regimen was repeated. Since no objective response was noted in these patients, another group of 28 patients was enrolled, to be treated with 21-day MTX [1.25 mg every 12 hr].	overall	None
Kalayoglu-Besisik, 2003 (Kalayoglu-Besisik, Caliskan et al. 2003)	Caucasian	11/42	CC: 8 CT: 3 TT: 0	CC: 17 CT: 20 TT: 5	C677T	hem. cell transplantation	Fifty-three patients received allogeneic HCT (bone marrow 16, peripheral blood 37) as primary and unmanipulated transplantation for hematologic malignancies. Median age 31.4 (range 16–51) years. 68% were males.	Short-term MTX (intravenously at day 1, 15 mg/m ² , and at days 3, 6, and 11, 10 mg/m ²)	mucositis	None

Kim, 2006 (Kim, Jun et al. 2006)	Asian	154/231	CC: 27 CT: 95 TT: 32	CC: 106 CT: 98 TT: 27	C677T	RA	The patients were 24 to 80 years of age and included 355 women.	The patients were taking low-dose MTX (5-20 mg/wk).	overall	All
Fijnheer, 1998 (Fijnheer, Roest et al. 1998)	Other	-	-	-	C677T	SLE	91.3 % were women. The median age was 36 (17-70 range).	No specific information provided.	thrombosis	None
Shimasaki, 2008 (Shimasaki, Mori et al. 2008)	Asian	14/6	CC: 9 CT: 3 TT: 1	CC: 2 CT: 4 TT: 0	C677T	ALL, lymphoma	The patients included in this study were Japanese children with ALL or LBL who had been enrolled in the Tokyo Children's Cancer Study Group's (TCCSG) protocols.	The scheduled dosage of oral MTX was 25 mg/m ² once a week.	hepatic, haematological	None
Kim, 2007 (Kim, Lee et al. 2007)	Asian	-	-	-	C677T, A1298C	hem. cell transplantation	72 patients receiving allogeneic HSCT at a single institution from 1998 to 2005. Median age 36 (range 16-52 years) years. 54% were males.	Four intravenous doses of MTX (day 1 at 15 mg/m ² and days 3, 6, and 11 at 10 mg/m ²)	hepatic, GVHD	None
Ranganathan 2008(Ranganathan, Culverhouse et al. 2008)	Caucasian	30/22	CC: 12 CT: 16 TT: 2	CC: 12 CT: 5 TT: 5	C677T	RA	The median age was 61 years (51-69) and the duration of RA 6.5 years (3.5-16.5).	The dose of MTX was determined by the treating rheumatologist. Concomitant medications included corticosteroids, sulfasalazine etc.	overall	All
Ranganathan 2008 (Ranganathan, Culverhouse et al. 2008)	Caucasian	15/76	CC: 7 CT: 4 TT: 4	CC: 28 CT: 41 TT: 7	C677T	RA	The median age was 42 years (48-67) and the duration of RA 7.5 years (4-19).	The dose of MTX was determined by the treating rheumatologist.	GI	All
Kishi, 2003 (Kishi, Griener et al. 2003)	Other	9/44	CC: 6 CT: 2 TT: 1	CC: 26 CT: 17 TT: 1	C677T	ALL	Of the 53 patients who were enrolled onto the protocol, 23 were male, 38 were white, and 15 were nonwhite. The median age was 6 years (range, 0 to 18 years).	Clinical protocol that included two courses of high-dose MTX (2.5 or 5.0 g/ m ² per day) as consolidation therapy.	nervous system, thrombosis	None

Jazbec, 2005 (Jazbec, Kitanovski et al. 2005)	Caucasian	15/30	CC: 7 CT: 6 TT: 2	CC: 11 CT: 15 TT: 4	C677T, A1298C	ALL	449 patients, aged ≤ 16 years at the time of diagnosis, were treated for ALL at the Pediatric Oncology Clinic at the University of Ljubljana and registered in the Cancer Registry of Slovenia.	High-dose methotrexate (5g/m^2) was included in all the protocols during consolidation and standard dose methotrexate (50mg/m^2 per week for 18 months) was used in maintenance.	secondary cancer	None
Chiusolo, 2002 (Chiusolo, Reddicono et al. 2002)	Caucasian	22/39	AA: 7 AC: 6 CC: 2 CC: 7 CT: 6 TT: 9	AA: 14 AC: 14 CC: 2 CC: 14 CT: 19 TT: 6	C677T	ALL	Patients with acute leukemia receiving maintenance chemotherapy including MTX. Median age was 32 (range 12–72) years. 56% were males.	MTX was administered weekly intramuscularly at a variable dosage of $15\text{--}30\text{ mg/m}^2$ in ALL, or 15 mg/m^2 in APL during maintenance for almost 2 years.	hepatic, haematological	None
Berkun, 2004 (Berkun, Levartovsky et al. 2004)	other	33/60	CC: 15 CT: 14 TT: 4	CC: 30 CT: 24 TT: 6	C677T, A1298C	RA	Ninety three consecutive patients with rheumatoid arthritis with mean age 58.74 ± 13.69 years. Females were 82.8%.	Duration of methotrexate treatment 3.70 ± 3.43 years. Methotrexate dose $11.93 \pm 3.81\text{ mg/week}$.	overall	Not all
Gemmati, 2007 (Gemmati, Ongaro et al. 2007)	Caucasian	18/50 for hepatic toxicity	AA: 22 AC: 8 CC: 3 CC: 5 CT: 7 TT: 6 for hepatic toxicity	AA: 28 AC: 12 CC: 20 CC: 17 CT: 29 TT: 4 for hepatic toxicity	C677T, A1298C	NHL	The whole group consisted of 110 patients with high-grade NHL. Among these 110 patients, 68 received MACOP-B treatment and 42 received CHOP treatment. Mean age was 56.8 ± 17.1 years. 61% were males.	No specific information provided.	hepatic, mucositis haematological	None
Ruiz-Arguelles, 2007 (Ruiz-Arguelles, Coconi-Linares et al. 2007)	Other	16/12	CC: 0 CT: 10 TT: 6	CC: 2 CT: 4 TT: 6	C677T	ALL	There were 13 patients less than 15 years; median age was 16 years, average 16.5 years, range 0–40 years. The group included 21 males and 7 females.	Patients were treated with a modification of the TOTAL XI schedule, which includes high-dose MTX rescued with folinic acid.	mucositis	All

Warren, 2009 (Warren, Smith et al. 2009)	Other	160/180	CC: 75 CT: 67 TT: 18	CC: 84 CT: 77 TT: 19	C677T, A1298C	psoriasis	Patients with chronic plaque psoriasis treated with MTX were included. Males 61% and mean age of onset 24 years.	No specific information provided.	overall	Not all
Kumagai, 2003 (Kumagai, Hiyama et al. 2003)	Asian	52/63	CC: 24 CT: 18 TT: 10	CC: 22 CT: 29 TT: 12	C677T, A1298C	RA	95 of the patients were females. The mean age was 60 years and the mean disease duration 11 years.	Patients had been prescribed various doses of MTX (2-12 mg/week) starting with the initial dose of 2-4 mg/week.	overall	Not all
Wessels, 2006 (Wessels, de Vries-Bouwstra et al. 2006)	Caucasian	68/130	AA: 36 AC: 14 CC: 2 AA: 19 AC: 42 CC: 7	AA: 44 AC: 15 CC: 4 AA: 63 AC: 52 CC: 15	A1298C	RA	The group of 247 patients enrolled in this study was a subcohort of 508 patients who participated in the BeSt trial. Mean age was 54.6 ±13.3. 69% were females	Patients received MTX at an initial dosage of 7.5mg/week, which was increased to 15mg/week and combined with folic acid (1 mg/day) after 4 weeks. If the Disease Activity Score in 44 joints (DAS44) was >2.4 at 3 months, MTX was increased to 25 mg/week.	GI	All
Murphy, 2006 (Murphy, Diviney et al. 2006)	other	158/29	CC: 78 CT+TT: 80	CC: 12 CT+TT: 17	C677T	hem. cell transplant ation	Variety of haematological conditions: AML, CML, lymphomas, myelodysplasia etc.	15 mg /m ² at day 1 and 10 mg/m ² on days 3, 6 and 11 after transplant.	GVHD	Not all
Shimsaki, 2006 (Shimasaki, Mori et al. 2006)	Asian	-	-	-	C677T	ALL, lymphoma	Forty-three courses of high-dose methotrexate chemotherapy were evaluated in the 15 patients. The median patient age was 6 years (range 1–14 years). Six patients were male and nine were female.	Patients received high-dose methotrexate over 12 hours in the first course and over 24 hours in the remaining courses intravenously with uniform supportive treatments followed by each protocol.	hepatic, haematological, GI, mucositis	None

Aplenc, 2005 (Aplenc, Thompson et al. 2005)	Other	-	-	-	C677T	ALL	We undertook a case-control study of patients treated on CCG-1891, a national intermediate risk ALL study run by CCG from 1989 to 1992 in 137 participating institutions. Intermediate risk was defined as age between 1 and 10 years at diagnosis, and initial WBC count of $<5 \times 10^4/\mu\text{L}$.	All treatment arms were prescribed the identical schedule and dose of age-adjusted I.T. and weekly p.o. 20 mg/m ² MTX during interim maintenance and maintenance therapy.	overall	None
Tsuchiya, 2008 (Tsuchiya, Inoue et al. 2008)	Asian	-	-	-	C677T	urothelial Ca	The patients consisted of 34 males and 6 females with a mean age of 64.4 years (range 46 to 80, median 67) at the start of chemotherapy. Primary cancer lesions were located in the bladder in 23 patients and the renal pelvis or ureter in 19.	Briefly classic MVAC consisted of 28-day cycles of 30 mg/m ² i.v. methotrexate on days 1, 15 and 22; 3 mg/m ² i.v. vinblastine on days 2, 15 and 22; 30 mg/m ² i.v. doxorubicin on day 2 and 70 mg/m ² cisplatin on day 2.	haematological	None
van Kooten, 2008 (van Kooten Niekerk, Schmiegelow et al. 2008)	Caucasian	-	-	-	C677T, A1298C	ALL	Ninety-nine children with non-B ALL diagnosed between January 1st 1992 and December 31st 2005 and treated at least once with HDMTX 5g/m ² according to the NOPHO ALL-92 or NOPHO ALL-2000 protocols at Aarhus University Hospital were eligible for the study.	Each patient was scheduled to receive three or four HDMTX courses in the consolidation phase and five courses in the maintenance phase, all at a dose of 5 g/m ² .	haematological	None
Hughes, 2006 (Hughes, Beasley et al. 2006)	Caucasian	-	-	-	A1298C	RA	No specific information provided.	The mean MTX dose was 19 mg a week.	overall	Not all
Weisman, 2006 (Weisman, Furst et al. 2006)	Other	-	-	-	C677T	RA	214 patients (n=53 at Cedars Sinai Medical Center, n=44 at UCLA, n=53 at the Florida site, and n=64 at the Albany site) who were undergoing MTX therapy for at least 1 month. Median age was 63 (range 52–71) years and 79% were females.	Dosage of MTX 12.5 (range 10–17.5) mg/week. No. of months receiving MTX 39 (range 11–89).	GI, nervous system, alopecia	Not all

Robien, 2006 (Robien, Bigler et al. 2006)	Other	158/192	CC: 127 CT: 16 TT: 15	CC: 131 CT: 21 TT: 40	C677T, A1298C	hem. cell transplant ation	Subjects in this retrospective cohort study were adults undergoing allogeneic HCT at the Fred Hutchinson Cancer Research Center. 40.4 (range 18-67) years. 56% were males.	Prophylactic MTX.	GVHD	None
Herrlinger, 2005 (Herrlinger, Cummings et al. 2005)	Caucasian	22/63	AA: 2 AC: 18 CC: 2	AA: 37 AC: 21 CC: 5	A1298C	IBD	Cohort of unrelated Caucasian IBD patients were recruited treated with MTX.	The MTX dose for IBD patients varied between 15–25 mg/week with a median weekly dose of 20 mg, patients treated for concomitant rheumatoid arthritis received lower doses.	overall	Not all

* The types of MTX adverse effects of the patients
 ** The folate supplementation of the patients included in each study
 *** The information is not provided. ORs are given instead.

Πίνακας Δ2. Τα αποτελέσματα της μετα-ανάλυσης για τον πολυμορφισμό C677T – πολυμεταβλητή μεθοδολογία.

Category	Method	Contrast	No of studies	Odds ratio	95% confidence interval	
overall toxicity more than 2 types	Logistic regression	CT vs. CC	8	1.208	0.735	1.984
		TT vs. CC	8	1.672	1.009	2.769
	Multivariate model	CT vs. CC	8	1.200	0.730	1.980
		TT vs. CC	8	1.670	1.000	2.770
overall toxicity more than 1 type	Logistic regression	CT vs. CC	20	1.228	0.868	1.739
		TT vs. CC	20	1.560	0.950	2.561
	Multivariate model	CT vs. CC	20	1.320	0.920	1.890
		TT vs. CC	20	1.480	0.920	2.390
hepatotoxicity	Logistic regression	CT vs. CC	5	1.463	0.674	3.176
		TT vs. CC	5	4.191	1.642	10.698
	Multivariate model	CT vs. CC	5	1.460	0.680	3.130
		TT vs. CC	5	4.260	1.750	10.400
haematological toxicity	Logistic regression	CT vs. CC	3	0.810	0.372	1.765
		TT vs. CC	3	2.596	0.951	7.084
	Multivariate model	CT vs. CC	3	0.800	0.370	1.760
		TT vs. CC	3	2.590	0.950	7.080
GVHD	Logistic regression	CT vs. CC	2	0.993	0.508	1.713
		TT vs. CC	2	0.446	0.242	0.821
	Multivariate model	CT vs. CC	2	0.930	0.500	1.700
		TT vs. CC	2	0.440	0.240	0.820
mucositis	Logistic regression	CT vs. CC	4	0.690	0.351	1.360
		TT vs. CC	4	1.069	0.541	2.114
	Multivariate model	CT vs. CC	4	0.670	0.320	1.360
		TT vs. CC	4	0.980	0.500	1.890
Rheumatoid arthritis	Logistic regression	CT vs. CC	6	1.565	0.755	3.243
		TT vs. CC	6	1.991	0.982	4.036
	Multivariate model	CT vs. CC	6	1.590	0.750	3.390
		TT vs. CC	6	2.010	0.960	4.200
Haematological disorders	Logistic regression	CT vs. CC	7	0.783	0.486	1.259
		TT vs. CC	7	1.752	0.847	3.626
	Multivariate model	CT vs. CC	7	0.780	0.480	1.250
		TT vs. CC	7	1.750	0.840	3.620
Hematopoietic cell transplantation	Logistic regression	CT vs. CC	3	0.798	0.457	1.392
		TT vs. CC	3	0.413	0.226	0.753
	Multivariate model	CT vs. CC	3	0.790	0.450	1.390
		TT vs. CC	3	0.410	0.220	0.750

Folate supplement: None	Logistic regression	CT vs. CC	11	0.800	0.539	1.187
		TT vs. CC	11	1.624	0.728	3.623
	Multivariate model	CT vs. CC	11	0.790	0.510	1.200
		TT vs. CC	11	1.390	0.620	3.110
Folate supplement: Not all	Logistic regression	CT vs. CC	3	0.903	0.629	1.298
		TT vs. CC	3	1.026	0.598	1.759
	Multivariate model	CT vs. CC	3	0.900	0.620	1.300
		TT vs. CC	3	1.020	0.590	1.750
Folate supplement: All	Logistic regression	CT vs. CC	6	2.736	1.347	5.559
		TT vs. CC	6	2.871	1.107	7.450
	Multivariate model	CT vs. CC	6	2.820	1.320	6.000
		TT vs. CC	6	2.920	1.110	7.770

Πίνακας Δ3. Τα αποτελέσματα της μετα-ανάλυσης για τον πολυμορφισμό C677T–μονομεταβλητή μεθοδολογία (τυχαίων επιδράσεων).

Sub-group: overall toxicity – more than two types of toxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	10	1.217	0.889	1.665	63.2
	TT vs. CT+CC	10	1.566	0.949	2.582	43.0
	TT+CT vs. CC	14	1.335	0.874	2.039	72.9
Caucasians	T vs. C allele	3	1.636	0.903	2.964	29.6
	TT vs. CT+CC	4	3.191	0.514	19.804	73.2
	TT+CT vs. CC	5	1.342	0.677	2.663	50.8
Asians	T vs. C allele	4	1.065	0.495	2.289	78.5
	TT vs. CT+CC	3	1.655	1.029	2.660	0.0
	TT+CT vs. CC	6	1.540	0.670	3.543	84.4
Other	T vs. C allele	3	1.021	0.802	1.299	0.0
	TT vs. CT+CC	3	1.131	0.665	1.921	0.0
	TT+CT vs. CC	3	0.976	0.690	1.380	0.0
HWE	T vs. C allele	7	1.335	0.894	1.996	71.4
	TT vs. CT+CC	7	1.604	1.063	2.419	14.1
	TT+CT vs. CC	7	1.225	0.640	2.344	77.9
* The sub-analysis concerning studies with patients who presented more than two types of side effects						
Sub-group: overall toxicity – more than one type of toxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	24	1.158	0.919	1.459	65.4
	TT vs. CT+CC	25	1.246	0.840	1.848	54.2
	TT+CT vs. CC	32	1.239	0.946	1.624	63.0
Caucasians	T vs. C allele	8	1.162	0.810	1.666	45.8
	TT vs. CT+CC	11	1.546	0.723	3.307	66.1
	TT+CT vs. CC	12	1.157	0.803	1.668	44.4

Asians	T vs. C allele	7	1.394	0.810	2.401	63.2
	TT vs. CT+CC	6	1.377	0.719	2.636	29.3
	TT+CT vs. CC	9	1.821	0.938	3.537	76.1
Other	T vs. C allele	9	1.019	0.737	1.408	64.0
	TT vs. CT+CC	8	0.917	0.529	1.590	35.3
	TT+CT vs. CC	11	0.909	0.658	1.256	28.0
HWE	T vs. C allele	18	1.284	0.985	1.674	61.1
	TT vs. CT+CC	18	1.575	1.100	2.254	26.1
	TT+CT vs. CC	18	1.189	0.808	1.752	62.8
* The sub-analysis concerning studies with patients who presented more than one type of side effects						
Sub-group: hepatotoxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	8	1.588	1.146	2.202	23.4
	TT vs. CT+CC	6	4.601	2.139	9.896	12.6
	TT+CT vs. CC	9	2.170	1.305	3.608	33.7
Caucasians	T vs. C allele	2	2.472	1.341	4.559	5.8
	TT vs. CT+CC	3	6.967	3.101	15.651	0.0
	TT+CT vs. CC	4	2.310	1.470	3.629	0.0
Asians	T vs. C allele	4	1.560	0.667	3.647	30.6
	TT vs. CT+CC	3	1.521	0.372	6.213	0.0
	TT+CT vs. CC	4	2.716	1.034	7.134	29.6
Other	T vs. C allele	2	1.307	0.963	1.776	0.0
	TT vs. CT+CC	0				
	TT+CT vs. CC	1	0.494	0.126	1.941	
HWE	T vs. C allele	5	2.004	1.074	3.738	38.4
	TT vs. CT+CC	5	3.690	1.605	8.483	5.7
	TT+CT vs. CC	5	1.841	0.880	3.853	0.0
* The sub-analysis concerning studies with patients who presented liver dysfunction						
Sub-group: haematological toxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	5	1.071	0.792	1.449	0.0
	TT vs. CT+CC	5	1.861	0.627	5.519	66.3
	TT+CT vs. CC	6	1.459	0.845	2.520	0.0
Caucasians	T vs. C allele	2	1.584	0.934	2.688	0.0
	TT vs. CT+CC	4	1.780	0.539	5.886	74.2
	TT+CT vs. CC	2	1.401	0.532	3.688	30.1
Asians	T vs. C allele	2	1.272	0.442	3.662	0.0
	TT vs. CT+CC	1	3.316	0.120	91.601	0.0
	TT+CT vs. CC	3	1.387	0.573	3.360	0.0
Other	T vs. C allele	1	0.844	0.570	1.249	
	TT vs. CT+CC	0				
	TT+CT vs. CC	1	1.789	0.461	6.948	

HWE	T vs. C allele	3	1.499	0.913	2.462	0.0
	TT vs. CT+CC	3	2.551	1.055	6.169	0.0
	TT+CT vs. CC	3	1.244	0.595	2.597	0.3
* The sub-analysis concerning studies with patients who presented one or more bone marrow toxicities						
Sub-group: neurotoxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	2	1.308	0.525	3.257	0.0
	TT vs. CT+CC	2	3.398	1.808	6.387	0.0
	TT+CT vs. CC	4	0.612	0.329	1.137	0.0
Asians	TT+CT vs. CC	2	0.474	0.216	1.041	0.0
Other	T vs. C allele	2	1.308	0.525	3.257	0.0
	TT vs. CT+CC	2	3.398	1.808	6.387	0.0
	TT+CT vs. CC	2	0.929	0.340	2.544	0.0
* The sub-analysis concerning studies with patients who presented one or more neurological disorders						
Sub-group: GVHD*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	2	0.836	0.247	2.826	84.6
	TT vs. CT+CC	4	0.441	0.253	0.771	0.0
	TT+CT vs. CC	4	0.773	0.426	1.401	36.0
Caucasians	TT vs. CT+CC	1	0.636	0.050	8.123	
Asians	T vs. C allele	1	1.667	0.690	4.024	
	TT vs. CT+CC	2	0.549	0.065	4.630	61.0
	TT+CT vs. CC	1	2.895	0.581	14.430	
Other	T vs. C allele	1	0.477	0.324	0.703	
	TT vs. CT+CC	1	0.399	0.211	0.753	
	TT+CT vs. CC	3	0.600	0.399	0.904	0.0
* The sub-analysis concerning studies with patients who presented GVHD after hematopoietic cell transplantation						
Sub-group: mucositis*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	6	1.015	0.706	1.459	20.4
	TT vs. CT+CC	5	1.115	0.463	2.687	43.5
	TT+CT vs. CC	5	1.770	0.416	1.425	27.5
Caucasians	T vs. C allele	3	0.849	0.517	1.394	41.2
	TT vs. CT+CC	4	0.979	0.306	3.135	54.2
	TT+CT vs. CC	3	0.675	0.337	1.355	44.8
Other	T vs. C allele	2	1.601	0.737	3.476	0.0
HWE	T vs. C allele	3	0.873	0.377	2.020	60.3
	TT vs. CT+CC	3	0.840	0.505	1.396	0.0

	TT+CT vs. CC	3	0.829	0.236	2.914	59.1
* The sub-analysis concerning studies with patients who presented oral mucositis						
Sub-group: Rheumatoid arthritis*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	7	1.297	0.807	2.085	67.7
	TT vs. CT+CC	7	1.422	0.852	2.375	27.5
	TT+CT vs. CC	13	1.546	0.964	2.479	76.0
Caucasians	T vs. C allele	2	1.069	0.602	1.898	0.0
	TT vs. CT+CC	2	0.986	0.071	13.756	82.2
	TT+CT vs. CC	5	1.335	0.667	2.673	67.5
Asians	T vs. C allele	3	1.202	0.523	2.762	83.2
	TT vs. CT+CC	2	1.578	0.846	2.944	31.8
	TT+CT vs. CC	5	1.843	0.783	4.337	86.0
Other	T vs. C allele	2	3.268	0.288	37.046	78.3
	TT vs. CT+CC	3	1.324	0.581	3.016	0.0
	TT+CT vs. CC	3	1.496	0.463	4.837	66.5
HWE	T vs. C allele	4	1.285	0.748	2.209	75.6
	TT vs. CT+CC	4	1.735	1.129	2.666	0.0
	TT+CT vs. CC	4	1.247	0.448	3.472	86.1
* The sub-analysis concerning studies with rheumatoid arthritis patients						
Sub-group: Haematological disorders*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	T vs. C allele	10	1.237	0.947	1.616	23.5
	TT vs. CT+CC	10	1.957	0.935	4.099	60.8
	TT+CT vs. CC	9	1.028	0.758	1.396	0.0
Caucasians	T vs. C allele	4	1.201	0.793	1.820	43.2
	TT vs. CT+CC	6	1.807	0.690	4.733	75.1
	TT+CT vs. CC	5	1.052	0.759	1.458	0.0
Asians	T vs. C allele	3	1.713	0.434	6.762	54.7
	TT vs. CT+CC	2	3.426	0.544	21.568	0.0
Other	T vs. C allele	3	1.271	0.903	1.790	0.0
	TT vs. CT+CC	2	2.152	0.562	8.238	0.0
	TT+CT vs. CC	3	1.135	0.435	2.960	0.0
HWE	T vs. C allele	7	1.154	0.845	1.577	18.1
	TT vs. CT+CC	7	1.904	0.888	4.081	47.5
	TT+CT vs. CC	7	0.979	0.708	1.354	0.0
* The sub-analysis concerning studies with patients suffering from blood diseases						

Sub-group: Hematopoietic cell transplantation*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval	<i>P</i> (%)	
All	T vs. C allele	3	0.634	0.254 1.583	73.6	
	TT vs. CT+CC	5	0.435	0.252 0.754	0.0	
	TT+CT vs. CC	5	0.681	0.385 1.205	35.5	
Caucasians	TT vs. CT+CC	2	0.460	0.067 3.182	0.0	
Asians	TT vs. CT+CC	2	0.549	0.065 4.630	61.0	
Other	TT+CT vs. CC	3	0.600	0.399 0.904	0.0	
HWE	T vs. C allele	2	0.736	0.131 4.143	79.5	
	TT vs. CT+CC	2	1.009	0.229 4.441	0.0	
	TT+CT vs. CC	2	0.839	0.078 9.070	79.2	
* The sub-analysis concerning studies with patients treated with hematopoietic cell transplantation						

Sub-group: Folate supplementation*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval	<i>P</i> (%)	
No folate supplementation	T vs. C allele	14	1.123	0.774 1.631	63.0	
	TT vs. CT+CC	14	1.437	0.662 3.119	66.4	
	TT+CT vs. CC	17	1.067	0.756 1.506	34.8	
Partial folate supplementation	T vs. C allele	4	0.978	0.785 1.218	0.0	
	TT vs. CT+CC	4	1.100	0.694 1.745	0.0	
	TT+CT vs. CC	8	1.225	0.773 1.940	69.0	
Full folate supplementation	T vs. C allele	6	1.511	0.897 2.546	77.2	
	TT vs. CT+CC	7	1.282	0.674 2.442	50.4	
	TT+CT vs. CC	7	1.714	0.809 3.632	79.8	
* The sub-analysis based on folate supplementation status of the patients						

Πίνακας Δ4. Τα αποτελέσματα της μετα-ανάλυσης για τον πολυμορφισμό A1298C– πολυμεταβλητή μεθοδολογία.

Category	Method	Contrast	No of studies	Odds ratio	95% confidence interval	
overall toxicity more than 2 types	Logistic regression	AC vs. AA	5	1.697	0.762 3.780	
		CC vs. AA	5	1.027	0.419 2.515	
	Multivariate model	AC vs. AA	5	1.690	0.680 4.150	
		CC vs. AA	5	1.150	0.430 3.060	
overall toxicity more than 1 type	Logistic regression	AC vs. AA	9	1.690	1.011 2.825	
		CC vs. AA	9	1.889	1.232 2.896	
	Multivariate model	AC vs. AA	9	1.490	0.860 2.580	
		CC vs. AA	9	1.420	0.730 2.780	
Rheumatoid arthritis	Logistic regression	AC vs. AA	3	1.490	0.812 2.735	
		CC vs. AA	3	0.539	0.185 1.570	
	Multivariate model	AC vs. AA	3	1.470	0.780 2.760	
		CC vs. AA	3	0.540	0.180 1.570	

Haematological disorders	Logistic regression	AC vs. AA	2	0.486	0.137	1.722
		CC vs. AA	2	4.038	0.771	21.144
	Multivariate model	AC vs. AA	2	0.520	0.170	1.530
		CC vs. AA	2	4.220	0.890	19.900
Folate supplement: None	Logistic regression	AC vs. AA	4	1.278	0.582	2.803
		CC vs. AA	4	4.131	1.864	9.155
	Multivariate model	AC vs. AA	4	1.420	0.840	2.390
		CC vs. AA	4	3.830	1.950	7.490
Folate supplement: Not all	Logistic regression	AC vs. AA	4	2.071	0.859	4.990
		CC vs. AA	4	1.494	0.733	3.047
	Multivariate model	AC vs. AA	4	2.140	0.810	5.600
		CC vs. AA	4	1.590	0.730	3.450

Πίνακας Δ5. Τα αποτελέσματα της μετα-ανάλυσης για τον πολυμορφισμό A1298C – μονομεταβλητή μεθοδολογία.

Sub-group: overall toxicity – more than two types of toxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	C vs. A allele	8	0.891	0.568	1.399	76.2
	CC vs. AC+AA	7	0.631	0.350	1.135	36.6
	AC+CC vs. AA	7	1.281	0.718	2.283	75.3
Caucasians	C vs. A allele	3	1.260	0.415	3.830	80.0
	CC vs. AC+AA	4	0.633	0.318	1.261	10.2
	AC+CC vs. AA	2	5.230	0.937	29.183	80.4
Asians	C vs. A allele	2	0.824	0.470	1.444	0.0
	AC+CC vs. AA	3	0.994	0.618	1.598	0.0
Other	C vs. A allele	3	0.678	0.403	1.142	71.0
	CC vs. AC+AA	2	0.540	0.092	3.177	82.7
	AC+CC vs. AA	2	0.693	0.356	1.348	51.3
HWE	C vs. A allele	4	1.373	0.871	2.164	68.7
	CC vs. AC+AA	4	1.041	0.624	1.737	0.0
	AC+CC vs. AA	4	1.981	0.821	4.781	83.5
* The sub-analysis concerning studies with patients who presented more than two types of side effects						
Sub-group: overall toxicity – more than one type of toxicity*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	C vs. A allele	12	1.099	0.737	1.640	77.6
	CC vs. AC+AA	12	0.915	0.476	1.759	67.2
	AC+CC vs. AA	12	1.415	0.913	2.193	67.3
Caucasians	C vs. A allele	5	1.433	0.757	2.711	62.1
	CC vs. AC+AA	7	0.843	0.357	1.992	59.6
	AC+CC vs. AA	4	2.111	0.722	6.169	70.0

Asians	C vs. A allele	3	0.948	0.578	1.553	0.0
	CC vs. AC+AA	2	0.751	0.162	3.482	0.0
	AC+CC vs. AA	4	1.082	0.689	1.699	0.0
Other	C vs. A allele	4	0.945	0.444	2.009	90.8
	CC vs. AC+AA	3	1.035	0.275	3.893	85.6
	AC+CC vs. AA	4	1.319	0.557	3.120	83.3
HWE	C vs. A allele	6	1.346	0.951	1.905	48.6
	CC vs. AC+AA	6	1.097	0.671	1.794	0.0
	AC+CC vs. AA	6	1.700	0.895	3.229	70.4
* The sub-analysis concerning studies with patients who presented more than one type of side effects						
Sub-group: Rheumatoid arthritis*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
All	C vs. A allele	5	0.645	0.304	1.365	77.7
	CC vs. AC+AA	4	0.506	0.273	0.936	9.5
	AC+CC vs. AA	5	1.049	0.588	1.871	61.4
Caucasians	C vs. A allele	2	0.390	0.018	8.512	84.5
	CC vs. AC+AA	2	0.655	0.326	1.317	0.0
Asians	C vs. A allele	2	0.824	0.470	1.444	0.0
	AC+CC vs. AA	3	0.995	0.614	1.611	0.0
HWE	C vs. A allele	2	1.286	0.820	2.017	27.8
	CC vs. AC+AA	2	0.803	0.349	1.847	0.0
	AC+CC vs. AA	2	1.638	0.709	3.782	63.3
* The sub-analysis concerning studies with rheumatoid arthritis patients						
Sub-group: Folate supplementation*						
Category	Contrast	Number of studies	Odds ratio (random effects)	95% confidence interval		P (%)
No folate supplementation	C vs. A allele	5	1.504	0.841	2.689	57.3
	CC vs. AC+AA	6	1.417	0.364	5.522	77.2
	AC+CC vs. AA	5	1.926	1.098	3.379	25.6
Partial folate supplementation	C vs. A allele	6	0.821	0.464	1.452	79.1
	CC vs. AC+AA	5	0.632	0.312	1.283	42.9
	AC+CC vs. AA	6	1.078	0.605	1.921	66.7
Full folate supplementation	C vs. A allele	1	1.520	0.988	2.336	-
	CC vs. AC+AA	1	0.880	0.340	2.273	-
	AC+CC vs. AA	1	2.425	1.289	4.560	-
* The sub-analysis based on folate supplementation status of the patients						

Παράρτημα Ε

Οι εκτιμήσεις της διασποράς βασίζονται στον Woolf estimator (Woolf 1955) και υπολογίστηκαν αρχικά χρησιμοποιώντας την προσέγγιση της διωνυμικής από την κανονική κατανομή για μεγάλο μέγεθος δείγματος. Οι διασπορές αυτές μπορούν να υπολογιστούν επίσης, μέσω μιας πρόσφατης μεθοδολογίας (Bagos 2012) θεωρώντας ότι οι ποσότητες των πινάκων συνάφειας είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με $E[n_{clpi}] = \text{var}[n_{clpi}] = m_{clpi}$ και τα logits της ευαισθησίας και ειδικότητας αποτελούν συγκρίσεις μεταξύ των log counts (Berrington and Cox 2003). Για τον υπολογισμό των συνδιακυμάνσεων εργαζόμαστε ως ακολούθως:

$$\begin{aligned}
 \text{cov}(\hat{y}_{1i}, \hat{y}_{3i}) &= \text{cov}\left(\log\left(\frac{TP_{1i}}{FN_{1i}}\right), \log\left(\frac{TP_{2i}}{FN_{2i}}\right)\right) \\
 &= \text{cov}(\log(TP_{1i}) - \log(FN_{1i}), \log(TP_{2i}) - \log(FN_{2i})) \\
 &= \text{cov}(\log(TP_{1i}), \log(TP_{2i})) - \text{cov}(\log(TP_{1i}), \log(FN_{2i})) \\
 &\quad - \text{cov}(\log(FN_{1i}), \log(TP_{2i})) + \text{cov}(\log(FN_{1i}), \log(FN_{2i})) \\
 &= \text{cov}(\log(n_{11+i}), \log(n_{1+1i})) - \text{cov}(\log(n_{11+i}), \log(n_{1+0i})) \\
 &\quad - \text{cov}(\log(n_{10+i}), \log(n_{1+1i})) + \text{cov}(\log(n_{10+i}), \log(n_{1+0i})) \quad (E1) \\
 &= \text{cov}(\log(n_{111i} + n_{110i}), \log(n_{111i} + n_{101i})) \\
 &\quad - \text{cov}(\log(n_{111i} + n_{110i}), \log(n_{110i} + n_{100i})) \\
 &\quad - \text{cov}(\log(n_{101i} + n_{100i}), \log(n_{111i} + n_{101i})) \\
 &\quad + \text{cov}(\log(n_{101i} + n_{100i}), \log(n_{110i} + n_{100i}))
 \end{aligned}$$

Παρατηρούμε ότι το δεξί μέρος της εξίσωσης (E1) περιλαμβάνει όρους της μορφής $\text{cov}[\log(S+U), \log(S+W)]$ (όπου οι S , U και W είναι ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με παραμέτρους s , u και w , αντίστοιχα), που πρέπει να υπολογιστούν. Ισχύει ότι (Bagos 2012):

$$\text{cov}[\log(S+U), \log(S+W)] = \frac{s}{(s+u)(s+w)} \quad (E2)$$

Τελικά, η εξίσωση E2, μπορεί να χρησιμοποιηθεί για την επαναδιατύπωση της εξίσωσης E1 διαμορφώνοντας έτσι μια προσεγγιστική εκτίμηση της συνδιακύμανσης:

$$\begin{aligned}
\text{cov}(\hat{y}_{1i}, \hat{y}_{3i}) &= \frac{n_{111i}}{(n_{111i} + n_{110i})(n_{111i} + n_{101i})} - \frac{n_{110i}}{(n_{111i} + n_{110i})(n_{110i} + n_{100i})} \\
&\quad - \frac{n_{101i}}{(n_{101i} + n_{100i})(n_{111i} + n_{101i})} + \frac{n_{100i}}{(n_{101i} + n_{100i})(n_{110i} + n_{100i})} \\
&= \sum_l \sum_p (-1)^{l-p} \left(\frac{n_{lpi}}{n_{l+i} n_{l+pi}} \right)
\end{aligned} \tag{E3}$$

Παρόμοια, για τον υπολογισμό της συνδιακύμανσης $\text{cov}(\hat{y}_{2i}, \hat{y}_{4i})$ ισχύει ότι:

$$\begin{aligned}
\text{cov}(\hat{y}_{2i}, \hat{y}_{4i}) &= \text{cov} \left(\log \left(\frac{FP_{1i}}{TN_{1i}} \right), \log \left(\frac{FP_{2i}}{TN_{2i}} \right) \right) \\
&= \text{cov}(\log(FP_{1i}), \log(FP_{2i})) - \text{cov}(\log(FP_{1i}), \log(TN_{2i})) \\
&\quad - \text{cov}(\log(TN_{1i}), \log(FP_{2i})) + \text{cov}(\log(TN_{1i}), \log(TN_{2i})) \\
&= \text{cov}(\log(n_{011i} + n_{010i}), \log(n_{011i} + n_{001i})) \\
&\quad - \text{cov}(\log(n_{011i} + n_{010i}), \log(n_{010i} + n_{000i})) \\
&\quad - \text{cov}(\log(n_{001i} + n_{000i}), \log(n_{011i} + n_{001i})) \\
&\quad + \text{cov}(\log(n_{001i} + n_{000i}), \log(n_{010i} + n_{000i})) \\
&= \frac{n_{011i}}{(n_{011i} + n_{010i})(n_{011i} + n_{001i})} - \frac{n_{010i}}{(n_{011i} + n_{010i})(n_{010i} + n_{000i})} \\
&\quad - \frac{n_{001i}}{(n_{001i} + n_{000i})(n_{011i} + n_{001i})} + \frac{n_{000i}}{(n_{001i} + n_{000i})(n_{010i} + n_{000i})} \\
&= \sum_l \sum_p (-1)^{l-p} \left(\frac{n_{lpi}}{n_{l+i} n_{l+pi}} \right)
\end{aligned} \tag{E4}$$

Για τον υπολογισμό της συνδιακύμανσης $\text{cov}(\hat{y}_{1i}, \hat{y}_{4i})$ ισχύει ότι:

$$\begin{aligned}
 \text{cov}(\hat{y}_{1i}, \hat{y}_{4i}) &= \text{cov}\left(\log\left(\frac{TP_{1i}}{FN_{1i}}\right), \log\left(\frac{FP_{2i}}{TN_{2i}}\right)\right) \\
 &= \text{cov}(\log(TP_{1i}), \log(FP_{2i})) - \text{cov}(\log(TP_{1i}), \log(TN_{2i})) \\
 &\quad - \text{cov}(\log(FN_{1i}), \log(FP_{2i})) + \text{cov}(\log(FN_{1i}), \log(TN_{2i})) \\
 &= \text{cov}(\log(n_{111i} + n_{110i}), \log(n_{011i} + n_{001i})) \\
 &\quad - \text{cov}(\log(n_{111i} + n_{110i}), \log(n_{010i} + n_{000i})) \\
 &\quad - \text{cov}(\log(n_{101i} + n_{100i}), \log(n_{011i} + n_{001i})) \\
 &\quad + \text{cov}(\log(n_{101i} + n_{100i}), \log(n_{010i} + n_{000i})) \\
 &= 0
 \end{aligned}
 \tag{E5}$$

Για τον υπολογισμό της συνδιακύμανσης $\text{cov}(\hat{y}_{3i}, \hat{y}_{2i})$ ισχύει ότι:

$$\begin{aligned}
 \text{cov}(\hat{y}_{3i}, \hat{y}_{2i}) &= \text{cov}\left(\log\left(\frac{TP_{2i}}{FN_{2i}}\right), \log\left(\frac{FP_{1i}}{TN_{1i}}\right)\right) \\
 &= \text{cov}(\log(TP_{2i}), \log(FP_{1i})) - \text{cov}(\log(TP_{2i}), \log(TN_{1i})) \\
 &\quad - \text{cov}(\log(FN_{2i}), \log(FP_{1i})) + \text{cov}(\log(FN_{2i}), \log(TN_{1i})) \\
 &= \text{cov}(\log(n_{111i} + n_{101i}), \log(n_{011i} + n_{010i})) \\
 &\quad - \text{cov}(\log(n_{111i} + n_{101i}), \log(n_{001i} + n_{000i})) \\
 &\quad - \text{cov}(\log(n_{110i} + n_{100i}), \log(n_{011i} + n_{010i})) \\
 &\quad + \text{cov}(\log(n_{110i} + n_{100i}), \log(n_{001i} + n_{000i})) \\
 &= 0
 \end{aligned}
 \tag{E6}$$

Οι δυο τελευταίες εξισώσεις E5-E6 είναι ίσες με μηδέν καθώς περιλαμβάνουν όρους της μορφής $\text{cov}[\log(S+U), \log(V+W)]$, οι οποίοι είναι ίσοι με το μηδέν.

Παράρτημα ΣΤ

Για να υπολογίσουμε τις ποσότητες των κελιών του Πίνακα 6.1 δοθέντων των γνωστών περιθωρίων αθροισμάτων και των conditional Odds Ratios, εργαζόμαστε ως ακολούθως. Με τις εξισώσεις ΣΤ1-10, απεικονίζονται τα γνωστά περιθώρια αθροίσματα καθώς επίσης και τα conditional Odds Ratios για τη συσχέτιση των δυο διαγνωστικών δοκιμασιών ανάλογα με την παρουσία ή απουσία της νόσου:

$$TP_{1i} = n_{11+i} = n_{111i} + n_{110i} \quad (\Sigma T1)$$

$$FN_{1i} = n_{10+i} = n_{101i} + n_{100i} \quad (\Sigma T2)$$

$$TN_{1i} = n_{00+i} = n_{001i} + n_{000i} \quad (\Sigma T3)$$

$$FP_{1i} = n_{01+i} = n_{011i} + n_{010i} \quad (\Sigma T4)$$

$$TP_{2i} = n_{11+i} = n_{111i} + n_{110i} \quad (\Sigma T5)$$

$$FN_{2i} = n_{10+i} = n_{110i} + n_{100i} \quad (\Sigma T6)$$

$$TN_{2i} = n_{00+i} = n_{010i} + n_{000i} \quad (\Sigma T7)$$

$$FP_{2i} = n_{01+i} = n_{011i} + n_{001i} \quad (\Sigma T8)$$

$$\widehat{OR}_{Y_i=1} = \frac{n_{111i}n_{100i}}{n_{101i}n_{110i}} \quad (\Sigma T9)$$

$$\widehat{OR}_{Y_i=0} = \frac{n_{011i}n_{00i}}{n_{010i}n_{001i}} \quad (\Sigma T10)$$

Έτσι, συνδυάζοντας τις εξισώσεις ΣΤ1-10 και λύνοντας δυο συστήματα 5 εξισώσεων με 4 αγνώστους, οι ποσότητες των κελιών της από κοινού κατανομής των δυο διαγνωστικών δοκιμασιών για τα άτομα που έχουν ή όχι τη νόσο, υπολογίζονται ως εξής:

$$n_{111i} = \frac{-\sqrt{\left(-TP_{1i}\widehat{OR}_{Y_i=1} - FN_{1i} - TP_{2i}\widehat{OR}_{Y_i=1} + TP_{2i}\right)^2 - 4TP_{1i}TP_{2i}\left(\widehat{OR}_{Y_i=1} - 1\right)\widehat{OR}_{Y_i=1}}}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} + \frac{\left(TP_{1i}\widehat{OR}_{Y_i=1} + FN_{1i} + TP_{2i}\widehat{OR}_{Y_i=1} - TP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} \quad (\Sigma T11)$$

$$n_{110i} = \frac{\sqrt{\left(-TP_{1i}\widehat{OR}_{Y_i=1} - FN_{1i} - TP_{2i}\widehat{OR}_{Y_i=1} + TP_{2i}\right)^2 - 4TP_{1i}TP_{2i}\left(\widehat{OR}_{Y_i=1} - 1\right)\widehat{OR}_{Y_i=1}}}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} + \frac{\left(TP_{1i}\widehat{OR}_{Y_i=1} - 2TP_{1i} - FN_{1i} - TP_{2i}\widehat{OR}_{Y_i=1} + TP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} \quad (\Sigma T12)$$

$$n_{101i} = \frac{\sqrt{\left(-TP_{1i}\widehat{OR}_{Y_i=1} - FN_{1i} - TP_{2i}\widehat{OR}_{Y_i=1} + TP_{2i}\right)^2 - 4TP_{1i}TP_{2i}\left(\widehat{OR}_{Y_i=1} - 1\right)\widehat{OR}_{Y_i=1}}}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} + \frac{\left(-TP_{1i}\widehat{OR}_{Y_i=1} - FN_{1i} + TP_{2i}\widehat{OR}_{Y_i=1} - TP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} \quad (\Sigma T13)$$

$$n_{100i} = \frac{-\sqrt{\left(-TP_{1i}\widehat{OR}_{Y_i=1} - FN_{1i} - TP_{2i}\widehat{OR}_{Y_i=1} + TP_{2i}\right)^2 - 4TP_{1i}TP_{2i}\left(\widehat{OR}_{Y_i=1} - 1\right)\widehat{OR}_{Y_i=1}}}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} + \frac{\left(TP_{1i}\widehat{OR}_{Y_i=1} + 2FN_{1i}\widehat{OR}_{Y_i=1} - FN_{1i} - TP_{2i}\widehat{OR}_{Y_i=1} + TP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=1} - 1\right)} \quad (\Sigma T14)$$

$$n_{011i} = \frac{-\sqrt{\left(-FP_{1i}\widehat{OR}_{Y_i=0} - TN_{1i} - FP_{2i}\widehat{OR}_{Y_i=0} + FP_{2i}\right)^2 - 4FP_{1i}FP_{2i}\left(\widehat{OR}_{Y_i=0} - 1\right)\widehat{OR}_{Y_i=0}}}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} + \frac{\left(FP_{1i}\widehat{OR}_{Y_i=0} + TN_{1i} + FP_{2i}\widehat{OR}_{Y_i=0} - FP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} \quad (\Sigma T15)$$

$$n_{010i} = \frac{\sqrt{\left(-FP_{1i} \widehat{OR}_{Y_i=0} - TN_{1i} - FP_{2i} \widehat{OR}_{Y_i=0} + FP_{2i}\right)^2 - 4FP_{1i}FP_{2i} \left(\widehat{OR}_{Y_i=0} - 1\right) \widehat{OR}_{Y_i=0}}}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} + \frac{\left(FP_{1i} \widehat{OR}_{Y_i=0} - 2FP_{1i} - TN_{1i} - FP_{2i} \widehat{OR}_{Y_i=0} + FP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} \quad (\Sigma T16)$$

$$n_{001i} = \frac{\sqrt{\left(-FP_{1i} \widehat{OR}_{Y_i=0} - TN_{1i} - FP_{2i} \widehat{OR}_{Y_i=0} + FP_{2i}\right)^2 - 4FP_{1i}FP_{2i} \left(\widehat{OR}_{Y_i=0} - 1\right) \widehat{OR}_{Y_i=0}}}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} + \frac{\left(-FP_{1i} \widehat{OR}_{Y_i=0} - TN_{1i} + FP_{2i} \widehat{OR}_{Y_i=0} - FP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} \quad (\Sigma T17)$$

$$n_{000i} = \frac{-\sqrt{\left(-FP_{1i} \widehat{OR}_{Y_i=0} - TN_{1i} - FP_{2i} \widehat{OR}_{Y_i=0} + FP_{2i}\right)^2 - 4FP_{1i}FP_{2i} \left(\widehat{OR}_{Y_i=0} - 1\right) \widehat{OR}_{Y_i=0}}}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} + \frac{\left(FP_{1i} \widehat{OR}_{Y_i=0} + 2TN_{1i} \widehat{OR}_{Y_i=0} - TN_{1i} - FP_{2i} \widehat{OR}_{Y_i=0} + FP_{2i}\right)}{2\left(\widehat{OR}_{Y_i=0} - 1\right)} \quad (\Sigma T18)$$

Όταν $\widehat{OR}_{Y_i=1}$ ή $\widehat{OR}_{Y_i=0}$ ισούται με 1, αντί για τις εξισώσεις ΣΤ11-18 χρησιμοποιούνται οι ακόλουθες σχέσεις:

$$n_{111i} = \frac{TP_{1i}TP_{2i}}{n_{1++i}} \quad (\Sigma T19)$$

$$n_{110i} = \frac{TP_{1i}(TP_{1i} + FN_{1i} - TP_{2i})}{n_{1++i}} \quad (\Sigma T20)$$

$$n_{101i} = \frac{FN_{1i}TP_{2i}}{n_{1++i}}$$

(ΣT21)

$$n_{100i} = \frac{TP_{1i}FN_{1i} + FN_{1i}^2 - FN_{1i}TP_{2i}}{n_{1++i}}$$

(ΣT22)

$$n_{011i} = \frac{FP_{1i}FP_{2i}}{n_{0++i}}$$

(ΣT23)

$$n_{010i} = \frac{FP_{1i}(FP_{1i} + TN_{1i} - FP_{2i})}{n_{0++i}}$$

(ΣT24)

$$n_{001i} = \frac{TN_{1i}FP_{2i}}{n_{0++i}}$$

(ΣT25)

$$n_{000i} = \frac{FP_{1i}TN_{1i} + TN_{1i}^2 - TN_{1i}FP_{2i}}{n_{0++i}}$$

(ΣT26)

Παράρτημα Z

Για να αξιολογήσουμε την προτεινόμενη μεθοδολογία πραγματοποιήσαμε μελέτες προσομοίωσης. Χρησιμοποιήσαμε ρεαλιστικές παραμέτρους κοντά σε αυτές που εκτιμώνται από πραγματικά δεδομένα ($\widehat{TPR}_1 = 0.8$, $\widehat{FPR}_1 = 0.2$, $\widehat{TPR}_2 = 0.8$ και 0.9 , $\widehat{FPR}_2 = 0.1$ και 0.2 , $\widehat{OR}_{Y=1} = 8$ και 10 , $\widehat{OR}_{Y=0} = 10$ και 12). Πραγματοποιήσαμε 500 μετα-αναλύσεις με 20 μελέτες σε κάθε μετα-ανάλυση. Για τις 20 αυτές μελέτες, οι 6 παρείχαν εκτιμήσεις και για τις δυο διαγνωστικές δοκιμασίες, ενώ από τις υπόλοιπες 14 οι μισές είχαν εκτιμήσεις μόνο για τη μια είτε μόνο για την άλλη δοκιμασία. Ο επιπολασμός της νόσου ήταν 50% με 200 άτομα σε κάθε μελέτη. Οι ποσότητες των κελιών του Πίνακα 6.1 προέρχονται από μια multinomial κατανομή με πιθανότητες που δίνονται από τις εξισώσεις ST11-18. Σε όλες τις αναλύσεις, η μεροληψία, η ακρίβεια, η 95% πιθανότητα κάλυψης και η ισχύς (bias, precision, Coverage of 95% CI, power) υπολογίστηκαν χρησιμοποιώντας την εντολή `simsum` στο Stata (White 2010).

Οι εκτιμήσεις των παραμέτρων του μοντέλου (β_1 , β_2 , β_3 και β_4), ήταν προσεγγιστικά αμερόληπτες τόσο για την προτεινόμενη μεθοδολογία όσο και για την υπάρχουσα διμεταβλητή προσέγγιση. Ωστόσο, η σχετική μεροληψία στις εκτιμήσεις των παραμέτρων ήταν συστηματικά μικρότερη στην προτεινόμενη πολυμεταβλητή προσέγγιση, με κάποια σενάρια να παρουσιάζει μείωση της τάξης του 60% ή ακόμα μεγαλύτερη. Το empirical standard error και mean-square error των εκτιμήσεων ήταν συγκρίσιμο και για τις δυο προσεγγίσεις. Η κάλυψη των εκτιμήσεων των παραμέτρων ($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ και $\hat{\beta}_4$) κυμαινόταν μεταξύ του 94% και 98% στα περισσότερα σενάρια που αξιολογήθηκαν και ήταν συχνά παρόμοια και για τις δυο προσεγγίσεις. Η διαφορά στις εκτιμήσεις των παραμέτρων ($\hat{\beta}_1 - \hat{\beta}_3$, $\hat{\beta}_2 - \hat{\beta}_4$) είναι επίσης σημαντική, καθώς αποτελεί έναν επίσημο τρόπο να συγκρίνονται τα αποτελέσματα των δυο διαγνωστικών δοκιμασιών. Το empirical standard error και mean-square error των διαφορών ήταν ελαφρώς μικρότερο στην πολυμεταβλητή προσέγγιση. Η ισχύς για να ανιχνευτούν διαφορές στις εκτιμήσεις των παραμέτρων ($\hat{\beta}_1 - \hat{\beta}_3$, $\hat{\beta}_2 - \hat{\beta}_4$) ήταν 100% και για τις δυο μεθόδους όταν οι πραγματικές διαφορές είναι μεγάλες (δηλαδή, $\widehat{TPR}_1 = 0.8$ και $\widehat{TPR}_2 = 0.9$), αλλά η πολυμεταβλητή προσέγγιση έχει πάντα μεγαλύτερη ισχύ να ανιχνεύσει μικρές διαφορές (1-5%) στην \widehat{TPR} και \widehat{FPR} , ένα κέρδος στην ισχύ που

φτάνει το 20%. Επιπλέον, η πολυμεταβλητή προσέγγιση έχει σχεδόν σε όλες τις περιπτώσεις σφάλμα τύπου I κοντά στο 5%, ενώ συστηματικά μικρότερες τιμές υπολογίζονται υπό τη διμεταβλητή προσέγγιση.

Αξιολογήθηκε επίσης η εκτίμηση των παραμέτρων της SROC curve (a_j, b_j) και της AUC . Για απλοποίηση, οι προσομοιώσεις πραγματοποιήθηκαν υποθέτοντας ομοιογένεια, και έτσι χρησιμοποιήσαμε την αναλυτική 6.26 έτσι ώστε να υπολογιστεί η AUC . Όπως έχει ήδη αναφερθεί, η εξίσωση 6.26 παρέχει μια καλή προσέγγιση της AUC ακόμα για την περίπτωση των ετερογενών μελετών. Οι παράμετροι της SROC curve (\hat{a}_j, \hat{b}_j) ήταν προσεγγιστικά αμερόληπτες τόσο για την προτεινόμενη πολυμεταβλητή όσο και για την υπάρχουσα διμεταβλητή προσέγγιση. Ωστόσο, η μεροληψία στην \widehat{AUC} ήταν συστηματικά χαμηλότερη στην πολυμεταβλητή προσέγγιση, για όλα τα σενάρια που αξιολογήθηκαν. Τα empirical standard error και mean-square error των εκτιμήσεων ήταν κοντινά και για τις δυο προσεγγίσεις. Η διαφορά των \widehat{AUC} των δυο διαγνωστικών δοκιμασιών είναι εξαιρετικά σημαντική καθώς αποτελεί έναν επίσημο τρόπο να συγκρίνουμε τις δυο καμπύλες, και έτσι τη συνολική ακρίβεια των δυο διαγνωστικών δοκιμασιών. Πιο συγκεκριμένα, το empirical standard error και the mean-square error των διαφορών, ήταν πάντα μικρότερο υπό την πολυμεταβλητή προσέγγιση. Η ισχύς για να ανιχνεύσουμε διαφορές στην \widehat{AUC} ($\widehat{AUC}_1 - \widehat{AUC}_2$) ήταν 100% και για τις δυο μεθόδους όταν οι πραγματικές διαφορές ήταν πάνω από 10%, αλλά η πολυμεταβλητή μέθοδος έχει πάντα μεγαλύτερη ισχύ να ανιχνεύσει μικρές διαφορές μεταξύ των \widehat{AUC}_1 και \widehat{AUC}_2 . Το κέρδος στην ισχύ ήταν πιο εμφανές για τιμές της \widehat{AUC} κοντά στο 0.6 και 0.7 όπου σε κάποιες περιπτώσεις φτάνει το 12% για να ανιχνεύσει πραγματικές διαφορές εύρους 3-5% (Εικόνα 6.1).

Όταν οι προσομοιώσεις επαναλήφθηκαν για μεγαλύτερο αριθμό ατόμων ανά μελέτη ($n=500, 1000$), και χωρίς ελλείπουσες παρατηρήσεις (δηλαδή όταν όλες οι μελέτες αναφέρουν αποτελέσματα και για τις δυο διαγνωστικές δοκιμασίες), οι διαφορές των δυο προσεγγίσεων ήταν λιγότερο εμφανείς, όπως θα ανέμενε κάποιος θεωρητικά. Αξιολογήσαμε επίσης τις συνέπειες του model misspecification όταν υπάρχει αλληλεπίδραση τρίτης τάξης αλλά χρησιμοποιείται ο αλγόριθμος IPF που δεν υποθέτει κάποια τέτοια δομή. Σε όλες τις αναλύσεις, το μη σωστά ορισμένο μοντέλο οδηγεί σε αυξημένη μεροληψία, empirical standard error και mean-square error των

εκτιμήσεων και ελαφρώς μειωμένη ισχύ σε σχέση με το ορθό μοντέλο, αλλά εξακολουθεί να είναι καλύτερο σε σχέση με τη διμεταβλητή προσέγγιση.

Πίνακας Z1. Μελέτη προσομοίωσης για την αξιολόγηση και σύγκριση της πολυμεταβλητής μεθοδολογίας μετα-ανάλυσης διαγνωστικών δοκιμασιών με την κλασσική διμεταβλητή προσέγγιση. Παρουσιάζονται 5 σενάρια ($\widehat{TPR}_1 = 0.8$, $\widehat{FPR}_1 = 0.2$, $\widehat{TPR}_2 = 0.8$, 0.85 και 0.9, $\widehat{FPR}_2 = 0.1$, 0.15 και 0.2, $\widehat{OR}_{Y=1} = 8$, $\widehat{OR}_{Y=0} = 12$). Πραγματοποιήσαμε 500 επαναλήψεις με 20 μελέτες σε κάθε μετα-ανάλυση. Ο επιπολασμός της νόσου ήταν 50% με 200 άτομα σε κάθε μελέτη.

Statistic	$\hat{\beta}_1$ Multivariate	$\hat{\beta}_1$ Bivariate	$\hat{\beta}_2$ Multivariate	$\hat{\beta}_2$ Bivariate	$\hat{\beta}_3$ Multivariate	$\hat{\beta}_3$ Bivariate	$\hat{\beta}_4$ Multivariate	$\hat{\beta}_4$ Bivariate	$\hat{\beta}_1 - \hat{\beta}_3$ Multivariate	$\hat{\beta}_1 - \hat{\beta}_3$ Bivariate	$\hat{\beta}_2 - \hat{\beta}_4$ Multivariate	$\hat{\beta}_2 - \hat{\beta}_4$ Bivariate
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.8$ $\widehat{FPR}_2 = 0.1$ $\widehat{OR}_{Y=1} = 8$ $\widehat{OR}_{Y=0} = 12$												
Bias in point estimate	-0.0056	-0.0083	-0.0005	0.0029	-0.0101	-0.0127	0.0292	0.0324	0.0045	0.0044	-0.0296	-0.0295
Empirical standard error	0.0697	0.0704	0.0679	0.0677	0.0688	0.0700	0.0885	0.0895	0.0866	0.0879	0.0963	0.0983
RMS model-based standard error	0.0790	0.0762	0.0781	0.0757	0.0780	0.0768	0.1008	0.0996	0.1000	0.1082	0.1158	0.1251
Coverage of nominal 95% confidence interval	96.6	96.2	97	97	96.8	95.6	95.8	95.2	96.6	98	97.8	97.2
Power of 5% level test	100	100	100	100	100	100	100	100	3.4	2	100	100
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.8$ $\widehat{FPR}_2 = 0.2$ $\widehat{OR}_{Y=1} = 8$ $\widehat{OR}_{Y=0} = 12$												
Bias in point estimate	-0.0055	-0.0075	0.0057	0.0068	-0.0081	-0.0101	0.0024	0.0042	0.0026	0.0027	0.0033	0.0026
Empirical standard error	0.0655	0.0659	0.0684	0.0701	0.0698	0.0695	0.0655	0.0651	0.0815	0.0816	0.0796	0.0810
RMS model-based standard error	0.0785	0.0765	0.0770	0.0755	0.0777	0.0759	0.0776	0.0757	0.0996	0.1078	0.0968	0.1069
Coverage of nominal 95% confidence interval	97	96.6	96.6	96	97.8	96.4	97.2	97.2	98	98.4	96.6	98
Power of 5% level test	100	100	100	100	100	100	100	100	2	1.6	3.4	2
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.9$ $\widehat{FPR}_2 = 0.1$ $\widehat{OR}_{Y=1} = 8$ $\widehat{OR}_{Y=0} = 12$												
Bias in point estimate	-0.0082	-0.0127	0.0065	0.0101	-0.0236	-0.0267	0.0251	0.0277	0.0154	0.0140	-0.0185	-0.0176
Empirical standard error	0.0689	0.0691	0.0689	0.0692	0.0932	0.0920	0.0859	0.0861	0.1085	0.1073	0.0974	0.0982
RMS model-based standard error	0.0800	0.0756	0.0791	0.0755	0.1025	0.1009	0.1024	0.1012	0.1213	0.1261	0.1181	0.1262

Coverage of nominal 95% confidence interval	96.8	95	98.2	97.2	95.2	94.8	97.4	97	95.8	98	98	98.8
Power of 5% level test	100	100	100	100	100	100	100	100	100	100	100	100
$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.9	$\widehat{FPR}_2 =$ 0.2	$\widehat{OR}_{Y=1} =$ 8	$\widehat{OR}_{Y=0} =$ 12							
Bias in point estimate	-0.0108	-0.0132	0.0076	0.0088	-0.0252	-0.0286	0.0101	0.0122	0.0143	0.0155	-0.0024	-0.0034
Empirical standard error	0.0711	0.0711	0.0676	0.0682	0.0917	0.0928	0.0720	0.0728	0.1018	0.1036	0.0874	0.0898
RMS model-based standard error	0.0788	0.0759	0.0793	0.0774	0.1022	0.1006	0.0786	0.0776	0.1196	0.1260	0.0983	0.1096
Coverage of nominal 95% confidence interval	95.8	94.6	97.4	96.6	96.4	94.8	96	96	98.2	98.4	97.2	98
Power of 5% level test	100	100	100	100	100	100	100	100	100	100	2.8	2
$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.85	$\widehat{FPR}_2 =$ 0.15	$\widehat{OR}_{Y=1} =$ 8	$\widehat{OR}_{Y=0} =$ 12							
Bias in point estimate	-0.0047	-0.0070	0.0120	0.0137	-0.0146	-0.0164	0.0143	0.0170	0.0100	0.0094	-0.0023	-0.0033
Empirical standard error	0.0698	0.0705	0.0697	0.0699	0.0724	0.0731	0.0774	0.0784	0.0903	0.0908	0.0897	0.0910
RMS model-based standard error	0.0788	0.0764	0.0789	0.0769	0.0869	0.0853	0.0862	0.0852	0.1067	0.1145	0.1039	0.1147
Coverage of nominal 95% confidence interval	97.8	96	97	95.6	97.4	96.2	97.4	95.8	98.2	99.2	97.4	98.2
Power of 5% level test	100	100	100	100	100	100	100	100	93.2	88.6	93.8	91.4

Πίνακας Z2. Μελέτη προσομοίωσης για την αξιολόγηση και σύγκριση της πολυμεταβλητής μεθοδολογίας μετα-ανάλυσης διαγνωστικών δοκιμασιών με την κλασσική διμεταβλητή προσέγγιση για την εκτίμηση της AUC . Παρουσιάζονται 5 σενάρια ($\widehat{TPR}_1 = 0.8$, $\widehat{FPR}_1 = 0.2$, $\widehat{TPR}_2 = 0.8$, 0.85 και 0.9 , $\widehat{FPR}_2 = 0.1$, 0.15 και 0.2 , $\widehat{OR}_{Y=1} = 8$, $\widehat{OR}_{Y=0} = 12$). Πραγματοποιήσαμε 500 επαναλήψεις με 20 μελέτες σε κάθε μετα-ανάλυση. Ο επιπολασμός της νόσου ήταν 50% με 200 άτομα σε κάθε μελέτη. Για απλούστευση, χρησιμοποιήσαμε την αναλυτική έκφραση 6.26 έτσι ώστε να υπολογίσουμε την \widehat{AUC} υποθέτοντας ότι οι μελέτες μας είναι ομοιογενείς.

Statistic	\widehat{AUC}_1	\widehat{AUC}_1	\widehat{AUC}_2	\widehat{AUC}_2	$\widehat{AUC}_1 - \widehat{AUC}_2$	$\widehat{AUC}_1 - \widehat{AUC}_2$
	Multivariate	Bivariate	Multivariate	Bivariate	Multivariate	Bivariate
$\widehat{TPR}_1 = 0.8$	$\widehat{FPR}_1 = 0.2$	$\widehat{TPR}_2 = 0.8$	$\widehat{FPR}_2 = 0.1$	$\widehat{AUC}_1 = 0.870$	$\widehat{AUC}_2 = 0.923$	
Bias in point estimate	-0.0015	-0.0021	-0.0019	-0.0021	0.0004	-0.0005
Empirical standard error	0.0082	0.0083	0.0061	0.0063	0.0090	0.0093
RMS model-based standard error	0.0092	0.0089	0.0069	0.0068	0.0104	0.0112
Coverage of nominal 95% confidence interval	97.4	95.8	96.8	95.6	96.4	97.6
Power of 5% level test	100	100	100	100	99.8	99.8
$\widehat{TPR}_1 = 0.8$	$\widehat{FPR}_1 = 0.2$	$\widehat{TPR}_2 = 0.8$	$\widehat{FPR}_2 = 0.2$	$\widehat{AUC}_1 = 0.870$	$\widehat{AUC}_2 = 0.870$	
Bias in point estimate	-0.0008	-0.0011	-0.0020	-0.0024	0.0011	0.0013
Empirical standard error	0.0079	0.0081	0.0077	0.0078	0.0097	0.0099
RMS model-based standard error	0.0089	0.0088	0.0089	0.0088	0.0112	0.0124
Coverage of nominal 95% confidence interval	98.2	96.6	95.6	95.6	97	98.4
Power of 5% level test	100	100	100	100	3	1.6

$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.9	$\widehat{FPR}_2 =$ 0.1	$\widehat{AUC}_1 =$ 0.870	$\widehat{AUC}_2 =$ 0.957	
Bias in point estimate	-0.0012	-0.0019	-0.0018	-0.0021	0.0006	0.0001
Empirical standard error	0.0078	0.0078	0.0042	0.0042	0.0084	0.0084
RMS model-based standard error	0.0092	0.0088	0.0048	0.0048	0.0097	0.0100
Coverage of nominal 95% confidence interval	97.4	96.6	97	96.2	96.4	96.8
Power of 5% level test	100	100	100	100	100	100
$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.9	$\widehat{FPR}_2 =$ 0.2	$\widehat{AUC}_1 =$ 0.870	$\widehat{AUC}_2 =$ 0.923	
Bias in point estimate	-0.0005	-0.0009	-0.0018	-0.0022	0.0014	0.0013
Empirical standard error	0.0077	0.0080	0.0064	0.0064	0.0087	0.0090
RMS model-based standard error	0.0091	0.0089	0.0069	0.0068	0.0104	0.0112
Coverage of nominal 95% confidence interval	97.4	96.4	96.4	94.6	97.6	97.6
Power of 5% level test	100	100	100	100	100	100
$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.85	$\widehat{FPR}_2 =$ 0.15	$\widehat{AUC}_1 =$ 0.870	$\widehat{AUC}_2 =$ 0.917	
Bias in point estimate	-0.0016	-0.0021	-0.0023	-0.0025	0.0007	0.0005
Empirical standard error	0.0080	0.0082	0.0059	0.0061	0.0088	0.0093
RMS model-based standard error	0.0091	0.0089	0.0070	0.0069	0.0104	0.0113
Coverage of nominal 95% confidence interval	96.6	96.2	97.2	96.4	97.6	98.6
Power of 5% level test	100	100	100	100	100	99.8

Πίνακας Z3. Μελέτη προσομοίωσης για την αξιολόγηση των συνεπειών του model misspecification όταν υπάρχουν αλληλεπιδράσεις τρίτης τάξης μεταξύ των μελετών αλλά δεν λαμβάνονται υπόψη με τη χρήση του αλγορίθμου IPF. Παρουσιάζονται 5 σενάρια ($\widehat{TPR}_1 = 0.8$, $\widehat{FPR}_1 = 0.2$, $\widehat{TPR}_2 = 0.8$, 0.85 και 0.9, $\widehat{FPR}_2 = 0.1$, 0.15 και 0.2, $\widehat{OR}_{Y=1} = 8$, $\widehat{OR}_{Y=0}=12$, $\widehat{OR}_{Y=1}^* = 10$, $\widehat{OR}_{Y=0}^*=10$). Πραγματοποιήσαμε 500 επαναλήψεις με 20 μελέτες σε κάθε μετα-ανάλυση. Ο επιπολασμός της νόσου ήταν 50% με 200 άτομα σε κάθε μελέτη. Με $\widehat{OR}_{Y=1}^*$ και $\widehat{OR}_{Y=0}^*$ ορίζονται τα conditional Odds Ratios υπό την υπόθεση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης.

Statistic	$\hat{\beta}_1$ Correct Model	$\hat{\beta}_1$ Misspecified Model	$\hat{\beta}_2$ Correct Model	$\hat{\beta}_2$ Misspecified Model	$\hat{\beta}_3$ Correct Model	$\hat{\beta}_3$ Misspecified Model	$\hat{\beta}_4$ Correct Model	$\hat{\beta}_4$ Misspecified Model	$\hat{\beta}_1 - \hat{\beta}_3$ Correct Model	$\hat{\beta}_1 - \hat{\beta}_3$ Misspecified Model	$\hat{\beta}_2 - \hat{\beta}_4$ Correct Model	$\hat{\beta}_2 - \hat{\beta}_4$ Misspecified Model
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.8$ $\widehat{FPR}_2 = 0.1$ $\widehat{OR}_{Y=1} = 8$ $\widehat{OR}_{Y=0} = 12$ $\widehat{OR}_{Y=1}^* = 10$ $\widehat{OR}_{Y=0}^* = 10$												
Bias in point estimate	-0.0068	-0.0072	0.0093	0.0104	-0.0109	-0.0107	0.0212	0.0198	0.0041	0.0035	-0.0118	-0.0094
Empirical standard error	0.0677	0.0670	0.0732	0.0734	0.0695	0.0696	0.0899	0.0897	0.0835	0.0831	0.1077	0.1075
RMS model-based standard error	0.0795	0.0795	0.0793	0.0796	0.0781	0.0782	0.1021	0.1025	0.1004	0.1002	0.1173	0.1186
Coverage of nominal 95% confidence interval	97.2	97.4	96	95.4	96.8	96.6	95.8	95.6	97	97.2	95.6	96.4
Power of 5% level test	100	100	100	100	100	100	100	100	3	2.8	100	100
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.8$ $\widehat{FPR}_2 = 0.2$ $\widehat{OR}_{Y=1} = 8$ $\widehat{OR}_{Y=0} = 12$ $\widehat{OR}_{Y=1}^* = 10$ $\widehat{OR}_{Y=0}^* = 10$												
Bias in point estimate	-0.0062	-0.0060	0.0017	0.0011	-0.0102	-0.0104	0.0117	0.0118	0.0040	0.0043	-0.0100	-0.0107
Empirical standard error	0.0699	0.0697	0.0693	0.0691	0.0687	0.0684	0.0669	0.0669	0.0867	0.0859	0.0826	0.0820
RMS model-based standard error	0.0781	0.0782	0.0767	0.0772	0.0777	0.0778	0.0768	0.0772	0.0991	0.0990	0.0955	0.0972

Coverage of nominal 95% confidence interval	96.8	96.6	96.8	96.8	96.6	96.4	96.8	97.4	96.4	96.4	97.4	97.6
Power of 5% level test	100	100	100	100	100	100	100	100	3.6	3.6	2.6	2.4
$\widehat{TPR}_1 = 0.8$	$\widehat{FPR}_1 = 0.2$	$\widehat{TPR}_2 = 0.9$	$\widehat{FPR}_2 = 0.1$	$\widehat{OR}_{Y=1} = 8$	$\widehat{OR}_{Y=0} = 12$	$\widehat{OR}_{Y=1}^* = 10$	$\widehat{OR}_{Y=0}^* = 10$					
Bias in point estimate	-0.0035	-0.0037	0.0097	0.0097	-0.0266	-0.0268	0.0253	0.0246	0.0231	0.0231	-0.0157	-0.0149
Empirical standard error	0.0664	0.0665	0.0690	0.0688	0.0877	0.0875	0.0912	0.0906	0.1065	0.1066	0.1013	0.1005
RMS model-based standard error	0.0795	0.0796	0.0803	0.0806	0.1029	0.1030	0.1020	0.1024	0.1202	0.1200	0.1178	0.1191
Coverage of nominal 95% confidence interval	97.4	97.4	97.2	97.4	96.8	96.8	96.6	96.8	96.2	96.2	97.4	97.4
Power of 5% level test	100	100	100	100	100	100	100	100	100	100	100	100
$\widehat{TPR}_1 = 0.8$	$\widehat{FPR}_1 = 0.2$	$\widehat{TPR}_2 = 0.9$	$\widehat{FPR}_2 = 0.2$	$\widehat{OR}_{Y=1} = 8$	$\widehat{OR}_{Y=0} = 12$	$\widehat{OR}_{Y=1}^* = 10$	$\widehat{OR}_{Y=0}^* = 10$					
Bias in point estimate	-0.0013	-0.0010	0.0021	0.0022	-0.0227	-0.0234	0.0073	0.0070	0.0214	0.0224	-0.0052	-0.0048
Empirical standard error	0.0705	0.0705	0.0607	0.0609	0.0945	0.0945	0.0657	0.0658	0.1065	0.1068	0.0777	0.0777
RMS model-based standard error	0.0801	0.0802	0.0784	0.0788	0.1024	0.1025	0.0775	0.0779	0.1212	0.1211	0.0968	0.0983
Coverage of nominal 95% confidence interval	97.4	97.6	99.4	99.2	95.8	95.6	96.6	97	96.4	97	99	98.8
Power of 5% level test	100	100	100	100	100	100	100	100	100	100	1	1.2
$\widehat{TPR}_1 = 0.8$	$\widehat{FPR}_1 = 0.2$	$\widehat{TPR}_2 = 0.85$	$\widehat{FPR}_2 = 0.15$	$\widehat{OR}_{Y=1} = 8$	$\widehat{OR}_{Y=0} = 12$	$\widehat{OR}_{Y=1}^* = 10$	$\widehat{OR}_{Y=0}^* = 10$					

Bias in point estimate	-0.0037	-0.0038	0.0056	0.0055	-0.0114	-0.0119	0.0058	0.0053	0.0077	0.0082	-0.0002	0.0003
Empirical standard error	0.0671	0.0669	0.0686	0.0686	0.0722	0.0721	0.0728	0.0725	0.0857	0.0856	0.0851	0.0843
RMS model-based standard error	0.0786	0.0787	0.0781	0.0784	0.0870	0.0871	0.0863	0.0867	0.1068	0.1066	0.1041	0.1055
Coverage of nominal 95% confidence interval	97.8	97.8	97.6	97.6	97.2	97	97.2	97.2	97.8	97.8	97.2	98.2
Power of 5% level test	100	100	100	100	100	100	100	100	93.6	94	95.2	95.8

Πίνακας Z4. Μελέτη προσομοίωσης για την αξιολόγηση των συνεπειών του model misspecification όταν υπάρχουν αλληλεπιδράσεις τρίτης τάξης μεταξύ των μελετών αλλά δεν λαμβάνονται υπόψη με τη χρήση του αλγορίθμου IPF για την εκτίμηση της AUC . Παρουσιάζονται 5 σενάρια ($\widehat{TPR}_1 = 0.8$, $\widehat{FPR}_1 = 0.2$, $\widehat{TPR}_2 = 0.8, 0.85$ και 0.9 , $\widehat{FPR}_2 = 0.1, 0.15$ και 0.2 , $\widehat{OR}_{Y=1} = 8$, $\widehat{OR}_{Y=0} = 12$, $\widehat{OR}_{Y=1}^* = 10$, $\widehat{OR}_{Y=0}^* = 10$). Πραγματοποιήσαμε 500 επαναλήψεις με 20 μελέτες σε κάθε μετα-ανάλυση. Ο επιπολασμός της νόσου ήταν 50% με 200 άτομα σε κάθε μελέτη. Για απλούστευση, χρησιμοποιήσαμε την αναλυτική έκφραση 6.26 έτσι ώστε να υπολογίσουμε την \widehat{AUC} υποθέτοντας ότι οι μελέτες μας είναι ομοιογενείς. Με $\widehat{OR}_{Y=1}^*$ και $\widehat{OR}_{Y=0}^*$ ορίζονται τα conditional Odds Ratios υπό την υπόθεση ότι δεν υπάρχει αλληλεπίδραση τρίτης τάξης.

Statistic	\widehat{AUC}_1 Correct Model	\widehat{AUC}_1 Misspecified Model	\widehat{AUC}_2 Correct Model	\widehat{AUC}_2 Misspecified Model	$\widehat{AUC}_1 -$ \widehat{AUC}_2 Correct Model	$\widehat{AUC}_1 -$ \widehat{AUC}_2 Misspecified Model
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.8$ $\widehat{FPR}_2 = 0.1$ $\widehat{AUC}_1 = 0.870$ $\widehat{AUC}_2 = 0.923$						
Bias in point estimate	-0.0015	-0.0016	-0.0019	-0.0018	0.0004	0.0002
Empirical standard error	0.0082	0.0082	0.0061	0.0061	0.0090	0.0089
RMS model-based standard error	0.0092	0.0092	0.0069	0.0069	0.0104	0.0105
Coverage of nominal 95% confidence interval	97.4	97.2	96.8	96.8	96.4	97
Power of 5% level test	100	100	100	100	99.8	99.8
$\widehat{TPR}_1 = 0.8$ $\widehat{FPR}_1 = 0.2$ $\widehat{TPR}_2 = 0.8$ $\widehat{FPR}_2 = 0.2$ $\widehat{AUC}_1 = 0.870$ $\widehat{AUC}_2 = 0.870$						
Bias in point estimate	-0.0008	-0.0008	-0.0020	-0.0020	0.0011	0.0012
Empirical standard error	0.0079	0.0079	0.0077	0.0077	0.0097	0.0096

RMS model-based standard error	0.0089	0.0089	0.0089	0.0090	0.0112	0.0113
Coverage of nominal 95% confidence interval	98.2	97.8	95.6	96.2	97	97.6
Power of 5% level test	100	100	100	100	3	2.4
$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.9	$\widehat{FPR}_2 =$ 0.1	$\widehat{AUC}_1 =$ 0.870	$\widehat{AUC}_2 =$ 0.957	
Bias in point estimate	-0.0012	-0.0013	-0.0018	-0.0018	0.0006	0.0006
Empirical standard error	0.0078	0.0078	0.0042	0.0042	0.0084	0.0083
RMS model-based standard error	0.0092	0.0092	0.0048	0.0048	0.0097	0.0097
Coverage of nominal 95% confidence interval	97.4	97.2	97	97	96.4	96.4
Power of 5% level test	100	100	100	100	100	100
$\widehat{TPR}_1 =$ 0.8	$\widehat{FPR}_1 =$ 0.2	$\widehat{TPR}_2 =$ 0.9	$\widehat{FPR}_2 =$ 0.2	$\widehat{AUC}_1 =$ 0.870	$\widehat{AUC}_2 =$ 0.923	
Bias in point estimate	-0.0005	-0.0004	-0.0018	-0.0018	0.0014	0.0014
Empirical standard error	0.0077	0.0077	0.0064	0.0064	0.0087	0.0087
RMS model-based standard error	0.0091	0.0091	0.0069	0.0069	0.0104	0.0105
Coverage of nominal 95% confidence interval	97.4	97.4	96.4	96.4	97.6	97.4
Power of 5% level test	100	100	100	100	100	100
$\widehat{TPR}_1 =$	$\widehat{FPR}_1 =$	$\widehat{TPR}_2 =$	$\widehat{FPR}_2 =$	$\widehat{AUC}_1 =$	$\widehat{AUC}_2 =$	

	0.8	0.2	0.85	0.15	0.870	0.917	
Bias in point estimate		-0.0016	-0.0016	-0.0023	-0.0023	0.0007	0.0007
Empirical standard error		0.0080	0.0081	0.0059	0.0059	0.0088	0.0088
RMS model-based standard error		0.0091	0.0092	0.0070	0.0070	0.0104	0.0104
Coverage of nominal 95% confidence interval		96.6	96.6	97.2	97	97.6	97.4
Power of 5% level test		100	100	100	100	100	100

Παράρτημα Η

Πίνακας Η1. Αληθώς θετικά (True positive-*TP*), ψευδώς θετικά (false positive-*FP*), αληθώς αρνητικά (true negative-*TN*) και ψευδώς αρνητικά (false negative-*FN*) ευρήματα των δυο διαγνωστικών δοκιμασιών (RF και αντίσωμα anti-CCP2) για κάθε μελέτη που περιλαμβάνεται στη μετα-ανάλυση για την ρευματοειδή αρθρίτιδα.

Συγγραφέας	Rheumatoid factor				Αντίσωμα Anti-CCP2											
	TP	FP	FN	TN	TP	FP	FN	TN	n ₁₁₁	n ₁₁₀	n ₁₀₁	n ₁₀₀	n ₀₀₁	n ₀₀₀	n ₀₁₀	n ₀₁₁
Quinn (Quinn, Gough et al. 2006)	115	53	67	63	147	10	35	106	107	8	40	27	5	58	48	5
Fernandez-Suarez (Fernandez-Suarez, Reneses et al. 2005)	30	2	23	73	31	0	22	75	23	7	8	15	0	73	2	0
Kwok (Kwok, Hui et al. 2005)	77	16	52	52	71	2	58	66	60	17	11	41
Greiner (Greiner, Plischke et al. 2005)	75	42	12	191	70	5	17	228	67	8	3	9	2	189	39	3
Sauerland (Sauerland, Becker et al. 2005)	161	89	70	380	171	26	60	443	151	10	20	50
Kamali (Kamali, Polat et al. 2005)	20	32	26	25	26	1	20	56	16	4	10	16
Choi (Choi, Lim et al. 2005)	261	54	63	197	236	20	88	231	221	40	15	48	8	189	42	12
Nell (Nell, Machold et al. 2005)	56	11	46	87	42	2	60	96
Raza (Raza, Breese et al. 2005)	22	2	20	80	24	3	18	79	19	3	5	15
De Rycke (De Rycke, Peene et al. 2004)	93	28	25	118	89	4	29	142
Girelli (Girelli, Foschi et al. 2004)	32	29	3	13	25	2	10	40	25	7	0	3	0	13	27	2
Grootenboer-Mignot (Grootenboer-Mignot, Nicaise-Roland et al. 2004)	64	18	29	73	167	8	98	88
Hitchon (Hitchon, Alex et al. 2004)	32	10	9	13	26	8	15	15
Lopez-Hoyos (Lopez-Hoyos, Ruiz de Alegria et al. 2004)	36	3	5	70	38	3	0	73
Bombardieri (Bombardieri, Alessandri et al. 2004)	27	6	3	33	23	0	7	39
Dubucquoi (Dubucquoi, Solau-Gervais et al. 2004)	84	41	56	90	90	2	50	129	69	15	12	44
Soderlin (Soderlin, Kastbom et al. 2004)	5	4	11	49	7	2	9	51
Vallbracht (Vallbracht, Rieber et al. 2004)	196	75	99	345	190	12	105	408	152	44	38	61	6	339	69	6

Vittecoq (Vittecoq, Incaugarat et al. 2004)	62	11	114	127	69	5	107	133	52	10	17	97	5	122	11	0
Lee (Lee and Schur 2003)	73	22	29	90	68	14	35	132	58	15	10	19	1	89	12	10
Rantapaa-Dahlqvist (Rantapaa-Dahlqvist, de Jong et al. 2003)	49	23	18	359	47	7	20	375	39	10	8	10	3	356	19	4
Suzuki (Suzuki, Sawada et al. 2003)	383	38	166	170	481	23	68	185	366	17	115	51
Aotsuka (Aotsuka, Okawa-Takatsuji et al. 2005)	115	17	16	73
Garcia-Berrocal (Garcia-Berrocal, Gonzalez et al. 2005)	69	8	18	38
van Gaalen (van Gaalen, Visser et al. 2005)	82	13	71	301
Correa (Correa, Tobon et al. 2004)	74	11	8	130
Kumagai (Kumagai, Nishimura et al. 2004)	64	14	15	293
Nielen (Nielen, van der Horst et al. 2005)	149	7	109	114
van Venrooij (van Venrooij, Vossenaar et al. 2004)	865	79	252	8	221
Bas (Bas, Genevay et al. 2003)	143	43	53	196
Saraux (Saraux, Berthelot et al. 2003)	35	8	51	149
Jansen (Jansen, van Schaardenburg et al. 2003)	130	8	128	113
Bizzaro (Bizzaro, Mazzanti et al. 2001)	61	36	37	196
Goldbach-Mansky (Goldbach-Mansky, Lee et al. 2000)	70	39	36	93
Anuradha (Anuradha and Chopra 2005)	482	2	82	153
Thammanichanond (Thammanichanond, Kunakorn et al. 2005)	57	25	6	111
Das (Das, Atsumi et al. 2004)	42	46	14	127
Spiritus (Spiritus, Verschueren et al. 2004)	57	9	33	93
Vasiliauskiene (Vasiliauskiene, Wiik et al. 2001)	75	21	21	106
Vittecoq (Vittecoq,	26	1	32	29

Salle et al. 2001)																			
Schellekens (Schellekens, Visser et al. 2000)	80	28	69	284
Aho (Aho, Palosuo et al. 1999)	64	16	27	153
Jonsson (Jonsson, Steinsson et al. 1998)	50	14	20	191
Swedler (Swedler, Wallman et al. 1997)	89	3	9	39
Young (Young, Sumar et al. 1991)	25	1	14	20
de Bois (de Bois, Arndt et al. 1996)	8	8	0	31
Cordonnier (Cordonnier, Meyer et al. 1996)	20	2	29	18
Visser (Visser, Gelinck et al. 1996)	157	287	78	1466
Berthelot (Berthelot, Maugars et al. 1995)	80	50	39	45
Saraux (Saraux, Valls et al. 1995)	8	8	31	91
Despres (Despres, Boire et al. 1994)	143	39	63	130
Gomes-Daudrix (Gomes-Daudrix, Sebbag et al. 1994)	48	1	40	99
Banchuin (Banchuin, Janyapoon et al. 1992)	36	6	41	313
Carpenter (Carpenter and Bartkowiak 1989)	60	8	20	119
Davis (Davis and Stein 1989)	18	3	31	25
Winkles (Winkles, Lunec et al. 1989)	113	19	29	481
van Leeuwen (van Leeuwen, Westra et al. 1988)	163	10	28	140

Παράρτημα Θ

Στην περίπτωση όπου κάποια μελέτη αναφέρει τα ευρήματα όταν και οι δυο διαγνωστικές δοκιμασίες είναι θετικές, ισχύει ότι:

$$Sensitivity_i' = \frac{n_{111i}}{(n_{111i} + n_{101i} + n_{110i} + n_{100i})} = m_i \quad (\Theta 1)$$

$$Specificity_i' = \frac{(n_{001i} + n_{010i} + n_{000i})}{(n_{001i} + n_{010i} + n_{000i} + n_{011i})} = k_i \quad (\Theta 2)$$

Οι εξισώσεις Θ1 και Θ2 συνδυάζονται με τα γνωστά περιθώρια αθροίσματα της συσχέτισης των δυο διαγνωστικών δοκιμασιών ανάλογα με την παρουσία ή απουσία της νόσου. Έτσι, υπολογίζονται οι ακόλουθες σχέσεις:

$$n_{111i} = m_i (TP_{1i} + FN_{1i}) \quad (\Theta 3)$$

$$n_{110i} = TP_{1i} - m_i (TP_{1i} + FN_{1i}) \quad (\Theta 4)$$

$$n_{101i} = TP_{2i} - (TP_{1i} + FN_{1i}) m_i \quad (\Theta 5)$$

$$n_{100i} = (TP_{1i} + FN_{1i}) m_i + FN_{1i} - TP_{2i} \quad (\Theta 6)$$

$$n_{001i} = (TN_{1i} + FP_{1i}) k_i - TN_{2i} \quad (\Theta 7)$$

$$n_{000i} = TN_{1i} + TN_{2i} - (TN_{1i} + FP_{1i}) k_i \quad (\Theta 8)$$

$$n_{010i} = (TN_{1i} + FP_{1i}) k_i - TN_{1i} \quad (\Theta 9)$$

$$n_{011i} = TN_{1i} + FP_{1i} - (FP_{1i} + TN_{1i}) k_i \quad (\Theta 10)$$

Στην περίπτωση όπου κάποια μελέτη αναφέρει τα ευρήματα όταν τουλάχιστον μια διαγνωστική δοκιμασία είναι θετική, ισχύει ότι:

$$Sensitivity_i'' = \frac{(n_{111i} + n_{110i} + n_{101i})}{(n_{111i} + n_{110i} + n_{101i} + n_{100i})} = g_i \quad (\Theta 11)$$

$$Specificity_i'' = \frac{n_{000i}}{(n_{000i} + n_{001i} + n_{010i} + n_{011i})} = h_i \quad (\Theta 12)$$

Ακολουθώντας την ίδια λογική, υπολογίζονται οι ακόλουθες σχέσεις:

$$n_{111i} = TP_{1i} + TP_{2i} - (TP_{1i} + FN_{1i}) g_i \quad (\Theta 13)$$

$$n_{110i} = (TP_{1i} + FN_{1i})g_i - TP_{2i} \quad (\Theta 14)$$

$$n_{101i} = (TP_{1i} + FN_{1i})g_i - TP_{1i} \quad (\Theta 15)$$

$$n_{100i} = TP_{1i} + FN_{1i} - (TP_{1i} + FN_{1i})g_i \quad (\Theta 16)$$

$$n_{001i} = TN_{1i} - (TN_{1i} + FP_{1i})h_i \quad (\Theta 17)$$

$$n_{000i} = h_i (TN_{1i} + FP_{1i}) \quad (\Theta 18)$$

$$n_{010i} = TN_{2i} - (TN_{1i} + FP_{1i})h_i \quad (\Theta 19)$$

$$n_{011i} = (TN_{1i} + FP_{1i})h_i + FP_{1i} - TN_{2i} \quad (\Theta 20)$$

Τέλος, στην περίπτωση όπου κάποια μελέτη αναφέρει τις εκτιμήσεις της ευαισθησίας, της ειδικότητας, της θετικής και αρνητικής προγνωστικής αξίας για το άτομα εκείνα με μια αρνητική διαγνωστική δοκιμασία, ισχύει ότι:

$$Sensitivity_i = \frac{n_{101i}}{n_{101i} + n_{100i}} = o_i \quad (\Theta 21)$$

$$Specificity_i = \frac{n_{000i}}{n_{000i} + n_{001i}} = t_i \quad (\Theta 22)$$

$$PPV_i = \frac{n_{101i}}{n_{101i} + n_{001i}} = q_i \quad (\Theta 23)$$

$$NPV_i = \frac{n_{000i}}{n_{000i} + n_{100i}} = r_i \quad (\Theta 24)$$

Για απλούστευση ισχύει ότι:

$$\frac{n_{101i}}{n_{100i}} = \frac{o_i}{1 - o_i} = O_i \quad (\Theta 25)$$

$$\frac{n_{000i}}{n_{001i}} = \frac{t_i}{1 - t_i} = T_i \quad (\Theta 26)$$

$$\frac{n_{101i}}{n_{001i}} = \frac{q_i}{1 - q_i} = Q_i \quad (\Theta 27)$$

$$\frac{n_{000i}}{n_{100i}} = \frac{r_i}{1 - r_i} = R_i \quad (\Theta 28)$$

Τα άτομα με μια αρνητική διαγνωστική δοκιμασία υπολογίζονται ως:

$$n_{101i} + n_{100i} + n_{000i} + n_{001i} = e_i \quad (\Theta 29)$$

Έτσι, εξάγονται οι ακόλουθες σχέσεις:

$$n_{101i} = \frac{Q_i e_i (O_i Q_i - T_i R_i)}{(T_i R_i - O_i Q_i - O_i + T_i - Q_i + R_i)(T_i + Q_i)} \quad (\Theta 30)$$

$$n_{100i} = \frac{(T_i^2 - Q_i^2) e_i}{(T_i R_i - O_i Q_i - O_i + T_i - Q_i + R_i)(T_i + Q_i)} \quad (\Theta 31)$$

$$n_{000i} = \frac{-T_i e_i (O_i Q_i - T_i R_i)}{(T_i R_i - O_i Q_i - O_i + T_i - Q_i + R_i)(T_i + Q_i)} \quad (\Theta 32)$$

$$n_{001i} = \frac{(R_i - O_i)(T_i + Q_i) e_i}{(T_i R_i - O_i Q_i - O_i + T_i - Q_i + R_i)(T_i + Q_i)} \quad (\Theta 33)$$

Βιβλιογραφία

- Ackerman, H., S. Usen, et al. (2005). "A comparison of case-control and family-based association methods: the example of sickle-cell and malaria." Ann Hum Genet **69**(Pt 5): 559-65.
- Aggarwal, P., S. Naik, et al. (2006). "Correlation between methotrexate efficacy & toxicity with C677T polymorphism of the methylenetetrahydrofolate gene in rheumatoid arthritis patients on folate supplementation." Indian J Med Res **124**(5): 521-6.
- Agresti, A. (1999). "On logit confidence intervals for the odds ratio with small samples." Biometrics **55**(2): 597-602.
- Agresti, A. (2002). Categorical Data Analysis. New York, John Wiley & Sons.
- Aho, K., T. Palosuo, et al. (1999). "Antifilaggrin antibodies in recent-onset arthritis." Scand J Rheumatol **28**(2): 113-6.
- Alfano, M., A. Crotti, et al. (2008). "New players in cytokine control of HIV infection." Curr HIV/AIDS Rep **5**(1): 27-32.
- Anderson, K. B., J. L. Guest, et al. (2004). "Hepatitis C virus coinfection increases mortality in HIV-infected patients in the highly active antiretroviral therapy era: data from the HIV Atlanta VA Cohort Study." Clin Infect Dis **39**(10): 1507-13.
- Ank, N., H. West, et al. (2006). "IFN-lambda: novel antiviral cytokines." J Interferon Cytokine Res **26**(6): 373-9.
- Anscombe, F. J. (1956). "On estimating binomial response relations." Biometrika **43**(3-4): 461-464.
- Anthopoulos, P. G., S. J. Hamodrakas, et al. (2010). "Apolipoprotein E polymorphisms and type 2 diabetes: a meta-analysis of 30 studies including 5423 cases and 8197 controls." Mol Genet Metab **100**(3): 283-91.
- Antonucci, G., E. Girardi, et al. (2005). "Role of hepatitis C virus (HCV) viremia and HCV genotype in the immune recovery from highly active antiretroviral therapy in a cohort of antiretroviral-naïve HIV-infected individuals." Clin Infect Dis **40**(12): e101-9.
- Anuradha, V. and A. Chopra (2005). "In the era of nephelometry, latex agglutination is still good enough to detect rheumatoid factor." J Rheumatol **32**(12): 2343-4.
- Aotsuka, S., M. Okawa-Takatsuji, et al. (2005). "A retrospective study of the fluctuation in serum levels of anti-cyclic citrullinated peptide antibody in patients with rheumatoid arthritis." Clin Exp Rheumatol **23**(4): 475-81.
- Aplenc, R., J. Thompson, et al. (2005). "Methylenetetrahydrofolate reductase polymorphisms and therapy response in pediatric acute lymphoblastic leukemia." Cancer Res **65**(6): 2482-7.
- Arends, L. R., T. H. Hamza, et al. (2008). "Bivariate random effects meta-analysis of ROC curves." Med Decis Making **28**(5): 621-38.
- Armitage, P. (1955). "Tests for Linear Trends in Proportions and Frequencies." Biometrics **11**(3): 375-386.
- Asensi, V., C. Rego, et al. (2008). "IL-1beta (+3954C/T) polymorphism could protect human immunodeficiency virus (HIV)-infected patients on highly active antiretroviral treatment (HAART) against lipodystrophic syndrome." Genet Med **10**(3): 215-23.

- Attia, J., A. Thakkinstian, et al. (2003). "Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology." J Clin Epidemiol **56**(4): 297-303.
- Aulchenko, Y. S., S. Ripke, et al. (2007). "GenABEL: an R library for genome-wide association analysis." Bioinformatics **23**(10): 1294-6.
- Avihingsanon, A., S. Jitmitraparp, et al. (2014). "Advanced liver fibrosis by transient elastography, fibrosis 4, and alanine aminotransferase/platelet ratio index among Asian hepatitis C with and without human immunodeficiency virus infection: role of vitamin D levels." J Gastroenterol Hepatol **29**(9): 1706-14.
- Azarpira, N., M. Ramzi, et al. (2007). "Methylenetetrahydrofolate reductase C677T genotypes and clinical outcome following hematopoietic cell transplant." Exp Clin Transplant **5**(2): 693-7.
- Baggott, J. E., W. H. Vaughn, et al. (1986). "Inhibition of 5-aminoimidazole-4-carboxamide ribotide transformylase, adenosine deaminase and 5'-adenylate deaminase by polyglutamates of methotrexate and oxidized folates and by 5-aminoimidazole-4-carboxamide riboside and ribotide." Biochem J **236**(1): 193-200.
- Bagos, P. G. (2008). "A unification of multivariate methods for meta-analysis of genetic association studies." Stat Appl Genet Mol Biol **7**(1): Article31.
- Bagos, P. G. (2008). "A unification of multivariate methods for meta-analysis of genetic association studies." Stat Appl Genet Mol Biol **7**: Article31.
- Bagos, P. G. (2011). "Meta-analysis of haplotype-association studies: comparison of methods and empirical evaluation of the literature." BMC genetics **12**(1): 8.
- Bagos, P. G. (2011). "Meta-analysis of haplotype-association studies: comparison of methods and empirical evaluation of the literature." BMC Genet **12**: 8.
- Bagos, P. G. (2012). "On the covariance of two correlated log-odds ratios." Stat Med.
- Bagos, P. G. (2012). "On the covariance of two correlated log-odds ratios." Stat Med **31**(14): 1418-31.
- Bagos, P. G., N. L. Dimou, et al. (2011). "Meta-analysis of family-based and case-control genetic association studies that use the same cases." Statistical Applications in Genetics and Molecular Biology **10**(1): 1-41.
- Bagos, P. G., A. L. Elefsinioti, et al. (2007). "The GNB3 C825T polymorphism and essential hypertension: a meta-analysis of 34 studies including 14,094 cases and 17,760 controls." J Hypertens **25**(3): 487-500.
- Bagos, P. G. and T. D. Liakopoulos (2010). "A multipoint method for meta-analysis of genetic association studies." Genet Epidemiol **34**(7): 702-15.
- Bagos, P. G. and G. K. Nikolopoulos (2007). "A method for meta-analysis of case-control genetic association studies using logistic regression." Stat Appl Genet Mol Biol **6**: Article17.
- Bagos, P. G. and G. K. Nikolopoulos (2009). "Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology." J Clin Epidemiol.
- Bagos, P. G. and G. K. Nikolopoulos (2009). "Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology." Journal of Clinical Epidemiology **in press**.
- Bagos, P. G. and G. K. Nikolopoulos (2009). "Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology." Journal of clinical epidemiology **62**(10): 1037-1044.

- Bagos, P. G. and G. K. Nikolopoulos (2009). "Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology." *J Clin Epidemiol* **62**(10): 1037-44.
- Bailey, J. A. and E. E. Eichler (2006). "Primate segmental duplications: crucibles of evolution, diversity and disease." *Nat Rev Genet* **7**(7): 552-64.
- Balding, D. J. (2006). "A tutorial on statistical methods for population association studies." *Nat Rev Genet* **7**(10): 781-91.
- Banchuin, N., K. Janyapoon, et al. (1992). "Re-evaluation of ELISA and latex agglutination test for rheumatoid factor detection in the diagnosis of rheumatoid arthritis." *Asian Pac J Allergy Immunol* **10**(1): 47-54.
- Barrowman, N. J., M. Fang, et al. (2003). "Identifying null meta-analyses that are ripe for updating." *BMC Medical research Methodology* **3**(1): 13.
- Bas, S., S. Genevay, et al. (2003). "Anti-cyclic citrullinated peptide antibodies, IgM and IgA rheumatoid factors in the diagnosis and prognosis of rheumatoid arthritis." *Rheumatology (Oxford)* **42**(5): 677-80.
- Bazrgar, M., M. Karimi, et al. (2007). "Apolipoprotein E gene polymorphism and left ventricular function in Iranian patients with thalassemia major." *Haematologica* **92**(2): 256-7.
- Becker, K. G., K. C. Barnes, et al. (2004). "The genetic association database." *Nat Genet* **36**(5): 431-2.
- Becker, T., S. Cichon, et al. (2005). "Multiple testing in the context of haplotype analysis revisited: application to case-control data." *Ann Hum Genet* **69**(Pt 6): 747-56.
- Becker, T. and M. Knapp (2004). "A powerful strategy to account for multiple testing in the context of haplotype analysis." *Am J Hum Genet* **75**(4): 561-70.
- Begg, C. B. and M. Mazumdar (1994). "Operating characteristics of a rank correlation test for publication bias." *Biometrics* **50**(4): 1088-101.
- Begg, C. B. and M. Mazumdar (1994). "Operating characteristics of a rank correlation test for publication bias." *Biometrics*: 1088-1101.
- Begg, C. B. and M. Mazumdar (1994). "Operating characteristics of a rank correlation test for publication bias." *Biometrics* **50**: 1088-1101.
- Begum, F., D. Ghosh, et al. (2012). "Comprehensive literature review and statistical considerations for GWAS meta-analysis." *Nucleic Acids Res.*
- Bellanti, F., G. Vendemiale, et al. (2012). "The impact of interferon lambda 3 gene polymorphism on natural course and treatment of hepatitis C." *Clin Dev Immunol* **2012**: 849373.
- Bentall, R. P. (2013). *Reconstructing schizophrenia*, Routledge.
- Benyamin, B., P. M. Visscher, et al. (2009). "Family-based genome-wide association studies." *Pharmacogenomics* **10**(2): 181-90.
- Berkey, C. S., D. C. Hoaglin, et al. (1998). "Meta-analysis of multiple outcomes by regression with random effects." *Stat Med* **17**(22): 2537-50.
- Berkun, Y., D. Levartovsky, et al. (2004). "Methotrexate related adverse effects in patients with rheumatoid arthritis are associated with the A1298C polymorphism of the MTHFR gene." *Ann Rheum Dis* **63**(10): 1227-31.
- Berlin, J. A., M. P. Longnecker, et al. (1993). "Meta-analysis of epidemiologic dose-response data." *Epidemiology* **4**(3): 218-28.
- Bernbeck, B., C. Mauz-Korholz, et al. (2003). "Methylenetetrahydrofolate reductase gene polymorphism and glucocorticoid intake in children with ALL and aseptic osteonecrosis." *Klin Padiatr* **215**(6): 327-31.

- Berrington, A. and D. R. Cox (2003). "Generalized least squares for the synthesis of correlated information." *Biostatistics* **4**(3): 423-31.
- Berthelot, J. M., Y. Maugars, et al. (1995). "Specificity of antiperinuclear factor for rheumatoid arthritis in rheumatoid factor-positive sera." *Br J Rheumatol* **34**(8): 716-20.
- Bhattacharjee, S., P. Rajaraman, et al. (2012). "A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits." *Am J Hum Genet* **90**(5): 821-35.
- Bigham, A. W., R. D. Mackelprang, et al. (2014). "Variants in host viral replication cycle genes are associated with heterosexual HIV-1 acquisition in Africans." *J Acquir Immune Defic Syndr* **66**(2): 127-34.
- Bizzaro, N., G. Mazzanti, et al. (2001). "Diagnostic accuracy of the anti-citrulline antibody assay for rheumatoid arthritis." *Clin Chem* **47**(6): 1089-93.
- Bohanec Grabar, P., D. Logar, et al. (2008). "Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism." *Eur J Clin Pharmacol* **64**(11): 1057-68.
- Bombardieri, M., C. Alessandri, et al. (2004). "Role of anti-cyclic citrullinated peptide antibodies in discriminating patients with rheumatoid arthritis from patients with chronic hepatitis C infection-associated polyarticular involvement." *Arthritis Res Ther* **6**(2): R137-41.
- Braitstein, P., C. Zala, et al. (2006). "Immunologic response to antiretroviral therapy in hepatitis C virus-coinfected adults in a population-based HIV/AIDS treatment program." *J Infect Dis* **193**(2): 259-68.
- Braun, J. and R. Rau (2009). "An update on methotrexate." *Curr Opin Rheumatol* **21**(3): 216-23.
- Breslow, N., B. Leroux, et al. (1998). "Approximate hierarchical modelling of discrete data in epidemiology." *Stat Methods Med Res* **7**(1): 49-62.
- Brookes, A. J. (1999). "The essence of SNPs." *Gene* **234**(2): 177-86.
- Burlone, M. E., A. Cerutti, et al. (2013). "IL28B polymorphism, blood interferon-alpha concentration, and disease stage of HCV mono-infected and HCV-HIV co-infected patients." *Curr HIV Res* **11**(1): 50-5.
- Burton, P. R., M. D. Tobin, et al. (2005). "Key concepts in genetic epidemiology." *The Lancet* **366**(9489): 941-951.
- Campalani, E., M. Arenas, et al. (2007). "Polymorphisms in folate, pyrimidine, and purine metabolism are associated with efficacy and toxicity of methotrexate in psoriasis." *J Invest Dermatol* **127**(8): 1860-7.
- Campillos, M., M. Kuhn, et al. (2008). "Drug target identification using side-effect similarity." *Science* **321**(5886): 263-266.
- Cao, A. and R. Galanello (2010). "Beta-thalassemia." *Genet Med* **12**(2): 61-76.
- Carlson, C. S., M. A. Eberle, et al. (2004). "Mapping complex disease loci in whole-genome association studies." *Nature* **429**(6990): 446-52.
- Carmo, R. A., M. D. Guimaraes, et al. (2008). "The influence of HCV coinfection on clinical, immunological and virological responses to HAART in HIV-patients." *Braz J Infect Dis* **12**(3): 173-9.
- Carpenter, A. B. and C. D. Bartkowiak (1989). "Rheumatoid factors determined by fluorescence immunoassay: comparison with qualitative and quantitative methods." *Clin Chem* **35**(3): 464-6.

- Castagna, A., L. Galli, et al. (2010). "Predicting the magnitude of short-term CD4+ T-cell recovery in HIV-infected patients during first-line highly active antiretroviral therapy." Antivir Ther **15**(2): 165-75.
- Chakravarti, A. (1998). "It's raining SNPs, hallelujah?" Nat Genet **19**(3): 216-7.
- Chalmers, T. C., J. Berrier, et al. (1987). "Meta-analysis of clinical trials as a scientific discipline. II: Replicate variability and comparison of studies that agree and disagree." Stat Med **6**(7): 733-44.
- Chan, E. S. and B. N. Cronstein (2002). "Molecular action of methotrexate in inflammatory diseases." Arthritis Res **4**(4): 266-73.
- Chang, J., H. M. Naif, et al. (1996). "The inhibition of HIV replication in monocytes by interleukin 10 is linked to inhibition of cell differentiation." AIDS Res Hum Retroviruses **12**(13): 1227-35.
- Chatterjee, A., A. Rathore, et al. (2009). "Association of IL-4 589 C/T promoter and IL-4Ralpha150V receptor polymorphism with susceptibility to HIV-1 infection in North Indians." J Med Virol **81**(6): 959-65.
- Chatterjee, A., A. Rathore, et al. (2009). "Genetic association of IL-10 gene promoter polymorphism and HIV-1 infection in North Indians." J Clin Immunol **29**(1): 71-7.
- Chen, J. and N. Chatterjee (2007). "Exploiting Hardy-Weinberg equilibrium for efficient screening of single SNP associations from case-control studies." Hum Hered **63**(3-4): 196-204.
- Chen, T. Y., E. L. Ding, et al. (2009). "Meta-analysis: increased mortality associated with hepatitis C in HIV-infected persons is unrelated to HIV disease progression." Clin Infect Dis **49**(10): 1605-15.
- Chen, Y., Y. Liu, et al. (2015). "A hybrid model for combining case-control and cohort studies in systematic reviews of diagnostic tests." J R Stat Soc Ser C Appl Stat **64**(3): 469-489.
- Chen, Y., Y. Liu, et al. (2014). "A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews." Stat Methods Med Res.
- Chen, Y., H. X. Xu, et al. (2012). "Meta-analysis: IL28B polymorphisms predict sustained viral response in HCV patients treated with pegylated interferon-alpha and ribavirin." Aliment Pharmacol Ther **36**(2): 91-103.
- Chen, Y. H. and H. W. Lin (2008). "Simple association analysis combining data from trios/sibships and unrelated controls." Genet Epidemiol **32**(6): 520-7.
- Chiusolo, P., G. Reddicono, et al. (2002). "Preponderance of methylenetetrahydrofolate reductase C677T homozygosity among leukemia patients intolerant to methotrexate." Ann Oncol **13**(12): 1915-8.
- Chiusolo, P., G. Reddicono, et al. (2007). "MTHFR polymorphisms' influence on outcome and toxicity in acute lymphoblastic leukemia patients." Leuk Res **31**(12): 1669-74.
- Choi, S. W., M. K. Lim, et al. (2005). "Diagnostic performances of anti-cyclic citrullinated peptides antibody and antifilaggrin antibody in Korean patients with rheumatoid arthritis." J Korean Med Sci **20**(3): 473-8.
- Chu, H., S. Chen, et al. (2009). "Random Effects Models in a Meta-Analysis of the Accuracy of Two Diagnostic Tests Without a Gold Standard." J Am Stat Assoc **104**(486): 512-523.
- Chu, H. and S. R. Cole (2006). "Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach." J Clin Epidemiol **59**(12): 1331-2; author reply 1332-3.

- Chu, H., L. Nie, et al. (2009). "Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection." *Stat Med* **28**(18): 2384-99.
- ChuanYuan, K., Z. Li, et al. (2011). "Association study of the frizzled 3 gene with Chinese Va schizophrenia." *Neuroscience letters* **505**(2): 196-199.
- Chung, R. T., S. R. Evans, et al. (2002). "Immune recovery is associated with persistent rise in hepatitis C virus RNA, infrequent liver test flares, and is not impaired by hepatitis C virus in co-infected subjects." *AIDS* **16**(14): 1915-23.
- Clayton, D. and M. Hills (1993). *Statistical Models in Epidemiology*, Oxford University Press.
- Clayton, D. G., N. M. Walker, et al. (2005). "Population structure, differential bias and genomic control in a large-scale, case-control association study." *Nat Genet* **37**(11): 1243-6.
- Cochran, W. G. (1954). "Some methods for strengthening the common chi-squared tests." *Biometrics* **10**(4): 417-451.
- Coghill, D. and T. Banaschewski (2009). "The genetics of attention-deficit/hyperactivity disorder." *Expert Rev Neurother* **9**(10): 1547-65.
- Collazos, J., V. Asensi, et al. (2010). "Association of HIV transmission categories with sociodemographic, viroimmunological and clinical parameters of HIV-infected patients." *Epidemiol Infect* **138**(7): 1016-24.
- Collins, F. S., M. S. Guyer, et al. (1997). "Variations on a theme: cataloging human DNA sequence variation." *Science* **278**(5343): 1580-1.
- Collins, F. S., A. Patrinos, et al. (1998). "New goals for the U.S. Human Genome Project: 1998-2003." *Science* **282**(5389): 682-9.
- Collins, R. and S. MacMahon (1994). "Blood pressure, antihypertensive drug treatment and the risks of stroke and of coronary heart disease." *Br Med Bull* **50**(2): 272-98.
- Conn, V. S., J. C. Valentine, et al. (2003). "Grey literature in meta-analyses." *Nurs Res* **52**(4): 256-61.
- Conrad, D. F., T. D. Andrews, et al. (2006). "A high-resolution survey of deletion polymorphism in the human genome." *Nat Genet* **38**(1): 75-81.
- Corchado, S., M. Marquez, et al. (2013). "Influence of Genetic Polymorphisms of Tumor Necrosis Factor Alpha and Interleukin 10 Genes on the Risk of Liver Cirrhosis in HIV-HCV Coinfected Patients." *PLoS One* **8**(6): e66619.
- Cordell, H. J. and D. G. Clayton (2005). "Genetic association studies." *Lancet* **366**(9491): 1121-31.
- Cordonnier, C., O. Meyer, et al. (1996). "Diagnostic value of anti-RA33 antibody, antikeratin antibody, antiperinuclear factor and antinuclear antibody in early rheumatoid arthritis: comparison with rheumatoid factor." *Br J Rheumatol* **35**(7): 620-4.
- Correa, P. A., G. J. Tobon, et al. (2004). "[Anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: relation with clinical features, cytokines and HLA-DRB1]." *Biomedica* **24**(2): 140-52.
- Cotsapas, C., B. F. Voight, et al. (2011). "Pervasive sharing of genetic effects in autoimmune disease." *PLoS Genet* **7**(8): e1002254.
- Crawley, E., R. Kay, et al. (1999). "Polymorphic haplotypes of the interleukin-10 5' flanking region determine variable interleukin-10 transcription and are associated with particular phenotypes of juvenile rheumatoid arthritis." *Arthritis Rheum* **42**(6): 1101-8.

- Cronstein, B. N. (1996). "Molecular therapeutics. Methotrexate and its mechanism of action." Arthritis Rheum **39**(12): 1951-60.
- Cronstein, B. N. (1997). "The mechanism of action of methotrexate." Rheum Dis Clin North Am **23**(4): 739-55.
- Cuadras, C. M. (2002). "On the Covariance between Functions." Journal of Multivariate Analysis **81**: 19-27.
- Curtis, D. (1997). "Use of siblings as controls in case-control association studies." Ann Hum Genet **61**(Pt 4): 319-33.
- Dalmaso, C., W. Carpentier, et al. (2008). "Distinct genetic loci control plasma HIV-RNA and cellular HIV-DNA levels in HIV-1 infection: the ANRS Genome Wide Association 01 study." PLoS One **3**(12): e3907.
- Daniels, M. J. and M. D. Hughes (1997). "Meta-analysis for the evaluation of potential surrogate markers." Stat Med **16**(17): 1965-82.
- Danta, M. and A. J. Rodger (2011). "Transmission of HCV in HIV-positive populations." Curr Opin HIV AIDS **6**(6): 451-8.
- Das, H., T. Atsumi, et al. (2004). "Diagnostic value of antiagalactosyl IgG antibodies in rheumatoid arthritis." Clin Rheumatol **23**(3): 218-22.
- Davis, P. and M. Stein (1989). "Evaluation of criteria for the classification of SLE in Zimbabwean patients." Br J Rheumatol **28**(6): 546-7.
- de Bakker, P. I., M. A. Ferreira, et al. (2008). "Practical aspects of imputation-driven meta-analysis of genome-wide association studies." Hum Mol Genet **17**(R2): R122-8.
- de Bois, M. H., J. W. Arndt, et al. (1996). "Technetium-99m labelled human immunoglobulin scintigraphy predicts rheumatoid arthritis in patients with arthralgia." Scand J Rheumatol **25**(3): 155-8.
- De Iorio, M., P. J. Newcombe, et al. (2011). "Bayesian semiparametric meta - analysis for genetic association studies." Genetic epidemiology **35**(5): 333-340.
- De Luca, A., R. Bugarini, et al. (2002). "Coinfection with hepatitis viruses and outcome of initial antiretroviral regimens in previously naive HIV-infected subjects." Arch Intern Med **162**(18): 2125-32.
- De Rycke, L., I. Peene, et al. (2004). "Rheumatoid factor and anticitrullinated protein antibodies in rheumatoid arthritis: diagnostic value, associations with radiological progression rate, and extra-articular manifestations." Ann Rheum Dis **63**(12): 1587-93.
- Dean, M., M. Carrington, et al. (1996). "Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study." Science **273**(5283): 1856-62.
- Deeks, J. J. (2001). "Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests." BMJ **323**(7305): 157-62.
- Delaporta, P., A. Kattamis, et al. (2013). "Correlation of NT-proBNP levels and cardiac iron concentration in patients with transfusion-dependent thalassemia major." Blood Cells Mol Dis **50**(1): 20-4.
- Deming, W. a. S. F. (1940). "On least square adjustment of sampled frequency tables when the expected marginal totals are known." Ann. Math. Statist. **6**: 427-444.
- Dendukuri, N., I. Schiller, et al. (2012). "Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference." Biometrics **68**(4): 1285-93.

- Deng, H. W., W. M. Chen, et al. (2002). "Transmission disequilibrium test with discordant sib pairs when parents are available." Hum Genet **110**(5): 451-61.
- DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials." Controlled clinical trials **7**(3): 177-188.
- DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials." Controlled Clinical Trials **7**: 177-188.
- DerSimonian, R. and N. Laird (1986). "Meta-analysis in clinical trials." Control Clin Trials **7**(3): 177-88.
- Despres, N., G. Boire, et al. (1994). "The Sa system: a novel antigen-antibody system specific for rheumatoid arthritis." J Rheumatol **21**(6): 1027-33.
- Devlin, B., S. A. Bacanu, et al. (2004). "Genomic Control to the extreme." Nat Genet **36**(11): 1129-30; author reply 1131.
- Devlin, B. and K. Roeder (1999). "Genomic control for association studies." Biometrics **55**(4): 997-1004.
- Devlin, B., K. Roeder, et al. (2003). "Analysis of multilocus models of association." Genet Epidemiol **25**(1): 36-47.
- Di Lello, F. A., A. Caruz, et al. (2013). "Effects of the genetic pattern defined by low-density lipoprotein receptor and IL28B genotypes on the outcome of hepatitis C virus infection." Eur J Clin Microbiol Infect Dis **32**(11): 1427-35.
- Di Pietrantonj, C. (2006). "Four-fold table cell frequencies imputation in meta analysis." Stat Med **25**(13): 2299-322.
- Dijksterhuis, J., J. Petersen, et al. (2014). "WNT/Frizzled signalling: receptor – ligand selectivity with focus on FZD – G protein signalling and its physiological relevance: IUPHAR Review 3." British journal of pharmacology **171**(5): 1195-1209.
- Dubucquoi, S., E. Solau-Gervais, et al. (2004). "Evaluation of anti-citrullinated filaggrin antibodies as hallmarks for the diagnosis of rheumatic diseases." Ann Rheum Dis **63**(4): 415-9.
- Duffy, S. W., T. E. Rohan, et al. (1989). "A method for combining matched and unmatched binary data. Application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy." Am J Epidemiol **130**(2): 371-8.
- Dunnett, C. W. (1955). "A Multiple Comparison Procedure for Comparing Several Treatments with a Control." Journal of the American Statistical Association **50**(272): 1096-1121.
- Duval, S. and R. Tweedie (2000). "A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis." JASA **95**(449): 89-98.
- Economou-Petersen, E., A. Aessopos, et al. (1998). "Apolipoprotein E epsilon4 allele as a genetic risk factor for left ventricular failure in homozygous beta-thalassemia." Blood **92**(9): 3455-9.
- Efthimiou, O., D. Mavridis, et al. (2013). "An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios." Stat Med **33**(13): 2275-87.
- Efthimiou, O., D. Mavridis, et al. (2014). "Joint synthesis of multiple correlated outcomes in networks of interventions." Biostatistics.
- Egger, M., G. Davey Smith, et al. (1997). "Bias in meta-analysis detected by a simple, graphical test." Bmj **315**(7109): 629-34.
- Egger, M., G. D. Smith, et al. (1997). "Bias in meta-analysis detected by a simple, graphical test." Bmj **315**(7109): 629-634.

- Ehret, G. B., P. B. Munroe, et al. (2011). "Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk." *Nature* **478**(7367): 103-9.
- Eichner, J. E., S. T. Dunn, et al. (2002). "Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review." *Am J Epidemiol* **155**(6): 487-95.
- El-Tagui, M. H., M. M. Hamdy, et al. (2013). "Apolipoprotein E gene polymorphism and the risk of left ventricular dysfunction among Egyptian beta-thalassemia major." *Gene* **524**(2): 292-5.
- Emokpae, M., E. Nwokedi, et al. (2008). "Impact of hepatitis C co-infection on CD4 T cell count in HIV infected subjects." *OJHAS* **7**: 1-4.
- Enoe, C., M. P. Georgiadis, et al. (2000). "Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown." *Prev Vet Med* **45**(1-2): 61-81.
- Epstein, M. P., C. D. Veal, et al. (2005). "Genetic association analysis using data from triads and unrelated subjects." *Am J Hum Genet* **76**(4): 592-608.
- Erikstrup, C., P. Kallestrup, et al. (2007). "Reduced mortality and CD4 cell loss among carriers of the interleukin-10 -1082G allele in a Zimbabwean cohort of HIV-1-infected adults." *AIDS* **21**(17): 2283-91.
- Evangelou, E., D. M. Maraganore, et al. (2007). "Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease." *PLoS One* **2**(2): e196.
- Evangelou, E., T. A. Trikalinos, et al. (2006). "Family-based versus unrelated case-control designs for genetic associations." *PLoS Genet* **2**(8): e123.
- Falk, C. T. and P. Rubinstein (1987). "Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations." *Ann Hum Genet* **51**(Pt 3): 227-33.
- Feinstein, A. R. (1987). "Quantitative ambiguities in matched versus unmatched analyses of the 2x2 table for a case-control study." *Int J Epidemiol* **16**(1): 128-34.
- Fellay, J. (2009). "Host genetics influences on HIV type-1 disease." *Antivir Ther* **14**(6): 731-8.
- Fellay, J., K. V. Shianna, et al. (2007). "A whole-genome association study of major determinants for host control of HIV-1." *Science* **317**(5840): 944-7.
- Fernandez-Suarez, A., S. Reneses, et al. (2005). "Efficacy of three ELISA measurements of anti-cyclic citrullinated peptide antibodies in the early diagnosis of rheumatoid arthritis." *Clin Chem Lab Med* **43**(11): 1234-9.
- Ferrara, M., S. M. Matarese, et al. (2001). "Role of apolipoprotein E (APOE) polymorphism on left cardiac failure in homozygous beta thalassaemic patients." *Br J Haematol* **114**(4): 959-60.
- Fienberg, S. E. (1970). "An Iterative Procedure for Estimation in Contingency Tables." *The Annals of Mathematical Statistics* **41**(3): 907-917.
- Fijnheer, R., M. Roest, et al. (1998). "Homocysteine, methylenetetrahydrofolate reductase polymorphism, antiphospholipid antibodies, and thromboembolic events in systemic lupus erythematosus: a retrospective cohort study." *J Rheumatol* **25**(9): 1737-42.
- Filippini, P., N. Coppola, et al. (2000). "Can HCV affect the efficacy of anti-HIV treatment?" *Arch Virol* **145**(5): 937-44.
- Fisher, M. C. and B. N. Cronstein (2009). "Metaanalysis of methylenetetrahydrofolate reductase (MTHFR) polymorphisms affecting methotrexate toxicity." *J Rheumatol* **36**(3): 539-45.

- Flint, J., R. M. Harding, et al. (1998). "1 The population genetics of the haemoglobinopathies." Baillière's clinical haematology **11**(1): 1-51.
- Franceschini, N., E. Fox, et al. (2013). "Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations." Am J Hum Genet **93**(3): 545-54.
- Franke, B., A. A. Vasquez, et al. (2010). "Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD." Neuropsychopharmacology **35**(3): 656-64.
- Frankenburg, F. R. (2007). "The role of one-carbon metabolism in schizophrenia and depression." Harv Rev Psychiatry **15**(4): 146-60.
- Freidlin, B., M. J. Podgor, et al. (1999). "Efficiency Robust Tests for Survival or Ordered Categorical Data." Biometrics **55**(3): 883-886.
- Freidlin, B., G. Zheng, et al. (2002). "Trend tests for case-control studies of genetic markers: power, sample size and robustness." Hum Hered **53**(3): 146-52.
- Freitas, F. B., S. S. Lima, et al. (2015). "Polymorphisms in the IFN γ , IL-10, and TGF β genes may be associated with HIV-1 infection." Dis Markers **2015**: 248571.
- Frosst, P., H. J. Blom, et al. (1995). "A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase." Nat Genet **10**(1): 111-3.
- Furman, L. M. (2008). "Attention-deficit hyperactivity disorder (ADHD): does new research support old concepts?" J Child Neurol **23**(7): 775-84.
- Galanello, R. and R. Origa (2010). "Beta-thalassemia." Orphanet J Rare Dis **5**: 11.
- Garcia-Berrocal, B., C. Gonzalez, et al. (2005). "Anti-cyclic citrullinated peptide autoantibodies in IgM rheumatoid factor-positive patients." Clin Chim Acta **354**(1-2): 123-30.
- Gardner, I. A., H. Stryhn, et al. (2000). "Conditional dependence between tests affects the diagnosis and surveillance of animal diseases." Prev Vet Med **45**(1-2): 107-22.
- Gart, J. J. and J. R. Zweifel (1967). "On the bias of various estimators of the logit and its variance with application to quantal bioassay." Biometrika **54**(1): 181-7.
- Gastwirth, J. L. (1985). "The Use of Maximin Efficiency Robust Tests in Combining Contingency Tables and Survival Analysis." Journal of the American Statistical Association **80**(390): 380-384.
- Gebregziabher, M., P. Guimaraes, et al. (2010). "A polytomous conditional likelihood approach for combining matched and unmatched case-control studies." Stat Med.
- Gemmati, D., A. Ongaro, et al. (2007). "Methylenetetrahydrofolate reductase C677T and A1298C gene variants in adult non-Hodgkin's lymphoma patients: association with toxicity and survival." Haematologica **92**(4): 478-85.
- Gervasini, G. (2009). "Polymorphisms in methotrexate pathways: what is clinically relevant, what is not, and what is promising." Curr Drug Metab **10**(6): 547-66.
- Ghodke, Y., A. Chopra, et al. (2008). "Are Thymidylate synthase and Methylene tetrahydrofolate reductase genes linked with methotrexate response (efficacy, toxicity) in Indian (Asian) rheumatoid arthritis patients?" Clin Rheumatol **27**(6): 787-9.
- Gibson, G. E., V. Haroutunian, et al. (2000). "Mitochondrial damage in Alzheimer's disease varies with apolipoprotein E genotype." Ann Neurol **48**(3): 297-303.

- Giordano, T. P., C. Hartman, et al. (2009). "Predictors of retention in HIV care among a national cohort of US veterans." HIV Clin Trials **10**(5): 299-305.
- Girelli, F., F. G. Foschi, et al. (2004). "Is Anti Cyclic citrullinated peptide a useful laboratory test for the diagnosis of rheumatoid arthritis?" Eur Ann Allergy Clin Immunol **36**(4): 127-30.
- Glas, A. S., J. G. Lijmer, et al. (2003). "The diagnostic odds ratio: a single indicator of test performance." J Clin Epidemiol **56**(11): 1129-35.
- Glaser, B. and P. Holmans (2009). "Comparison of methods for combining case-control and family-based association studies." Hum Hered **68**(2): 106-16.
- Glass, G. (1976). "Primary, secondary and meta-analysis of research." Educ Res **5**: 3-8.
- Glass, G. (1976). "Primary, secondary and meta-analysis of research." Educational Research **5**: 3-8.
- Gleser, L. J. and I. Olkin (1994). Stochastically dependent effect sizes. The Handbook of Research Synthesis H. M. Cooper and L. V. Hedges. New York, Russell Sage Foundation: 339-355.
- Goldbach-Mansky, R., J. Lee, et al. (2000). "Rheumatoid arthritis associated autoantibodies in patients with synovitis of recent onset." Arthritis Res **2**(3): 236-43.
- Golder, S., Y. K. Loke, et al. (2011). "Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview." PLoS medicine **8**(5): 717.
- Gomes-Daudrix, V., M. Sebbag, et al. (1994). "Immunoblotting detection of so-called 'antikeratin antibodies': a new assay for the diagnosis of rheumatoid arthritis." Ann Rheum Dis **53**(11): 735-42.
- Goncalves, L. S., S. M. Ferreira, et al. (2009). "Influence of IL-1 gene polymorphism on the periodontal microbiota of HIV-infected Brazilian individuals." Braz Oral Res **23**(4): 452-9.
- Gonzalez, J. R., J. L. Carrasco, et al. (2008). "Maximizing association statistics over genetic models." Genet Epidemiol **32**(3): 246-54.
- Gonzalez, V. D., K. Falconer, et al. (2009). "High levels of chronic immune activation in the T-cell compartments of patients coinfectd with hepatitis C virus and human immunodeficiency virus type 1 and on highly active antiretroviral therapy are reverted by alpha interferon and ribavirin treatment." J Virol **83**(21): 11407-11.
- Goyette, P., J. S. Sumner, et al. (1994). "Human methylenetetrahydrofolate reductase: isolation of cDNA mapping and mutation identification." Nat Genet **7**(4): 551.
- Grady, B. P., M. Prins, et al. (2015). "BMI, male sex and IL28B genotype associated with persistently high hepatitis C virus RNA levels among chronically infected drug users up to 23 years following seroconversion." J Viral Hepat **22**(3): 263-71.
- Grebely, J., M. Hellard, et al. (2012). "Virological responses during treatment for recent hepatitis C virus: potential benefit for ribavirin use in HCV/HIV co-infection." AIDS **26**(13): 1653-61.
- Greene, N. D., P. Stanier, et al. (2009). "Genetics of human neural tube defects." Hum Mol Genet **18**(R2): R113-29.
- Greenland, S. (1998). Meta-analysis. Modern Epidemiology. K. J. Rothman and S. Greenland, Lippincott Williams & Wilkins: 643-673.

- Greenland, S. and M. P. Longnecker (1992). "Methods for trend estimation from summarized dose-response data, with applications to meta-analysis." Am J Epidemiol **135**(11): 1301-9.
- Greenland, S. and J. M. Robins (1985). "Estimation of a common effect parameter from sparse follow-up data." Biometrics **41**(1): 55-68.
- Greiner, A., H. Plischke, et al. (2005). "Association of anti-cyclic citrullinated peptide antibodies, anti-citrullin antibodies, and IgM and IgA rheumatoid factors with serological parameters of disease activity in rheumatoid arthritis." Ann N Y Acad Sci **1050**: 295-303.
- Gress, T. W., F. J. Nieto, et al. (2000). "Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. Atherosclerosis Risk in Communities Study." N Engl J Med **342**(13): 905-12.
- Greub, G., B. Ledergerber, et al. (2000). "Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus coinfection: the Swiss HIV Cohort Study." Lancet **356**(9244): 1800-5.
- Grootenboer-Mignot, S., P. Nicaise-Roland, et al. (2004). "Second generation anti-cyclic citrullinated peptide (anti-CCP2) antibodies can replace other anti-filaggrin antibodies and improve rheumatoid arthritis diagnosis." Scand J Rheumatol **33**(4): 218-20.
- Guo, C. Y., K. L. Lunetta, et al. (2009). "Combined haplotype relative risk (CHRR): a general and simple genetic association test that combines trios and unrelated case-controls." Genet Epidemiol **33**(1): 54-62.
- Hafler, D. A., A. Compston, et al. (2007). "Risk alleles for multiple sclerosis identified by a genomewide study." N Engl J Med **357**(9): 851-62.
- Hahalis, G., D. Alexopoulos, et al. (2005). "Heart failure in beta-thalassemia syndromes: a decade of progress." Am J Med **118**(9): 957-67.
- Hajarizadeh, B., B. Grady, et al. (2015). "Patterns of hepatitis C virus RNA levels during acute infection: the InC3 study." PLoS One **10**(4): e0122232.
- Haldane, J. B. (1956). "The estimation and significance of the logarithm of a ratio of frequencies." Ann Hum Genet **20**(4): 309-11.
- Hamajima, N., K. Hirose, et al. (1994). "Case-control studies: matched controls or all available controls?" J Clin Epidemiol **47**(9): 971-5.
- Hamza, T. H., J. B. Reitsma, et al. (2008). "Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches." Med Decis Making **28**(5): 639-49.
- Hamza, T. H., H. C. van Houwelingen, et al. (2008). "The binomial distribution of meta-analysis was preferred to model within-study variability." J Clin Epidemiol **61**(1): 41-51.
- Hancock, D. B., N. C. Gaddis, et al. (2015). "Associations of common variants in the BST2 region with HIV-1 acquisition in African American and European American people who inject drugs." AIDS **29**(7): 767-77.
- HapMap. "<http://hapmap.ncbi.nlm.nih.gov/>." HapMap homepage.
- Harbord, R. M., J. J. Deeks, et al. (2007). "A unification of models for meta-analysis of diagnostic accuracy studies." Biostatistics **8**(2): 239-51.
- Hardy, R. J. and S. G. Thompson (1996). "A likelihood approach to meta-analysis with random effects." Stat Med **15**(6): 619-29.
- Harten, P. (2005). "[Reducing toxicity of methotrexate with folic acid]." Z Rheumatol **64**(5): 353-8.

- Hashimoto, R., T. Suzuki, et al. (2005). "Association study of the frizzled-3 (FZD3) gene with schizophrenia and mood disorders." Journal of neural transmission **112**(2): 303-307.
- Hayakawa, T., T. Takamura, et al. (2007). "Association of the C825T polymorphism of the G-protein beta3 subunit gene with hypertension, obesity, hyperlipidemia, insulin resistance, diabetes, diabetic complications, and diabetic therapies among Japanese." Metabolism **56**(1): 44-8.
- Henze, N. and B. Zirkler (1990). "A class of invariant consistent tests for multivariate normality." Communications in Statistics - Theory and Methods **19**(10): 3595–3617.
- Herrlinger, K. R., J. R. Cummings, et al. (2005). "The pharmacogenetics of methotrexate in inflammatory bowel disease." Pharmacogenet Genomics **15**(10): 705-11.
- Herrlinger, K. R. and D. P. Jewell (2006). "Review article: interactions between genotype and response to therapy in inflammatory bowel diseases." Aliment Pharmacol Ther **24**(10): 1403-12.
- Hider, S. L., W. Thomson, et al. (2008). "Polymorphisms within the adenosine receptor 2a gene are associated with adverse events in RA patients treated with MTX." Rheumatology (Oxford) **47**(8): 1156-9.
- Higgins, J. and A. Whitehead (1996). "Borrowing strength from external trials in a meta - analysis." Statistics in medicine **15**(24): 2733-2749.
- Higgins, J. P., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." BMJ **327**(7414): 557-60.
- Higgins, J. P. and A. Whitehead (1996). "Borrowing strength from external trials in a meta-analysis." Stat Med **15**(24): 2733-49.
- Higgins, J. P., A. Whitehead, et al. (2001). "Meta-analysis of continuous outcome data from individual patients." Stat Med **20**(15): 2219-41.
- Hindorff, L. A., J. MacArthur, et al. (2009). "A Catalog of Published Genome-Wide Association Studies." Retrieved 2011, from www.genome.gov/gwastudies.
- Hirschhorn, J. N., K. Lohmueller, et al. (2002). "A comprehensive review of genetic association studies." Genet Med **4**(2): 45-61.
- Hitchon, C. A., P. Alex, et al. (2004). "A distinct multicytokine profile is associated with anti-cyclical citrullinated peptide antibodies in patients with early untreated inflammatory arthritis." J Rheumatol **31**(12): 2336-46.
- Horberg, M., M. Silverberg, et al. (2008). "Influence of prior antiretroviral experience on adherence and responses to new highly active antiretroviral therapy regimens." AIDS Patient Care STDS **22**(4): 301-12.
- Horvath, S. and N. M. Laird (1998). "A discordant-sibship test for disequilibrium and linkage: no need for parental data." Am J Hum Genet **63**(6): 1886-97.
- Hothorn, L. A. and T. Hothorn (2009). "Order-restricted scores test for the evaluation of population-based case-control studies when the genetic model is unknown." Biom J **51**(4): 659-69.
- Hou, W., X. Wang, et al. (2009). "Lambda interferon inhibits human immunodeficiency virus type 1 infection of macrophages." J Virol **83**(8): 3834-42.
- Hsu, L., J. R. Starr, et al. (2009). "On combining triads and unrelated subjects data in candidate gene studies: an application to data on testicular cancer." Hum Hered **67**(2): 88-103.

- Huang, J., A. D. Johnson, et al. (2011). "PRIME: a method for characterization and evaluation of pleiotropic regions from multiple genome-wide association studies." *Bioinformatics* **27**(9): 1201-6.
- Hughes, L. B., T. M. Beasley, et al. (2006). "Racial or ethnic differences in allele frequencies of single-nucleotide polymorphisms in the methylenetetrahydrofolate reductase gene and their influence on response to methotrexate in rheumatoid arthritis." *Ann Rheum Dis* **65**(9): 1213-8.
- Hui, S. L. and X. H. Zhou (1998). "Evaluation of diagnostic tests without gold standards." *Stat Methods Med Res* **7**(4): 354-70.
- Hunt, P. W., J. N. Martin, et al. (2003). "T cell activation is associated with lower CD4+ T cell gains in human immunodeficiency virus-infected patients with sustained viral suppression during antiretroviral therapy." *J Infect Dis* **187**(10): 1534-43.
- Ide, M., T. Muratake, et al. (2004). "Genetic and expression analyses of *FZD3* in schizophrenia." *Biological psychiatry* **56**(6): 462-465.
- Imanishi, H., N. Okamura, et al. (2007). "Genetic polymorphisms associated with adverse events and elimination of methotrexate in childhood acute lymphoblastic leukemia and malignant lymphoma." *J Hum Genet* **52**(2): 166-71.
- Infante-Rivard, C., L. Mirea, et al. (2009). "Combining case-control and case-trio data from the same population in genetic association analyses: overview of approaches and illustration with a candidate gene study." *Am J Epidemiol* **170**(5): 657-64.
- Ingiliz, P., I. Krznaric, et al. (2014). "Multiple hepatitis C virus (HCV) reinfections in HIV-positive men who have sex with men: no influence of HCV genotype switch or interleukin-28B genotype on spontaneous clearance." *HIV Med* **15**(6): 355-61.
- Ioannidis, J. P., C. D. Mulrow, et al. (2006). "Adverse events: the more you search, the more you find." *Annals of internal medicine* **144**(4): 298-300.
- Ioannidis, J. P., E. E. Ntzani, et al. (2001). "Replication validity of genetic association studies." *Nat Genet* **29**(3): 306-9.
- Ioannidis, J. P., P. S. Rosenberg, et al. (2001). "Effects of CCR5-Delta32, CCR2-64I, and SDF-1 3'A alleles on HIV-1 disease progression: An international meta-analysis of individual-patient data." *Ann Intern Med* **135**(9): 782-95.
- Ioannidis, J. P. and T. A. Trikalinos (2005). "Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials." *J Clin Epidemiol* **58**(6): 543-9.
- Ioannidis, J. P., T. A. Trikalinos, et al. (2003). "Genetic associations in large versus small studies: an empirical assessment." *Lancet* **361**(9357): 567-71.
- Jablonowska, E., A. Kolacinska, et al. (2010). "Interleukin-6 and the IL-6 (-174) C/G polymorphism in breast pathologies and in HIV-infected patients." *Arch Med Sci* **6**(6): 860-5.
- Jablonowska, E., A. Piekarska, et al. (2012). "Sustained virologic response and IL28B single-nucleotide polymorphisms in patients with chronic hepatitis C treated with pegylated interferon alfa and ribavirin." *Acta Biochim Pol* **59**(3): 333-7.
- Jackson, D., R. Riley, et al. (2011). "Multivariate meta-analysis: Potential and promise." *Stat Med*.
- Jackson, D., I. R. White, et al. (2010). "Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses." *Stat Med* **29**(12): 1282-97.

- Jansen, L. M., D. van Schaardenburg, et al. (2003). "The predictive value of anti-cyclic citrullinated peptide antibodies in early arthritis." J Rheumatol **30**(8): 1691-5.
- Janssen, R., L. Bont, et al. (2007). "Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes." Journal of Infectious Diseases **196**(6): 826-834.
- Janssen, R., L. Bont, et al. (2007). "Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes." J Infect Dis **196**(6): 826-34.
- Janssens, A. C., A. M. Gonzalez-Zuloeta Ladd, et al. (2009). "An empirical comparison of meta-analyses of published gene-disease associations versus consortium analyses." Genet Med **11**(3): 153-62.
- Jazbec, J., L. Kitanovski, et al. (2005). "No evidence of association of methylenetetrahydrofolate reductase polymorphism with occurrence of second neoplasms after treatment of childhood leukemia." Leuk Lymphoma **46**(6): 893-7.
- Jennes, W., B. Vuylsteke, et al. (2004). "HIV-specific T helper responses and frequency of exposure among HIV-exposed seronegative female sex workers in Abidjan, Cote d'Ivoire." J Infect Dis **189**(4): 602-10.
- Jeong, S. H., E. J. Joo, et al. (2006). "Investigation of genetic association between human Frizzled homolog 3 gene (FZD3) and schizophrenia: results in a Korean population and evidence from meta-analysis." Psychiatry Res **143**(1): 1-11.
- Johnson, E. O., D. B. Hancock, et al. (2015). "Novel genetic locus implicated for HIV-1 acquisition with putative regulatory links to HIV replication and infectivity: a genome-wide association study." PLoS One **10**(3): e0118149.
- Jonsson, T., K. Steinsson, et al. (1998). "Combined elevation of IgM and IgA rheumatoid factor has high diagnostic specificity for rheumatoid arthritis." Rheumatol Int **18**(3): 119-22.
- Joo, J., M. Kwak, et al. (2009). "A robust genome-wide scan statistic of the Wellcome Trust Case-Control Consortium." Biometrics **65**(4): 1115-22.
- Joo, J., M. Kwak, et al. (2010). "Improving power for testing genetic association in case-control studies by reducing the alternative space." Biometrics **66**(1): 266-76.
- Joo, J., X. Tian, et al. (2007). "Joint analysis of case-parents trio and unrelated case-control designs in large scale association studies." BMC Proc **1 Suppl 1**: S28.
- Kalayoglu-Besisik, S., Y. Caliskan, et al. (2003). "Methylenetetrahydrofolate reductase C677T polymorphism and toxicity in allogeneic hematopoietic cell transplantation." Transplantation **76**(12): 1775-7.
- Kallas, E., K. Huik, et al. (2015). "Influence of interleukin 10 polymorphisms -592 and -1082 to the HIV, HBV and HCV serostatus among intravenous drug users." Infect Genet Evol **30**: 175-80.
- Kamali, S., N. G. Polat, et al. (2005). "Anti-CCP and antikeratin antibodies in rheumatoid arthritis, primary Sjogren's syndrome, and Wegener's granulomatosis." Clin Rheumatol **24**(6): 673-6.
- Kaplan, N. and R. Morris (2001). "Issues concerning association studies for fine mapping a susceptibility gene for a complex disease." Genetic epidemiology **20**(4): 432-457.
- Kato, N., T. Sugiyama, et al. (1998). "G protein beta3 subunit variant and essential hypertension in Japanese." Hypertension **32**(5): 935-8.

- Katsu, T., H. Ujike, et al. (2003). "The human frizzled-3 (< i> FZD3</i>) gene on chromosome 8p21, a receptor gene for Wnt ligands, is associated with the susceptibility to schizophrenia." Neuroscience letters **353**(1): 53-56.
- Kaur, G. and N. Mehra (2009). "Genetic determinants of HIV-1 infection and progression to AIDS: susceptibility to HIV infection." Tissue Antigens **73**(4): 289-301.
- Kazeem, G. R. and M. Farrall (2005). "Integrating case-control and TDT studies." Ann Hum Genet **69**(Pt 3): 329-35.
- Keller, T. and I. Olkin (2004). Combining correlated unbiased estimators of the mean of a normal distribution. A Festschrift for Herman Rubin. A. DasGupta, Institute of Mathematical Statistics. **45**: 218-227.
- Kim, I., K. H. Lee, et al. (2007). "Polymorphisms of the methylenetetrahydrofolate reductase gene and clinical outcomes in HLA-matched sibling allogeneic hematopoietic stem cell transplantation." Ann Hematol **86**(1): 41-8.
- Kim, S. K., J. B. Jun, et al. (2006). "Cost-effectiveness analysis of MTHFR polymorphism screening by polymerase chain reaction in Korean patients with rheumatoid arthritis receiving methotrexate." J Rheumatol **33**(7): 1266-74.
- Kirkham, J. J., R. D. Riley, et al. (2012). "A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews." Stat Med **31**(20): 2179-95.
- Kishi, S., J. Griener, et al. (2003). "Homocysteine, pharmacogenetics, and neurotoxicity in children with leukemia." J Clin Oncol **21**(16): 3084-91.
- Klein, M. B., R. G. Lalonde, et al. (2003). "The impact of hepatitis C virus coinfection on HIV progression before and after highly active antiretroviral therapy." J Acquir Immune Defic Syndr **33**(3): 365-72.
- Klein, R. J., C. Zeiss, et al. (2005). "Complement factor H polymorphism in age-related macular degeneration." Science **308**(5720): 385-9.
- Konenkov, V. I. and M. V. Smol'nikova (2002). "Polymorphism of promotor sites of interleukins-4 and -10 and tumor necrosis factor-alpha genes in HIV-infected patients." Bull Exp Biol Med **133**(4): 389-91.
- Kooloos, W. M., J. A. Wessels, et al. "Functional polymorphisms and methotrexate treatment outcome in recent-onset rheumatoid arthritis." Pharmacogenomics **11**(2): 163-75.
- Korner, C., B. Kramer, et al. (2009). "Effects of HCV co-infection on apoptosis of CD4+ T-cells in HIV-positive patients." Clin Sci (Lond) **116**(12): 861-70.
- Kremastinos, D. T., D. Farmakis, et al. (2010). "β-Thalassemia Cardiomyopathy History, Present Considerations, and Future Perspectives." Circulation: Heart Failure **3**(3): 451-458.
- Kremer, J. M. (1998). "Methotrexate and emerging therapies." Rheum Dis Clin North Am **24**(3): 651-8.
- Kremer, J. M. (1999). "Methotrexate and emerging therapies." Clin Exp Rheumatol **17**(6 Suppl 18): S43-6.
- Kremer, J. M. (2004). "Toward a better understanding of methotrexate." Arthritis Rheum **50**(5): 1370-82.
- Kruglyak, L. (1997). "The use of a genetic map of biallelic markers in linkage studies." Nat Genet **17**(1): 21-4.
- Krull, K. R., P. Brouwers, et al. (2008). "Folate pathway genetic polymorphisms are related to attention disorders in childhood leukemia survivors." J Pediatr **152**(1): 101-5.

- Kuhner, M. K., P. Beerli, et al. (2000). "Usefulness of single nucleotide polymorphism data for estimating population parameters." Genetics **156**(1): 439-47.
- Kumagai, K., K. Hiyama, et al. (2003). "Polymorphisms in the thymidylate synthase and methylenetetrahydrofolate reductase genes and sensitivity to the low-dose methotrexate therapy in patients with rheumatoid arthritis." Int J Mol Med **11**(5): 593-600.
- Kumagai, S., K. Nishimura, et al. (2004). "[Topics on immunological tests for rheumatoid arthritis]." Rinsho Byori **52**(10): 836-43.
- Kurzwawska, G., A. Seremak-Mrozikiewicz, et al. (2009). "[Genetic conditioned changes in activity of 5,10-methylenetetrahydrofolate reductase (MTHFR) and recurrent miscarriages]." Ginek Pol **80**(10): 762-7.
- Kwa, D., R. P. van Rij, et al. (2003). "Association between an interleukin-4 promoter polymorphism and the acquisition of CXCR4 using HIV-1 variants." AIDS **17**(7): 981-5.
- Kwok, J. S., K. H. Hui, et al. (2005). "Anti-cyclic citrullinated peptide: diagnostic and prognostic values in juvenile idiopathic arthritis and rheumatoid arthritis in a Chinese population." Scand J Rheumatol **34**(5): 359-66.
- Langefeld, C. D. and T. E. Fingerlin (2007). "Association methods in human genetics." Methods Mol Biol **404**: 431-60.
- Langhans, B., B. Kupfer, et al. (2011). "Interferon-lambda serum levels in hepatitis C." J Hepatol **54**(5): 859-65.
- Laskus, T., M. Radkowski, et al. (2004). "Human immunodeficiency virus facilitates infection/replication of hepatitis C virus in native human macrophages." Blood **103**(10): 3854-9.
- Lasky-Su, J., R. J. Anney, et al. (2008). "Genome-wide association scan of the time to onset of attention deficit hyperactivity disorder." Am J Med Genet B Neuropsychiatr Genet **147B**(8): 1355-8.
- Lasky-Su, J., S. Won, et al. (2010). "On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls." Am J Hum Genet **86**(4): 573-80.
- Last, J. (2001). "A dictionary of epidemiology." New York: Oxford University Press.
- Lau, J., E. M. Antman, et al. (1992). "Cumulative meta-analysis of therapeutic trials for myocardial infarction." New England Journal of Medicine **327**(4): 248-254.
- Lau, J., E. M. Antman, et al. (1992). "Cumulative meta-analysis of therapeutic trials for myocardial infarction." N Engl J Med **327**(4): 248-54.
- Lau, J., C. H. Schmid, et al. (1995). "Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care." J Clin Epidemiol **48**(1): 45-57; discussion 59-60.
- Lau, J., C. H. Schmid, et al. (1995). "Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care." Journal of clinical epidemiology **48**(1): 45-57.
- le Cessie, S., N. Nagelkerke, et al. (2008). "Combining matched and unmatched control groups in case-control studies." Am J Epidemiol **168**(10): 1204-10.
- Le Clerc, S., S. Limou, et al. (2009). "Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03)." J Infect Dis **200**(8): 1194-201.
- Lee, D. M. and P. H. Schur (2003). "Clinical utility of the anti-CCP assay in patients with rheumatic diseases." Ann Rheum Dis **62**(9): 870-4.

- Lee, W. C. (2003). "Genetic association studies of adult-onset diseases using the case-spouse and case-offspring designs." Am J Epidemiol **158**(11): 1023-32.
- Lee, W. C. (2003). "Searching for disease-susceptibility loci by testing for Hardy-Weinberg disequilibrium in a gene bank of affected individuals." Am J Epidemiol **158**(5): 397-400.
- Lettre, G., C. Lange, et al. (2007). "Genetic model testing and statistical power in population-based association studies of quantitative traits." Genet Epidemiol **31**(4): 358-62.
- Levy, J. A. (2009). "HIV pathogenesis: 25 years of progress and persistent challenges." AIDS **23**(2): 147-60.
- Lewis, C. M. (2002). "Genetic association studies: design, analysis and interpretation." Brief Bioinform **3**(2): 146-53.
- Lewis, C. M. and J. Knight (2012). "Introduction to genetic association studies." Cold Spring Harb Protoc **2012**(3): 297-306.
- Li, J. and J. P. Fine (2011). "Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis." Biostatistics **12**(4): 710-22.
- Li, M., X. Liu, et al. (2009). "Interferon-lambdas: the modulators of antiviral, antitumor, and immune responses." J Leukoc Biol **86**(1): 23-32.
- Li, Q., G. Zheng, et al. (2008). "Efficient approximation of P-value of the maximum of correlated tests, with applications to genome-wide association studies." Ann Hum Genet **72**(Pt 3): 397-406.
- Lichtenstein, P., B. H. Yip, et al. (2009). "Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study." The Lancet **373**(9659): 234-239.
- Limou, S., O. Delaneau, et al. (2012). "Multicohort genomewide association study reveals a new signal of protection against HIV-1 acquisition." J Infect Dis **205**(7): 1155-62.
- Lin, D. Y. and P. F. Sullivan (2009). "Meta-analysis of genome-wide association studies with overlapping subjects." Am J Hum Genet **85**(6): 862-72.
- Lin, D. Y. and D. Zeng (2010). "Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data." Genet Epidemiol **34**(1): 60-6.
- Lin, Y. and D. Zeng (2010). "On the relative efficiency of using summary statistics versus individual-level data in meta-analysis." Biometrika **97**(2): 321-332.
- Lincoln, D., K. Petoumenos, et al. (2003). "HIV/HBV and HIV/HCV coinfection, and outcomes following highly active antiretroviral therapy." HIV Med **4**(3): 241-9.
- Littenberg, B. and L. E. Moses (1993). "Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method." Med Decis Making **13**(4): 313-21.
- Liu, C. M., H. G. Hwu, et al. (2005). "Linkage evidence of schizophrenia to loci near neuregulin 1 gene on chromosome 8p21 in Taiwanese families." American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **134**(1): 79-83.
- Liu, Y., Y. Chen, et al. (2015). "A unification of models for meta-analysis of diagnostic accuracy studies without a gold standard." Biometrics **71**(2): 538-47.
- Livak, K. J., J. Marmaro, et al. (1995). "Towards fully automated genome-wide polymorphism screening." Nat Genet **9**(4): 341-2.

- Loke, Y. K., S. P. Golder, et al. (2011). "Comprehensive evaluations of the adverse effects of drugs: importance of appropriate study selection and data sources." Therapeutic advances in drug safety **2**(2): 59-68.
- Longo, F., G. Zecchina, et al. (1999). "The influence of hemochromatosis mutations on iron overload of thalassemia major." Haematologica **84**(9): 799-803.
- Lopez-Hoyos, M., C. Ruiz de Alegria, et al. (2004). "Clinical utility of anti-CCP antibodies in the differential diagnosis of elderly-onset rheumatoid arthritis and polymyalgia rheumatica." Rheumatology (Oxford) **43**(5): 655-7.
- Isa, S. E., L. N. Gwamzhi, et al. (2010). "A prospective cohort study of immunologic and virologic outcomes in patients with HIV/AIDS and hepatitis virus co-infection in Jos, Nigeria." Niger J Med **19**(3): 279-85.
- Lu, G. and A. E. Ades (2004). "Combination of direct and indirect evidence in mixed treatment comparisons." Stat Med **23**(20): 3105-24.
- Lumley, T. (2002). "Network meta - analysis for indirect treatment comparisons." Statistics in medicine **21**(16): 2313-2324.
- Luo, Z., G. Q. Zhang, et al. (2013). "Mining patterns of adverse events using aggregated clinical trial results." AMIA Summits on Translational Science Proceedings **2013**: 112.
- Ma, X., Y. Chen, et al. (2014). "A hybrid Bayesian hierarchical model combining cohort and case-control studies for meta-analysis of diagnostic tests: Accounting for partial verification bias." Stat Methods Med Res.
- Ma, X., L. Nie, et al. (2013). "Statistical methods for multivariate meta-analysis of diagnostic tests: An overview and tutorial." Stat Methods Med Res.
- Macaskill, P. (2004). "Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis." J Clin Epidemiol **57**(9): 925-32.
- Macaskill, P., S. D. Walter, et al. (2001). "A comparison of methods to detect publication bias in meta-analysis." Stat Med **20**(4): 641-54.
- Macias, J., J. A. Pineda, et al. (1998). "Influence of hepatitis C virus infection on the mortality of antiretroviral-treated patients with HIV disease." Eur J Clin Microbiol Infect Dis **17**(3): 167-70.
- Macias, J., J. A. Pineda, et al. (2003). "Impaired recovery of CD4+ cell counts following highly active antiretroviral therapy in drug-naïve patients coinfecting with human immunodeficiency virus and hepatitis C virus." Eur J Clin Microbiol Infect Dis **22**(11): 675-80.
- Magi, R. and A. P. Morris (2010). "GWAMA: software for genome-wide association meta-analysis." BMC Bioinformatics **11**: 288.
- Mahley, R. W. and S. C. Rall, Jr. (2000). "Apolipoprotein E: far more than a lipid transport protein." Annu Rev Genomics Hum Genet **1**: 507-37.
- Malbon, C. C. (2004). "Frizzleds: new members of the superfamily of G-protein-coupled receptors." Frontiers in bioscience: a journal and virtual library **9**: 1048-1058.
- Malta, M., S. A. Strathdee, et al. (2008). "Adherence to antiretroviral therapy for human immunodeficiency virus/acquired immune deficiency syndrome among drug users: a systematic review." Addiction **103**(8): 1242-57.
- Mandrekar, J. N. (2010). "Receiver operating characteristic curve in diagnostic test assessment." J Thorac Oncol **5**(9): 1315-6.
- Manolio, T. A. (2010). "Genomewide association studies and assessment of the risk of disease." N Engl J Med **363**(2): 166-76.

- Marchini, J., D. Cutler, et al. (2006). "A comparison of phasing algorithms for trios and unrelated individuals." The American Journal of Human Genetics **78**(3): 437-450.
- Marchini, J., P. Donnelly, et al. (2005). "Genome-wide strategies for detecting multiple loci that influence complex diseases." Nat Genet **37**(4): 413-7.
- Marshall, A. W. and I. Olkin (1985). "A Family of Bivariate Distributions Generated by the Bivariate Bernoulli Distribution." Journal of the American Statistical Association **80**(390): 332-338.
- Martin, M. P., Y. Qi, et al. (2010). "IL28B polymorphism does not determine outcomes of hepatitis B virus or HIV infection." J Infect Dis **202**(11): 1749-53.
- Mathew, T. and K. Nordstrom (1999). "On the equivalence of meta-analysis using literature and using individual patient data." Biometrics **55**(4): 1221-3.
- Mavridis, D. and G. Salanti (2012). "A practical introduction to multivariate meta-analysis." Stat Methods Med Res.
- Max, J. E., A. E. Lansing, et al. (2004). "Attention deficit hyperactivity disorder in children and adolescents following traumatic brain injury." Dev Neuropsychol **25**(1-2): 159-77.
- McCullagh, P. and J. A. Nelder (1989). Generalized Linear Models. London, Chapman & Hall.
- McKeigue, P. M. (1998). "Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture." Am J Hum Genet **63**(1): 241-51.
- Mehta, S. H., G. Lucas, et al. (2007). "Early immunologic and virologic responses to highly active antiretroviral therapy and subsequent disease progression among HIV-infected injection drug users." AIDS Care **19**(5): 637-45.
- Melis, M. A., M. Cau, et al. (2002). "H63D mutation in the HFE gene increases iron overload in beta-thalassemia carriers." Haematologica **87**(3): 242-5.
- Merrill, J. T., C. Shen, et al. (1997). "Adenosine A1 receptor promotion of multinucleated giant cell formation by human monocytes: a mechanism for methotrexate-induced nodulosis in rheumatoid arthritis." Arthritis Rheum **40**(7): 1308-15.
- Miller, M. F., C. Haley, et al. (2005). "Impact of hepatitis C virus on immune restoration in HIV-infected patients who start highly active antiretroviral therapy: a meta-analysis." Clin Infect Dis **41**(5): 713-20.
- Minelli, C., J. R. Thompson, et al. (2005). "Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies." Stat Med **24**(24): 3845-61.
- Minelli, C., J. R. Thompson, et al. (2005). "The choice of a genetic model in the meta-analysis of molecular association studies." Int J Epidemiol **34**(6): 1319-28.
- Minelli, C., J. R. Thompson, et al. (2008). "How should we use information about HWE in the meta-analyses of genetic association studies?" Int J Epidemiol **37**(1): 136-46.
- Minelli, C., J. R. Thompson, et al. (2004). "An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization." Am J Epidemiol **160**(5): 445-52.
- Mitchell, L. E. (2000). "Relationship between case-control studies and the transmission/disequilibrium test." Genet Epidemiol **19**(3): 193-201.
- Miyata, M. and J. D. Smith (1996). "Apolipoprotein E allele-specific antioxidant activity and effects on cytotoxicity by oxidative insults and beta-amyloid peptides." Nat Genet **14**(1): 55-61.

- Miyazawa, M., L. Lopalco, et al. (2009). "The 'immunologic advantage' of HIV-exposed seronegative individuals." *AIDS* **23**(2): 161-75.
- Mocroft, A., S. Madge, et al. (1999). "A comparison of exposure groups in the EuroSIDA study: starting highly active antiretroviral therapy (HAART), response to HAART, and survival." *J Acquir Immune Defic Syndr* **22**(4): 369-78.
- Modi, W. S., T. R. O'Brien, et al. (2003). "Haplotype diversity in the interleukin-4 gene is not associated with HIV-1 transmission and AIDS progression." *Immunogenetics* **55**(3): 157-64.
- Moher, D., A. Liberati, et al. (2010). "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." *Int J Surg* **8**(5): 336-41.
- Mohsen, A. H., S. Murad, et al. (2005). "Prevalence of hepatitis C in an ethnically diverse HIV-1-infected cohort in south London." *HIV Med* **6**(3): 206-15.
- Moreno, V., M. L. Martin, et al. (1996). "Combined analysis of matched and unmatched case-control studies: comparison of risk estimates from different studies." *Am J Epidemiol* **143**(3): 293-300.
- Morrow, M. P., P. Pankhong, et al. (2009). "Comparative ability of IL-12 and IL-28B to regulate Treg populations and enhance adaptive cellular immunity." *Blood* **113**(23): 5868-77.
- Moses, L. E., D. Shapiro, et al. (1993). "Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations." *Stat Med* **12**(14): 1293-316.
- Mosser, D. M. and X. Zhang (2008). "Interleukin-10: new perspectives on an old cytokine." *Immunol Rev* **226**: 205-18.
- Murphy, A., T. W. S., et al. (2010). "Two-stage testing strategies for genome-wide association studies in family-based designs." *Methods Mol Biol* **620**: 485-96.
- Murphy, A., S. T. Weiss, et al. (2008). "Screening and replication using the same data set: testing strategies for family-based studies in which all probands are affected." *PLoS Genet* **4**(9): e1000197.
- Murphy, N., M. Diviney, et al. (2006). "Donor methylenetetrahydrofolate reductase genotype is associated with graft-versus-host disease in hematopoietic stem cell transplant patients treated with methotrexate." *Bone Marrow Transplant* **37**(8): 773-9.
- Nagelkerke, N. J., B. Hoebee, et al. (2004). "Combining the transmission disequilibrium test and case-control methodology using generalized logistic regression." *Eur J Hum Genet* **12**(11): 964-70.
- Naggie, S., A. Osinusi, et al. (2012). "Dysregulation of innate immunity in hepatitis C virus genotype 1 IL28B-unfavorable genotype patients: impaired viral kinetics and therapeutic response." *Hepatology* **56**(2): 444-54.
- Naicker, D. D., L. Werner, et al. (2009). "Interleukin-10 promoter polymorphisms influence HIV-1 susceptibility and primary HIV-1 pathogenesis." *J Infect Dis* **200**(3): 448-52.
- Nakayama, E. E., Y. Hoshino, et al. (2000). "Polymorphism in the interleukin-4 promoter affects acquisition of human immunodeficiency virus type 1 syncytium-inducing phenotype." *J Virol* **74**(12): 5452-9.
- Nakayama, E. E., L. Meyer, et al. (2002). "Protective effect of interleukin-4 -589T polymorphism on human immunodeficiency virus type 1 disease progression: relationship with virus load." *J Infect Dis* **185**(8): 1183-6.

- Nattermann, J., M. Vogel, et al. (2007). "Effect of the interleukin-6 C174G gene polymorphism on treatment of acute and chronic hepatitis C in human immunodeficiency virus coinfecting patients." Hepatology **46**(4): 1016-25.
- Neale, B. M., J. Lasky-Su, et al. (2008). "Genome-wide association scan of attention deficit hyperactivity disorder." Am J Med Genet B Neuropsychiatr Genet **147B**(8): 1337-44.
- Neale, B. M., S. Medland, et al. (2010). "Case-control genome-wide association study of attention-deficit/hyperactivity disorder." J Am Acad Child Adolesc Psychiatry **49**(9): 906-20.
- Neale, B. M., S. E. Medland, et al. (2010). "Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder." J Am Acad Child Adolesc Psychiatry **49**(9): 884-97.
- Nell, V. P., K. P. Machold, et al. (2005). "Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis." Ann Rheum Dis **64**(12): 1731-6.
- Nelson, P. K., B. M. Mathers, et al. (2011). "Global epidemiology of hepatitis B and hepatitis C in people who inject drugs: results of systematic reviews." Lancet **378**(9791): 571-83.
- Neukam, K., P. Barreiro, et al. (2013). "Pegylated interferon plus ribavirin is suboptimal in IL28B CC carriers without rapid response." J Infect **67**(1): 59-64.
- Neves, C., R. Jorge, et al. (2009). "[The network of methotrexate toxicity]." Acta Reumatol Port **34**(1): 11-34.
- Newcombe, P. J., C. Verzilli, et al. (2009). "Multilocus Bayesian meta-analysis of gene-disease associations." Am J Hum Genet **84**(5): 567-80.
- Nickerson, D. A., C. Whitehurst, et al. (1992). "Identification of clusters of biallelic polymorphic sequence-tagged sites (pSTSs) that generate highly informative and automatable markers for genetic linkage mapping." Genomics **12**(2): 377-87.
- Nielen, M. M., A. R. van der Horst, et al. (2005). "Antibodies to citrullinated human fibrinogen (ACF) have diagnostic and prognostic value in early arthritis." Ann Rheum Dis **64**(8): 1199-204.
- Nikolopoulos, G. K., P. G. Bagos, et al. (2011). "Developing the evidence base for cancer chemoprevention: use of meta-analysis." Curr Drug Targets **12**(13): 1989-97.
- Nishimura, K., D. Sugiyama, et al. (2007). "Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis." Ann Intern Med **146**(11): 797-808.
- Niu, T. (2004). "Algorithms for inferring haplotypes." Genetic epidemiology **27**(4): 334-347.
- Normand, S.-L. T. (1999). "Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting." Statistics in medicine **18**(3): 321-359.
- Normand, S. L. (1999). "Meta-analysis: formulating, evaluating, combining, and reporting." Stat Med **18**(3): 321-59.
- O'Brien, P. C. (1984). "Procedures for comparing samples with multiple endpoints." Biometrics **40**(4): 1079-87.
- Olkin, I. and A. Sampson (1998). "Comparison of meta-analysis versus analysis of variance of individual patient data." Biometrics **54**(1): 317-22.
- Operskalski, E. A. and A. Kovacs (2011). "HIV/HCV co-infection: pathogenesis, clinical complications, treatment, and new therapeutic technologies." Curr HIV/AIDS Rep **8**(1): 12-22.

- Ortego, C., T. B. Huedo-Medina, et al. (2011). "Adherence to highly active antiretroviral therapy (HAART): a meta-analysis." *AIDS Behav* **15**(7): 1381-96.
- Ouyang, W., S. Rutz, et al. (2011). "Regulation and functions of the IL-10 family of cytokines in inflammation and disease." *Annu Rev Immunol* **29**: 71-109.
- Palmer, L. J. and W. O. Cookson (2001). "Using single nucleotide polymorphisms as a means to understanding the pathophysiology of asthma." *Respir Res* **2**(2): 102-12.
- Pan, Z., T. A. Trikalinos, et al. (2005). "Local literature bias in genetic epidemiology: an empirical evaluation of the chinese literature." *PLoS Med* **2**(12): e334.
- Panaccione, I., F. Napoletano, et al. (2013). "Neurodevelopment in Schizophrenia: The Role of the Wnt Pathways." *Current neuropharmacology* **11**(5): 535.
- Pemberton, L. A., E. Stone, et al. (2008). "The relationship between ApoE, TNFA, IL1a, IL1b and IL12b genes and HIV-1-associated dementia." *HIV Med* **9**(8): 677-80.
- Pereyra, F., X. Jia, et al. (2010). "The major genetic determinants of HIV-1 control affect HLA class I peptide presentation." *Science* **330**(6010): 1551-7.
- Perneger, T. V. (2008). "Estimating the relative hazard by the ratio of logarithms of event-free proportions." *Contemp Clin Trials* **29**(5): 762-6.
- Petiti, D. B. (1994). *Meta-analysis Decision Analysis and Cost-Effectiveness Analysis*, Oxford University Press.
- Pfeiffer, R. M., D. Pee, et al. (2008). "On combining family and case-control studies." *Genet Epidemiol* **32**(7): 638-46.
- Pierik, M., P. Rutgeerts, et al. (2006). "Pharmacogenetics in inflammatory bowel disease." *World J Gastroenterol* **12**(23): 3657-67.
- Pihusch, M., P. Lohse, et al. (2004). "Impact of thrombophilic gene mutations and graft-versus-host disease on thromboembolic complications after allogeneic hematopoietic stem-cell transplantation." *Transplantation* **78**(6): 911-8.
- Piperno, A., R. Mariani, et al. (2000). "Haemochromatosis in patients with beta-thalassaemia trait." *Br J Haematol* **111**(3): 908-14.
- Platt, R. W., B. G. Leroux, et al. (1999). "Generalized linear mixed models for meta-analysis." *Stat Med* **18**(6): 643-54.
- Pociot, F., J. Molvig, et al. (1992). "A TaqI polymorphism in the human interleukin-1 beta (IL-1 beta) gene correlates with IL-1 beta secretion in vitro." *Eur J Clin Invest* **22**(6): 396-402.
- Pokorny, J., N. T. Luan, et al. (1976). "Changes of sensory value by interaction of alkanals with amino acids and proteins." *Nahrung* **20**(3): 267-72.
- Poli, G., A. L. Kinter, et al. (1994). "Interleukin 1 induces expression of the human immunodeficiency virus alone and in synergy with interleukin 6 in chronically infected U1 cells: inhibition of inductive effects by the interleukin 1 receptor antagonist." *Proc Natl Acad Sci U S A* **91**(1): 108-12.
- Pomp, E. R., K. J. Van Stralen, et al. (2010). "Experience with multiple control groups in a large population-based case-control study on genetic and environmental risk factors." *Eur J Epidemiol*.
- Pontillo, A., T. M. Oshiro, et al. (2010). "Polymorphisms in inflammasome' genes and susceptibility to HIV-1 infection." *J Acquir Immune Defic Syndr* **59**(2): 121-5.
- Pontillo, A., T. M. Oshiro, et al. (2012). "Polymorphisms in inflammasome' genes and susceptibility to HIV-1 infection." *J Acquir Immune Defic Syndr* **59**(2): 121-5.
- Price, A. L., N. J. Patterson, et al. (2006). "Principal components analysis corrects for stratification in genome-wide association studies." *Nat Genet* **38**(8): 904-9.

- Price, P., D. M. Calder, et al. (1999). "Periodontal attachment loss in HIV-infected patients is associated with the major histocompatibility complex 8.1 haplotype (HLA-A1,B8,DR3)." *Tissue Antigens* **54**(4): 391-9.
- Price, P., G. Morahan, et al. (2002). "Polymorphisms in cytokine genes define subpopulations of HIV-1 patients who experienced immune restoration diseases." *AIDS* **16**(15): 2043-7.
- Puhan, M. A., M. L. Van Natta, et al. (2010). "Excess mortality in patients with AIDS in the era of highly active antiretroviral therapy: temporal changes and risk factors." *Clin Infect Dis* **51**(8): 947-56.
- Pulido, F., A. Hill, et al. (2012). "Impact of hepatitis C co-infection on response to antiretroviral treatment." *AIDS Rev* **14**(2): 124-31.
- Pulver, A. E., V. K. Lasseter, et al. (1995). "Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes." *American journal of medical genetics* **60**(3): 252-260.
- Purcell, S., B. Neale, et al. (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." *Am J Hum Genet* **81**(3): 559-75.
- Quinn, M. A., A. K. Gough, et al. (2006). "Anti-CCP antibodies measured at disease onset help identify seronegative rheumatoid arthritis and predict radiological and functional outcome." *Rheumatology (Oxford)* **45**(4): 478-80.
- Rabe-Hesketh, S., A. Skrondal, et al. (2002). "Reliable estimation of generalized linear mixed models using adaptive quadrature." *The Stata Journal* **2**: 1-21.
- Rabe-Hesketh, S., A. Skrondal, et al. (2005). "Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects." *Journal of Econometrics* **128**(2): 301-323.
- Rachmilewitz, E. A. and P. J. Giardina (2011). "How I treat thalassemia." *Blood* **118**(13): 3479-88.
- Rallon, N. I., C. Restrepo, et al. (2011). "Interleukin-28B gene polymorphisms do not influence the susceptibility to HIV-infection or CD4 cell decline." *AIDS* **25**(2): 269-71.
- Ramaseri Sunder, S., S. R. Hanumanth, et al. (2012). "IL-10 high producing genotype predisposes HIV infected individuals to TB infection." *Hum Immunol* **73**(6): 605-11.
- Ramezani, A., E. Kalantar, et al. (2015). "Lack of Association between Interleukin-10 Gene Promoter Polymorphisms with HIV Susceptibility and Progression to AIDS." *Iran J Pathol* **10**(2): 141-8.
- Ranganathan, P., R. Culverhouse, et al. (2008). "Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis." *J Rheumatol* **35**(4): 572-9.
- Rantapaa-Dahlqvist, S., B. A. de Jong, et al. (2003). "Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis." *Arthritis Rheum* **48**(10): 2741-9.
- Raza, K., M. Breese, et al. (2005). "Predictive value of antibodies to cyclic citrullinated peptide in patients with very early inflammatory arthritis." *J Rheumatol* **32**(2): 231-8.
- Reif, A., M. Melchers, et al. (2007). *FZD3 is not a risk gene for schizophrenia: a case-control study in a Caucasian sample*, Springer.
- Reitsma, J. B., A. S. Glas, et al. (2005). "Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews." *J Clin Epidemiol* **58**(10): 982-90.

- Riley, R. D. (2009). "Multivariate meta-analysis: the effect of ignoring within-study correlation." Journal of the Royal Statistical Society: Series A (Statistics in Society) **172**(4): 789-811.
- Riley, R. D., K. R. Abrams, et al. (2007). "An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes." Stat Med **26**(1): 78-97.
- Riley, R. D., K. R. Abrams, et al. (2007). "Bivariate random-effects meta-analysis and the estimation of between-study correlation." BMC Med Res Methodol **7**: 3.
- Riley, R. D., P. C. Lambert, et al. (2008). "Meta-analysis of continuous outcomes combining individual patient data and aggregate data." Stat Med **27**(11): 1870-93.
- Riley, R. D., J. R. Thompson, et al. (2008). "An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown." Biostatistics **9**(1): 172-86.
- Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-7.
- Risch, N. J. (2000). "Searching for genetic determinants in the new millennium." Nature **405**(6788): 847-56.
- Robien, K., J. Bigler, et al. (2006). "Methylenetetrahydrofolate reductase and thymidylate synthase genotypes and risk of acute graft-versus-host disease following hematopoietic cell transplantation for chronic myelogenous leukemia." Biol Blood Marrow Transplant **12**(9): 973-80.
- Rockstroh, J. K., A. Mocroft, et al. (2005). "Influence of hepatitis C virus infection on HIV-1 disease progression and response to highly active antiretroviral therapy." J Infect Dis **192**(6): 992-1002.
- Rozen, R. (1996). "Molecular genetics of methylenetetrahydrofolate reductase deficiency." J Inherit Metab Dis **19**(5): 589-94.
- Rubinsztein, D. C. and D. F. Easton (1999). "Apolipoprotein E genetic variation and Alzheimer's disease. a meta-analysis." Dement Geriatr Cogn Disord **10**(3): 199-209.
- Ruiz-Arguelles, G. J., L. N. Coconi-Linares, et al. (2007). "Methotrexate-induced mucositis in acute leukemia patients is not associated with the MTHFR 677T allele in Mexico." Hematology **12**(5): 387-91.
- Rund, D. and E. Rachmilewitz (2005). "Beta-thalassemia." N Engl J Med **353**(11): 1135-46.
- Rutter, C. M. and C. A. Gatsonis (2001). "A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations." Stat Med **20**(19): 2865-84.
- Sacks, H. S., J. Berrier, et al. (1987). "Meta-analyses of randomized controlled trials." N Engl J Med **316**(8): 450-5.
- Sajadi, M. M., N. Shakeri, et al. (2011). "IL28B genotype does not correlate with HIV control in African Americans." Clin Transl Sci **4**(4): 282-4.
- Salanti, G., G. Amountza, et al. (2005). "Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power." Eur J Hum Genet **13**(7): 840-8.
- Salanti, G. and J. P. Higgins (2008). "Meta-analysis of genetic association studies under different inheritance models using data reported as merged genotypes." Stat Med **27**(5): 764-77.
- Salas, M., A. Hofman, et al. (1999). "Confounding by indication: an example of variation in the use of epidemiologic terminology." Am J Epidemiol **149**(11): 981-3.

- Samson, M., F. Libert, et al. (1996). "Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene." Nature **382**(6593): 722-5.
- Sandberg, J. K., K. Falconer, et al. (2010). "Chronic immune activation in the T cell compartment of HCV/HIV-1 co-infected patients." Virulence **1**(3): 177-9.
- Santin, M., M. Mestre, et al. (2008). "Impact of hepatitis C virus coinfection on immune restoration during successful antiretroviral therapy in chronic human immunodeficiency virus type 1 disease." Eur J Clin Microbiol Infect Dis **27**(1): 65-73.
- Saraux, A., J. M. Berthelot, et al. (2003). "Value of antibodies to citrulline-containing peptides for diagnosing early rheumatoid arthritis." J Rheumatol **30**(12): 2535-9.
- Saraux, A., I. Valls, et al. (1995). "How useful are tests for rheumatoid factors, antiperinuclear factors, antikeratin antibody, and the HLA DR4 antigen for the diagnosis of rheumatoid arthritis?" Rev Rhum Engl Ed **62**(1): 16-20.
- Sasieni, P. D. (1997). "From genotypes to genes: doubling the sample size." Biometrics **53**(4): 1253-61.
- Sato, Y., H. Suganami, et al. (2006). "The confidence interval of allelic odds ratios under the Hardy-Weinberg disequilibrium." J Hum Genet **51**(9): 772-80.
- Sauerland, U., H. Becker, et al. (2005). "Clinical utility of the anti-CCP assay: experiences with 700 patients." Ann N Y Acad Sci **1050**: 314-8.
- Saumoy, M., M. Lopez-Dupla, et al. (2008). "The IL-6 system in HIV-1-infection and in HAART-related fat redistribution syndromes." AIDS **22**(7): 893-6.
- Schaid, D. J. and S. S. Sommer (1994). "Comparison of statistics for candidate-gene association studies using cases and parents." Am J Hum Genet **55**(2): 402-9.
- Schellekens, G. A., H. Visser, et al. (2000). "The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide." Arthritis Rheum **43**(1): 155-63.
- Schmeling, H., D. Biber, et al. (2005). "Influence of methylenetetrahydrofolate reductase polymorphisms on efficacy and toxicity of methotrexate in patients with juvenile idiopathic arthritis." J Rheumatol **32**(9): 1832-6.
- Schmiegelow, K. (2009). "Advances in individual prediction of methotrexate toxicity: a review." Br J Haematol **146**(5): 489-503.
- Seaberg, E. C., M. D. Witt, et al. (2015). "Spontaneous Clearance of the Hepatitis C Virus Among Men Who Have Sex With Men." Clin Infect Dis **61**(9): 1381-8.
- Seidemann, K., M. Book, et al. (2006). "MTHFR 677 (C-->T) polymorphism is not relevant for prognosis or therapy-associated toxicity in pediatric NHL: results from 484 patients of multicenter trial NHL-BFM 95." Ann Hematol **85**(5): 291-300.
- Seminari, E., C. Tinelli, et al. (2010). "Hepatitis C infection on immune recovery in HIV-positive patients on successful HAART: the role of genotype 3." Curr HIV Res **8**(3): 186-93.
- Serra, C., A. Biolchini, et al. (2008). "Type III and I interferons increase HIV uptake and replication in human cells that overexpress CD4, CCR5, and CXCR4." AIDS Res Hum Retroviruses **24**(2): 173-80.
- Sheppard, P., W. Kindsvogel, et al. (2003). "IL-28, IL-29 and their class II cytokine receptor IL-28R." Nat Immunol **4**(1): 63-8.
- Shimasaki, N., T. Mori, et al. (2006). "Effects of methylenetetrahydrofolate reductase and reduced folate carrier 1 polymorphisms on high-dose methotrexate-

- induced toxicities in children with acute lymphoblastic leukemia or lymphoma." J Pediatr Hematol Oncol **28**(2): 64-8.
- Shimasaki, N., T. Mori, et al. (2008). "Influence of MTHFR and RFC1 polymorphisms on toxicities during maintenance chemotherapy for childhood acute lymphoblastic leukemia or lymphoma." J Pediatr Hematol Oncol **30**(5): 347-52.
- Shin, H. D., C. Winkler, et al. (2000). "Genetic restriction of HIV-1 pathogenesis to AIDS by promoter alleles of IL10." Proc Natl Acad Sci U S A **97**(26): 14467-72.
- Shoemaker, C. A., M. Pungliya, et al. (2001). "Computational methods for single-point and multipoint analysis of genetic variants associated with a simulated complex disorder in a general population." Genet Epidemiol **21 Suppl 1**: S738-45.
- Shriner, D. (2012). "Moving toward System Genetics through Multiple Trait Analysis in Genome-Wide Association Studies." Front Genet **3**: 1.
- Siadat, M. S., J. T. Philbrick, et al. (2004). "Repeated-measures modeling improved comparison of diagnostic tests in meta-analysis of dependent studies." J Clin Epidemiol **57**(7): 698-711.
- Skol, A. D., L. J. Scott, et al. (2006). "Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies." Nat Genet **38**(2): 209-13.
- Skrondal, A. and S. Rabe-Hesketh (2003). "Multilevel logistic regression for polytomous data and rankings." Psychometrika **68**(2): 267-287.
- Sladek, R., G. Rocheleau, et al. (2007). "A genome-wide association study identifies novel risk loci for type 2 diabetes." Nature **445**(7130): 881-5.
- Smolnikova, M., M. Freidin, et al. (2001). "Genetic polymorphism in the interleukin-4 promoter region in human immunodeficiency virus." Eur J Immunogenet **28**: 243.
- So, H. C. and P. C. Sham (2011). "Robust association tests under different genetic models, allowing for binary or quantitative traits and covariates." Behav Genet **41**(5): 768-75.
- Sobti, R. C., N. Berhane, et al. (2010). "Polymorphisms of IL-6 174 G/C, IL-10 -592 C/A and risk of HIV/AIDS among North Indian population." Mol Cell Biochem **337**(1-2): 145-52.
- Sobti, R. C., N. Berhane, et al. (2012). "Polymorphisms of IL-6 174 G/C, IL-10 -592 C/A and risk of HIV/AIDS among North Indian population." Mol Cell Biochem **337**(1-2): 145-52.
- Sobti, R. C., A. M. Salih, et al. (2010). "Insights into the role of IL-12B and IFN-gamma cytokine gene polymorphisms in HIV-1/AIDS infection." Folia Biol (Praha) **56**(3): 110-5.
- Soderlin, M. K., A. Kastbom, et al. (2004). "Antibodies against cyclic citrullinated peptide (CCP) and levels of cartilage oligomeric matrix protein (COMP) in very early arthritis: relation to diagnosis and disease activity." Scand J Rheumatol **33**(3): 185-8.
- Solomon, S. S., A. K. Srikrishnan, et al. (2008). "High prevalence of HIV, HIV/hepatitis C virus coinfection, and risk behaviors among injection drug users in Chennai, India: a cause for concern." J Acquir Immune Defic Syndr **49**(3): 327-32.
- Solovieff, N., C. Cotsapas, et al. (2013). "Pleiotropy in complex traits: challenges and strategies." Nat Rev Genet **14**(7): 483-95.

- Song, K. and R. C. Elston (2006). "A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine-mapping in case-control studies." Stat Med **25**(1): 105-26.
- Soriano, V., E. Vispo, et al. (2010). "Viral hepatitis and HIV co-infection." Antiviral Res **85**(1): 303-15.
- Spielman, R. S. and W. J. Ewens (1996). "The TDT and other family-based tests for linkage disequilibrium and association." Am J Hum Genet **59**(5): 983-9.
- Spielman, R. S. and W. J. Ewens (1998). "A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test." Am J Hum Genet **62**(2): 450-8.
- Spielman, R. S., R. E. McGinnis, et al. (1993). "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)." Am J Hum Genet **52**(3): 506-16.
- Spiritus, T., P. Verschueren, et al. (2004). "Diagnostic characteristics of a gelatin based Waaler-Rose assay (Serodia-RA) for the detection of rheumatoid factor." Ann Rheum Dis **63**(9): 1169-71.
- Stallings, R. L., A. F. Ford, et al. (1991). "Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes." Genomics **10**(3): 807-15.
- Stapleton, J. T., K. Bennett, et al. (2007). "Effect of antiretroviral therapy and hepatitis c co-infection on changes in lipid levels in HIV-Infected patients 48 weeks after initiation of therapy." HIV Clin Trials **8**(6): 429-36.
- Stenkvist, J., A. Sonnerborg, et al. (2013). "HCV RNA decline in chronic HCV genotype 2 and 3 during standard of care treatment according to IL28B polymorphism." J Viral Hepat **20**(3): 193-9.
- Sterne, J. A., D. Gavaghan, et al. (2000). "Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature." J Clin Epidemiol **53**(11): 1119-29.
- Stevens, J. (1986). Applied Multivariate Statistics for the Social Sciences.
- Stoumpos, S., S. Hamodrakas, et al. (2012). "The association between apolipoprotein E gene polymorphisms and essential hypertension: a meta-analysis of 45 studies including 13 940 cases and 16 364 controls." Journal of human hypertension **27**(4): 245-255.
- Stram, D. O. (1996). "Meta-analysis of published data using a linear mixed-effects model." Biometrics **52**(2): 536-44.
- Strasser, H. and C. Weber (1999). "On the Asymptotic Theory of Permutation Statistics." Mathematical Methods of Statistics **8**(2): 220-250.
- Stroup, D. F., J. A. Berlin, et al. (2000). "Meta-analysis of observational studies in epidemiology: a proposal for reporting." Jama **283**(15): 2008-2012.
- Stroup, D. F., J. A. Berlin, et al. (2000). "Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group." Jama **283**(15): 2008-12.
- Sugimoto, K., M. Murata, et al. (2008). "Decreased risk of acute graft-versus-host disease following allogeneic hematopoietic stem cell transplantation in patients with the 5,10-methylenetetrahydrofolate reductase 677TT genotype." Int J Hematol **87**(5): 451-8.
- Sulkowski, M. S., R. D. Moore, et al. (2002). "Hepatitis C and progression of HIV disease." JAMA **288**(2): 199-206.
- Sullivan, P. F., K. S. Kendler, et al. (2003). "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies." Archives of general psychiatry **60**(12): 1187-1192.

- Sun, F., W. D. Flanders, et al. (1999). "Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT." Am J Epidemiol **150**(1): 97-104.
- Suzuki, K., T. Sawada, et al. (2003). "High diagnostic performance of ELISA detection of antibodies to citrullinated antigens in rheumatoid arthritis." Scand J Rheumatol **32**(4): 197-204.
- Suzuki, S., T. Moro-oka, et al. (2004). "The conditional relative odds ratio provided less biased results for comparing diagnostic test accuracy in meta-analyses." J Clin Epidemiol **57**(5): 461-9.
- Swedler, W., J. Wallman, et al. (1997). "Routine measurement of IgM, IgG, and IgA rheumatoid factors: high sensitivity, specificity, and predictive value for rheumatoid arthritis." J Rheumatol **24**(6): 1037-44.
- Taioli, E., M. A. Garza, et al. (2009). "Meta- and pooled analyses of the methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism and colorectal cancer: a HuGE-GSEC review." Am J Epidemiol **170**(10): 1207-21.
- Taniguchi, A., W. Urano, et al. (2007). "Validation of the associations between single nucleotide polymorphisms or haplotypes and responses to disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a proposal for prospective pharmacogenomic study in clinical practice." Pharmacogenet Genomics **17**(6): 383-90.
- Teo, Y. Y. (2008). "Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure." Curr Opin Lipidol **19**(2): 133-43.
- Terwilliger, J. D. and J. Ott (1992). "A haplotype-based 'haplotype relative risk' approach to detecting allelic associations." Hum Hered **42**(6): 337-46.
- Thakkinian, A., P. McElduff, et al. (2005). "A method for meta-analysis of molecular association studies." Stat Med **24**(9): 1291-306.
- Thammanichanond, D., M. Kunakorn, et al. (2005). "Raising rheumatoid factor cutoff helps distinguish rheumatoid arthritis." Asian Pac J Allergy Immunol **23**(2-3): 165-8.
- Thomas, D. L., J. Astemborski, et al. (2000). "The natural history of hepatitis C virus infection: host, viral, and environmental factors." JAMA **284**(4): 450-6.
- Thomas, D. L., C. L. Thio, et al. (2009). "Genetic variation in IL28B and spontaneous clearance of hepatitis C virus." Nature **461**(7265): 798-801.
- Thompson, B. (1990). "Multinor: A fortran program that assists in evaluating multivariate normality." Educ Psychol Measurement **50**: 845-8.
- Thompson, J. R., C. Minelli, et al. (2005). "Meta-analysis of genetic studies using Mendelian randomization--a multivariate approach." Stat Med **24**(14): 2241-54.
- Thompson, S. G. and S. J. Sharp (1999). "Explaining heterogeneity in meta-analysis: a comparison of methods." Stat Med **18**(20): 2693-708.
- Thompson, S. G. and S. J. Sharp (1999). "Explaining heterogeneity in meta - analysis: a comparison of methods." Statistics in medicine **18**(20): 2693-2708.
- Thompson, S. G., R. M. Turner, et al. (2001). "Multilevel models for meta-analysis, and their application to absolute risk differences." Stat Methods Med Res **10**(6): 375-92.
- Thompson, W. D., J. L. Kelsey, et al. (1982). "Cost and efficiency in the choice of matched and unmatched case-control study designs." Am J Epidemiol **116**(5): 840-51.

- Tian, H. and B. N. Cronstein (2007). "Understanding the mechanisms of action of methotrexate: implications for the treatment of rheumatoid arthritis." Bull NYU Hosp Jt Dis **65**(3): 168-73.
- Toffoli, G., A. Russo, et al. (2003). "Effect of methylenetetrahydrofolate reductase 677C-->T polymorphism on toxicity and homocysteine plasma level after chronic methotrexate treatment of ovarian cancer patients." Int J Cancer **103**(3): 294-9.
- Trabetti, E. (2008). "Homocysteine, MTHFR gene polymorphisms, and cardiovascular risk." J Appl Genet **49**(3): 267-82.
- Trevino, L. R., N. Shimasaki, et al. (2009). "Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects." J Clin Oncol **27**(35): 5972-8.
- Trikalinos, T. A., D. C. Hoaglin, et al. (2014). "Methods for the joint meta-analysis of multiple tests." Res Synth Methods **5**(4): 294-312.
- Trikalinos, T. A. and I. Olkin (2008). "A method for the meta-analysis of mutually exclusive binary outcomes." Stat Med **27**(21): 4279-300.
- Trikalinos, T. A., G. Salanti, et al. (2006). "Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations." Am J Epidemiol **163**(4): 300-9.
- Trikalinos, T. A., G. Salanti, et al. (2008). "Meta-analysis methods." Adv Genet **60**: 311-34.
- Tsuchiya, N., T. Inoue, et al. (2008). "Drug related genetic polymorphisms affecting adverse reactions to methotrexate, vinblastine, doxorubicin and cisplatin in patients with urothelial cancer." J Urol **180**(6): 2389-95.
- Turner, J., L. Bansi, et al. (2010). "The prevalence of hepatitis C virus (HCV) infection in HIV-positive individuals in the UK - trends in HCV testing and the impact of HCV on HIV treatment outcomes." J Viral Hepat **17**(8): 569-77.
- Turner, R. M., R. Z. Omar, et al. (2000). "A multilevel model framework for meta-analysis of clinical trials with binary outcomes." Stat Med **19**(24): 3417-32.
- Uccellini, L., F. C. Tseng, et al. (2012). "HCV RNA levels in a multiethnic cohort of injection drug users: human genetic, viral and demographic associations." Hepatology **56**(1): 86-94.
- Urano, W., A. Taniguchi, et al. (2002). "Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses." Pharmacogenetics **12**(3): 183-90.
- Vallbracht, I., J. Rieber, et al. (2004). "Diagnostic and clinical value of anti-cyclic citrullinated peptide antibodies compared with rheumatoid factor isotypes in rheumatoid arthritis." Ann Rheum Dis **63**(9): 1079-84.
- van der Put, N. M., F. Gabreels, et al. (1998). "A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects?" Am J Hum Genet **62**(5): 1044-51.
- van der Sluis, S., D. Posthuma, et al. (2013). "TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies." PLoS Genet **9**(1): e1003235.
- van Ede, A. E., R. F. Laan, et al. (1998). "Methotrexate in rheumatoid arthritis: an update with focus on mechanisms involved in toxicity." Semin Arthritis Rheum **27**(5): 277-92.

- van Ede, A. E., R. F. Laan, et al. (2001). "The C677T mutation in the methylenetetrahydrofolate reductase gene: a genetic risk factor for methotrexate-related elevation of liver enzymes in rheumatoid arthritis patients." *Arthritis Rheum* **44**(11): 2525-30.
- van Gaalen, F. A., H. Visser, et al. (2005). "A comparison of the diagnostic accuracy and prognostic value of the first and second anti-cyclic citrullinated peptides (CCP1 and CCP2) autoantibody tests for rheumatoid arthritis." *Ann Rheum Dis* **64**(10): 1510-2.
- van Houwelingen, H. C., L. R. Arends, et al. (2002). "Advanced methods in meta-analysis: multivariate approach and meta-regression." *Stat Med* **21**(4): 589-624.
- van Houwelingen, H. C., K. H. Zwinderman, et al. (1993). "A bivariate approach to meta-analysis." *Stat Med* **12**(24): 2273-84.
- van Kooten Niekerk, P. B., K. Schmiegelow, et al. (2008). "Influence of methylene tetrahydrofolate reductase polymorphisms and coadministration of antimetabolites on toxicity after high dose methotrexate." *Eur J Haematol* **81**(5): 391-8.
- van Leeuwen, M. A., J. Westra, et al. (1988). "Quantitation of IgM, IgA and IgG rheumatoid factors by ELISA in rheumatoid arthritis and other rheumatic disorders." *Scand J Rheumatol Suppl* **75**: 25-31.
- van Manen, D., A. B. van 't Wout, et al. (2012). "Genome-wide association studies on HIV susceptibility, pathogenesis and pharmacogenomics." *Retrovirology* **9**: 70.
- van Venrooij, W. J., E. R. Vossenaar, et al. (2004). "Anti-CCP antibodies: the new rheumatoid factor in the serology of rheumatoid arthritis." *Autoimmun Rev* **3 Suppl 1**: S17-9.
- Vasiliauskiene, L., A. Wiik, et al. (2001). "Prevalence and clinical significance of antikeratin antibodies and other serological markers in Lithuanian patients with rheumatoid arthritis." *Ann Rheum Dis* **60**(5): 459-66.
- Verzilli, C., T. Shah, et al. (2008). "Bayesian meta-analysis of genetic association studies with different sets of markers." *Am J Hum Genet* **82**(4): 859-72.
- Vimaleswaran, K. S., J. Luan, et al. (2008). "The Gly482Ser genotype at the PPARGC1A gene and elevated blood pressure: a meta-analysis involving 13,949 individuals." *J Appl Physiol* (1985) **105**(4): 1352-8.
- Visser, H., L. B. Gelinck, et al. (1996). "Diagnostic and prognostic characteristics of the enzyme linked immunosorbent rheumatoid factor assays in rheumatoid arthritis." *Ann Rheum Dis* **55**(3): 157-61.
- Vittecoq, O., B. Incauragarat, et al. (2004). "Autoantibodies recognizing citrullinated rat filaggrin in an ELISA using citrullinated and non-citrullinated recombinant proteins as antigens are highly diagnostic for rheumatoid arthritis." *Clin Exp Immunol* **135**(1): 173-80.
- Vittecoq, O., V. Salle, et al. (2001). "Autoantibodies to the 27 C-terminal amino acids of calpastatin are detected in a restricted set of connective tissue diseases and may be useful for diagnosis of rheumatoid arthritis in community cases of very early arthritis." *Rheumatology (Oxford)* **40**(10): 1126-34.
- Vucic, N., M. Frleta, et al. (2009). "[Thrombophilia, preeclampsia and other pregnancy complications]." *Acta Med Croatica* **63**(4): 297-305.
- Walter, S. D. (2002). "Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data." *Stat Med* **21**(9): 1237-56.

- Wang, C., W. Song, et al. (2004). "Cytokine and chemokine gene polymorphisms among ethnically diverse North Americans with HIV-1 infection." J Acquir Immune Defic Syndr **35**(5): 446-54.
- Warren, R. B., R. L. Smith, et al. (2009). "Outcomes of methotrexate therapy for psoriasis and relationship to genetic polymorphisms." Br J Dermatol **160**(2): 438-41.
- Weatherall, D. J. (2001). "Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias." Nat Rev Genet **2**(4): 245-55.
- Wei, J. and G. P. Hemmings (2004). "Lack of a genetic association between the frizzled-3 gene and schizophrenia in a British population." Neuroscience letters **366**(3): 336-338.
- Wei, Y. and J. P. Higgins (2013). "Estimating within-study covariances in multivariate meta-analysis with multiple outcomes." Stat Med **32**(7): 1191-205.
- Weisberg, I., P. Tran, et al. (1998). "A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity." Mol Genet Metab **64**(3): 169-72.
- Weisman, M. H., D. E. Furst, et al. (2006). "Risk genotypes in folate-dependent enzymes and their association with methotrexate-related side effects in rheumatoid arthritis." Arthritis Rheum **54**(2): 607-12.
- Weissman, D., G. Poli, et al. (1994). "Interleukin 10 blocks HIV replication in macrophages by inhibiting the autocrine loop of tumor necrosis factor alpha and interleukin 6 induction of virus." AIDS Res Hum Retroviruses **10**(10): 1199-206.
- Wessels, J. A., J. K. de Vries-Bouwstra, et al. (2006). "Efficacy and toxicity of methotrexate in early rheumatoid arthritis are associated with single-nucleotide polymorphisms in genes coding for folate pathway enzymes." Arthritis Rheum **54**(4): 1087-95.
- West, A., K. Langley, et al. (2002). "Evidence to suggest biased phenotypes in children with Attention Deficit Hyperactivity Disorder from completely ascertained trios." Mol Psychiatry **7**(9): 962-6.
- White, I. R. (2009). "Multivariate random-effects meta-analysis." The Stata Journal **9**(1): 40-56.
- White, I. R. (2009). "Multivariate random-effects meta-analysis." Stata Journal **9**: 40-56.
- White, I. R. (2010). "simsum: Analyses of simulation studies including Monte Carlo error." The Stata Journal **10**(3): 369-385.
- Wichukchinda, N., E. E. Nakayama, et al. (2006). "Protective effects of IL4-589T and RANTES-28G on HIV-1 disease progression in infected Thai females." AIDS **20**(2): 189-96.
- Willer, C. J., Y. Li, et al. (2010). "METAL: fast and efficient meta-analysis of genomewide association scans." Bioinformatics **26**(17): 2190-1.
- Williams, R. L. (2000). "A note on robust variance estimation for cluster-correlated data." Biometrics **56**(2): 645-6.
- Winkles, J. W., J. Lunec, et al. (1989). "Automated enhanced latex agglutination assay for rheumatoid factors in serum." Clin Chem **35**(2): 303-7.
- Wisaksana, R., A. K. Indrati, et al. (2010). "Response to first-line antiretroviral treatment among human immunodeficiency virus-infected patients with and without a history of injecting drug use in Indonesia." Addiction **105**(6): 1055-61.

- Wittke-Thompson, J. K., A. Pluzhnikov, et al. (2005). "Rational inferences about departures from Hardy-Weinberg equilibrium." Am J Hum Genet **76**(6): 967-86.
- Woolf, B. (1955). "On estimating the relationship between blood group and disease." Human Genet **19**: 251-253.
- WTCCC (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-78.
- Xu, H., X. Wu, et al. (2004). "Comparison of haplotype inference methods using genotypic data from unrelated individuals." Human heredity **58**(2): 63-68.
- Xu, X., L. Tian, et al. (2003). "Combining dependent tests for linkage or association across multiple phenotypic traits." Biostatistics **4**(2): 223-9.
- Yang, B., R. C. Chan, et al. (2007). "A meta-analysis of association studies between the 10-repeat allele of a VNTR polymorphism in the 3'-UTR of dopamine transporter gene and attention deficit hyperactivity disorder." Am J Med Genet B Neuropsychiatr Genet **144B**(4): 541-50.
- Yang, J., T. Si, et al. (2003). "Association study of the human FZD3 locus with schizophrenia." Biological psychiatry **54**(11): 1298-1301.
- Yang, Q. and Y. Wang (2012). "Methods for Analyzing Multivariate Phenotypes in Genetic Association Studies." J Probab Stat **2012**: 652569.
- Yang, Q., H. Wu, et al. (2010). "Analyze multivariate phenotypes in genetic association studies by combining univariate association tests." Genet Epidemiol **34**(5): 444-54.
- Young, A., N. Sumar, et al. (1991). "Agalactosyl IgG: an aid to differential diagnosis in early synovitis." Arthritis Rheum **34**(11): 1425-9.
- Yuan, H., B. Adams-Huet, et al. (2012). "A single nucleotide polymorphism in IL28B affects viral evolution of hepatitis C quasispecies after pegylated interferon and ribavirin therapy." J Med Virol **84**(12): 1913-9.
- Yusuf, S., R. Peto, et al. (1985). "Beta blockade during and after myocardial infarction: an overview of the randomized trials." Prog Cardiovasc Dis **27**(5): 335-71.
- Zang, Y., W. K. Fung, et al. (2010). "Simple algorithms to calculate asymptotic null distribution for robust tests in case-control genetic association studies in R." Journal of Statistical Software **33**(8).
- Zaykin, D. V. and D. O. Kozbur (2010). "P-value based analysis for shared controls design in genome-wide association studies." Genet Epidemiol **34**(7): 725-38.
- Zaykin, D. V. and D. O. Kozbur (2010). "P - value based analysis for shared controls design in genome - wide association studies." Genetic epidemiology **34**(7): 725-738.
- Zeggini, E. and J. P. Ioannidis (2009). "Meta-analysis in genome-wide association studies." Pharmacogenomics **10**(2): 191-201.
- Zeng, Q. Y., Y. K. Wang, et al. (2008). "Pharmacogenetic study of 5,10-methylenetetrahydrofolate reductase C677T and thymidylate synthase 3R/2R gene polymorphisms and methotrexate-related toxicity in Chinese Han patients with inflammatory arthritis." Ann Rheum Dis **67**(8): 1193-4.
- Zeremski, M., R. B. Dimova, et al. (2013). "IL28B polymorphism, pretreatment CXCL10, and HCV RNA levels predict treatment response in racially diverse HIV/HCV coinfecting and HCV mono-infected patients." J Acquir Immune Defic Syndr **63**(1): 9-16.

- Zhang, Y., X. Yu, et al. (2004). "Positive association of the human frizzled 3 (FZD3) gene haplotype with schizophrenia in Chinese Han population." American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **129**(1): 16-19.
- Zhao, J., Z.-J. Yu, et al. (2011). "A retrospective study: The influence of human immunodeficiency virus co-infection with hepatitis C virus or hepatitis B virus on the efficacy with HAART in China AIDS area." Life Science Journal **8**: 27-31.
- Zheng, G., B. Freidlin, et al. (2006). "Comparison of Robust Tests for Genetic Association Using Case-Control Studies." IMS Lecture Notes-Monograph Series **49**(Optimality: The Second Erich L. Lehmann Symposium): 253-265.
- Zheng, G., B. Freidlin, et al. (2006). "Robust genomic control for association studies." Am J Hum Genet **78**(2): 350-6.
- Zheng, G., B. Freidlin, et al. (2003). "Choice of Scores in Trend Tests for Case-Control Studies of Candidate-Gene Associations." Biometrical Journal **45**(3): 335-348.
- Zheng, G., J. Joo, et al. (2009). "Pearson's test, trend test, and MAX are all trend tests with different types of scores." Ann Hum Genet **73**(2): 133-40.
- Zheng, G. and H. K. Ng (2008). "Genetic model selection in two-phase analysis for case-control association studies." Biostatistics **9**(3): 391-9.
- Zhou, B., J. Shi, et al. (2011). "Optimal methods for meta-analysis of genome-wide association studies." Genet Epidemiol **35**(7): 581-91.
- Zhou, J., G. J. Dore, et al. (2007). "Hepatitis B and C virus coinfection in The TREAT Asia HIV Observational Database." J Gastroenterol Hepatol **22**(9): 1510-8.
- Zhu, X., T. Feng, et al. (2015). "Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension." Am J Hum Genet **96**(1): 21-36.
- Ziegler, A., I. R. Konig, et al. (2008). "Biostatistical aspects of genome-wide association studies." Biom J **50**(1): 8-28.
- Zondervan, K. T. and L. R. Cardon (2004). "The complex interplay among factors that influence allelic association." Nature Reviews Genetics **5**(2): 89-100.

Δημοσιεύσεις

Statistical Applications in Genetics and Molecular Biology

Volume 10, Issue 1

2011

Article 19

Meta-Analysis of Family-Based and Case- Control Genetic Association Studies that Use the Same Cases

Pantelis G. Bagos, *University of Central Greece*

Niki L. Dimou, *University of Central Greece*

Theodore D. Liakopoulos, *University of Central Greece*

Georgios K. Nikolopoulos, *Hellenic Centre for Diseases
Control and Prevention, and University of Central Greece*

Recommended Citation:

Bagos, Pantelis G.; Dimou, Niki L.; Liakopoulos, Theodore D.; and Nikolopoulos, Georgios K. (2011) "Meta-Analysis of Family-Based and Case-Control Genetic Association Studies that Use the Same Cases," *Statistical Applications in Genetics and Molecular Biology*: Vol. 10: Iss. 1, Article 19.

DOI: 10.2202/1544-6115.1640

Available at: <http://www.bepress.com/sagmb/vol10/iss1/art19>

©2011 Berkeley Electronic Press. All rights reserved.

Meta-Analysis of Family-Based and Case-Control Genetic Association Studies that Use the Same Cases

Pantelis G. Bagos, Niki L. Dimou, Theodore D. Liakopoulos, and Georgios K. Nikolopoulos

Abstract

In many cases in genetic epidemiology, the investigators in an effort to control for different sources of confounding and simultaneously to increase the power perform a family-based and a population-based case-control study within the same population, using the same or largely overlapping, set of cases. Various methods have been proposed for performing a combined analysis, but they all require access to individual data that are difficult to gather in a meta-analysis. Here, we propose a simple and efficient summary-based method for performing the meta-analysis. The key point, contrary to the methods presented earlier that need individual data, is the calculation of the covariance between the study estimates (log-Odds Ratios), using only data derived from the literature in the form of a 2x2 contingency table. Afterwards, the studies can easily be combined either in a two-step procedure using traditional methods for univariate meta-analysis or in a single-step approach using hierarchical models. In any case, the meta-analysis can be performed using standard software and because of the increased sample size the statistical power of the meta-analysis is increased whereas the procedure allows performing several diagnostics (publication bias, cumulative meta-analysis, sensitivity analysis). The method is evaluated on a dataset of 356 Single Nucleotide polymorphisms (SNPs) which were evaluated for their potential association with Respiratory Syncytial Virus Bronchiolitis (RSV) and subsequently is applied in a meta-analysis concerning the association of the 10-Repeat Allele of a VNTR Polymorphism in the 3'-UTR of Dopamine Transporter Gene with Attention Deficit Hyperactivity Disorder (ADHD), as well as in a genome-wide association study for Multiple Sclerosis. Implementation of the method is straightforward and in the Appendix, a Stata program is given for implementing the methods presented here.

KEYWORDS: meta-analysis, GWAS, case-control, family-based, TDT, HHRR, random effects, covariance

Author Notes: The authors would like to thank the associate editor and two anonymous reviewers whose comments and constructive criticism helped in improving the quality of the manuscript.

1. Introduction

The rapidly evolving field of genetic epidemiology, which brings together traditional genetics and epidemiology, focuses on the genetic component of diseases' aetiology and on the joint effects of genetic and environmental factors in large populations (Burton et al, 2005). Within the context of genetic association studies, there are two discrete lines of research, the family-based approach, most usually adopted by geneticists, and the population-based studies in which epidemiologists are mostly involved (Davey Smith et al, 2005; Hopper et al, 2005). Family-based genetic association studies, in which cases' genotypes are compared against their family members (parents, siblings, cousins etc) are not vulnerable to the confounding effect of population stratification, but they have greater genotyping and sampling requirements and the adequate data are not easily gathered (Cardon & Palmer, 2003). On the other hand, the typical population-based case-control design, which is familiar to epidemiologists, offers the advantage of easily collecting larger samples of cases and unrelated controls, improving thus considerably the power to discern a statistically significant effect.

The continuously increasing number of published genetic association studies (Becker et al, 2004; Hirschhorn et al, 2002), has made imperative the need for synthesizing the available information for a particular gene-disease association and providing a quantitative overall estimate in a procedure known as meta-analysis (Glass, 1976; Petiti, 1994). In medical literature, meta-analysis was initially applied in the field of randomized clinical trials (Chalmers et al, 1987; Sacks et al, 1987), but nowadays is considered a valuable tool for the statistical combination of observational research (Stroup et al, 2000) and for gene-disease association studies in particular, for which specialized methodology has been developed (Bagos, 2008; Bagos & Nikolopoulos, 2007; Minelli et al, 2005; Minelli et al, 2004; Salanti & Higgins, 2008; Salanti et al, 2007; Salanti et al, 2006; Thakkestian et al, 2005; Trikalinos et al, 2008). Besides the problems encountered in the meta-analysis of observational studies and clinical trials, special care is needed when dealing with meta-analyses of genetic factors speculated to be involved in the aetiopathogenesis of a particular disease (Attia et al, 2003; Ioannidis, 2005a; Ioannidis, 2005b; Ioannidis & Trikalinos, 2005; Salanti et al, 2005b).

The above-mentioned approaches (family-based and population-based) have attracted considerable interest in the past and various efforts have been attempted to derive a common measure that will allow their direct comparison (Mitchell, 2000). Moreover, the synthesis of estimates of family-based and population-based studies has been recently illustrated using summary data methods in a meta-analysis setting (Kazeem & Farrall, 2005). Empirical evaluations have also shown that these two study designs are comparable

(Evangelou et al, 2006) and thus, integrating the results in meta-analyses using summary data derived from the literature would increase the power of detecting weak genotype effects.

In many cases, the investigators in an effort to control for different sources of confounding and simultaneously to increase the power, perform a family-based and a population-based case-control study within the same population using the same, or largely overlapping, set of cases (Ackerman et al, 2005; Pomp et al, 2010). In such a situation, the different estimates derived from the same study are stochastically dependent (correlated) and thus, pooling them in order to derive a single estimate requires specialized methodology. During the last years, various methods have been proposed in the literature for addressing the particular problem allowing a combined analysis (Chen & Lin, 2008; Epstein et al, 2005; Gebregziabher et al, 2010; Glaser & Holmans, 2009; Guo et al, 2009; Hsu et al, 2009; Joo et al, 2007; Nagelkerke et al, 2004; Pfeiffer et al, 2008). However, these methods depend on the availability of individual data and thus, are not suitable for meta-analysis of published studies from which the researcher has usually access only to summary data (genotype or allele counts). Recently, the different approaches have been presented and compared with respect to their advantages and drawbacks (Infante-Rivard et al, 2009). We should emphasize here that the distinction of population-based and family-based studies, resembles closely the well-known in epidemiology distinction of unmatched and matched case-controls studies whose pros and cons have been extensively discussed in the past (Feinstein, 1987; Hamajima et al, 1994; Thompson et al, 1982). Although methods for combining such studies have also been proposed for years, they utilize individual data and require that the study populations are independent (Duffy et al, 1989; Moreno et al, 1996).

Table 1. The 2x2 contingency table corresponding to a population-based case-control study in which allele *B* is considered the susceptibility allele. The total number of *B* and *A* alleles are compared between cases and controls. For brevity we denote $n_{01}=2BB_0+AB_0$, $n_{00}=2AA_0+AB_0$, $n_{10}=2AA_1+AB_1$ and $n_{11}=2BB_1+AB_1$. The total number of cases' alleles is n_1 and controls' n_0 (i.e. the total number of cases is $n_1/2$ and that of controls $n_0/2$).

		<i>Allele</i>		
		<i>B</i>	<i>A</i>	Total
<i>Status</i>	<i>Cases</i>	n_{11}	n_{10}	n_1
	<i>Controls</i>	n_{01}	n_{00}	n_0
Total				$n_0 + n_1$

Lately, a general method that uses the sandwich variance estimator or the bootstrap has been presented (le Cessie et al, 2008), which is applicable to both situations (genetic association studies and traditional epidemiological analyses).

The method was successfully applied also for combining the results obtained from family-based and population-based case-control analyses (Janssen et al, 2007). Nevertheless, this method requires access to individual data that are not easily obtained in a meta-analysis.

In this work, we propose a simple summary-based method for pooling genetic association studies, which perform both family-based and population-based analyses and use overlapping sets of cases. The key point in this method is the calculation of the covariance between the family-based and the population-based study estimate that are obtained from the same cases, using though only the data derived from the literature in the form of a 2x2 contingency table. Afterwards, the studies can be combined either in a two-step procedure using traditional methods for univariate meta-analysis or in a single-step approach using hierarchical models. In any case, the meta-analysis can be performed using standard software and because of the increased sample size the statistical power of the meta-analysis becomes larger.

The method is applied in three real datasets. Firstly, we analysed the dataset of 356 Single Nucleotide polymorphisms (SNPs) presented by Janssen et al (2007) in order to compare the accuracy in estimating the covariance and subsequently, we analysed an already published meta-analysis concerning the association of the 10-Repeat Allele of a VNTR Polymorphism in the 3'-UTR of Dopamine Transporter Gene and Attention Deficit Hyperactivity Disorder (ADHD) (Yang et al, 2007), as well as the data from a recent genome-wide association study that was conducted in order to identify susceptibility loci for Multiple Sclerosis (Hafler et al, 2007).

2. Methods

2.1 Population-based case-control studies

Consider a biallelic locus A,B with B denoting the susceptibility allele. In a traditional population-based case-control association study, cases ($j=1$) and unrelated controls ($j=0$) are sampled independently and their genotypes (AA_j, AB_j, BB_j) are compared in a 3x2 contingency table (Sasieni, 1997). Collapsing the table in a 2x2 format, results in testing different genetic models of inheritance (dominant, recessive, co-dominant) (Sasieni, 1997); however, for reasons which will be obvious later, we will restrict our attention to the co-dominant model of inheritance (additive). In such a case, the total number of B ($2BB_j+AB_j$) and A ($2AA_j+AB_j$) alleles are compared between cases and controls. The data are therefore in the form of Table 1, in which for brevity we denote $n_{01}=2BB_0+AB_0$, $n_{00}=2AA_0+AB_0$, $n_{10}=2AA_1+AB_1$ and $n_{11}=2BB_1+AB_1$. Assuming a meta-analysis of

k case-control studies, the logarithm of the Odds Ratio (OR) for the i^{th} study ($i=1,2,\dots,k$), will be denoted by y_i^{CC} and will be given by:

$$y_i^{CC} = \log \left(\frac{n_{11}n_{00}}{n_{01}n_{10}} \right) \quad (1)$$

The logOR is asymptotically normally distributed (Woolf, 1955) with variance given by:

$$\text{var}(y_i^{CC}) = \frac{1}{n_{11}} + \frac{1}{n_{00}} + \frac{1}{n_{01}} + \frac{1}{n_{10}} \quad (2)$$

The variance is easily computed by treating the counts in the 2x2 table as independent Poisson variables (U), with $E[U]=\text{var}[U]=u$ and the logORs as contrasts among the log counts. Then, using the delta-method (Agresti, 2002) we will have:

$$\text{var}[\log(U)] \approx \text{var}(U) \left(\frac{\partial \log(E[U])}{\partial (E[U])} \right)^2 = u \left(\frac{1}{u} \frac{\partial U}{\partial U} \right)^2 = \frac{1}{u} \quad (3)$$

As we have already noticed, the case-control design, albeit being powerful and easily implemented, it is prone to a special form of confounding occurring in genetic association studies, known as population stratification. The effect of population admixture emerges when there is systematic difference in allele frequencies between subgroups within a population, possibly due to different ancestry (Cardon & Palmer, 2003). Thus, in that case the random recruitment of cases and controls may result in sampling from different sub-populations, leading to spurious gene-disease associations. Although there are methods for detecting and adjusting for this form of bias, they typically require genotyping additional unlinked markers, increasing thus the overall cost of the study (Devlin & Roeder, 1999; Epstein et al, 2007; Freedman et al, 2004; Pritchard & Rosenberg, 1999). An alternative is to use the family members of cases as their controls, resulting by this way in family-based studies. These designs are largely immune to population stratification, but generally lack sufficient statistical power due to difficulties in recruiting the patients' family members (more usually their unaffected parents).

2.2 Family-based case-control studies

Even though there are several family-based designs, the most frequently encountered approach in association studies is the *family-trio*, which consists of the affected offspring and its unaffected parents as controls. Within this context, the most commonly endorsed methods of analysis are the *transmission disequilibrium test* (Spielman & Ewens, 1996; Spielman et al, 1993) and the *haplotype-based haplotype relative risk* (Terwilliger & Ott, 1992).

2.2.1. The transmission disequilibrium test

The transmission disequilibrium test (TDT) states that under the null hypothesis of no association, the two parental genotypes are independent. In this case, each of the n_1 cases' parents can be classified according to which allele is transmitted and which is not. The data are presented in the form of Table 2 from which is obvious that only discordant pairs contribute to the analysis. Thus, the TDT is another formulation of the well-known McNemar's chi-square statistic for matched observations. The score (A) derived from Table 2 is given by:

$$A = (b - c)/2 \quad (4)$$

with a corresponding variance (V) given by:

$$V = (b + c)/4 \quad (5)$$

Thus, the statistical significance can be tested using $z = A/\sqrt{V} \sim N(0,1)$.

Table 2. Presentation of the data in a family-based study using the Transmission Disequilibrium Test (TDT). The transmitted alleles are contrasted against the non-transmitted ones and the OR is given by the ratio of the discordant pairs (b/c). For comparison with Table 1 we denote $a+b=w=n_{11}$ and $c+d=x=n_{10}$.

<i>Transmitted Allele</i>	<i>Non-transmitted allele</i>			
		<i>B</i>	<i>A</i>	Total
	<i>B</i>	a	b	w
	<i>A</i>	c	d	x
	Total	y	z	n_1

For the purposes of meta-analysis however, it is more convenient to use the appropriate OR derived from Table 2, the logarithm of which is given by:

$$y_i^{TDT} = \log\left(\frac{b}{c}\right) \quad (6)$$

Similarly to Eq. (2) the asymptotic variance will be given by:

$$\text{var}(y_i^{TDT}) = \frac{1}{b} + \frac{1}{c} \quad (7)$$

2.2.2. The haplotype-based haplotype relative risk

The haplotype-based haplotype relative risk (HHRR), in an effort to increase power (i.e. to decrease the variance), uses the unmatched version of Table 2, since, under the null hypothesis, the two alleles of each parent are independent. The transition to the unmatched analysis is given in Table 3. Another way of looking at the HHRR is to consider that the control for each case is a "pseudo-

control” having alleles the non-transmitted parental alleles (one from each parent).

The haplotype-based haplotype relative risk test was originally presented as a chi-squared statistic but similarly to the TDT we use here the equivalent OR derived from Table 3:

$$y_i^{HHRR} = \log \left(\frac{wz}{xy} \right) \quad (8)$$

Similarly, the asymptotic variance will be given by:

$$\text{var} \left(y_i^{HHRR} \right) = \frac{1}{w} + \frac{1}{z} + \frac{1}{x} + \frac{1}{y} \quad (9)$$

Table 3. Presentation of the data of a family-based study under the Haplotype-based Haplotype Relative Risk (HHRR). The transmitted alleles are contrasted against the non-transmitted alleles of parents that form a “pseudo-control” population. The OR is given by wz/xy . To make the connection with the data in Table 1 we have to notice that the first rows of the tables are identical ($n_{11}=w$ and $n_{10}=x$)

	Allele		Total
	B	A	
<i>Transmitted</i>	w	x	n_1
<i>Non-transmitted</i>	y	z	n_1
<i>Total</i>	w+y	x+z	$2n_1$

From a historical point of view, it is worth-noting that Falk and Rubinstein were the first to propose the use of untransmitted alleles to form a single pseudo-control genotype (Falk & Rubinstein, 1987). Later, Terwilliger and Ott extended this idea and they discussed, for the first time, the use of McNemar’s test, although they concluded that it was less powerful than the unmatched analysis that corresponds to the HHRR test (Terwilliger & Ott, 1992). Few years later, the McNemar’s statistic was reformulated and presented as the TDT test that is now widely used (Spielman & Ewens, 1996; Spielman et al, 1993).

2.2.3. Extensions of the transmission disequilibrium test

During the years, the TDT has become the method of choice for the analysis of family-based data and several extensions have been proposed in order to handle some special cases. These extensions are more suitable for adult-onset diseases when parental information may be missing and include situations where only one parent (Sun et al, 1999), or only siblings are available (Curtis, 1997; Spielman & Ewens, 1998), situations where discordant sibs are used (Deng et al, 2002;

Horvath & Laird, 1998) and finally case-spouse and case-offspring designs (Lee, 2003).

Although these tests are not usually derived directly from a corresponding 2x2 contingency table (such as Table 2), they all use some kind of score test such as the one given in Eq. (4) and Eq. (5). Thus, we can calculate the test statistics as originally suggested in each case and assume that such a 2x2 table exists. Then, we can approximate the respective (hypothetical) values of b' and c' that would have given rise to such a table solving the system of Eqs. (4) and (5). Thus we could obtain:

$$b' = 2V + A \quad (10)$$

and

$$c' = 2V - A \quad (11)$$

Therefore, we can use these values to calculate a logOR and its corresponding variance using Eq. (6) and Eq. (7) and subsequently proceed as if the standard TDT statistic was used in the first place. One should keep in mind however, that Eq. (10) and Eq. (11) are approximate and they are not applicable when the respective test is performed by adjusting for covariates, especially when these are correlated with both disease status and allele.

2.3 Methods for meta-analysis

If the eligible studies in a meta-analysis report exclusively only one of the ORs presented earlier, it is straightforward to perform a standard fixed or random effects meta-analysis with inverse variance weights (Normand, 1999) as suggested by (Kazeem & Farrall, 2005). It should be noted that among the many alternative genetic contrasts available in population-based case-control studies, the additive model derived from the allelic association (Table 1) is the one that is comparable with the various family-based designs. In such a meta-analysis which involves a summary estimate y_i ($i=1,2,3,\dots,k$) from k studies, the traditional random-effects model can be written (van Houwelingen et al, 2002) in the form of a random-intercept linear model:

$$\begin{aligned} y_i &= \beta + \beta_i + \varepsilon_i \\ \beta_i &\sim N(0, \tau^2), \varepsilon_i \sim N(0, \Sigma_i) \end{aligned} \quad (12)$$

In this model, β is the overall mean effect that needs to be estimated, β_i is the random effect (random intercept) associated with study i , τ^2 is the between studies variance and ε_i is the residual error terms which are assumed fixed and known quantities with:

$$\Sigma_i = \begin{bmatrix} \text{var}(y_1) & 0 & \dots & 0 \\ 0 & \text{var}(y_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \text{var}(y_k) \end{bmatrix} \quad (13)$$

However, if we are combining the results of a family and a population-based case-control study that use the same (or largely overlapping) set of cases (Pomp et al, 2010), we will have $j=1,2$ correlated estimates from each of the i studies. One easily implemented approach, is to perform a two-step analysis in which we can combine the results in order to obtain a pooled, weighted, estimate for study i (le Cessie et al, 2008):

$$y_i^{\text{pooled}} = \lambda_i y_i^{\text{CC}} + (1 - \lambda_i) y_i^{\text{TDT}} \quad (14)$$

The optimal weight λ is the one that minimizes the pooled variance:

$$\begin{aligned} \text{var}(y_i^{\text{pooled}}) &= \lambda_i^2 \text{var}(y_i^{\text{CC}}) + (1 - \lambda_i)^2 \text{var}(y_i^{\text{TDT}}) \\ &\quad + 2\lambda_i(1 - \lambda_i) \text{cov}(y_i^{\text{CC}}, y_i^{\text{TDT}}) \end{aligned} \quad (15)$$

Minimizing now Eq. (15) we obtain:

$$\lambda_i = \frac{\text{var}(y_i^{\text{TDT}}) - \text{cov}(y_i^{\text{CC}}, y_i^{\text{TDT}})}{\text{var}(y_i^{\text{CC}}) + \text{var}(y_i^{\text{TDT}}) - 2\text{cov}(y_i^{\text{CC}}, y_i^{\text{TDT}})} \quad (16)$$

Following this procedure, from each study i in the meta-analysis we will be given a single estimate for the allelic OR and we can easily perform a standard meta-analysis with inverse variance weights using the model of Eq. (12) (Normand, 1999). Usually, random-effects models are preferred and among the various estimators, the non-iterative method of moments proposed by DerSimonian and Laird (DerSimonian & Laird, 1986) is the most widely used, although other iterative methods like maximum likelihood (ML) or restricted maximum likelihood (REML) are also available (Thompson & Sharp, 1999).

Alternatively, we could directly include into the model's variance-covariance matrix the covariance terms for the estimates of a study derived from the same set of cases:

$$\begin{aligned} y_{ij} &= \beta + \beta_i + \varepsilon_{ij} \\ \beta_i &\sim N(0, \tau^2), \varepsilon_{ij} \sim N(0, \Sigma_{ij}) \end{aligned} \quad (17)$$

Thus, the variance-covariance matrix would be:

$$\Sigma_{ij} = \begin{bmatrix} \text{var}(y_1^{CC}) & \text{cov}(y_1^{CC}, y_1^{TDT}) & \dots & 0 & 0 \\ \text{cov}(y_1^{CC}, y_1^{TDT}) & \text{var}(y_1^{TDT}) & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \text{var}(y_k^{CC}) & \text{cov}(y_k^{CC}, y_k^{TDT}) \\ 0 & 0 & \dots & \text{cov}(y_k^{CC}, y_k^{TDT}) & \text{var}(y_k^{TDT}) \end{bmatrix} \quad (18)$$

Instead of modelling directly the correlations, an alternative would be to consider a sandwich-type estimator of the variance (Williams, 2000) in order to account for the correlated estimates. In this case, the procedure is more closely related (in a meta-analytic sense) to the approach followed by le Cessie et al, (2008) for combining the results within a single study.

Finally, we could also formulate the problem in terms of hierarchical models fitted within the general linear mixed model framework (Breslow et al, 1998; Platt et al, 1999; Stram, 1996; Thompson et al, 2001). Thus, we will now have sub-groups (j) nested within studies (i) with their respective random terms:

$$y_{ij} = \beta + \beta_i + \beta_{ij} + \varepsilon_{ij} \quad (19)$$

$$\beta_i \sim N(0, \tau_s^2), \beta_{ij} \sim N(0, \tau_g^2), \varepsilon_{ij} \sim N(0, \Sigma_{ij})$$

Here, τ_s^2 is the between-studies variance and τ_g^2 is the variance between-groups within a study. Similarly to the model of Eq. (17) the variance-covariance matrix Σ_{ij} could be of the form of Eq. (13), in which case the sandwich-type estimator of variance (Williams, 2000) should be used to account for the correlated estimates; alternatively the covariance can be used directly through Eq. (18).

2.4 Calculation of the covariance

A necessary condition (if we don't use the sandwich-type variance estimate) is to calculate the within-studies covariance of Eq. (18). In this section we provide estimates of the covariance that can be computed using only the summary counts available in published reports. The problem reduces to a special case of a general framework that is to be discussed elsewhere. Briefly, we need to calculate analytically the covariance between the two logORs, for instance from case-control and HHRR studies:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HRR}) &= \text{cov}\left[\log\left(\frac{n_{11}n_{00}}{n_{01}n_{10}}\right), \log\left(\frac{wz}{xy}\right)\right] \\ &= \text{cov}\left[\log(n_{11}n_{00}) - \log(n_{01}n_{10}), \log(wz) - \log(xy)\right]\end{aligned}\quad (20)$$

which after trivial calculations making use of the standard properties of the covariance function (Appendix I), reduces to:

$$\text{cov}(y_i^{CC}, y_i^{HRR}) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \quad (21)$$

In case a study reports results from a case-control and a TDT analysis, the covariance will also be given by (Appendix I):

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \quad (22)$$

We should emphasize that Eq. (21) and Eq. (22) apply also to situations where some of the cases (denoted by n_{11-} and n_{10-}) did not participate for some reason in the family-based analysis (i.e. when their parents are not available or the genotyping has failed). In such case, $n_{11} = n_{11*} + n_{11-}$ and $n_{10} = n_{10*} + n_{10-}$ and thus, Eq. (21) and Eq. (22) remain unchanged. Similar arguments can also be made when a number of cases participate only in the family-based analysis and not in the case-control analysis. In the more general case when we have sets of cases that overlap only partially, the equations need to be modified. For instance, if we assume that the two designs share $m/2$ cases that contribute m alleles (with $m < n_1$) the respective counts in Eq. (21) need to be decomposed into two disjoint subsets denoted by subscripts s and d (where s stands for shared and d for distinct). Thus, Eq. (21) will become:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HRR}) &= \text{cov}\left[\log(n_{11s} + n_{11d}), \log(w_s + w_d)\right] \\ &\quad + \text{cov}\left[\log(n_{10s} + n_{10d}), \log(x_s + x_d)\right]\end{aligned}\quad (23)$$

In Appendix I we show that this covariance finally reduces to:

$$\text{cov}(y_i^{CC}, y_i^{HRR}) = \frac{n_{11s}}{n_{11}w} + \frac{n_{10s}}{n_{10}x} \quad (24)$$

Similarly (Appendix I), the extended version of Eq. (22) becomes:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \frac{b_s}{b} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) \quad (25)$$

A practical problem with Eq. (24) can arise however, since it is not expected that the studies will report the specific counts (i.e. n_{11s} , n_{11d} and so on). In such situations, one can use the original estimate of Eq. (21) which assumes total overlap producing this way a larger covariance that will lead to anti-

conservative results. Another alternative will be to assume that the overlap will be constant (i.e. $n_{11s}/n_{11} = n_{10s}/n_{10}$). This is a reasonable assumption that leads to an approximate covariance given by:

$$\text{cov}(y_i^{CC}, y_i^{HRR}) \approx \frac{m}{n_1} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) = \frac{w_s + x_s}{w + x} \left(\frac{1}{w} + \frac{1}{x} \right) \quad (26)$$

Similar arguments hold for the covariance of Eq. (25), which becomes:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) \approx \frac{m}{n_1} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) \quad (27)$$

Obviously, the two estimates can be used consecutively in a sensitivity analysis, with the estimates of Eq. (21) and Eq. (22) being the upper bounds of the covariance. Interestingly, similar equations have been presented in a different framework, for instance, for comparing continuous variables against a common control group (Dunnett, 1955) and recently, in genome-wide association studies for comparing differences in allele frequencies (Zaykin & Kozbur, 2010).

By substituting the estimates for the covariance obtained by Eq. (21) and Eq. (22), into the expression for the variance of the pooled estimate in Eq. (15), we get an expression entirely in terms of the observed counts. Then, it is easily verified the generally known result (Keller & Olkin, 2004), i.e. that pooling increases the power by decreasing the variance:

$$\text{var}(y_i^{\text{pooled}}) \leq \min \{ \text{var}(y_i^{CC}), \text{var}(y_i^{TDT}) \} \quad (28)$$

In the particular case the closed-form expression for the covariance implies that the inequality is strict and given that all tests considered here are asymptotically normal, we can easily conclude that the *Asymptotic Relative Efficiency* (*ARE*) of the pooled estimate compared to the estimate with the smaller variance (usually the case-control estimator) is always greater than one:

$$ARE = \frac{\min \{ \text{var}(y_i^{CC}), \text{var}(y_i^{TDT}) \}}{\text{var}(y_i^{\text{pooled}})} > 1 \quad (29)$$

Assuming an equal number of cases and controls and a common odds ratio for the case-control and TDT analysis, it can be shown that *ARE* is a function of only three parameters: the common odds ratio (*OR*), the risk allele frequency in controls (*f*) and the total sample size (*n*). In Figure 1, we plot the *ARE* for a range of plausible values of *f* and *OR* (the influence of *n* is negligible for $n > 50$, as one would expect). The results illustrate that the pooled estimate will be 12-47% more efficient compared to the usual case-control estimator. Similar results can also be obtained by combining the results of HRR and case-control studies.

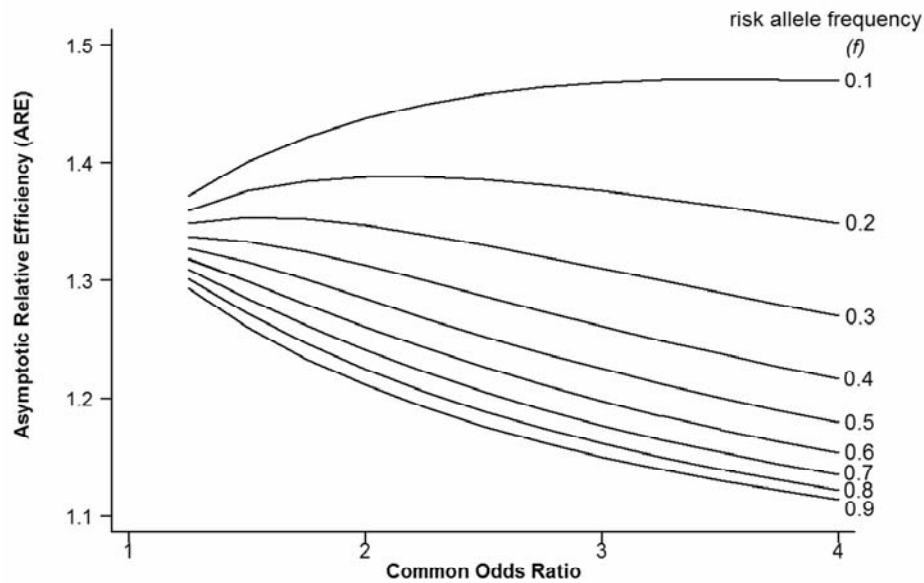


Figure 1. The ARE of the combined estimate compared to the case-control estimate, plotted for a range of possible values of f and OR . For the most commonly expected values of the OR (1.25 to 2), the ARE ranges from 1.25 to 1.4. Even in the extreme case of $f=0.9$ and $OR=4$, the ARE is larger than 1.12 suggesting that the pooled estimate will be 12% more efficient (it will have 11% smaller variance). The influence of n is negligible for $n>50$.

A final remark should be made concerning studies that report the results of both TDT and HHRR analyses. In such cases, the estimates are redundant and only one should be used. For instance, the covariance between y_i^{TDT} and y_i^{HHRR} , can be easily shown to be equal to $\text{var}(y_i^{HHRR})$ and consequently, Eq. (15) is minimized when $y_i^{pooled} = y_i^{HHRR}$. This can also be easily understood, if we consider that the two tests use the same data and HHRR has smaller variance compared to TDT.

2.5 Implementation

The approaches described above have a varying degree of difficulty during the implementation process. For instance, the two-step approach can be performed even in a spreadsheet since it employs non-iterative techniques. Moreover, the two-step approach has the advantage of allowing the conduct of several well-known and informative meta-analytic diagnostics. Such diagnostics include the tests for publication bias (Begg & Mazumdar, 1994; Egger et al, 1997), the cumulative meta-analysis (Bagos & Nikolopoulos, 2009; Lau et al, 1992) and many sensitivity or sub-group analyses. On the other hand, the linear mixed model approaches require the use of some advanced statistical package like Stata,

R, or SAS. Stata lacks a command for fitting models using an arbitrary variance/covariance matrix like the one presented in Eq. (18).

In Appendix II, a Stata program for fitting the two-step approach is given using the covariance estimates of Eq. (21) and (27). In this program, we present a standard meta-analysis with inverse variance weights employing the non-iterative method of moments proposed by DerSimonian and Laird (DerSimonian & Laird, 1986); however, we also show the applicability of other estimation methods such as maximum likelihood and restricted maximum likelihood (Thompson & Sharp, 1999). Additionally, we illustrate how the models of Eq. (17) and (19) can be fitted using a sandwich-type estimator of variance (Williams, 2000). These models were fitted with the `gllamm` module for Stata (Rabe-Hesketh et al, 2002; Rabe-Hesketh et al, 2005), which uses numerical integration by adaptive quadrature to integrate the latent variables and obtain the marginal log-likelihood. Afterwards, the log-likelihood is maximized by Newton-Raphson using numerical first and second derivatives.

3. Results

In this section we apply the methods developed in this work in three real datasets. Initially, the method was applied in the dataset of 356 SNPs, which were evaluated for their potential association with Respiratory Syncytial Virus (RSV) Bronchiolitis (Janssen et al, 2007). In the particular study, 470 children hospitalized for RSV Bronchiolitis, their parents, and 1008 independent controls were analyzed using the method reported by (le Cessie et al, 2008). The particular dataset was selected because it reports a large number of results that could be used to empirically evaluate our method against the technique proposed earlier (le Cessie et al, 2008), which uses individual data.

We also applied the method in a genomewide-association study (GWAS) that was conducted in order to identify susceptibility loci for Multiple Sclerosis. In the particular study, 931 family trios were initially genotyped and tested using the TDT and subsequently, the same cases were contrasted against 2,431 unrelated controls (Hafler et al, 2007). Usually in GWAS, the investigators are interested in replicating their initial significant findings and a combined analysis of screening and replication stage results has been shown to be more powerful than relying on the replication stage alone (Skol et al, 2006). Even though in the particular study several additional datasets were used, we focused our analysis only on the above-mentioned samples.

We finally applied the newly developed method in a published meta-analysis concerning the association of the 10-Repeat Allele of a VNTR Polymorphism in the 3'-UTR of Dopamine Transporter (DAT1) Gene with ADHD (Yang et al, 2007). The particular meta-analysis includes 26 studies, of

which two provided data for both the TDT and case-control approach using the same cases and one study contained information for HHRR and case-control analysis. One of the studies, reported both HHRR and TDT results, but as we discussed earlier only one of them could be used (we chose the TDT approach). The meta-analysis also includes 15 studies that used solely the TDT design, 3 studies that used only the case-control methodology, and 5 studies focused exclusively on the HHRR design. In the original publication, the data were analysed separately, but here we were able to perform the analysis integrating all the available information.

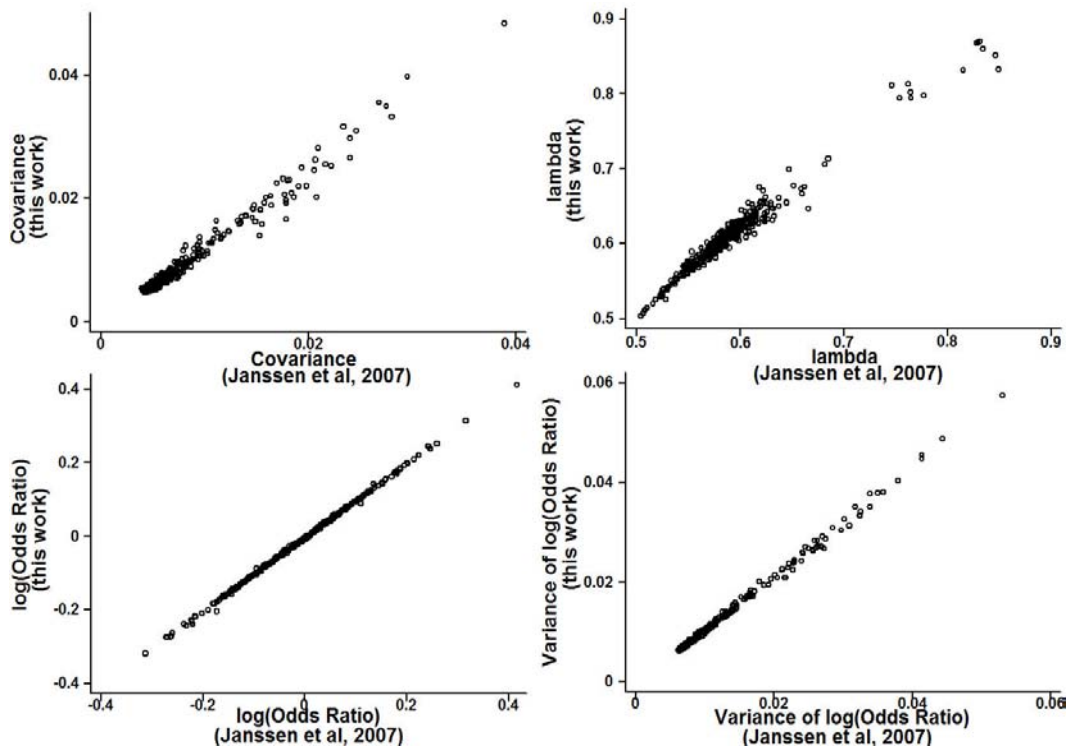


Figure 2. Comparison of the results obtained by applying the methods proposed in this work against the estimates produced by applying the sandwich estimator on the data of 356 SNPs that were reported by Janssen et al, (2007). We present results for the covariance, the estimate of lambda, the combined log Odds Ratio and its variance. We used reconstructed (imputed) data for the 2x2 tables and in all cases the R^2 was greater than 0.96.

In the analysis of the data concerning the potential association of polymorphisms with RSV Bronchiolitis (Janssen et al, 2007) as well as the GWAS for Multiple Sclerosis (Hafler et al, 2007), an additional complication stemmed from the fact that the detailed 2x2 contingency tables were not given in the original publications. Instead, only the estimates (OR and 95% C.I.s) were provided for the case-control and the TDT methods (the data of Janssen et al, provided also the results of the combined analysis using the sandwich estimator that were used for comparison). Therefore, we employed a previously proposed method (Di Pietrantonj, 2006) for reconstructing the data of case-control analysis (the TDT analysis can be easily reconstructed since we only need 2 counts). Since the reconstruction method provides two plausible solutions (Di Pietrantonj, 2006), we kept the one that produces allele frequencies closer to the ones reported by the Hapmap project for European populations. In the dataset of Janssen et al, six out of the 356 SNPs that were initially analysed, provided very small allele counts and thus they were excluded and the same happened for three out of the 110 SNPs analysed in the GWAS of Multiple Sclerosis (in the initial analyses of individual data the 95% C.I.s were also very wide).

When Eq. (22) was applied to the reconstructed data of the 356 SNPs reported by Janssen and coworkers, there was a remarkable agreement with the original results that were obtained using the sandwich estimator. The results clearly indicate that although we used reconstructed data (an approach with an inherent degree of uncertainty) the obtained covariances are quite close to those produced initially by the sandwich estimator. Similar findings are also obtained by comparing the λ that was estimated by Eq. (16). Most importantly, the final estimate of the combined log-Odds Ratio and its variance were also in very good agreement with the published estimates (Figure 1). The method developed here, produces marginally larger estimates for the covariance and this eventually leads to slightly more conservative estimates compared to the sandwich estimator using individual data. Since our method is based on the approximate distribution of the counts of the contingency table, it is obvious that it will deviate more in case the counts in the contingency table are small or even close to zero. Nevertheless, the slight upward bias, is known in the literature and it can be reduced by simply adding a small constant (i.e. $c=1/2$) to the observed counts (Anscombe, 1956; Gart & Zweifel, 1967; Haldane, 1956).

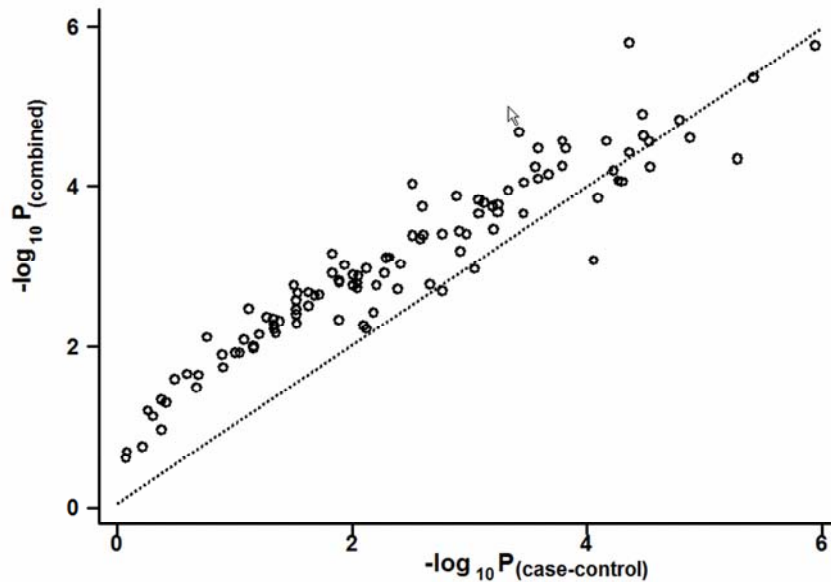


Figure 3. Application of the method in the GWAS for Multiple Sclerosis (Hafler et al, 2007). We plot the negative logarithm (base 10) of the p -values obtained from the combined analysis, against the respective value obtained by the case-control analysis alone.

When the method was applied to the top most significantly associated SNPs (110 in total) reported in the GWAS of Multiple Sclerosis (Hafler et al, 2007), the results clearly highlight the usefulness of our approach (Figure 3). Compared to the case-control analysis, the combined estimate produces smaller p -values for the majority of analysed SNPs, suggesting an increased power to detect weak genotype effects. Even though none of the SNPs reached the commonly accepted strict limit of $p < 10^{-7}$, the combined analysis yields 9 additional SNPs with $p < 10^{-4}$, compared to the original analysis based on case-control data. These results clearly illustrate the gain in power obtained by using the proposed method.

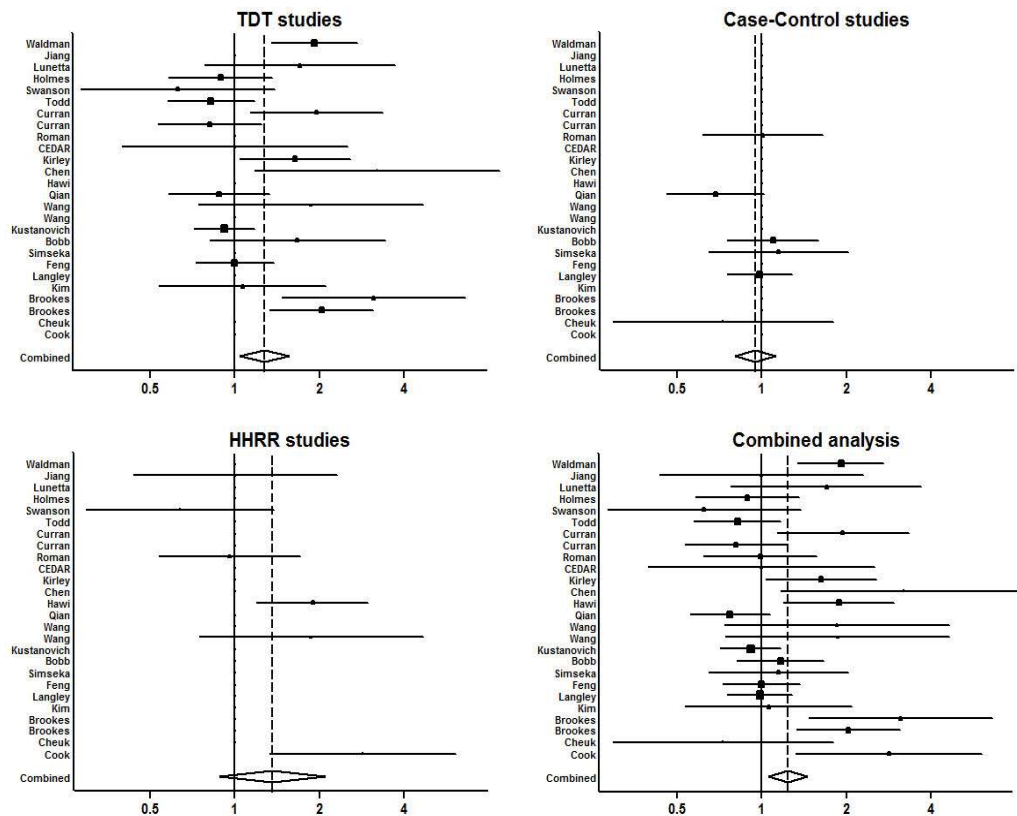


Figure 4. Forest plots for the separate meta-analyses (for the TDT, Case-Control and HHRR studies) as well as for the combined analysis using the two-step approach. The former analyses are nearly identical to the ones obtained in the original publication (Yang et al, 2007), whereas the latter can be obtained only with the methods developed here. The heterogeneity is rather large and in the various sub-analyses, there is a remarkable disagreement between population-based case-control studies and the family-based studies (TDT and HHRR).

In the meta-analysis for the association of the 10-Repeat Allele of the VNTR Polymorphism in the 3'-UTR of Dopamine Transporter Gene with ADHD (Yang et al, 2007), the separate analyses produced results similar with those obtained by the original analysis which point to a significant association, only among the 17 TDT studies with an OR of 1.27 (95% C.I.: 1.04, 1.56). Analyzing only the available data obtained by the 6 case-control studies yields an OR of 0.95 (95% C.I.: 0.80, 1.12), whereas the analysis of the 6 HHRR studies provided an OR of 1.36 (95% C.I.: 0.88, 2.09). Clearly these results suggest that there are differences between the family-based analyses (TDT and HHRR) and the case-control approach, but the obtained estimates could not be formally compared or pooled since the simple analysis does not account for the within studies correlation. If we try to perform a traditional meta-analysis keeping from each study that reported two groups only the case-control one, the results point to a significant overall association (OR=1.23; 95% C.I.: 1.05, 1.45). When the same procedure was repeated keeping only the family-based analyses, the results are similar even though with slightly larger confidence intervals (OR=1.27, 95% C.I.: 1.08, 1.49). Nevertheless, such an analysis was not performed by the authors in the original publication.

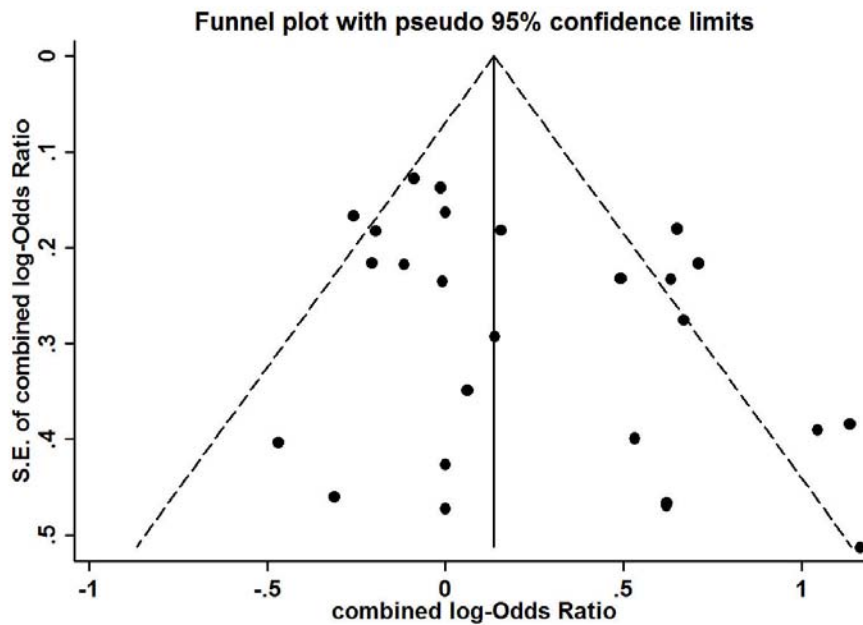


Figure 5. Funnel plot for detecting the asymmetry of the combined estimate. There is only a weak indication of publication or other small study-related bias. The test of Begg and Mazumbar do not provide evidence but the other tests (Egger's test, random effects regression test and the "trim and fill" test) provide some evidence for the existence of bias (see text for details).

When we applied the methods developed in this work using the two-step analysis, the pooled estimate reached a statistical significance with an OR of 1.24 (95% C.I.: 1.06, 1.45). This analysis produced slightly narrower confidence intervals compared to all the previous approaches. Even though the results are negligible from a practical point of view, one should keep in mind that if additional studies reported several sub-groups, the difference would be noticeable. Figure 4 presents the results of the three separate analyses, as well as the combined analysis in the form of the forest plot. The between-studies heterogeneity from the combined analysis is rather high with a τ^2 equal to 0.0922 ($I^2=62.7\%$), but this cannot be attributed solely to the study design or the racial descent of the participants (for instance, the I^2 among the TDT studies is similarly equal to 66.1%). The ML and REML estimates are nearly identical (data not shown). Fitting the hierarchical models of Eq. (17) and Eq. (19) with ML and the sandwich estimator of the variance yields also very similar results, since the overall OR was found equal to 1.23 (95% C.I.: 1.03, 1.46) in both cases. The last result also indicates that the between-groups variance within a study is negligible ($\tau_g^2 < 10^{-8}$), but this could also be attributed to the small number of studies that reported such sub-groups. The between-studies variance τ^2 from both the models of Eq. (17) and Eq. (19) is equal to 0.0994, which also implies that the two-step and the one-step approaches yield nearly identical results.

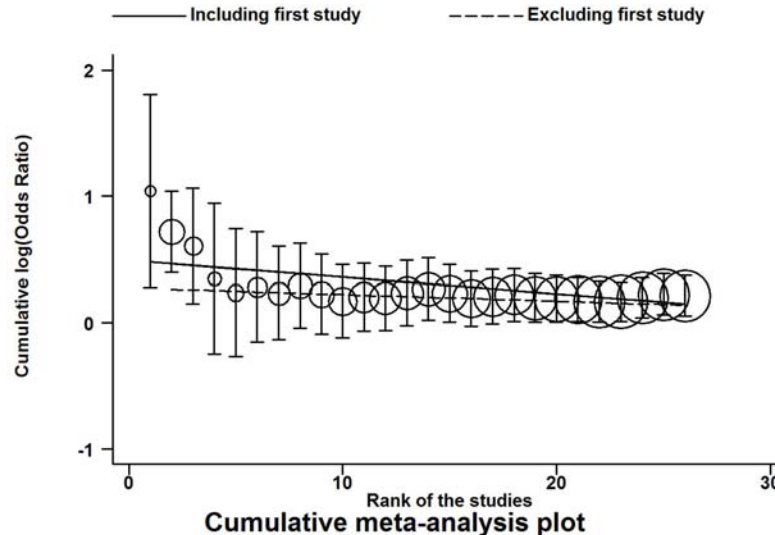


Figure 6. Cumulative meta-analysis plot for detecting trend of the combined estimate over time. The regression lines (obtained by including or excluding the first study) were fitted using the GLS regression-based test (Bagos and Nikolopoulos, 2009). Both the regression-based test and the first vs. subsequent comparison, corroborate to the existence of the Proteus phenomenon since the first published study exhibits a rather large Odds Ratio compared to the subsequent ones. If the first study is removed from the analysis (dashed line) the trend disappears but the overall significance of the combined estimate is not altered (see text for details).

The combined analysis allows us to examine several well-known diagnostics incorporating information from all the available data. For instance, it is now possible to perform a series of available tests for publication bias (Macaskill et al, 2001). In this dataset, the method of Begg and Mazumbar (Begg & Mazumdar, 1994) excludes the possibility of publication bias, but the regression-based technique of Egger (Egger et al, 1997) and its random-effects analogue (Thompson & Sharp, 1999), provided moderate evidence of publication or other small study-related bias (p -values equal to 0.042 and 0.060, respectively). Similar findings were also obtained using the non-parametric trim and fill method (Duval & Tweedie, 2000). The visual inspection of the funnel plot however, provides only some weak evidence for asymmetry of the estimates (Figure 5). Nevertheless, the large heterogeneity could also be responsible for the behaviour of these tests.

The cumulative meta-analysis plot (Lau et al, 1992), revealed that the first published study on the topic, exhibited a rather large Odds Ratio compared to the subsequently published studies (Figure 6). The formal statistical tests, namely, the GLS regression-based test (Bagos & Nikolopoulos, 2009) and the first vs. subsequent comparison (Ioannidis & Trikalinos, 2005), provided also similar conclusions (p -values equal to 0.031 and 0.001 respectively). After excluding the first published study, the overall estimate remains statistically significant, even though to a lesser extent (OR=1.21; 95% C.I.: 1.03, 1.41) and the trend almost disappeared, indicating that the first published study was over-estimating the suggested risk conferred by the polymorphism. Interestingly, the first study used solely the HHRR statistic and showed a rather large effect (OR=2.84; 95% C.I.: 1.32, 6.09). These results indicate the presence of the so-called “Proteus phenomenon” (Ioannidis & Trikalinos, 2005), also known as the “winner’s curse”, which is empirically known to threaten the replication validity of genetic-association studies (Ioannidis et al, 2001). Clearly, the method developed here allowed for the identification of such a bias that was missed by the authors of the original publication.

4. Discussion

During the last years, there is an increased interest in the combination of the results of family-based and population-based case-control studies that involve the same patients group (cases). Various methods have been proposed for addressing the particular problem allowing a combined analysis (Epstein et al, 2005; Gebregziabher et al, 2010; Glaser & Holmans, 2009; Guo et al, 2009; Hsu et al, 2009; Joo et al, 2007; Lasky-Su et al, 2010; Nagelkerke et al, 2004; Pfeiffer et al, 2008). However, these techniques require individual information to be available

and thus, are not suitable for the meta-analysis of published studies in which case the researcher has access, most of the times, only to summary data (genotype or allele counts).

In this work we presented an easily implemented and efficient method of analysis utilizing only summary data, usually available in the published literature. The approach follows the general method for dealing with correlated estimates (Gleser & Olkin, 1994). Other situations in which the correlation can be computed analytically include studies with multiple treatment arms (Higgins & Whitehead, 1996; Lu & Ades, 2004), clinical trials with mutually exclusive outcomes (Trikalinos & Olkin, 2008), dose-response models (Berrington & Cox, 2003; Greenland & Longnecker, 1992), multivariate modeling of log-odds ratios in genetic association studies (Bagos, 2008), genetic association studies with overlapping sets of controls (Lin & Sullivan, 2009) and multipoint meta-analysis of genetic association studies (Bagos & Liakopoulos, 2010).

The statistical theory behind the proposed methodology is very simple and is based on standard large sample approximations (Agresti, 2002; Clayton & Hills, 1993), that are routinely used in meta-analyses of published data (summary-based meta-analysis). By comparing the results against those obtained by the general individual-based method proposed by le Cessie et al (2008), we showed that our method is quite accurate. The accuracy is probably even higher if we consider also the fact that we did not use the actual data, but we imputed the respective 2x2 contingency table. It should be noticed that the covariance derived in Eq. (21) is essentially the approximate variance of the log-counts of cases. These approximations have been shown to perform quite well even for small sample sizes (Agresti, 1999) justifying this way their popularity. Theoretically, the method is expected to produce biased results when the sample size is very small or when there are small, or even zero cell counts. However, in the majority of the practical applications the particular counts needed for calculating the covariance should be sufficiently large for the summary-based analysis to be valid in the first place. For instance, this could happen when the risk allele frequency is very small (<1%) or very large (>99%), in which case the covariance will be biased and the use of summary-data methods is also questionable. Nevertheless, we should mention that the slight bias observed in this work compared to the estimators that use individual data, is well known in the literature and it can be minimized by simply adding a small constant (i.e. $c=1/2$) to the observed counts (Anscombe, 1956; Gart & Zweifel, 1967; Haldane, 1956).

Apart from accuracy of the large-sample approximation, other limitations which are consequences of performing a summary-based meta-analysis should be considered. First of all, as we already discussed in the introduction, the particular method pools the estimates assuming a co-dominant model of inheritance (i.e. allele-based). Other models of inheritance can be estimated from published

reports of case-control association studies (Bagos, 2008), but when it comes to family-based analyses the classic TDT/HRR approach allows only the allele-based comparison. Other approaches that allow different models of inheritance, even though available, are not widely used (Schaid & Sommer, 1994). Thus, since the majority of published family-based association studies usually report only the number of transmitted and untransmitted alleles, pooling the estimates in an allele-based comparison is the only reasonable choice. Another limitation is that the allele-based comparison in case-control studies makes the assumption of Hardy-Weinberg equilibrium (HWE) in the population. HWE is a debating issue in genetic association studies (Minelli et al, 2008; Salanti et al, 2005a; Trikalinos et al, 2006) and from our point of view, it is sufficient to say that in case of deviations, the variance equations for the allelic odds ratio are biased and suitable corrections could be used (Sato et al, 2006; Trikalinos et al, 2006).

We should emphasize however once again, that the proposed method, is the only currently available technique that allows the pooling of effect estimates using only published information and thus we expect to be useful when individual data are missing or hard to retrieve (as usually is the case in a meta-analysis). We have also shown that even when the original data are not provided in the form of 2x2 contingency tables, the method can be quite accurate, since the tables can be reconstructed using standard methods (Di Pietrantonj, 2006). The only problem that can potentially arise in such a case is when the published statistics (odds ratios and confidence intervals) have been calculated by adjusting for potential confounders, especially when these are correlated with both disease status and polymorphism.

We have also shown with an application in real data, that the particular methodology, which provides a closed-form expression for the covariance function could be easily used in genome-wide association studies in which the TDT design is also applicable (Benyamin et al, 2009; Murphy et al, 2010; Murphy et al, 2008). Meta-analysis of GWAS is also a common approach to increase statistical power and detect weak genotype effects (Neale et al, 2010b) and thus, our method could be a useful alternative in combining the results of family-based and population-based studies (Lasky-Su et al, 2010) or in performing replication studies utilizing the same cases (Murphy et al, 2010; Murphy et al, 2008). Moreover, recent evidence suggests that using time-consuming individual data methods for meta-analysis of GWAS does not necessarily results in increased efficiency (Lin & Zeng, 2010), making this way the particular method very attractive in terms of computational time.

Using the newly developed method, we re-analyzed the data from a published meta-analysis concerning the association of the 10-Repeat Allele of a VNTR Polymorphism in the 3'-UTR of Dopamine Transporter Gene with ADHD (Yang et al, 2007). The combined analysis allowed the calculation of several

diagnostics for publication bias and the conduct of cumulative meta-analysis that were not possible to be performed with the traditional approach. We were able to provide a combined estimate that point to a significant association of the 3'-UTR polymorphism with ADHD. Moreover, we were able to show the existence of the Proteus phenomenon, since the first published study produced a larger Odds Ratio compared to the subsequent ones. Thus, it is reasonable for the combined estimate to be quite smaller and subsequently, there was a need for a large number of studies in order to establish statistical significance. Finally, we have shown that there is a considerable asymmetry in the funnel plot, possibly indicating publication bias.

Although the current analysis does not aim to perform a comprehensive meta-analysis on the subject, but only to illustrate the usefulness of our method, the discrepancy observed between the results of case-control and family-based studies (TDT and HHRR) need to be discussed. ADHD is considered a complex disease where multiple genes (Coghill & Banaschewski, 2009), operate in parallel or in synergy with environmental factors (Max et al, 2004). Concerning ADHD phenotype, recent evidence suggests that ADHD is unlikely to exist as an identifiable disease (Furman, 2008). Therefore, it seems that disparities between case-control and TDT/HHRR studies including dissimilarities in the definition of the disease under investigation or of its subtypes, in the clinical criteria adopted and in the genealogy of the selected population could indeed lead to different results. Moreover, the differences between population-based case-control and TDT/HHRR studies can be also partially attributed to the frequently observed missing genotype data of TDT/HHRR analyses. In many family studies, the analysis focuses on complete parent-proband trios, which can produce biased results because the remaining samples might not be representative of the total population. In a relatively recent study, there was indeed evidence that children with ADHD whose fathers' samples were missing were phenotypically different from children for whom genotype data was derived from both parents (West et al, 2002). Moreover, in a family-based GWAS concerning ADHD, a systemic bias in the transmission of the major vs. the minor alleles was identified. The authors however, based on the quality control analysis suggested that this was probably due to the effect of differential rates of missing genotypes (Neale et al, 2008). In recent case-control GWAS, polymorphisms of DAT1 were not found to have any significant association (Neale et al, 2010a), but in a family-based GWAS a polymorphism of the same gene (rs2937639) showed a nominal significant association ($p=0.004$) with late onset ADHD (Lasky-Su et al, 2008). Interestingly, a large multicenter study provided another possible explanation of these discrepancies, pointing to a differential association of DAT1 with ADHD in children and in adults (Franke et al, 2010).

In conclusion, we presented an efficient method for combining the results of family-based and population-based case-control studies that use the same cases. The theoretical innovation lies in the derivation of closed form formula for the covariance of the correlated estimates. Implementation of the method is straightforward and in Appendix II, a Stata program is given for putting the methods presented here into practice.

Appendix I

For calculating the covariance of estimates derived from case-control and HHRR studies, we will have:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HHRR}) &= \text{cov}\left[\log\left(\frac{n_{11}n_{00}}{n_{01}n_{10}}\right), \log\left(\frac{wz}{xy}\right)\right] \\ &= \text{cov}[\log(n_{11}n_{00}) - \log(n_{01}n_{10}), \log(wz) - \log(xy)]\end{aligned}\quad (\text{A1})$$

which after trivial calculations, making use of the standard properties of the covariance function, reduces to:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{HHRR}) &= \text{cov}[\log(n_{11}), \log(w)] + \text{cov}[\log(n_{10}), \log(x)] \\ &= \text{var}[\log(n_{11})] + \text{var}[\log(n_{10})] \\ &= \frac{1}{n_{11}} + \frac{1}{n_{10}}\end{aligned}\quad (\text{A2})$$

The first equality is derived by noticing from Table 3 that $n_{11}=w$ and $n_{10}=x$ (since the same cases are used), whereas the other two are simple properties of the large-sample behaviour of the counts in the contingency table.

In case a study i reports results from a case-control and a TDT analysis, the covariance will be:

$$\begin{aligned}\text{cov}(y_i^{CC}, y_i^{TDT}) &= \text{cov}\left[\log\left(\frac{n_{11}n_{00}}{n_{01}n_{10}}\right), \log\left(\frac{b}{c}\right)\right] \\ &= \text{cov}[\log(n_{11}n_{00}) - \log(n_{01}n_{10}), \log(b) - \log(c)]\end{aligned}\quad (\text{A3})$$

which, after trivial calculations reduces to:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \text{cov}[\log(n_{11}), \log(b)] + \text{cov}[\log(n_{10}), \log(c)]\quad (\text{A4})$$

Given that $n_{11}=w=a+b$ and $n_{10}=x=c+d$, we observe that the right-hand side of Eq. (A4) includes terms of the form $\text{cov}[\log(S+U), \log(S)]$. We observe now that S and

U are independent Poisson random variables with $E[S] = \text{var}[S] = s$ and $E[U] = \text{var}[U] = u$ and thus, $Z = S + U$ is also Poisson random variable with $E[Z] = \text{var}[Z] = z = s + u$. Consequently, (Z, S) will follow a bivariate Poisson distribution (Marshall & Olkin, 1985) and the covariance between Z and S would then be:

$$\text{cov}[S + U, S] = \text{cov}[S, S] = \text{var}[S] = s \quad (\text{A5})$$

The covariance between functions of random variables Z and S , requires knowledge of the joint and the marginal cumulative distribution functions (Cuadras, 2002). However, we can easily approximate it using the delta method (Agresti, 2002) with a first order Taylor expansion around the means:

$$\text{cov}[f(Z), g(S)] \approx \frac{\partial f(E[Z])}{\partial Z} \frac{\partial g(E[S])}{\partial S} \text{cov}(Z, S) \quad (\text{A6})$$

Thus, we will have:

$$\begin{aligned} \text{cov}[\log(Z), \log(S)] &= \text{cov}(Z, S) \frac{\partial \log Z}{\partial Z} \frac{\partial \log S}{\partial S} \\ &= \frac{s}{z} = \frac{1}{s + u} \end{aligned} \quad (\text{A7})$$

Using Eq. (A7), we finally obtain:

$$\text{cov}(y_i^{CC}, y_i^{TDT}) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \quad (\text{A8})$$

In the more general case when we have sets of cases that overlap only partially, the equations need to be modified. For instance, if we assume that the two designs share $m/2$ cases that contribute m alleles (with $m < n_1$) the respective counts in Eq. (A2) need to be decomposed into two disjoint subsets denoted by subscripts s and d (where s stands for shared and d for distinct). Thus, Eq. (A2) will become:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{HRR}) &= \text{cov}[\log(n_{11s} + n_{11d}), \log(w_s + w_d)] \\ &\quad + \text{cov}[\log(n_{10s} + n_{10d}), \log(x_s + x_d)] \end{aligned} \quad (\text{A9})$$

By observing that $n_{11s} = w_s$ and $n_{10s} = x_s$ and $m = n_{11s} + n_{10s} = w_s + x_s$ and subsequently applying Eq. (A7) in order to calculate $\text{cov}[\log(S+U), \log(S+W)]$, we now have:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{HRR}) &= \frac{n_{11s}}{(n_{11s} + n_{11d})(n_{11s} + w_d)} + \frac{n_{10s}}{(n_{10s} + n_{10d})(n_{10s} + x_d)} \\ &= \frac{n_{11s}}{n_{11}w} + \frac{n_{10s}}{n_{10}x} \end{aligned} \quad (\text{A10})$$

Similarly by observing that $n_{11s} = w_s = a_s + b_s$ and $n_{10s} = x_s = c_s + d_s$, the extended version of Eq. (A8) becomes:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{TDT}) &= \text{cov}[\log(n_{11s} + n_{11d}), \log(b_s + b_d)] \\ &+ \text{cov}[\log(n_{10s} + n_{10d}), \log(b_s + b_d)] \end{aligned} \quad (\text{A11})$$

which in turn simplifies to:

$$\begin{aligned} \text{cov}(y_i^{CC}, y_i^{TDT}) &= \frac{b_s}{(n_{11s} + n_{11d})(b_s + b_d)} + \frac{b_s}{(n_{10s} + n_{10d})(b_s + b_d)} \\ &= \frac{b_s}{b} \left(\frac{1}{n_{11}} + \frac{1}{n_{10}} \right) \end{aligned} \quad (\text{A12})$$

Appendix II

** the data are from the meta-analysis for the association of
 ** the 10-Repeat Allele of a VNTR Polymorphism in the
 ** 3'-UTR of Dopamine Transporter Gene with Attention Deficit
 ** Hyperactivity Disorder (Yang et al, 2007)

input id str20 author year b c n11 n10 n01 n00 w x y z

```
1 Waldman 1998 90 47 . . . . .
2 Swanson 2000 10 16 . . . . 60 20 66 14
3 Lunetta 2000 17 10 . . . . .
4 Holmes 2000 40 45 . . . . .
5 Todd 2001 55 67 . . . . .
6 Curran 2001 39 20 . . . . .
7 Curran 2001 39 48 . . . . .
8 Kirley 2002 49 30 . . . . .
9 CEDAR 2002 9 9 . . . . .
10 Chen 2003 16 5 . . . . .
11 Qian 2004 43 49 578 86 392 40 . . . .
12 Kustanovich 2004 119 130 . . . . .
13 Wang 2004 13 7 . . . . .
14 Kim 2005 17 16 . . . . .
15 Feng 2005 76 76 . . . . .
16 Bobb 2005 20 12 88 238 65 193 . . . .
17 Brookes 2006 65 32 . . . . .
18 Brookes 2006 28 9 . . . . .
19 Cheuk 2006 . . 116 12 119 9 . . . .
20 Langley 2005 . . 387 139 424 150 . . . .
21 Roman 2001 . . 98 34 166 58 105 30 106 29
22 Simseka 2005 . . 59 33 67 43 . . . .
23 Hawi 2003 . . . . . 145 42 121 66
24 Wang 2004 . . . . . 100 8 94 14
25 Cook 1995 . . . . . 72 12 57 27
26 Jiang 1999 . . . . . 136 12 136 12
end
```

** calculation of the population-based allelic OR and its variance - Eq.
 (1) and (2)

```
gen logorcc=log( (n11* n00)/( n10*n01))
gen varcc=1/ n11+1/ n10+1/ n01+1/ n00
```

```

gen secc=sqrt(varcc)

** calculation of the tdt-based OR and its variance - Eq. (6) and (7)
gen logortdt=log(b/c)
gen vartdt=1/b+1/c
gen setdt=sqrt(vartdt)

** calculation of the HHRR and its variance - Eq. (8) and (9)
gen logorhhrr=log((w*z)/(x*y))
gen varhhrr=1/x+1/z+1/y+1/w
gen sehrr=sqrt(varhhrr)

** code the different types of studies
gen type=1 if logortdt!=. & logorcc==.&logorhhrr==.
replace type=2 if logorcc!=. & logortdt==.&logorhhrr==.
replace type=3 if logorhhrr!=. & logortdt==.&logorcc==.
replace type=4 if logorcc!=. & logortdt!=.&logorhhrr==.
replace type=5 if logorcc!=. &logorhhrr !=.&logortdt==.
replace type=6 if logorhhrr!=. & logortdt!=.&logorcc==.

*calculation of the covariances - Eq. (21) and (22)
gen cov_tdt_cc=1/n11+1/n10 if type==4
gen cov_hhrr_cc=1/n11+1/n10 if type ==5

** generate the combined logOR
gen logor= logortdt if type==1
replace logor= logorcc if type==2
replace logor= logorhhrr if type==3

** generate the variance of the combined logOR
gen var=vartdt if type==1
replace var=varcc if type==2
replace var=varhhrr if type==3

**calculate lambda from Eq. (16)
gen lambda=(vartdt-cov_tdt_cc)/(vartdt+varcc-2*cov_tdt_cc) if type==4
replace lambda=(varhhrr-cov_hhrr_cc)/(varhhrr+varcc-2*cov_hhrr_cc) if
type== 5

** compute the combined logOR - Eq. (14)
replace logor=lambda*logorcc+(1-lambda)*logortdt if type==4
replace logor=lambda*logorcc+(1-lambda)*logorhhrr if type==5

** compute the variance of the combined logOR - Eq. (15)
replace var=lambda^2*varcc+(1-lambda)^2*vartdt+2*lambda*(1-
lambda)*cov_tdt_cc if type ==4
replace var=lambda^2*varcc +(1-lambda)^2*varhhrr+2*lambda*(1-
lambda)*cov_hhrr_cc if type ==5

replace logor=logortdt if type==6
replace var=vartdt if type ==6

gen se=sqrt(var)

label variable logor "combined log-Odds Ratio"
label variable se "S.E. of combined log-Odds Ratio"

```



```
** perform the meta-analysis with the method of DerSimonian and Laird
label define type 1 "TDT" 2 "CC" 3 "HHRR" 4 "CC+TDT" 5 "CC+HHRR" 6
      "TDT+HHRR"
metan logor se,randomi by(type) label(namevar=author,yearvar=year) eform
      xlab(0.5, 1, 2, 4)

** meta-analysis using the ML or REML
metareg logor,wsse(se) bse(ml)
metareg logor,wsse(se) bse(reml)

**tests for publication bias
metabias logor se

** cumulative meta-analysis
sort year
metacum logor se, id(author) effect(r) gr xlab
metatrend logor se

**display the combined results graphically
replace setdt=0 if setdt ==.
replace secc=0 if secc ==.
replace sehrr=0 if sehrr==.
replace logortdt =0 if logortdt ==.
replace logorhhrr=0 if logorhhrr==.
replace logorcc=0 if logorcc==.

** this is due to a bug in metagraph command
rename y y00

metan logortdt setdt,randomi eform nograph
metagraph logortdt setdt, id(author ) combined(`exp(r(ES))'
      `exp(r(ci_low))' `exp(r(ci_upp))') x(0.5, 1, 2, 4) eform nodraw
      name(graph1) title(TDT studies)
metan logorcc secc,randomi eform nograph
metagraph logorcc secc, id(author ) combined(`exp(r(ES))'
      `exp(r(ci_low))' `exp(r(ci_upp))') x(0.5, 1, 2, 4) eform nodraw
      name(graph2) title(Case-Control studies)
metan logorhhrr sehrr,randomi eform nograph
metagraph logorhhrr sehrr, id(author ) combined(`exp(r(ES))'
      `exp(r(ci_low))' `exp(r(ci_upp))') x(0.5, 1, 2, 4) eform nodraw
      name(graph3) title(HHRR studies)
metan logor se,randomi eform nograph
metagraph logor se, id(author ) combined(`exp(r(ES))' `exp(r(ci_low))'
      `exp(r(ci_upp))') x(0.5, 1, 2, 4) eform nodraw name(graph4)
      title(Combined analysis)
graph combine graph1 graph2 graph3 graph4 , cols(2) xcommon altshrink

** data rearrangements in order to use a linear mixed model
drop b c n11 n10 n01 n00 w x y00 z secc setdt sehrr logor var
rename id study
rename logortdt logor1
rename vartdt var1
rename logorcc logor2
rename varcc var2
```

```

rename logorhhrr logor3
rename varhhrr var3
reshape long logor var, i(study) j(newtype)
keep if logor~=.
drop type
rename newtype type
gen id=_n

** gllamm needs to have the logarithm of the standard error
gen s=log(sqrt(var))
eq wgt: s
constraint define 1 [lnsl]s=1

**fitting the model of Eq. (17) using the sandwich variance estimator
gllamm logor ,i(study) s(wgt) constraint(1) adapt nip(8) cluster(study)

**fitting the model of Eq. (19) using the sandwich variance estimator
gllamm logor ,i(id study) s(wgt) constraint(1) adapt nip(8)
    cluster(study)

```

References

- Ackerman H, Usen S, Jallow M, Sisay-Joof F, Pinder M, Kwiatkowski DP (2005) A comparison of case-control and family-based association methods: the example of sickle-cell and malaria. *Ann Hum Genet* **69**: 559-565
- Agresti A (1999) On logit confidence intervals for the odds ratio with small samples. *Biometrics* **55**: 597-602
- Agresti A (2002) *Categorical Data Analysis*, 2nd edn. New York: John Wiley & Sons.
- Anscombe FJ (1956) On estimating binomial response relations. *Biometrika* **43**: 461-464
- Attia J, Thakkestian A, D'Este C (2003) Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol* **56**: 297-303
- Bagos PG (2008) A unification of multivariate methods for meta-analysis of genetic association studies. *Stat Appl Genet Mol Biol* **7**: Article31
- Bagos PG, Liakopoulos TD (2010) A multipoint method for meta-analysis of genetic association studies. *Genet Epidemiol* **34**: 702-715

- Bagos PG, Nikolopoulos GK (2007) A method for meta-analysis of case-control genetic association studies using logistic regression. *Stat Appl Genet Mol Biol* **6**: Article17
- Bagos PG, Nikolopoulos GK (2009) Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology. *J Clin Epidemiol* **62**: 1037-1044
- Becker KG, Barnes KC, Bright TJ, Wang SA (2004) The genetic association database. *Nat Genet* **36**: 431-432
- Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* **50**: 1088-1101
- Benyamin B, Visscher PM, McRae AF (2009) Family-based genome-wide association studies. *Pharmacogenomics* **10**: 181-190
- Berrington A, Cox DR (2003) Generalized least squares for the synthesis of correlated information. *Biostatistics* **4**: 423-431
- Breslow N, Leroux B, Platt R (1998) Approximate hierarchical modelling of discrete data in epidemiology. *Stat Methods Med Res* **7**: 49-62
- Burton PR, Tobin MD, Hopper JL (2005) Key concepts in genetic epidemiology. *Lancet* **366**: 941-951
- Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet* **361**: 598-604
- Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R (1987) Meta-analysis of clinical trials as a scientific discipline. II: Replicate variability and comparison of studies that agree and disagree. *Stat Med* **6**: 733-744
- Chen YH, Lin HW (2008) Simple association analysis combining data from trios/sibships and unrelated controls. *Genet Epidemiol* **32**: 520-527
- Clayton D, Hills M (1993) *Statistical Models in Epidemiology*: Oxford University Press.

- Coghill D, Banaschewski T (2009) The genetics of attention-deficit/hyperactivity disorder. *Expert Rev Neurother* **9**: 1547-1565
- Cuadras CM (2002) On the Covariance between Functions. *Journal of Multivariate Analysis* **81**: 19-27
- Curtis D (1997) Use of siblings as controls in case-control association studies. *Ann Hum Genet* **61**: 319-333
- Davey Smith G, Ebrahim S, Lewis S, Hansell AL, Palmer LJ, Burton PR (2005) Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* **366**: 1484-1498
- Deng HW, Chen WM, Recker RR (2002) Transmission disequilibrium test with discordant sib pairs when parents are available. *Hum Genet* **110**: 451-461
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**: 177-188
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* **55**: 997-1004
- Di Pietrantonj C (2006) Four-fold table cell frequencies imputation in meta analysis. *Stat Med* **25**: 2299-2322
- Duffy SW, Rohan TE, Altman DG (1989) A method for combining matched and unmatched binary data. Application to randomized, controlled trials of photocoagulation in the treatment of diabetic retinopathy. *Am J Epidemiol* **130**: 371-378
- Dunnett CW (1955) A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association* **50**: 1096-1121
- Duval S, Tweedie R (2000) A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *JASA* **95**: 89-98
- Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *Bmj* **315**: 629-634

- Epstein MP, Allen AS, Satten GA (2007) A simple and improved correction for population stratification in case-control studies. *Am J Hum Genet* **80**: 921-930
- Epstein MP, Veal CD, Trembath RC, Barker JN, Li C, Satten GA (2005) Genetic association analysis using data from triads and unrelated subjects. *Am J Hum Genet* **76**: 592-608
- Evangelou E, Trikalinos TA, Salanti G, Ioannidis JP (2006) Family-based versus unrelated case-control designs for genetic associations. *PLoS Genet* **2**: e123
- Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* **51**: 227-233
- Feinstein AR (1987) Quantitative ambiguities in matched versus unmatched analyses of the 2x2 table for a case-control study. *Int J Epidemiol* **16**: 128-134
- Franke B, Vasquez AA, Johansson S, Hoogman M, Romanos J, Boreatti-Hummer A, Heine M, Jacob CP, Lesch KP, Casas M, Ribases M, Bosch R, Sanchez-Mora C, Gomez-Barros N, Fernandez-Castillo N, Bayes M, Halmoy A, Halletland H, Landaas ET, Fasmer OB, Knappskog PM, Heister AJ, Kiemeny LA, Kooij JJ, Boonstra AM, Kan CC, Asherson P, Faraone SV, Buitelaar JK, Haavik J, Cormand B, Ramos-Quiroga JA, Reif A (2010) Multicenter analysis of the SLC6A3/DAT1 VNTR haplotype in persistent ADHD suggests differential involvement of the gene in childhood and persistent ADHD. *Neuropsychopharmacology* **35**: 656-664
- Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. *Nat Genet* **36**: 388-393
- Furman LM (2008) Attention-deficit hyperactivity disorder (ADHD): does new research support old concepts? *J Child Neurol* **23**: 775-784
- Gart JJ, Zweifel JR (1967) On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika* **54**: 181-187

- Gebregziabher M, Guimaraes P, Cozen W, Conti DV (2010) A polytomous conditional likelihood approach for combining matched and unmatched case-control studies. *Stat Med*
- Glaser B, Holmans P (2009) Comparison of methods for combining case-control and family-based association studies. *Hum Hered* **68**: 106-116
- Glass G (1976) Primary, secondary and meta-analysis of research. *Educ Res* **5**: 3-8
- Gleser LJ, Olkin I (1994) Stochastically dependent effect sizes. In *The Handbook of Research Synthesis* Cooper HM, Hedges LV (eds), pp 339-355. New York: Russell Sage Foundation
- Greenland S, Longnecker MP (1992) Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *Am J Epidemiol* **135**: 1301-1309
- Guo CY, Lunetta KL, DeStefano AL, Cupples LA (2009) Combined haplotype relative risk (CHRR): a general and simple genetic association test that combines trios and unrelated case-controls. *Genet Epidemiol* **33**: 54-62
- Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, De Jager PL, de Bakker PI, Gabriel SB, Mirel DB, Ivinson AJ, Pericak-Vance MA, Gregory SG, Rioux JD, McCauley JL, Haines JL, Barcellos LF, Cree B, Oksenberg JR, Hauser SL (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* **357**: 851-862
- Haldane JB (1956) The estimation and significance of the logarithm of a ratio of frequencies. *Ann Hum Genet* **20**: 309-311
- Hamajima N, Hirose K, Inoue M, Takezaki T, Kuroishi T, Tajima K (1994) Case-control studies: matched controls or all available controls? *J Clin Epidemiol* **47**: 971-975
- Higgins JP, Whitehead A (1996) Borrowing strength from external trials in a meta-analysis. *Stat Med* **15**: 2733-2749
- Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. *Genet Med* **4**: 45-61

- Hopper JL, Bishop DT, Easton DF (2005) Population-based family studies in genetic epidemiology. *Lancet* **366**: 1397-1406
- Horvath S, Laird NM (1998) A discordant-sibship test for disequilibrium and linkage: no need for parental data. *Am J Hum Genet* **63**: 1886-1897
- Hsu L, Starr JR, Zheng Y, Schwartz SM (2009) On combining triads and unrelated subjects data in candidate gene studies: an application to data on testicular cancer. *Hum Hered* **67**: 88-103
- Infante-Rivard C, Mirea L, Bull SB (2009) Combining case-control and case-trio data from the same population in genetic association analyses: overview of approaches and illustration with a candidate gene study. *Am J Epidemiol* **170**: 657-664
- Ioannidis JP (2005a) Molecular bias. *Eur J Epidemiol* **20**: 739-745
- Ioannidis JP (2005b) Why most published research findings are false. *PLoS Med* **2**: e124
- Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. *Nat Genet* **29**: 306-309
- Ioannidis JP, Trikalinos TA (2005) Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* **58**: 543-549
- Janssen R, Bont L, Siezen CL, Hodemaekers HM, Ermers MJ, Doornbos G, van 't Slot R, Wijmenga C, Goeman JJ, Kimpen JL, van Houwelingen HC, Kimman TG, Hoebee B (2007) Genetic susceptibility to respiratory syncytial virus bronchiolitis is predominantly associated with innate immune genes. *J Infect Dis* **196**: 826-834
- Joo J, Tian X, Zheng G, Stylianou M, Lin JP, Geller NL (2007) Joint analysis of case-parents trio and unrelated case-control designs in large scale association studies. *BMC Proc* **1 Suppl 1**: S28
- Kazeem GR, Farrall M (2005) Integrating case-control and TDT studies. *Ann Hum Genet* **69**: 329-335

- Keller T, Olkin I (2004) Combining correlated unbiased estimators of the mean of a normal distribution. In *A Festschrift for Herman Rubin*, DasGupta A (ed), Vol. 45, pp 218-227. Institute of Mathematical Statistics
- Lasky-Su J, Anney RJ, Neale BM, Franke B, Zhou K, Maller JB, Vasquez AA, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, Gill M, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Sonuga-Barke E, Steinhausen HC, Taylor E, Daly M, Laird N, Lange C, Faraone SV (2008) Genome-wide association scan of the time to onset of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* **147B**: 1355-1358
- Lasky-Su J, Won S, Mick E, Anney RJ, Franke B, Neale B, Biederman J, Smalley SL, Loo SK, Todorov A, Faraone SV, Weiss ST, Lange C (2010) On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls. *Am J Hum Genet* **86**: 573-580
- Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC (1992) Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* **327**: 248-254
- le Cessie S, Nagelkerke N, Rosendaal FR, van Stralen KJ, Pomp ER, van Houwelingen HC (2008) Combining matched and unmatched control groups in case-control studies. *Am J Epidemiol* **168**: 1204-1210
- Lee WC (2003) Genetic association studies of adult-onset diseases using the case-spouse and case-offspring designs. *Am J Epidemiol* **158**: 1023-1032
- Lin DY, Sullivan PF (2009) Meta-analysis of genome-wide association studies with overlapping subjects. *Am J Hum Genet* **85**: 862-872
- Lin DY, Zeng D (2010) Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol* **34**: 60-66
- Lu G, Ades AE (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* **23**: 3105-3124
- Macaskill P, Walter SD, Irwig L (2001) A comparison of methods to detect publication bias in meta-analysis. *Stat Med* **20**: 641-654

- Marshall AW, Olkin I (1985) A Family of Bivariate Distributions Generated by the Bivariate Bernoulli Distribution. *Journal of the American Statistical Association* **80**: 332-338
- Max JE, Lansing AE, Koele SL, Castillo CS, Bokura H, Schachar R, Collings N, Williams KE (2004) Attention deficit hyperactivity disorder in children and adolescents following traumatic brain injury. *Dev Neuropsychol* **25**: 159-177
- Minelli C, Thompson JR, Abrams KR, Thakkestian A, Attia J (2005) The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol* **34**: 1319-1328
- Minelli C, Thompson JR, Abrams KR, Thakkestian A, Attia J (2008) How should we use information about HWE in the meta-analyses of genetic association studies? *Int J Epidemiol* **37**: 136-146
- Minelli C, Thompson JR, Tobin MD, Abrams KR (2004) An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *Am J Epidemiol* **160**: 445-452
- Mitchell LE (2000) Relationship between case-control studies and the transmission/disequilibrium test. *Genet Epidemiol* **19**: 193-201
- Moreno V, Martin ML, Bosch FX, de Sanjose S, Torres F, Munoz N (1996) Combined analysis of matched and unmatched case-control studies: comparison of risk estimates from different studies. *Am J Epidemiol* **143**: 293-300
- Murphy A, S TW, Lange C (2010) Two-stage testing strategies for genome-wide association studies in family-based designs. *Methods Mol Biol* **620**: 485-496
- Murphy A, Weiss ST, Lange C (2008) Screening and replication using the same data set: testing strategies for family-based studies in which all probands are affected. *PLoS Genet* **4**: e1000197
- Nagelkerke NJ, Hoebee B, Teunis P, Kimman TG (2004) Combining the transmission disequilibrium test and case-control methodology using generalized logistic regression. *Eur J Hum Genet* **12**: 964-970

- Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB, Vasquez AA, Asherson P, Chen W, Banaschewski T, Buitelaar J, Ebstein R, Gill M, Miranda A, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Sonuga-Barke E, Mulas F, Taylor E, Laird N, Lange C, Daly M, Faraone SV (2008) Genome-wide association scan of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* **147B**: 1337-1344
- Neale BM, Medland S, Ripke S, Anney RJ, Asherson P, Buitelaar J, Franke B, Gill M, Kent L, Holmans P, Middleton F, Thapar A, Lesch KP, Faraone SV, Daly M, Nguyen TT, Schafer H, Steinhausen HC, Reif A, Renner TJ, Romanos M, Romanos J, Warnke A, Walitza S, Freitag C, Meyer J, Palmason H, Rothenberger A, Hawi Z, Sergeant J, Roeyers H, Mick E, Biederman J (2010a) Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**: 906-920
- Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP, Faraone SV, Nguyen TT, Schafer H, Holmans P, Daly M, Steinhausen HC, Freitag C, Reif A, Renner TJ, Romanos M, Romanos J, Walitza S, Warnke A, Meyer J, Palmason H, Buitelaar J, Vasquez AA, Lambregts-Rommelse N, Gill M, Anney RJ, Langely K, O'Donovan M, Williams N, Owen M, Thapar A, Kent L, Sergeant J, Roeyers H, Mick E, Biederman J, Doyle A, Smalley S, Loo S, Hakonarson H, Elia J, Todorov A, Miranda A, Mulas F, Ebstein RP, Rothenberger A, Banaschewski T, Oades RD, Sonuga-Barke E, McGough J, Nisenbaum L, Middleton F, Hu X, Nelson S (2010b) Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* **49**: 884-897
- Normand SL (1999) Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* **18**: 321-359
- Petiti DB (1994) *Meta-analysis Decision Analysis and Cost-Effectiveness Analysis*, Vol. 24: Oxford University Press.
- Pfeiffer RM, Pee D, Landi MT (2008) On combining family and case-control studies. *Genet Epidemiol* **32**: 638-646

- Platt RW, Leroux BG, Breslow N (1999) Generalized linear mixed models for meta-analysis. *Stat Med* **18**: 643-654
- Pomp ER, Van Stralen KJ, Le Cessie S, Vandenbroucke JP, Rosendaal FR, Doggen CJ (2010) Experience with multiple control groups in a large population-based case-control study on genetic and environmental risk factors. *Eur J Epidemiol*
- Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* **65**: 220-228
- Rabe-Hesketh S, Skrondal A, Pickles A (2002) Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal* **2**: 1-21.
- Rabe-Hesketh S, Skrondal A, Pickles A (2005) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* **128**: 301-323
- Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC (1987) Meta-analyses of randomized controlled trials. *N Engl J Med* **316**: 450-455
- Salanti G, Amountza G, Ntzani EE, Ioannidis JP (2005a) Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet* **13**: 840-848
- Salanti G, Higgins JP (2008) Meta-analysis of genetic association studies under different inheritance models using data reported as merged genotypes. *Stat Med* **27**: 764-777
- Salanti G, Higgins JP, Trikalinos TA, Ioannidis JP (2007) Bayesian meta-analysis and meta-regression for gene-disease associations and deviations from Hardy-Weinberg equilibrium. *Stat Med* **26**: 553-567
- Salanti G, Higgins JP, White IR (2006) Bayesian synthesis of epidemiological evidence with different combinations of exposure groups: application to a gene-gene-environment interaction. *Stat Med* **25**: 4147-4163
- Salanti G, Sanderson S, Higgins JP (2005b) Obstacles and opportunities in meta-analysis of genetic association studies. *Genet Med* **7**: 13-20

- Sasieni PD (1997) From genotypes to genes: doubling the sample size. *Biometrics* **53**: 1253-1261
- Sato Y, Suganami H, Hamada C, Yoshimura I, Sakamoto H, Yoshida T, Yoshimura K (2006) The confidence interval of allelic odds ratios under the Hardy-Weinberg disequilibrium. *J Hum Genet* **51**: 772-780
- Schaid DJ, Sommer SS (1994) Comparison of statistics for candidate-gene association studies using cases and parents. *Am J Hum Genet* **55**: 402-409
- Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* **38**: 209-213
- Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* **59**: 983-989
- Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet* **62**: 450-458
- Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**: 506-516
- Stram DO (1996) Meta-analysis of published data using a linear mixed-effects model. *Biometrics* **52**: 536-544
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB (2000) Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama* **283**: 2008-2012
- Sun F, Flanders WD, Yang Q, Khoury MJ (1999) Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. *Am J Epidemiol* **150**: 97-104
- Terwilliger JD, Ott J (1992) A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum Hered* **42**: 337-346

- Thakkinstian A, McElduff P, D'Este C, Duffy D, Attia J (2005) A method for meta-analysis of molecular association studies. *Stat Med* **24**: 1291-1306
- Thompson SG, Sharp SJ (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* **18**: 2693-2708
- Thompson SG, Turner RM, Warn DE (2001) Multilevel models for meta-analysis, and their application to absolute risk differences. *Stat Methods Med Res* **10**: 375-392
- Thompson WD, Kelsey JL, Walter SD (1982) Cost and efficiency in the choice of matched and unmatched case-control study designs. *Am J Epidemiol* **116**: 840-851
- Trikalinos TA, Olkin I (2008) A method for the meta-analysis of mutually exclusive binary outcomes. *Stat Med* **27**: 4279-4300
- Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JP (2006) Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. *Am J Epidemiol* **163**: 300-309
- Trikalinos TA, Salanti G, Zintzaras E, Ioannidis JP (2008) Meta-analysis methods. *Adv Genet* **60**: 311-334
- van Houwelingen HC, Arends LR, Stijnen T (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* **21**: 589-624
- West A, Langley K, Hamshere ML, Kent L, Craddock N, Owen MJ, O'Donovan M, Thapar A (2002) Evidence to suggest biased phenotypes in children with Attention Deficit Hyperactivity Disorder from completely ascertained trios. *Mol Psychiatry* **7**: 962-966
- Williams RL (2000) A note on robust variance estimation for cluster-correlated data. *Biometrics* **56**: 645-646
- Woolf B (1955) On estimating the relationship between blood group and disease. *Human Genet* **19**: 251-253

- Yang B, Chan RC, Jing J, Li T, Sham P, Chen RY (2007) A meta-analysis of association studies between the 10-repeat allele of a VNTR polymorphism in the 3'-UTR of dopamine transporter gene and attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* **144B**: 541-550
- Zaykin DV, Kozbur DO (2010) P-value based analysis for shared controls design in genome-wide association studies. *Genet Epidemiol* **34**: 725-738

A multivariate method for meta-analysis and comparison of diagnostic tests

Niki L. Dimou, Maria Adam and Pantelis G. Bagos^{*†}

We present here an extension of the classic bivariate random effects meta-analysis for the log-transformed sensitivity and specificity that can be applied for two or more diagnostic tests. The advantage of this method is that a closed-form expression is derived for the calculation of the within-studies covariances. The method allows the direct calculation of sensitivity and specificity, as well as, the diagnostic odds ratio, the area under curve and the parameters of the summary receiver operator's characteristic curve, along with the means for a formal comparison of these quantities for different tests. There is no need for individual patient data or the simultaneous evaluation of both diagnostic tests in all studies. The method is simple and fast; it can be extended for several diagnostic tests and can be fitted in nearly all statistical packages. The method was evaluated in simulations and applied in a meta-analysis for the comparison of anti-cyclic citrullinated peptide antibody and rheumatoid factor for discriminating patients with rheumatoid arthritis, with encouraging results. Simulations suggest that the method is robust and more powerful compared with the standard bivariate approach that ignores the correlation between tests. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: meta-analysis; diagnostic tests; SROC method

1. Introduction

Laboratory diagnostic tests are routinely used in medical research to screen for, diagnose, grade and monitor the progression of disease. The most common way to describe the performance of a diagnostic test is the 2×2 table, which gives a number of positive and negative test results among the subjects with and without the disease. Diagnostic accuracy is commonly measured using true positive rate (*TPR*) and false positive rate (*FPR*). An equivalent parameterization is in terms of Sensitivity (*Se*) and Specificity (*Sp*). Thus, *TPR* is equivalent to Sensitivity and *FPR* to 1-Specificity of a diagnostic test. A well-established method of summarizing the performance of a diagnostic test is the receiver operating characteristic (ROC) curve, which indicates the relationship between the *TPR* and *FPR* at different diagnostic thresholds [1].

Meta-analysis constitutes a particular type of research, in which a set of original studies is synthesized and the potential diversity across them is explored using specific statistical methods [2–5]. Meta-analytic techniques have been proposed for combining the results of different studies that evaluated the accuracy of a given diagnostic test [6–10]. The major difference of diagnostic studies is that a pair of estimates (*TPR* and *FPR*) is usually reported rather than a single statistic and thus specialized (i.e. bivariate) methodology needs to be utilized.

The simplest (and definitely not recommended) method for meta-analysis of diagnostic tests suggests that the numbers of true positives (*TP*), false positives (*FP*), true negatives (*TN*) and false negatives (*FN*) from each study are summed resulting in a non-stratified meta-analysis design. A separate meta-analysis of sensitivity and specificity has been proposed based on logit transformations. This approach allows for heterogeneity in sensitivity and specificity but ignores their between-studies correlation (i.e. the correlation of the random effects). Moreover, separate meta-analysis of positive and negative likelihood ratios has been proposed [11], which also ignores the correlation between these two parameters. Other authors have proposed that an estimate of the diagnostic odds ratio (*DOR*) could be derived from each study and

Department of Computer Science and Biomedical Informatics, University of Thessaly, Papasiopoulou 2-4, Lamia 35100, Greece

^{*}Correspondence to: Dr Pantelis G. Bagos, Department of Computer Science and Biomedical Informatics, University of Thessaly, Papasiopoulou 2-4, Lamia 35100, Greece.

E-mail: pbagos@compgen.org

combined in a standard meta-analysis [12], which can be performed using either fixed or random effects models. In the latter case, the between studies heterogeneity is taken into account but we ignore the fact that in each study a different threshold may have been used; additionally, calculation of a pooled estimate of sensitivity and specificity is not feasible in this case. The most commonly used approach is the summary ROC (SROC) method [7,13], which uses logit-transforms of TPR and FPR , and it is based on simple linear regression of their difference (D , which is equal to $\log DOR$) on their sum (S). A summary ROC curve can then be derived from the fitted regression line. This method incorporates the use of a different threshold in each study, but even if a weighted analysis for the sum of the parameters is performed, the measurement error in S is not taken into account.

The hierarchical SROC method [8] expresses the logit-transformed TPR and FPR in terms of two parameters (accuracy and threshold). This method allows for between studies heterogeneity by modelling the accuracy parameter as a random effects term. The model was simplified by using Empirical Bayes estimates, and the results were close to those obtained using the Bayesian analysis [9] and was extended to account for the situation where no gold standard test is available [14]. Recently, a standard bivariate random effects meta-analysis of logit-transformed TPR and FPR has been proposed. This method allows for between studies heterogeneity and for the correlation of TPR and FPR [10,15] involving a hierarchical structure where the within-study variation refers to the variation in the repeated sampling of the studies' results if they were replicated and the between-study variation refers to any variation in the studies' true underlying estimates of TPR and FPR . Methods that directly model the binomial structure data are, in general, recommended [10,16]. Harbord and co-workers, showed that the various previously mentioned multivariate methods for meta-analysis of diagnostic tests are essentially equivalent models that use a different parameterization [6]. Recently, a composite likelihood method for bivariate meta-analysis has been proposed, which overcomes the nonconvergence problem of the standard likelihood estimation methods when the number of studies is small and is more robust than the standard likelihood inference to misspecifications of the joint distributions assumptions [17]. When the disease prevalence is known, as in cohort studies, a trivariate modelling of the disease prevalence, sensitivity and specificity has been proposed [18]. This approach accounts for the dependence of sensitivity and specificity on disease prevalence, which is more obvious when a continuous trait is used as a classifier. This dependence was assessed using a Pearson-type correlation coefficient [19]. A shortcoming of this trivariate modelling approach is that it can only include cohort studies with information estimating study-specific disease prevalence. Thus, two main alternative models have been introduced, including a novel Bayesian hierarchical model for combining cohort and case-control studies and correcting partial verification bias [20] and a hybrid model, where an alternative inference procedure based on composite likelihood is used [21]. Ma and coworkers have presented a recent overview of the existing multivariate methods for meta-analysis of diagnostic studies [22].

When one wants to compare two or more diagnostic tests, the situation is complicated. Several methods, which make use of the DOR , have been proposed. The traditional method of meta-analysis for the comparison of two diagnostic tests extracts D values (i.e. the difference of logit-transformed TPR and FPR) from each study (or the estimated values using the SROC model) and then summarizes them. The diagnostic tests can then be compared using the difference of summary D s. Another approach is based on the relative OR as the relative accuracy of one test against the other and makes the assumption that the two tests were performed on 'paired' subjects within each study. If sufficient data are available (which however, rarely is the case), the calculation of the conditional relative OR ($CROR$) can be used based on the discordant results of diagnostic tests [23]. Furthermore, statistical methodology for repeated measurements has also been adopted in order to combine several studies of diagnostic tests, where each study reports on more than one test [24]. In situations when a gold standard is not available, a multivariate random effects model has been used to model simultaneously the sensitivity, the specificity and the prevalence of the disease [25]. Mixed effects models or Bayesian hierarchical models can be used for the estimation of the parameters. When the conditional independence assumption does not hold [26], a correlation among the diagnostic tests can be imposed. The aforementioned approach and the hierarchical SROC framework by Dendukuri and coworkers [14] are closely related and some of their submodels are equivalent [27]. Finally, Trikalinos and coworkers introduced a Bayesian model that incorporates the relations between TPR and FPR across two or more tests when those are applied to the same patients [28].

In this work, we present a simple yet powerful approach for performing multivariate meta-analysis of diagnostic studies. The model we propose is a direct extension of the multivariate meta-analysis method of diagnostic tests [10,15]. The key element of our approach is the calculation of the within studies covariance of the parameters necessary for the multivariate meta-analysis method [29]. Because the

model is based on the general model for multivariate meta-analysis, it has some very important features: it can easily incorporate studies reporting only one of the tests under a missing at random assumption and, thus, allows for borrowing strength from external studies [30]; it can be easily fitted using standard software used for multivariate meta-analysis [31,32]. It allows after appropriate transformations the recovery of other important metrics such as the DOR, the SROC curve and the area under the curve (AUC), and finally, it allows the direct comparison of these parameters of the different tests using formal techniques. In Section 2, we introduce the multivariate meta-analysis model for two or more diagnostic tests; we derive formulae for within-study covariance and recover the SROC method and the AUC. In Section 3, we perform a simulation study to investigate the properties of our method. In Section 4, we apply our method to data from a meta-analysis of diagnostic tests for rheumatoid arthritis. We close with discussion in Section 5.

2. Methods

2.1. The multivariate model

Let Y_i , X_{1i} and X_{2i} denote three categorical random variables with two levels (i.e. 0, 1) that are used to classify n_i individuals for study i ($i = 1, 2, \dots, k$). Usually, Y_i denotes the disease status and X_{1i} , X_{2i} the test result (positive or negative) for study i ($i = 1, 2, \dots, k$). In the general case, the data would be presented in the form of a three-dimensional ($2 \times 2 \times 2$) contingency table (Table I) where we denote the counts as n_{clpi} with $c, l, p \in \{0, 1\}$ and this table is referred to as partial contingency table. Usually, the results of two diagnostic tests are represented in the form of Table II, which, in the terminology of contingency tables, is a marginal table because for each test, the outcome of the other test is ignored. The logit transformations of $\widehat{TPR}_{ji}(\widehat{Se}_{ji})$ and $\widehat{FPR}_{ji}(1 - \widehat{Sp}_{ji})$ for test j ($j = 1, 2$) and for study i ($i = 1, 2, \dots, k$) are given by the following:

$$\hat{y}_{1i} = \text{logit}(\widehat{TPR}_{1i}) = \text{logit}(\widehat{Se}_{1i}) = \log\left(\frac{TP_{1i}}{FN_{1i}}\right) = \log\left(\frac{n_{11+i}}{n_{10+i}}\right) \quad (1)$$

$$\hat{y}_{2i} = \text{logit}(\widehat{FPR}_{1i}) = \text{logit}(1 - \widehat{Sp}_{1i}) = \log\left(\frac{FP_{1i}}{TN_{1i}}\right) = \log\left(\frac{n_{01+i}}{n_{00+i}}\right) \quad (2)$$

Table I. The table that defines the joint distribution for the association of the two diagnostic tests according to the disease status for study i ($i = 1, 2, \dots, k$).

		Y_i			
		$Y_i = 1$		$Y_i = 0$	
X_{1i}	$X_{1i} = 1$	$X_{2i} = 1$	$X_{2i} = 0$	$X_{2i} = 1$	$X_{2i} = 0$
	$X_{1i} = 0$	n_{111i}	n_{110i}	n_{011i}	n_{010i}
		n_{101i}	n_{100i}	n_{001i}	n_{000i}

We denote the counts as n_{clpi} , with c being the indicator for Y_i , l the indicator for X_{1i} and p the indicator for X_{2i} . This table is usually referred to as «partial contingency table». See also [29,39].

Table II. Classification of the findings of two diagnostic tests according to the test result (positive or negative) and the disease status for study i ($i = 1, 2, \dots, k$).

		X_{1i}		X_{2i}	
		$X_{1i} = 1$	$X_{1i} = 0$	$X_{2i} = 1$	$X_{2i} = 0$
Y_i	$Y_i = 1$	$n_{11+i} (TP_{1i})$	$n_{10+i} (FN_{1i})$	$n_{1+1i} (TP_{2i})$	$n_{1+0i} (FN_{2i})$
	$Y_i = 0$	$n_{01+i} (FP_{1i})$	$n_{00+i} (TN_{1i})$	$n_{0+1i} (FP_{2i})$	$n_{0+0i} (TN_{2i})$

We denote the counts as n_{clpi} , with c being the indicator for Y_i , l the indicator for X_{1i} and p the indicator for X_{2i} . Thus, the interior cells of the table consist of the marginals of Table I (i.e. $n_{cl+i} = n_{cl0i} + n_{cl1i}$ and $n_{c+pi} = n_{c0pi} + n_{c1pi}$). In terms of contingency tables, these two tables are named «marginal tables». See also [29,39].

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

$$\hat{y}_{3i} = \text{logit}(\widehat{TPR}_{2i}) = \text{logit}(\widehat{Se}_{2i}) = \log\left(\frac{TP_{2i}}{FN_{2i}}\right) = \log\left(\frac{n_{1+1i}}{n_{1+0i}}\right) \quad (3)$$

$$\hat{y}_{4i} = \text{logit}(\widehat{FPR}_{2i}) = \text{logit}(1 - \widehat{Sp}_{2i}) = \log\left(\frac{FP_{2i}}{TN_{2i}}\right) = \log\left(\frac{n_{0+1i}}{n_{0+0i}}\right) \quad (4)$$

with approximate variances estimated by the following:

$$s_{1i}^2 = \frac{1}{TP_{1i}} + \frac{1}{FN_{1i}} = \frac{1}{n_{11+i}} + \frac{1}{n_{10+i}} \quad (5)$$

$$s_{2i}^2 = \frac{1}{FP_{1i}} + \frac{1}{TN_{1i}} = \frac{1}{n_{01+i}} + \frac{1}{n_{00+i}} \quad (6)$$

$$s_{3i}^2 = \frac{1}{TP_{2i}} + \frac{1}{FN_{2i}} = \frac{1}{n_{1+1i}} + \frac{1}{n_{1+0i}} \quad (7)$$

$$s_{4i}^2 = \frac{1}{FP_{2i}} + \frac{1}{TN_{2i}} = \frac{1}{n_{0+1i}} + \frac{1}{n_{0+0i}} \quad (8)$$

It should be noted that in the traditional approach for bivariate meta-analysis for a single diagnostic test [6,10], a bivariate meta-analysis is performed for the pairs of outcomes (i.e. \hat{y}_{1i} , \hat{y}_{2i} or \hat{y}_{3i} , \hat{y}_{4i}). When we want to perform a joint analysis, following the general framework for multivariate meta-analysis [33,34], we will denote by \mathbf{y}_i the vector containing the four different estimates and by $\boldsymbol{\beta}$, the vector of the overall means given by the following:

$$\mathbf{y}_i = \begin{pmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \hat{y}_{3i} \\ \hat{y}_{4i} \end{pmatrix}, \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \quad (9)$$

In a multivariate random-effects setting, we assume that the four \hat{y}_i 's are distributed following a multivariate normal distribution:

$$\begin{bmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \hat{y}_{3i} \\ \hat{y}_{4i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{bmatrix}, [\mathbf{C}_i] \right\} \quad (10)$$

where \mathbf{C}_i is the within-studies covariance matrix:

$$\mathbf{C}_i = \begin{pmatrix} s_{1i}^2 & \rho_{w12i}s_{1i}s_{2i} & \rho_{w13i}s_{1i}s_{3i} & \rho_{w14i}s_{1i}s_{4i} \\ \rho_{w12i}s_{1i}s_{2i} & s_{2i}^2 & \rho_{w23i}s_{2i}s_{3i} & \rho_{w24i}s_{2i}s_{4i} \\ \rho_{w13i}s_{1i}s_{3i} & \rho_{w23i}s_{2i}s_{3i} & s_{3i}^2 & \rho_{w34i}s_{3i}s_{4i} \\ \rho_{w14i}s_{1i}s_{4i} & \rho_{w24i}s_{2i}s_{4i} & \rho_{w34i}s_{3i}s_{4i} & s_{4i}^2 \end{pmatrix} \quad (11)$$

The means $(\beta_{1i}\beta_{2i}\beta_{3i}\beta_{4i})$ are considered random terms, distributed similarly as follows:

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \beta_{3i} \\ \beta_{4i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, [\boldsymbol{\Sigma}] \right\} \quad (12)$$

where $\boldsymbol{\Sigma}$ is the between-studies covariance matrix, given by the following:

$$\Sigma = \begin{pmatrix} \tau_1^2 & \rho_{B12}\tau_1\tau_2 & \rho_{B13}\tau_1\tau_3 & \rho_{B14}\tau_1\tau_4 \\ \rho_{B12}\tau_1\tau_2 & \tau_2^2 & \rho_{B23}\tau_2\tau_3 & \rho_{B24}\tau_2\tau_4 \\ \rho_{B13}\tau_1\tau_3 & \rho_{B23}\tau_2\tau_3 & \tau_3^2 & \rho_{B34}\tau_3\tau_4 \\ \rho_{B14}\tau_1\tau_4 & \rho_{B24}\tau_2\tau_4 & \rho_{B34}\tau_3\tau_4 & \tau_4^2 \end{pmatrix} \quad (13)$$

Thus, the final marginal model on which we base the inference is as follows:

$$y_i \sim MVN(\beta, \Sigma + C_i) \quad (14)$$

or we could rewrite the marginal model described in Eqn 14 as follows:

$$\begin{bmatrix} \hat{y}_{1i} \\ \hat{y}_{2i} \\ \hat{y}_{3i} \\ \hat{y}_{4i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \tau_1^2 & \rho_{w12i}s_{1i}s_{2i} + \rho_{B12}\tau_1\tau_2 & \rho_{w13i}s_{1i}s_{3i} + \rho_{B13}\tau_1\tau_3 & \rho_{w14i}s_{1i}s_{4i} + \rho_{B14}\tau_1\tau_4 \\ \rho_{w12i}s_{1i}s_{2i} + \rho_{B12}\tau_1\tau_2 & s_{2i}^2 + \tau_2^2 & \rho_{w23i}s_{2i}s_{3i} + \rho_{B23}\tau_2\tau_3 & \rho_{w24i}s_{2i}s_{4i} + \rho_{B24}\tau_2\tau_4 \\ \rho_{w13i}s_{1i}s_{3i} + \rho_{B13}\tau_1\tau_3 & \rho_{w23i}s_{2i}s_{3i} + \rho_{B23}\tau_2\tau_3 & s_{3i}^2 + \tau_3^2 & \rho_{w34i}s_{3i}s_{4i} + \rho_{B34}\tau_3\tau_4 \\ \rho_{w14i}s_{1i}s_{4i} + \rho_{B14}\tau_1\tau_4 & \rho_{w24i}s_{2i}s_{4i} + \rho_{B24}\tau_2\tau_4 & \rho_{w34i}s_{3i}s_{4i} + \rho_{B34}\tau_3\tau_4 & s_{4i}^2 + \tau_4^2 \end{bmatrix} \right\} \quad (15)$$

The diagonal elements of C_i are the study-specific estimates of the variance that are assumed known [33] and are given in Eqns 5–9. Assuming s_i^2 terms to be known instead of to be estimated has little impact on the results, which is the basis for the traditional meta-analysis [35]. The off-diagonal elements of C_i correspond to the pairwise within-studies covariances, for instance $\rho_{w12i}s_{1i}s_{2i} = \text{cov}(\hat{y}_{1i}, \hat{y}_{2i})$ (with ρ_{w12i} being the within-studies correlation of the *TPR* and *FPR* of the first diagnostic test for study i that has to be known beforehand). Because $\text{cov}(\hat{y}_{1i}, \hat{y}_{2i}) = 0$ and $\text{cov}(\hat{y}_{3i}, \hat{y}_{4i}) = 0$ the remaining covariances need to be calculated (see next section). The diagonal elements of Σ are the between studies variances, whereas the off-diagonal elements correspond to the between studies covariances that are to be estimated from the data during the fitting procedure (for instance ρ_{B12} is the between studies correlation of the *TPR* and *FPR* of the first diagnostic test). We propose an unstructured specification for Σ or we could eliminate the parameters that have to be estimated by imposing a structured variant Σ when the number of tests increases (for instance we could set between-studies variances and correlations among tests to be equal).

The parameters estimated from the multivariate model of Eqn 14 can be directly compared for instance, using a Wald test or a bivariate test based on a chi-square distribution. Other measures such as the DOR, the AUC and the parameters of the SROC curve can be easily constructed [10] and compared, as we shall show in the next sections. A direct extension of the method for more than two diagnostic tests is straightforward.

2.2. Calculation of the within-studies covariances

The method used here is a special case of a general methodology that has been recently proposed [29]. Thus, we may proceed with some probability calculations (Appendix A) using the properties of the covariance function in order to derive the following:

$$\begin{aligned} \text{cov}(\hat{y}_{1i}, \hat{y}_{3i}) &= \frac{n_{111i}}{(n_{111i} + n_{110i})(n_{111i} + n_{101i})} - \frac{n_{110i}}{(n_{111i} + n_{110i})(n_{110i} + n_{100i})} \\ &\quad - \frac{n_{101i}}{(n_{101i} + n_{100i})(n_{111i} + n_{101i})} + \frac{n_{100i}}{(n_{101i} + n_{100i})(n_{110i} + n_{100i})} \\ &= \sum_l \sum_p (-1)^{l-p} \left(\frac{n_{1lpi}}{n_{1l+i}n_{1+pi}} \right) \end{aligned} \quad (16)$$

$$\begin{aligned} \text{cov}(\hat{y}_{2i}, \hat{y}_{4i}) &= \frac{n_{011i}}{(n_{011i} + n_{010i})(n_{011i} + n_{001i})} - \frac{n_{010i}}{(n_{011i} + n_{010i})(n_{010i} + n_{000i})} \\ &\quad - \frac{n_{001i}}{(n_{001i} + n_{000i})(n_{011i} + n_{001i})} + \frac{n_{000i}}{(n_{001i} + n_{000i})(n_{010i} + n_{000i})} \\ &= \sum_l \sum_p (-1)^{l-p} \left(\frac{n_{0lpi}}{n_{0l+i}n_{0+pi}} \right) \end{aligned} \quad (17)$$

$$\text{cov}(\hat{y}_{1i}, \hat{y}_{4i}) = \text{cov}(\hat{y}_{2i}, \hat{y}_{3i}) = 0 \quad (18)$$

Attention should be paid when the number of the participants is not equal for the two diagnostic tests for a given published study. In such a case, we can extrapolate the counts of the diagnostic test with the smaller sum to match the larger one and proceed based on the extrapolated counts.

2.3. Handling missing information

As it is apparent from Eqns 16 and 17, the exact counts of Table I need to be known for the calculation of the within studies covariances. In the simplest case, where, for each study i , the counts of Table I are known, the calculation of the covariances is straightforward. In practical applications however, we do not expect all studies to report such information.

When this information is available only for a subset of the studies, we can proceed as follows. First, for these «complete» studies, we can calculate, the *conditional ORs*, which reflect the association of the two diagnostic tests for a given disease status:

$$\widehat{OR}_{Y_i=1} = \frac{n_{111i}n_{100i}}{n_{101i}n_{110i}} \quad (19)$$

$$\widehat{OR}_{Y_i=0} = \frac{n_{011i}n_{000i}}{n_{010i}n_{001i}} \quad (20)$$

Standard random effects meta-analysis can be used in order to derive a combined estimate of the *conditional ORs* for the diseased and the non-diseased individuals, across the «complete» subset of studies. After obtaining these combined estimates, we can use them on the «incomplete studies» (those for which the interior cells of Table I are not given) along with the knowledge of the marginal counts of Table I for each disease status and derive expressions of the counts for each cell solving a system of equations (Appendix B).

If the *conditional ORs* do not differ according to disease status or if we have some prior knowledge that this is the case, the iterative proportional fitting (IPF) algorithm [36] can be used in order to obtain the interior cells, even if no study reports the full partial table (Table I). It should be noted that the IPF algorithm assumes that there is no three-way interaction in the $2 \times 2 \times 2$ contingency table. In other words, it assumes that the two ORs of Eqns 19 and 20 are equal. Thus, this approach (although not perfect) is more general than the assumption of conditional independence that is usually employed in analysis [37,38] and meta-analysis of diagnostic tests with no gold standard [14,25]

2.4. Recovering the summary receiver operating characteristic method and the area under the curve

Under the SROC approach [7,13], the logit-transformed TPR_{ji} and FPR_{ji} are used and a linear regression of their difference (D_{ji}) on their sum (S_{ji}) is performed. The parameters of the regression of D_{1i} on S_{1i} for the first test can be expressed using the parameters of the bivariate model with τ_1^2 and τ_2^2 denoting the between studies variances of y_{1i} and y_{2i} respectively, as well as $\rho_{B12}\tau_1\tau_2$ the between studies covariances of y_{1i} and y_{2i} . In particular, from Eqn 12, it turns out that the covariance of D_{1i} and S_{1i} is equal to $\tau_1^2 - \tau_2^2$, and the variance of S_{1i} is equal to $\tau_1^2 + \tau_2^2 + 2\rho_{B12}\tau_1\tau_2$. Thus, the slope is $b_1 = (\tau_1^2 - \tau_2^2)/(\tau_1^2 + \tau_2^2 + 2\rho_{B12}\tau_1\tau_2)$;

the intercept is $a_1 = \beta_1 - \beta_2 - b(\beta_1 + \beta_2)$, and the residual variance of the regression is given by $\sigma_{D_1|S_1}^2 =$

$(\tau_1^2 + \tau_2^2 - 2\rho_{B12}\tau_1\tau_2) - \frac{(\tau_1^2 - \tau_2^2)^2}{\tau_1^2 + \tau_2^2 + 2\rho_{B12}\tau_1\tau_2}$ [10]. These can easily be estimated by plugging in the model parameter estimates; the standard errors are calculated using the Delta method [39]. Similar expressions can be derived for the second diagnostic test by using the appropriate estimates of the between studies variance–covariance matrix (i.e. τ_3^2 , τ_4^2 , $\rho_{B34}\tau_3\tau_4$). It is worth-mentioning that the coefficient b_j represents the dependence of the test accuracy on threshold. If $b_j \approx 0$, then the studies are homogeneous and can be summarized by an overall DOR_j noting that $a_j = \ln(OR_j)$.

When the parameters a_j and b_j are estimated, the relationship between TPR_{ji} and FPR_{ji} can be deduced:

$$TPR_{ji} = \frac{\exp\left(\frac{a_j}{1-b_j}\right) \left(\frac{FPR_{ji}}{1-FPR_{ji}}\right)^{(1+b_j)/(1-b_j)}}{1 + \exp\left(\frac{a_j}{1-b_j}\right) \left(\frac{FPR_{ji}}{1-FPR_{ji}}\right)^{(1+b_j)/(1-b_j)}} \quad (21)$$

Equation 21 gives TPR_{ji} at any given value of FPR_{ji} and hence the entire SROC curve. While there may be some interest in identifying particular points on the curve, it is often useful to have an overall

summary measure of the behaviour of the curve. Perhaps, the most appropriate such measure is the AUC, which can be calculated as [40]:

$$AUC_j = \int_0^1 \frac{\exp\left(\frac{a_j}{1-b_j}\right) \left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)}}{1 + \exp\left(\frac{a_j}{1-b_j}\right) \left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)}} dx \quad (22)$$

Using the delta method, an approximate variance for \hat{AUC}_j is as follows:

$$\begin{aligned} \text{var}(\hat{AUC}_j) &= \left(\frac{\partial AUC_j}{\partial a_j}\right)^2 \text{var}(\hat{a}_j) + \left(\frac{\partial AUC_j}{\partial b_j}\right)^2 \text{var}(\hat{b}_j) \\ &\quad + 2\left(\frac{\partial AUC_j}{\partial a_j}\right) \left(\frac{\partial AUC_j}{\partial b_j}\right) \text{cov}(\hat{a}_j, \hat{b}_j) \end{aligned} \quad (23)$$

where:

$$\frac{\partial(AUC_j)}{\partial a_j} = \left(\frac{1}{1-b_j}\right) \exp\left(\frac{a_j}{1-b_j}\right) \int_0^1 \frac{\left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)}}{\left(1 + \left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)} \exp\left(\frac{a_j}{1-b_j}\right)\right)^2} dx \quad (24)$$

$$\frac{\partial(AUC_j)}{\partial b_j} = \left(\frac{1}{1-b_j}\right)^2 \exp\left(\frac{a_j}{1-b_j}\right) \int_0^1 \frac{\left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)} (a_j + 2\ln\left(\frac{x}{1-x}\right))}{\left(1 + \left(\frac{x}{1-x}\right)^{(1+b_j)/(1-b_j)} \exp\left(\frac{a_j}{1-b_j}\right)\right)^2} dx \quad (25)$$

In the homogeneous case, $b_j=0$, and the general expression (Eqn 22) becomes [40]:

$$AUC_{\text{hom}_j} = \frac{DOR_j}{(DOR_j - 1)^2} [(DOR_j - 1) - \ln(DOR_j)] \quad (26)$$

where AUC_{hom_j} indicates the AUC_j for homogeneous studies and $DOR_j = \exp(a_j)$. If $a_j=0$, then $AUC_{\text{hom}_j} = \frac{1}{2}$. Although only valid for homogeneous studies, Eqn 26 is a useful upper bound and provides a good approximation for AUC_j in heterogeneous studies. Using the delta method, we may also obtain:

$$SE(\hat{AUC}_{\text{hom}_j}) = \frac{1}{(DOR_j - 1)^3} [(DOR_j + 1)\ln DOR_j - 2(DOR_j - 1)] SE(\hat{DOR}_j) \quad (27)$$

Finally, a formal test for the equality of the AUC for the two tests (i.e. $AUC_1 = AUC_2$) could be derived using the properties of the AUC [40] and the Delta method. This would lead to the formulation of the null hypothesis $H_0: d = AUC_1 - AUC_2 = 0$, $H_a: d \neq 0$. The variance of \hat{d} will be equal to \mathbf{GVG}' where \mathbf{V} is the estimated variance-covariance matrix and \mathbf{G} is the derivative matrix of d with respect to the vector of estimated coefficients a_j, b_j $\left[\frac{\partial(d)}{\partial a_1} \quad \frac{\partial(d)}{\partial b_1} \quad \frac{\partial(d)}{\partial a_2} \quad \frac{\partial(d)}{\partial b_2}\right]$. Thus, \mathbf{GVG}' will be equal to

$$\begin{aligned} \mathbf{GVG}' &= \frac{\partial(d)}{\partial a_1} \left\{ \frac{\partial(d)}{\partial a_1} \text{var}(\hat{a}_1) + \frac{\partial(d)}{\partial b_1} \text{cov}(\hat{a}_1, \hat{b}_1) + \frac{\partial(d)}{\partial a_2} \text{cov}(\hat{a}_1, \hat{a}_2) + \frac{\partial(d)}{\partial b_2} \text{cov}(\hat{a}_1, \hat{b}_2) \right\} \\ &\quad + \frac{\partial(d)}{\partial b_1} \left\{ \frac{\partial(d)}{\partial a_1} \text{cov}(\hat{a}_1, \hat{b}_1) + \frac{\partial(d)}{\partial b_1} \text{var}(\hat{b}_1) + \frac{\partial(d)}{\partial a_2} \text{cov}(\hat{a}_2, \hat{b}_1) + \frac{\partial(d)}{\partial b_2} \text{cov}(\hat{b}_1, \hat{b}_2) \right\} \\ &\quad + \frac{\partial(d)}{\partial a_2} \left\{ \frac{\partial(d)}{\partial a_1} \text{cov}(\hat{a}_1, \hat{a}_2) + \frac{\partial(d)}{\partial b_1} \text{cov}(\hat{a}_2, \hat{b}_1) + \frac{\partial(d)}{\partial a_2} \text{var}(\hat{a}_2) + \frac{\partial(d)}{\partial b_2} \text{cov}(\hat{a}_2, \hat{b}_2) \right\} \\ &\quad + \frac{\partial(d)}{\partial b_2} \left\{ \frac{\partial(d)}{\partial a_1} \text{cov}(\hat{a}_1, \hat{b}_2) + \frac{\partial(d)}{\partial b_1} \text{cov}(\hat{b}_1, \hat{b}_2) + \frac{\partial(d)}{\partial a_2} \text{cov}(\hat{a}_2, \hat{b}_2) + \frac{\partial(d)}{\partial b_2} \text{var}(\hat{b}_2) \right\} \end{aligned} \quad (28)$$

with the partial derivatives for diagnostic test j ($j=1,2$) given by the following:

$$\frac{\partial(d)}{\partial a_j} = (-1)^{j+1} \left(\frac{1}{1-b_j} \right) \exp \left(\frac{a_j}{1-b_j} \right) \int_0^1 \frac{\left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)}}{\left(1 + \left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)} \exp \left(\frac{a_j}{1-b_j} \right) \right)^2} dx \quad (29)$$

$$\frac{\partial(d)}{\partial b_j} = (-1)^{j+1} \left(\frac{1}{1-b_j} \right)^2 \exp \left(\frac{a_j}{1-b_j} \right) \int_0^1 \frac{\left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)} (a_j + 2 \ln \left(\frac{x}{1-x} \right))}{\left(1 + \left(\frac{x}{1-x} \right)^{(1+b_j)/(1-b_j)} \exp \left(\frac{a_j}{1-b_j} \right) \right)^2} dx \quad (30)$$

In the homogeneous case (i.e. $b_j=0$), the general expression (Eqn 28) for estimating the variance of \hat{d} is simplified because the terms including b_j 's are cancelled out. In such a case, the variance of \hat{d}_{hom} is:

$$\text{var}(\hat{d}_{\text{hom}}) = \left(\frac{\partial d_{\text{hom}}}{\partial a_1} \right)^2 \text{var}(\hat{a}_1) + \left(\frac{\partial d_{\text{hom}}}{\partial a_2} \right)^2 \text{var}(\hat{a}_2) + 2 \left(\frac{\partial d_{\text{hom}}}{\partial a_1} \right) \left(\frac{\partial d_{\text{hom}}}{\partial a_2} \right) \text{cov}(\hat{a}_1, \hat{a}_2) \quad (31)$$

and the partial derivatives for diagnostic test j ($j=1,2$) can be calculated as follows:

$$\frac{\partial(d_{\text{hom}})}{\partial a_j} = (-1)^{j+1} \frac{\exp(a_j) [\exp(a_j)(a_j - 2) + a_j + 2]}{[\exp(a_j) - 1]^3} \quad (32)$$

3. Simulation study

To evaluate the proposed methodology, we conducted simulation studies. Section C of our supplementary material provides the simulation procedures and the results from this study. We compare the proposed multivariate method with the standard bivariate approach. Briefly, our simulations suggest that in the scenarios with 20 studies and 200 participants per study, the multivariate method produces

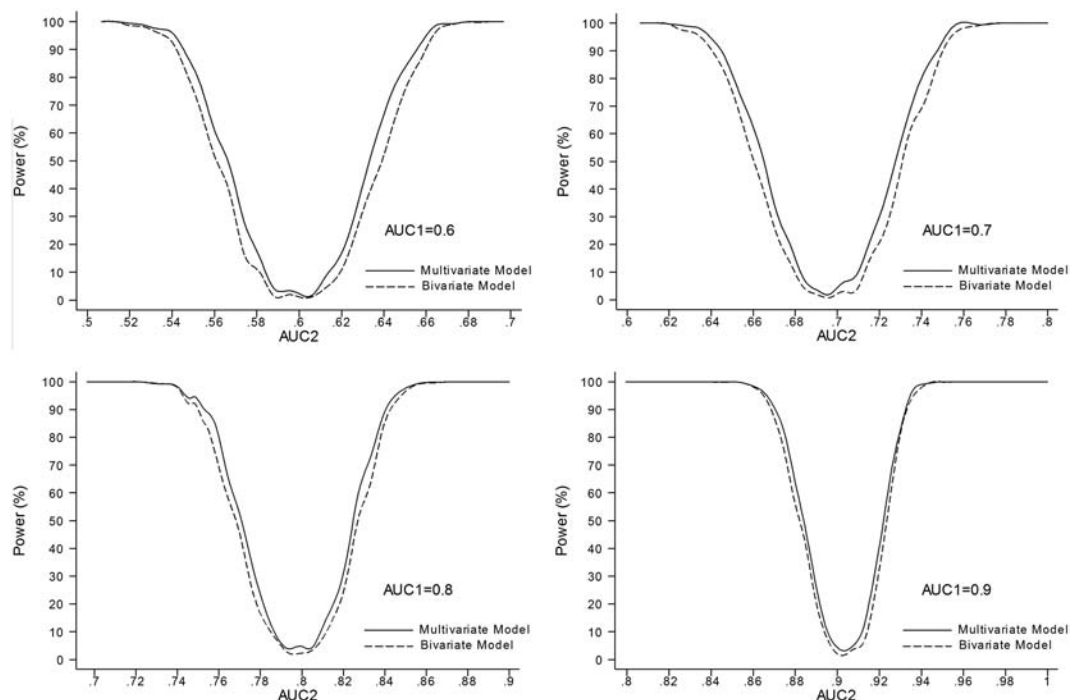


Figure 1. Power calculations to detect differences in $\widehat{AUC}_1 - \widehat{AUC}_2$ under the multivariate and bivariate method for meta-analysis of diagnostic studies. We conducted 500 replications with 20 studies in each meta-analysis. The disease prevalence was set at 50% with 200 participants in each study. For simplicity, we used the exact analytical expression (Eqn 26) in order to calculate the area under the curve (AUC) assuming that our studies are homogeneous.

unbiased estimates and has greater power to detect small differences in sensitivity or specificity preserving the nominal type I error rate when compared with the bivariate approach. Moreover, the multivariate method has always greater power (up to 12%) to detect small differences in the \widehat{AUC} (Figure 1). The gain of our method was less pronounced in the scenarios with large number of participants (>500) per study that report results of both tests (no missing data) and when there are large differences in either sensitivity or specificity.

4. Application of the method

The method was applied in the data obtained from a meta-analysis that aimed to determine whether anti-cyclic citrullinated peptide antibody (anti-CCP antibody) identifies more accurately patients with rheumatoid arthritis than does of rheumatoid factor (RF) [41] (Appendix D). As far as anti-CCP antibodies are concerned, we included studies using the CCP2 assay. In particular, a total of 50 and 29 studies provided information concerning only RF or anti-CCP2 antibody, respectively. Twenty two studies assessed both RF and anti-CCP2 antibody for diagnosing rheumatoid arthritis.

From the 22 studies, which assessed simultaneously both diagnostic tests, detailed data for the counts of Table I were available for two studies. A total of six studies reported estimates for sensitivity and specificity when both diagnostic tests were positive or when at least one of the tests is positive. In such a case, we show that the counts of Table I can be accurately reconstructed (Appendix E). One study provided information for sensitivity, specificity, positive and negative predictive value for those individuals who had a negative test result for one of the tests (the RF diagnostic test in particular). Similarly, a set of corresponding equations can be derived for this case, and the counts for Table I were also calculated (Appendix E). The total number of the participants was not equal for both diagnostic tests in one study, so we extrapolated the counts of the smaller sum to match the larger one.

Finally, detailed data for the counts of Table I were available for six studies for diseased participants only. There was lack of data for a total of seven studies that evaluated both tests, and these numbers needed to be imputed. A combined estimate of the *conditional OR* for the association of the two diagnostic tests was calculated for non-diseased and diseased individuals ($\widehat{OR}_{Y_i=0} = 6.537$ with $I^2 = 52.7\%$ and $\tau^2 = 0.7817$, $\widehat{OR}_{Y_i=1} = 12.112$ with $I^2 = 52.8\%$ and $\tau^2 = 0.2298$, respectively) using a standard random effects meta-analysis [42]. These estimates differ significantly, indicating that there is a three-way interaction in the contingency table. Consequently, our best guess is to assume that these estimates are close to the true values for the respective *ORs* in the studies that did not provide sufficient information. Then, using the equations illustrated in Appendix B, we can derive the counts of Table I. We used *mymeta* command [43] in Stata (Appendix F) under an unstructured specification for Σ . We need to emphasize that the largest portion of the code refers to the estimation of the counts for the studies that did not provide sufficient information, the calculation of the parameters of the regression of D_{ji} on S_{ji} and the calculation of the AUC_j . For comparison, the parameters of the bivariate model were also estimated.

The \widehat{TPR} of the two diagnostic tests does not differ significantly (Wald test: $z = -0.20$; p -value = 0.834, Table III). Anti-CCP2 antibody, however, had lower \widehat{FPR} and thus higher specificity than RF (Wald test: $z = 5.57$; p -value < 0.001) (Figure 2). From the 22 studies, which assessed simultaneously both tests, the within-studies correlation of the \widehat{TPR} of the two tests (i.e. $\hat{\rho}_{w13i}$) varied from 0.148–0.684 with a mean value equal to 0.463 while the correlation of the \widehat{FPR} of the two tests (i.e. $\hat{\rho}_{w24i}$) varied from 0–0.669 with

Table III. Estimates of the multivariate and bivariate model of \widehat{TPR}_{ji} and \widehat{FPR}_{ji} (logits) for the two diagnostic tests ($\hat{\beta}_1$ – $\hat{\beta}_2$ for rheumatoid factor and $\hat{\beta}_3$ – $\hat{\beta}_4$ for anti-cyclic citrullinated peptide 2 antibody).

Multivariate model				Bivariate model		
	Mean (95% CI)	z	$\hat{\tau}^2$	Mean (95% CI)	z	$\hat{\tau}^2$
$\hat{\beta}_1$	0.750 (0.538, 0.962)	6.94	0.537	0.708 (0.491, 0.925)	6.39	0.534
$\hat{\beta}_2$	−1.845 (−2.141, −1.550)	−12.23	0.977	−1.876 (−2.173, −1.579)	−12.37	0.990
$\hat{\beta}_3$	0.774 (0.534, 1.015)	6.31	0.449	0.823 (0.571, 1.075)	6.41	0.406
$\hat{\beta}_4$	−2.986 (−3.310, −2.663)	−18.08	0.519	−2.909 (−3.227, −2.590)	−17.90	0.527
$\hat{\rho}_{B12} = 0.212, \hat{\rho}_{B13} = 0.779, \hat{\rho}_{B14} = 0.386, \hat{\rho}_{B23}$ $= 0.194, \hat{\rho}_{B24} = 0.410, \hat{\rho}_{B34} = 0.487$				$\hat{\rho}_{B12} = 0.220, \hat{\rho}_{B34} = 0.477$		

TPR, true positive rate; FPR, false positive rate.

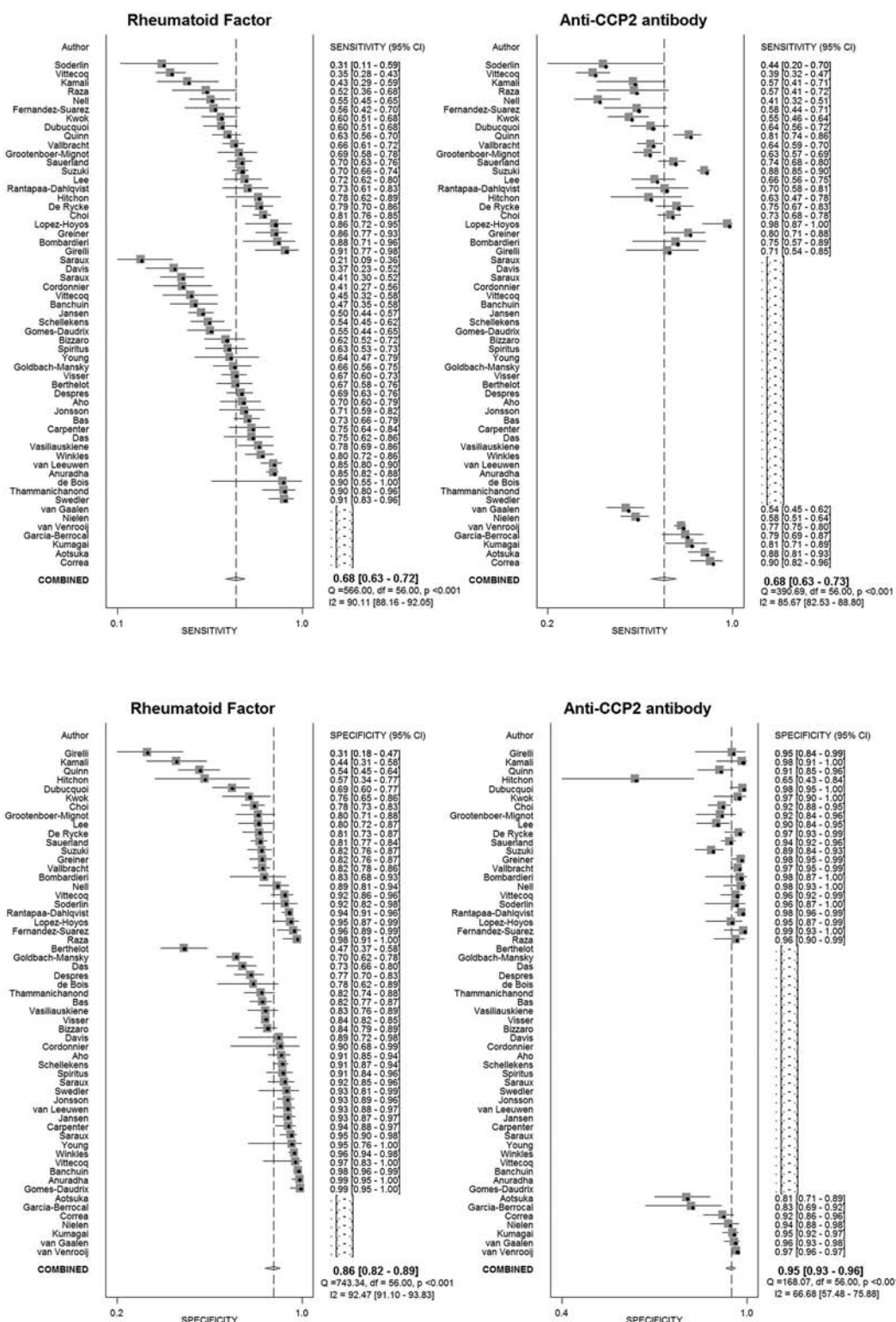


Figure 2. Forest plots of the sensitivity and specificity of the two diagnostic tests, obtained by the multivariate meta-analysis. CCP, cyclic citrullinated peptide.

a mean value equal to 0.197. The between-studies correlation of the \widehat{TPR} and \widehat{FPR} of the two tests were estimated as $\hat{\rho}_{B13} = 0.779$ and $\hat{\rho}_{B24} = 0.410$. Furthermore, using the estimates of the between studies covariances, the parameters of the regression of D_{ji} on S_{ji} were calculated as well as the SROC curve [10] with $\hat{b}_1 = -0.241$ [95% confidence interval (CI): $-0.505, 0.022$], $\hat{b}_2 = -0.049$ (95% CI: $-0.344,$

0.246), $\hat{a}_1 = 2.331$ (95% CI: 1.901, 2.761) and $\hat{a}_2 = 3.653$ (95% CI: 2.984, 4.323) (Figure 3). \widehat{AUC} for RF and Anti-CCP2 antibody was calculated as 0.825 and 0.927 respectively with these two values showing a statistically significant difference, suggesting that in overall, the anti-CCP2 antibody is better compared with RF ($z = -3.54$; $p\text{-value} < 0.001$) (Figure 4). We should also mention that model misspecification (i.e. ignoring the differences in the conditional ORs in patients and controls) results in no significant changes in the pooled estimates or in the conclusions regarding the superiority of anti-CCP2 antibody.

5. Discussion

We described here a simple yet powerful method for meta-analysis and comparison of diagnostic tests by extending the bivariate approach [10,15]. The key point of this method is that the within-studies covariances can be calculated via a closed form expression using a recently published method [29]. The method inherits all the advantages derived from the bivariate model; that is, the between-studies correlation among sensitivity and specificity is accounted for. Point estimates and confidence intervals for these two parameters can be calculated, and a construction of SROC curve is also feasible. Parameters such as the *DOR*, the *AUC* and the likelihood ratios can also be calculated and compared for the two tests. Most importantly, the method can easily incorporate (under a missing at random assumption) studies

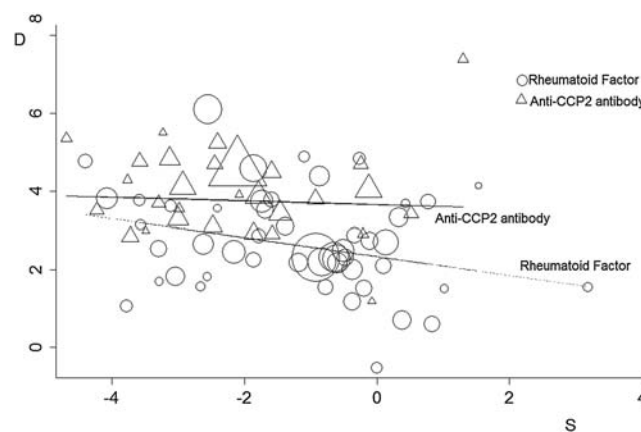


Figure 3. Plot of the sum of logit-transforms of \widehat{TPR}_{ji} and $\widehat{FPR}_{ji}(\hat{S}_{ji})$ on their difference (\widehat{D}_{ji}) of the two diagnostic tests under the summary receiver operating characteristic method. The regression coefficients are obtained by the multivariate meta-analysis model as described in the text. CCP, cyclic citrullinated peptide; TPR, true positive rate; FPR, false positive rate.

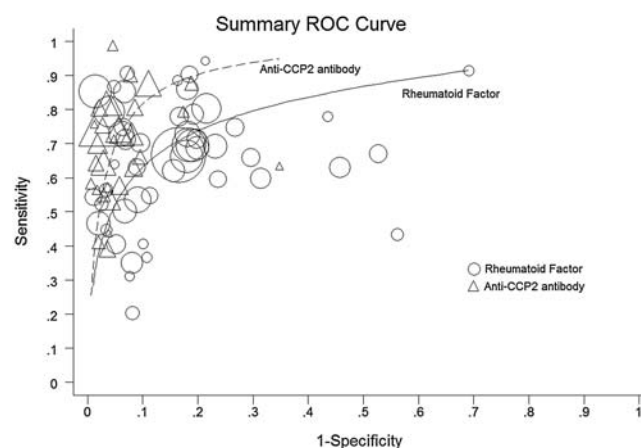


Figure 4. Summary receiver operating characteristic (SROC) curve obtained from the multivariate model of the two diagnostic tests ($AUC = 0.825$ and 0.927 for rheumatoid factor (RF) and anti-cyclic citrullinated peptide (CCP) 2, respectively). *AUC*, area under the curve.

reporting only one of the tests and thus allows borrowing strength from external studies [30]; it can be easily fitted using standard software used for multivariate meta-analysis [31,32], and it allows for the direct formal comparison of the different tests using formal techniques (Wald test). As a standard multivariate model, it can also incorporate covariates, which may be correlated with the disease process, the diagnostic testing procedure or account for strata with different disease prevalence. This approach thus yields a random-effects meta-regression [44], where various study level characteristics can be used as linear predictors, in order to estimate the extent to which these covariates explain the observed heterogeneity.

Our simulations reveal that in the realistic scenario that we evaluated, our method is more likely to produce unbiased estimates and to preserve the nominal type-I error rate, whereas it has greater power to detect small differences in sensitivity or specificity. When the meta-analysis includes a large number of studies, each having a large number of participants, that report the results of both tests and when these tests have large differences in either sensitivity or specificity, we expect that the gain of using the method will be negligible. However, we do not expect in any case the method to be worse compared with standard bivariate meta-analysis. Moreover, simulations under a misspecified model (i.e. assuming that there is no three-way interaction and using the IPF algorithm) showed that the method is still better compared with the standard bivariate approach. These conclusions are in agreement with previous simulation results concerning the superiority of bivariate meta-analysis over univariate one but also have theoretical justification [45–47]. Algebraic calculations reveal that for \widehat{TPR} and $\widehat{FPR} > 0.5$, the covariance is always positive, and thus, we expect the variance of the differences $(\widehat{\beta}_1 - \widehat{\beta}_3, \widehat{\beta}_2 - \widehat{\beta}_4)$ to be always smaller when we take this covariance into account. However, the covariance is rather small compared with the variance, and its contribution decreases with increased sample size or increased between studies heterogeneity. All these taken together, explain the fact that we do not see remarkably better results (at least for the range of values that we are interested in) but also provide the assurance that the multivariate method will be at least as accurate as the bivariate one, under any circumstances. The same rationale holds also for the power to detect differences in the \widehat{AUC} . When the real difference between the tests is large, both methods have 100% power to detect it, but the multivariate method has always greater power (up to 12%) to detect small differences that may be important in clinical practice.

Some other methods for comparing in a meta-analysis of two diagnostic tests have been proposed in the literature. As far as diagnostic *ORs* are concerned [12], the simple comparison of their log-transformations or those derived using SROC method ignores the correlation of the two diagnostic tests, in cases where the same patients are measured. When the tests are applied to different populations (which does not usually happen), the test is valid, but still, it involves only the *DOR* and not sensitivity or specificity. The CROR method [23] is simple but requires individual patient data for both tests which are not always available. A very important side-effect of our work that needs to be emphasized is that the methods for reconstructing the tables that we presented in the Appendix E can be directly used with the CROR method. Thus, a method originally developed for use mainly with individual data can be used using summary data collected from the literature. The particular approach, after the imputation procedure, has the advantage of being easy to use and requires no more than a standard software for univariate meta-analysis. As a proof of principle, from the original data, nine studies reported data sufficient to calculate CROR. Combining these studies, yield an estimate of CROR of 0.144 (95% C.I.: 0.066, 0.317) for the RF over the anti-CCP2 antibody. However, using the equations of the Appendix E, we were able to reconstruct the data for an additional 13 studies, and this enables us to calculate a more precise estimate of CROR of 0.152 (95% C.I.: 0.090, 0.256), suggesting a superiority of the anti-CCP2 antibody over RF. Nevertheless, this method has the additional disadvantage that it cannot incorporate studies that report only one test, whereas a direct comparison of sensitivity, specificity or AUC is not feasible and neither is it the construction of a summary SROC curve. We also need to comment on the Bayesian approach proposed by Trikalinos and coworkers [28], an approach that shares many common features with this work. The main difference besides the Bayesian formalism is the fact that Trikalinos and coworkers model the correlation of the two tests as a random parameter, whereas in this work, we use directly the actual study-specific correlations. When all studies report the cross-tabulation table, we expect that the two approaches will yield nearly identical results. Moreover, the simulations we conducted clearly showed that the model is robust even under severe misspecification. Thus, the main advantage of our approach is the simplicity and the fact that it can be fitted with standard software, without however losing in accuracy. One additional advantage as we already noted is that it inherits all the advantages derived from the standard bivariate model and additionally we made a lot of effort to derive tests for the parameters of the SROC curve and AUC.

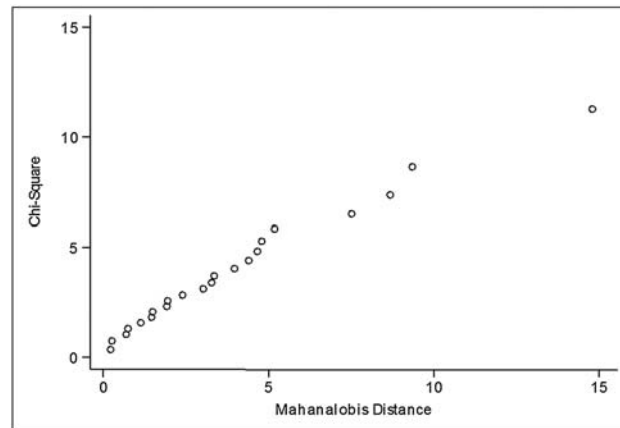


Figure 5. Plot check for multivariate normality of $\text{logit}(\widehat{TPR}_{ji})$ and $\text{logit}(\widehat{FPR}_{ji})$ for the 22 studies, which assessed simultaneously both diagnostic tests for discriminating patients with rheumatoid arthritis. The estimates are approximately normal as the graph approximates a 45° line.

We have to emphasize that our method is designed to be used for comparison of diagnostic tests in situations when a gold standard exists. On the contrary, some other multivariate random effects methods have been proposed to model simultaneously the sensitivity and the specificity of the two tests along with the prevalence of the disease, in situations when a gold standard is not available [14,25]. These methods are based on the conditional independence assumption [26,37,38], which is rather strong, but it is not needed in our approach. When this assumption is not met, a correlation among the diagnostic tests can be imposed, but nevertheless, our approach is more direct because it calculates the correlation directly from the joint distribution of the two tests and the disease status. Assumptions about the missing data mechanism become crucial as the amount of missing data increases when multiple diagnostic tests are used and the missing at random assumption may hold only in special situations, excluding of course cases where test ordering depends on health status. Because we used the classical model for multivariate meta-analysis, the method can handle data missing completely at random or missing at random [32,47]. In general, simple diagnostics for the case of informative missing mechanisms can be constructed, for instance by performing subgroup analyses for the studies that report both tests versus studies that report only one of the tests.

Our method uses summary data available on the published reports. The statistical theory behind the proposed methodology is very simple and is based on standard large sample approximations and normality assumptions [39,48] that are in everyday use by researchers performing meta-analyses of published data. In our illustrative example with the two diagnostic tests used for discriminating patients with Rheumatoid Arthritis, Henze–Zirkler’s test [49] provided by the `multnorm` command in Stata [50,51], provided no indication that the multivariate normality assumption is violated (Figure 5). In some extreme cases, some practical issues may arise using the approximate likelihood inference [52,53]. Theoretically, the method is expected to fail when the sample size is very small or when there are several studies with small, or even zero, cell counts. In such situations, relying on the usual continuity correction that consists of adding $\frac{1}{2}$ to the cell counts is the only option that additionally seems to perform quite well [54]. Another option could be to use multinomial distributions but to the authors’ knowledge of multinomial likelihood for the particular problem cannot be fit using `xtmelogit` routine in Stata or `lmer` in R. Optimizing the likelihood for performing multivariate meta-analysis using the multinomial likelihood involves calculating complicated integrals numerically, which is outside the scope of the present work.

Overall, we presented a simple and powerful method for performing meta-analysis and comparison of diagnostic tests. The method can be fitted in nearly all statistical packages. In Appendix F, we give illustrative code in Stata, and we hope that this method will be widely used in future studies.

Acknowledgements

The authors would like to thank the two anonymous reviewers and the associate editor whose constructive criticism helped in improving the quality of the manuscript. The authors are financially supported by the project ‘Integration of Data from Multiple Sources’ (IntDaMus), which is implemented under the ‘ARISTEIA II’ Action

of the 'OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING' and is co-funded by the European Social Fund (ESF) and National Resources.

References

1. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology* 2010; **5**(9):1315–1316.
2. Glass G. Primary, secondary and meta-analysis of research. *Educational Research* 1976; **5**:3–8.
3. Greenland S. In *Meta-Analysis, in Modern Epidemiology*, Rothman KJ, Greenland S (eds). Lippincott Williams & Wilkins, 1998; 643–673.
4. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine* 1999; **18**(3):321–359.
5. Petiti DB. Monographs in epidemiology and biostatistics. In *Meta-Analysis Decision Analysis and Cost-Effectiveness Analysis*, Vol. **24**. Oxford University Press, 1994.
6. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007; **8**(2):239–251.
7. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in Medicine* 1993; **12**(14):1293–1316.
8. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine* 2001; **20**(19):2865–2884.
9. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology* 2004; **57**(9):925–932.
10. Arends LR, Hamza TH, van Houwelingen JC, Heijtenbroek-Kal MH, Hunink MG, Stijnen T. Bivariate random effects meta-analysis of ROC curves. *Medical Decision Making* 2008; **28**(5):621–638.
11. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; **323**(7305):157–162.
12. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology* 2003; **56**(11):1129–1135.
13. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Medical Decision Making* 1993; **13**(4):313–321.
14. Dendukuri N, Schiller I, Joseph L, Pai M. Bayesian meta-analysis of the accuracy of a test for tuberculous pleuritis in the absence of a gold standard reference. *Biometrics* 2012; **68**(4):1285–1293.
15. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; **58**(10):982–990.
16. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology* 2006; **59**(12):1331–1332.
17. Chen Y, Liu Y, Ning J, Nie L, Zhu H, Chu H. A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews. *Statistical Methods in Medical Research* 2014. [Epub ahead of print].
18. Chu H, Nie L, Cole SR, Poole C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: alternative parameterizations and model selection. *Statistics in Medicine* 2009; **28**(18):2384–2399.
19. Li J, Fine JP. Assessing the dependence of sensitivity and specificity on prevalence in meta-analysis. *Biostatistics* 2011; **12**(4):710–722.
20. Ma X, Chen Y, Cole SR, Chu H. A hybrid Bayesian hierarchical model combining cohort and case-control studies for meta-analysis of diagnostic tests: accounting for partial verification bias. *Statistical Methods in Medical Research* 2014. [Epub ahead of print].
21. Chen Y, Liu Y, Ning J, Cormier J, Chu H. A hybrid model for combining case-control and cohort studies in systematic reviews of diagnostic tests. *Journal of the Royal Statistical Society: Series c: Applied Statistics* 2015; **64**(3):469–489.
22. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. *Statistical Methods in Medical Research* 2013. [Epub ahead of print].
23. Suzuki S, Moro-oka T, Choudhry NK. The conditional relative odds ratio provided less biased results for comparing diagnostic test accuracy in meta-analyses. *Journal of Clinical Epidemiology* 2004; **57**(5):461–469.
24. Siadat MS, Philbrick JT, Heim SW, Schectman JM. Repeated-measures modeling improved comparison of diagnostic tests in meta-analysis of dependent studies. *Journal of Clinical Epidemiology* 2004; **57**(7):698–711.
25. Chu H, Chen S, Louis TA. Random effects models in a meta-analysis of the accuracy of two diagnostic tests without a gold standard. *Journal of the American Statistical Association* 2009; **104**(486):512–523.
26. Gardner IA, Stryhn H, Lind P, Collins MT. Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Preventive Veterinary Medicine* 2000; **45**(1–2):107–122.
27. Liu Y, Chen Y, Chu H. A unification of models for meta-analysis of diagnostic accuracy studies without a gold standard. *Biometrics* 2015; **71**(2):538–547.
28. Trikalinos TA, Hoaglin DC, Small KM, Terrin N, Schmid CH. Methods for the joint meta-analysis of multiple tests. *Research Synthesis Methods* 2014; **5**(4):294–312.
29. Bagos PG. On the covariance of two correlated log-odds ratios. *Statistics in Medicine* 2012; **31**(14):1418–1431.
30. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996; **15**(24):2733–2749.
31. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Statistical Methods in Medical Research* 2012. [Epub ahead of print].
32. Jackson D, Riley R, White IR. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine* 2011; **30**(20):2481–2498.

33. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine* 2002; **21**(4):589–624.
34. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F, Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine* 1998; **17**(22):2537–2550.
35. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**(6):619–629.
36. Deming WaSF. On least square adjustment of sampled frequency tables when the expected marginal totals are known. *The Annals of Mathematical Statistics* 1940; **6**:427–444.
37. Enoe C, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive Veterinary Medicine* 2000; **45**(1-2):61–81.
38. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 1998; **7**(4):354–370.
39. Agresti A. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. In *Categorical Data Analysis* (2nd edn). John Wiley & Sons: New York, 2002.
40. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine* 2002; **21**(9):1237–1256.
41. Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, Saigo K, Morinobu A, Koshiba M, Kuntz KM, Kamae I, Kumagai S. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Annals of Internal Medicine* 2007; **146**(11):797–808.
42. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.
43. White IR. Multivariate random-effects meta-analysis. *Stata Journal* 2009; **9**:40–56.
44. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**(20):2693–2708.
45. Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series a (Statistics in Society)* 2009; **172**(4):789–811.
46. Riley RD, Abrams KR, Lambert PC, Sutton AJ, Thompson JR. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine* 2007; **26**(1):78–97.
47. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology* 2007; **7**:3.
48. Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford University Press, 1993.
49. Henze N, Zirkler B. A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods* 1990; **19**(10):3595–3617.
50. Stevens J. In *Applied Multivariate Statistics for the Social Sciences*, Hillsdale P (ed). : L Erlbaum Assoc., 1986.
51. Thompson B. Multinor: a fortran program that assists in evaluating multivariate normality. *Educ Psychol Measurement* 1990; **50**:845–848.
52. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. *Medical Decision Making* 2008; **28**(5):639–649.
53. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology* 2008; **61**(1):41–51.
54. Agresti A. On logit confidence intervals for the odds ratio with small samples. *Biometrics* 1999; **55**(2):597–602.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Methylene tetrahydrofolate reductase gene polymorphisms and their association with methotrexate toxicity: a meta-analysis

Kalliopi P. Spyridopoulou^a, Niki L. Dimou^b, Stavros J. Hamodrakas^a and Pantelis G. Bagos^b

Objective A systematic review and a meta-analysis were conducted, to investigate the possible association of methylene tetrahydrofolate reductase (MTHFR) gene polymorphisms with adverse effects related to methotrexate (MTX).

Methods A systematic literature search in PubMed retrieved a total of 44 studies (42 unique articles). Two polymorphisms were included in the meta-analysis: C677T and A1298C. Random effect models were used in the analysis. Odds ratios along with their 95% confidence intervals were computed to compare the distribution of alleles and genotypes between cases and controls.

Results The analysis highlighted a significant association of C677T polymorphism with overall MTX toxicity, hepatotoxicity, hematological toxicity, and neurotoxicity. It also revealed an association with MTX toxicity in patients with rheumatoid arthritis. In contrast, a protective effect of C677T MTHFR polymorphism on acute graft-versus-host disease and on patients treated with hematopoietic cell transplantation was found. As for the A1298C polymorphism, a statistically significant

association with overall MTX toxicity and a protective role of the polymorphism in rheumatoid arthritis patients was detected.

Conclusion These results indicate the association of MTHFR polymorphisms with MTX toxicity. However, further studies are needed to reveal the underlying biological mechanism of the association. *Pharmacogenetics and Genomics* 00:000–000 © 2011 Wolters Kluwer Health | Lippincott Williams & Wilkins.

Pharmacogenetics and Genomics 2011, 00:000–000

Keywords: gene polymorphisms, meta-analysis, methotrexate toxicity, methylene tetrahydrofolate reductase

^aDepartment of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens and ^bDepartment of Computer Science and Biomedical Informatics, University of Central Greece, Papasiopoulou, Lamia, Greece

Correspondence to Dr Pantelis G. Bagos, Department of Computer Science and Biomedical Informatics, University of Central Greece, Papasiopoulou 2-4, Lamia 35100, Greece
Tel: +30 223 106 6914; fax: +30 223 106 6915 e-mail: pbagos@ucg.gr

Received 20 November 2011 Accepted 7 October 2011

Introduction

Methotrexate (MTX) is a widely used drug in the treatment of a variety of diseases such as cancer, acute lymphoblastic leukemia or lymphoma, hematopoietic cell transplantation, rheumatoid arthritis, psoriasis, inflammatory bowel disease, and other diseases [1,2]. As a structural analogue of folic acid, MTX reversibly inhibits dihydrofolate reductase, an enzyme that participates in the tetrahydrofolate synthesis [3–5]. MTX is also known to inhibit other enzymes in the folic acid cycle, such as thymidylate synthase and methylene tetrahydrofolate reductase (MTHFR) [6,7].

As an inhibitor of DNA and protein synthesis, MTX interferes with and inhibits the growth of normal cells and tissues besides the target abnormal cells. It can, therefore, cause a significant number of side effects that may lead to therapy discontinuation. The most common side effects include oral mucositis, nausea, diarrhoea, and alopecia. Furthermore, it may also cause serious toxicities in liver, bone marrow, pulmonary, and the central nervous system [8,9]. A special kind of toxicity of MTX in

patients treated with hematopoietic cell transplantation is acute graft-versus-host disease (GVHD).

MTHFR catalyses the conversion of 5,10-methylene-tetrahydrofolate to 5-methyltetrahydrofolate in folic acid cycle, and thus, it plays an essential role in nucleotide biosynthesis. In contrast, 5-methyltetrahydrofolate acts as a methyl donor for the remethylation of homocysteine to methionine. There are two common single-nucleotide polymorphisms (SNPs) of MTHFR: C677T and A1298C. The C677T polymorphism changes alanine to valine at the 222 position of the amino acid sequence. The enzyme activity of the TT homozygote is decreased compared with that of the CC homozygote [10,11]. The A1298C polymorphism is responsible for the substitution of glutamate by alanine at the 429 position of the protein [12,13]. This polymorphism has been associated with a reduced MTHFR activity *in vitro*. The reduced enzyme activity leads to high blood homocysteine levels and this disorder is connected to coronary and peripheral artery diseases [14].

Several research groups have investigated the relevance between MTHFR SNPs and MTX toxicity with conflicting results. Moreover, many individual studies have inadequate power to reveal mild gene effects because of their small sample sizes. Therefore, we conducted a meta-analysis along with many subgroup analyses to summarize the results of separate studies concerning this scientific topic and to explore the potential impact of between-studies heterogeneity on the summary estimates.

Methods

Retrieval of published studies

A systematic, computerized literature search (using PubMed, Scopus, and Google Scholar) for papers published before December 2008 was conducted by two different reviewers. In addition, the two reviewers independently examined the retrieved articles to extract the information needed using a data collection form. The following keywords were utilized: ‘Methotrexate’, combined with ‘MTHFR’ or ‘Methylene Tetrahydrofolate Reductase’ and ‘polymorphism’ or ‘variant’. After an initial screening of titles and abstracts, only relevant articles remained. The qualifying full-text publications and their reference lists were carefully examined to ascertain their relevance to the topic of interest. Furthermore, to eliminate bias resulting from the ‘grey literature’, studies published in conference proceedings or as short abstracts were also considered. Manuscripts in languages other than English were also considered for review, to avoid the local literature bias. However, only articles published in English were finally retrieved and found to be relevant to the topic of interest.

Inclusion and exclusion criteria

Population-based studies were included in the meta-analysis if (a) they examined the association of MTHFR polymorphisms with MTX toxicity and (b) they provided adequate information (i.e., the genotype or allele counts) for estimating a measure of association [odds Ratio (OR)]. To limit selection bias, language or quality restrictions were avoided [15,16]. Furthermore, to eliminate bias resulting from the ‘grey literature’ [17], studies published in conference proceedings or as short abstracts were also considered. Fig. 1 demonstrates the paper elimination process of the meta-analysis.

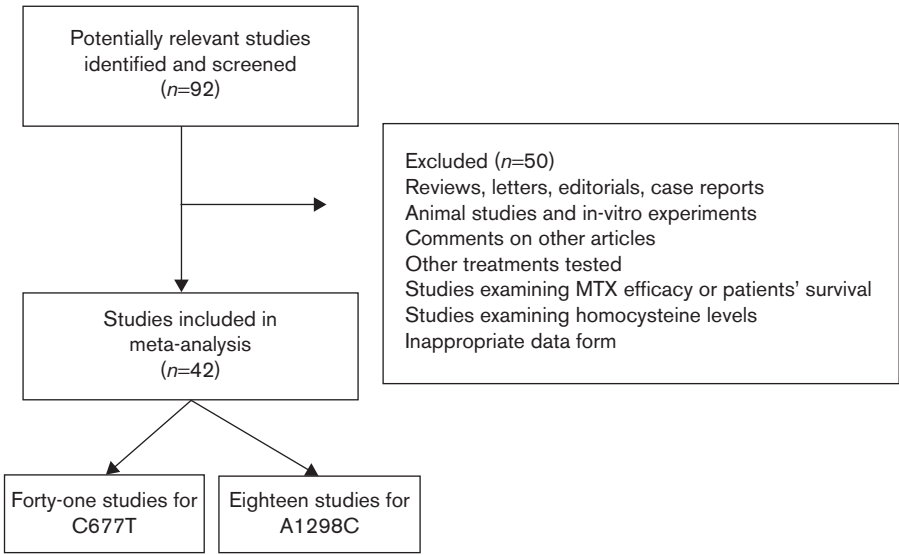
Data extraction

The following data were recorded from each study: (a) first author’s name, journal, year of publication, ethnicity of participants; (b) number of eligible genotyped participants; (c) the polymorphism under investigation and the disease form; (d) the type of toxicities of MTX and the folate supplementation of patients; (e) the distribution of genotypes and alleles in all the participants.

Statistical analysis

The OR was used to compare the distribution of alleles and genotypes between participants with and without MTX toxicity. Initially, we used a recently proposed multivariate random-effects method of meta-analysis for comparing the effect of other genotypes (TT and CT for C677T and CC and CA for A1298C) against the wild-type genotype CC and AA, respectively, for C677T and A1298C [18]. We also used a recently published methodology for meta-analysis of gene-disease association studies that is based on random-effects logistic regression [19].

Fig. 1



Flow chart showing the meta-analysis paper selection process and exclusion reasons. MTX, methotrexate.

Both methods avoid multiple comparisons and test the genetic model of inheritance directly, whereas the logistic regression method can operate even in the presence of zero cell counts. Their disadvantage is that they require the full genotype distribution (i.e., the TT, CT, and CC counts) from a given study, which may not be available in all cases. For these reasons, we also used the traditional univariate approach based on computing the ORs for three different genetic contrasts (T vs. C allele, TT vs. CT + CC genotypes, and TT + CT vs. CC genotypes for C677T; C vs. A allele, CC vs. AC + AA genotypes, CC + AC vs. AA genotypes for A1298C). The between-studies heterogeneity was evaluated using the Cochran's Q statistic [20] and the inconsistency index I^2 [21]. The combined ORs and their 95% confidence intervals (CIs) were estimated applying the random-effects methodology [22], which is appropriate in case of significant heterogeneity.

In secondary analyses, the summary ORs were stratified by the racial descent of the participants and the Hardy-Weinberg equilibrium (HWE) status of the individuals without MTX toxicity. Deviations from the HWE were calculated by the χ^2 method. Besides the analysis of the overall toxicity, we performed subgroup meta-analyses for six types of toxicities and three different forms of diseases. The analysis for the overall toxicity (which was defined as the appearance of any type of adverse effect during MTX therapy, mild or severe, hematological or nonhematological) comprised of two categories including (a) studies with at least one type of MTX toxicity and (b) studies with patients who presented at least two different MTX adverse effects. A separate subanalysis according to the folic acid supplementation of patients was performed in order to investigate the role and a probable interaction of the folate supplement with MTX toxicity. In accordance to this, the studies were divided into three categories: (a) those in which none of the patients was supplied with folic acid, (b) those in which some of the patients were supplied with folic acid, and (c) those in which all the patients were treated with folic acid.

The presence of publication-related or other small-study-related bias was assessed using the rank correlation method of Begg and Mazumdar [23], the Egger *et al.*'s regression method [24] and its random-effects analogue [25]. Influential studies were detected by checking the effect of removing an individual study on the overall significance of the estimate or on the heterogeneity statistic. Cumulative meta-analysis [26,27] was applied to identify any trend of the effect estimate with accruing evidence [28]. For the detection of the time-trend, we used the standard cumulative meta-analysis approach, which consists of inspecting the plot visually and a recently proposed regression-based method [29].

For all analyses, the statistical package STATA 10 (Stata Corporation, College Station, Texas, USA) was used.

Results with a *P* value of less than 0.05 were considered to be statistically significant.

Results

In total, 44 studies concerning the association of MTHFR gene polymorphisms with MTX toxicity were populated in our meta-analysis (Table 1). In particular, 26 studies addressed the C677T polymorphism, three studies examined the A1298C polymorphism, and the remaining 15 studies provided data for both polymorphisms. One of the studies [53] included information about three independent populations, and thus, it was counted as three studies in the meta-analysis.

MTHFR C677T polymorphism (rs1801133)

Overall toxicity: more than two adverse effects

In total, eighteen studies [31–35,37,42–44,47,49,51,53,56,57,60,61,69] examined the association of C677T polymorphism with the appearance of at least two adverse effects of MTX. Among them, six studies were conducted in Caucasian populations, seven in Asian populations, and five studies in other ethnicities. Seven studies were found to be in HWE [47,49,51,56,57,60,61]. The logistic regression model was applied in eight studies [47,49,51,53,56,57,60,61] and provided a statistically significant effect in the TT versus CC contrast (OR: 1.672, 95% CI: 1.009, 2.769). The multivariate methodology (included eight studies [47,49,51,53,56,57,60,61]) also yielded a very similar estimate in TT versus CC genotype (OR: 1.670, 95% CI: 1.000, 2.770). The λ value was 0.39, which indicates a possible codominant model of inheritance.

The conventional meta-analysis was conducted in 18 studies [31–35,37,42–44,47,49,51,53,56,57,60,61,69]. The per-allele OR of the T variant was 1.217 (95% CI: 0.889, 1.665). A much higher, but marginally not significant effect was detected in the TT versus CT + CC contrast (OR: 1.566, 95% CI: 0.949, 2.582). However, the analysis revealed a significant result in the TT versus CT + CC comparison in studies concerning Asian populations (OR: 1.655, 95% CI: 1.029, 2.660) and in studies in HWE (OR: 1.604, 95% CI: 1.063, 2.419). Heterogeneity was not observed in most calculations, and there was no evidence for publication bias. Sensitivity analysis revealed a single influential study [49]. After removing this study from the analysis, the OR value decreased essentially (T vs. C OR: 1.043, 95% CI: 0.856, 1.269). The cumulative analysis showed a decrease in the effect over years (data not shown).

Overall toxicity: more than one adverse effect

Forty-one articles investigated the association of C677T polymorphism and the appearance [30–68] of at least one MTX adverse event. Fifteen studies enrolled Caucasian populations, 12 studies enrolled Asian populations, and 14 were applied in other ethnicities. We found 18 studies

Table 1 Characteristics of studies included in the meta-analysis

Author, year	Descent	No. of cases/controls	Toxicity genotypes	Nontoxicity genotypes	Polymorphism	Disease	Patient's characteristics	Treatment	Toxicities ^a	Folate status ^b
Van Ede <i>et al.</i> , [30]	Caucasian	30/206	CC: 10 CT+TT: 30	CC: 112 CT+TT: 94	C677T	RA	Of the 236 patients evaluated, 171 were females, 190 were positive for rheumatoid factor (RF), and the mean age was 55 years.	Patients were randomly assigned to receive either MTX alone, MTX plus folic acid (1.0 mg/day), or MTX plus folinic acid (2.5 mg/week). The initial MTX dosage was 7.5 mg/week and could be increased to a maximum of 25 mg/week if necessary. The dosages of folic and folinic acid were doubled when the MTX dosage reached \$15 mg/week.	Hepatic	Not all
Bohanec Grabar <i>et al.</i> , [31]	Caucasian	113/37	CC: 60 CT+TT: 53 AA+ CA: 102 CC: 11	CC: 14 CT+TT: 23 AA+ CA: 30 CC: 7	C677T, A1298C	RA	The study population consisted of 213 unrelated adult RA patients, 25 males (11.7%) and 188 females (88.3%), median age was 61 years (range: 51–69 years).	One-hundred fifty-seven out of 213 patients were treated with MTX with a median dose of 10 mg/week (range: 10.0–12.5 mg/week). The remaining 56 of 213 patients were treated with a median dose of 10 mg of MTX per week (range: 10.0–14.4 mg) for a median time of 24 months (range: 8–48 months), but discontinued MTX before entry into the study.	Overall	Not all
Aggarwal <i>et al.</i> , [32]	Other	30/120	CC: 19 CT+TT: 11	CC: 68 CT+TT: 52	C677T	RA	Of the 150 patients evaluated, 130 were females, 84.3% were seropositive for rheumatoid factor (RF), and the mean age was 42.9 ± 11.1 years with a mean disease duration of 7.65 ± 5.2 years.	Mean duration of MTX treatment was 26.1 ± 20.6 months and the mean MTX dose was 11.1 ± 3.1 mg/week.	Hepatic, hematological, GI, pulmonary	All
Zeng, <i>et al.</i> , [33]	Asian	71/122	CC: 39 CT+TT: 32	CC: 60 CT+TT: 62	C677T	RA	One-hundred ninety-three Chinese Han consecutive outpatients with inflammatory arthritis who received low-dose MTX treatment (5–15) mg/week.	Low-dose MTX treatment [5–15 mg/week, mean (SD) 9.4 (1.9) mg/week].	GI, nervous system	None
Urano <i>et al.</i> , [34]	Asian	21/81	CC: 3 CT+TT: 18 AA: 15 AC+CC: 6	CC: 32 CT+TT: 49 AA: 56 AC+CC: 25	C677T, A1298C	RA	One-hundred and six RA patients (all Japanese nationals; 93 females, 13 males) who had been administered MTX were selected.	MTX treatment had been started at initial doses of 2.5 mg–7.5 mg per week.	Overall	Not all
Taniguchi <i>et al.</i> , [35]	Asian	43/113	CC: 10 CT+TT: 33 AA: 28 AC+CC: 15	CC: 56 CT+TT: 57 AA: 74 AC+CC: 39	C677T, A1298C	RA	A total of 384 patients with RA were randomly selected from the outpatient clinic at the Institute of Rheumatology, Tokyo Women's Medical University.	MTX treatment was started at initial doses of 4–6 mg/week in almost all patients.	Overall	Not all
Krull <i>et al.</i> , [36]	Other	11/37	CC: 6 CT+TT: 5 AA: 3 AC+CC: 8	CC: 18 CT+TT: 19 AA: 22 AC+CC: 15	C677T, A1298C	ALL	Mean age at diagnosis for the genotyped sample was 4.1 years (range: 0.7–9.1 years), and median time off therapy was 3.8 years (range: 1.5–9.3 years) for the study population. Thirty-two of the patients (67%) were males.	Patients were treated with various Pediatric Oncology Group protocols (8617, 8698, 8699, 9005, 9006, 9297, 9405, 9406, 9411, 9605).	Nervous system	None
Schmeling <i>et al.</i> , [37]	Caucasian	26/32	CC: 9 CT+TT: 17	CC: 20 CT+TT: 12	C677T	Arthritis	Fifty-eight patients (43 girls, 15 boys). Twenty-one with seronegative polyarticular onset, 14 with oligoarticular onset, seven with enthesitis-related arthritis, four with psoriatic arthritis, two systemic onset, and two unclassified arthritis.	The treatment included MTX orally for at least 3 months at a mean dosage of 12.8 ± 2.6 mg/m ² (range: 8.4–20.5 mg/m ²).	Hepatic, GI	None

Bernbeck <i>et al.</i> , [38]	Caucasian	14/73	CC: 4 CT + TT: 10	CC: 32 CT + TT: 41	C677T	ALL	Fourteen of 209 patients with ALL BME and/or AON where diagnosed by MRI while they suffered from increasing pain of the affected region, with an incidence of 6.7%.	The therapy included oral MTX (20 mg/m ² /week)	Bones	None
Azarpira <i>et al.</i> , [39]	Other	9/21	CC: 6 CT + TT: 3	CC: 15 CT + TT: 6	C677T	Hem. cell transplantation	Participants in this retrospective cohort study were 30 patients (20 males, 10 females; median age: 29 years; range: 14–48 years) receiving an allogeneic hematopoietic cell transplant at the Bone Marrow Transplant Center of the Shiraz University of Medical Sciences in Shiraz, Iran, between 2006 and 2007.	Prophylactic MTX	GVHD	None
Sugimoto <i>et al.</i> , [40]	Asian	33/123	CC + CT: 32 TT: 1	CC + CT: 104 TT: 19	C677T	Hem. cell transplantation	The study population included adult patients who had received bone marrow or peripheral blood stem cell transplantation from an HLA-identical sibling donor. Median age was 38 years (range: 15–62 years). Fifty-six percent were males.	Prophylactic MTX	GVHD	None
Imanishi <i>et al.</i> , [41]	Asian	5/13	CC + CT: 2 TT: 3	CC + CT: 10 TT: 3	C677T	ALL, lymphoma	Blood samples were obtained from 26 patients with ALL ($n=20$) or ML ($n=6$). The age range was 10 months–15 years, (average: 6.7 ± 4.7).	The patients were administered a high dose of MTX (3 g/m ²) followed by leucovorin rescue treatment in the event of complete remission.	Hepatic	None
Campalani <i>et al.</i> , [42]	Other	103/73	C allele: 131 T allele: 75 A allele: 158 C allele: 48	C allele: 92 T allele: 54 A allele: 103 C allele: 43	C677T, A1298C	Psoriasis	Two-hundred three patients with psoriasis (192 with chronic plaque, eight with palmoplantar pustular, two with erythrodermic, and one with guttate psoriasis), who had previously or currently been treated with oral MTX were recruited in this study.	Patients had previously or currently been treated with oral MTX.	Hepatic	Not all
Ghodke <i>et al.</i> , [43]	Asian	13/21	C allele: 22 T allele: 4 A allele: 158 C allele: 48	C allele: 34 T allele: 8 A allele: 103 C allele: 43	C677T, A1298C	RA	Thirty-four naive RA patients (27 females), and American College of Rheumatology (ACR) classified.	Six months of supervised MTX therapy (7.5–17.5 mg weekly).	Overall	None
Chiusolo <i>et al.</i> , [44]	Caucasian	31/24	CC + CT: 19 TT: 12 AA + CA: 31 CC: 0	CC + CT: 23 TT: 1 AA + CA: 20 CC: 4	C677T, A1298C	ALL	Patients with ALL receiving chemotherapy.	Homogeneous maintenance therapy with MTX administered weekly intramuscularly at a variable dosage of 15–30 mg/m ² for almost 2 years.	Hepatic, hematological	None
Pihusch <i>et al.</i> , [45]	Caucasian	28/4	CC + CT: 26 TT: 2	CC + CT: 3 TT: 1	C677T	Hem. cell transplantation	Median age was 43 years (range: 14–62 years). 62% were males.	MTX, intravenous MTX (15 mg/m ² on day 1, 10 mg/m ² on days 3 and 6).	GVHD, mucositis	All
Seidemann <i>et al.</i> , [46]	Caucasian	– ^c	–	–	C677T	NHL	Seven-hundred ninety-one eligible patients up to 18 years of age with newly diagnosed NHL were enrolled in multicenter therapy trial NHL-BFM 95.	High-dose of MTX infusion regimens were randomized (4-h vs. 24-h infusion) in patients with B-cell lymphoma; MTX was applied as a 24-h infusion in all patients with lymphoblastic lymphoma and anaplastic large cell lymphoma.	GI, mucositis, infection, secondary cancer	All
Toffoli <i>et al.</i> , [47]	Caucasian	33/10	CC: 9 CT: 11 TT: 13	CC: 4 CT: 6 TT: 0	C677T	Ovarian cancer	Forty-three patients with ovarian cancer (age: 18–75 years)	The first 15 patients were treated with oral MTX 1.25 mg every 12 h for 21 days. After a 2-week rest, this regimen was repeated. As no objective response was noted in these patients, another group of 28 patients were enrolled, to be treated with a 21-day MTX (1.25 mg every 12 h).	Overall	None

Table 1 (continued)

Author, year	Descent	No. of cases/controls	Toxicity genotypes	Nontoxicity genotypes	Polymorphism	Disease	Patient's characteristics	Treatment	Toxicities ^a	Folate status ^b
Kalayoglu-Besisik <i>et al.</i> , [48]	Caucasian	11/42	CC: 8 CT: 3 TT: 0	CC: 17 CT: 20 TT: 5	C677T	Hem. cell transplantation	Fifty-three patients received allogeneic HCT (bone marrow: 16, peripheral blood: 37) as primary and unmanipulated transplantation for hematologic malignancies. Median age was 31.4 years (range: 16–51 years). Sixty-eight percent were males.	Short-term MTX (intravenously at day 1, 15 mg/m ² , and at days 3, 6, and 11, 10 mg/m ²)	Mucositis	None
Kim <i>et al.</i> , [49]	Asian	154/231	CC: 27 CT: 95 TT: 32	CC: 106 CT: 98 TT: 27	C677T	RA	The patients were of 24–80 years of age and included 355 women.	The patients were taking low-dose MTX (5–20 mg/week).	Overall	All
Fijnheer <i>et al.</i> , [50]	Other	–	–	–	C677T	SLE	Had 91.3% of women. The median age was 36 years (range: 17–70 years).	No specific information provided.	Thrombosis	None
Shimasaki <i>et al.</i> , [51]	Asian	14/6	CC: 9 CT: 3 TT: 1	CC: 2 CT: 4 TT: 0	C677T	ALL, lymphoma	The patients included in this study were Japanese children with ALL or LBL who had been enrolled in the Tokyo Children's Cancer Study Group's protocols.	The scheduled dosage of oral MTX was 25 mg/m ² once a week.	Hepatic, haematological	None
Kim <i>et al.</i> , [52]	Asian	–	–	–	C677T, A1298C	Hem. cell transplantation	Seventy-two patients receiving allogeneic HSCT at a single institution from 1998 to 2005. Median age was 36 years (range: 16–52 years). Fifty-four percent were males.	Four intravenous doses of MTX (day 1 at 15 mg/m ² and days 3, 6, and 11 at 10 mg/m ²)	Hepatic, GVHD	None
Ranganathan <i>et al.</i> , [53]	Caucasian	30/22	CC: 12 CT: 16 TT: 2	CC: 12 CT: 5 TT: 5	C677T	RA	The median age was 61 years (51–69 years) and the duration of RA was 6.5 years (range: 3.5–16.5 years).	The dose of MTX was determined by the treating rheumatologist. Concomitant medications included corticosteroids, sulfasalazine, etc.	Overall	All
Ranganathan <i>et al.</i> , [53]	Caucasian	15/76	CC: 7 CT: 4 TT: 4	CC: 28 CT: 41 TT: 7	C677T	RA	The median age was 42 years (range: 48–67 years) and the duration of RA was 7.5 years (range: 4–19 years).	The dose of MTX was determined by the treating rheumatologist.	GI	All
Kishi <i>et al.</i> , [54]	Other	9/44	CC: 6 CT: 2 TT: 1	CC: 26 CT: 17 TT: 1	C677T	ALL	Of the 53 patients who were enrolled onto the protocol, 23 were males, 38 were Whites, and 15 were Nonwhites. The median age was 6 years (range: 0–18 years).	Clinical protocol that included two courses of high-dose MTX (2.5 or 5.0 g/m ² per day) as consolidation therapy.	Nervous system, thrombosis	None
Jazbec <i>et al.</i> , [55]	Caucasian	15/30	CC: 7 CT: 6 TT: 2 AA: 7 AC: 6 CC: 2	CC: 11 CT: 15 TT: 4 AA: 14 AC: 14 CC: 2	C677T, A1298C	ALL	Four-hundred and forty-nine patients, aged ≤ 16 years at the time of diagnosis, were treated for ALL at the Pediatric Oncology Clinic at the University of Ljubljana and registered in the Cancer Registry of Slovenia.	High-dose of MTX (5 g/m ²) was included in all the protocols during consolidation and standard dose MTX (50 mg/m ² per week for 18 months) was used in maintenance.	Secondary cancer	None
Chiusolo <i>et al.</i> , [56]	Caucasian	22/39	CC: 7 CT: 6 TT: 9	CC: 14 CT: 19 TT: 6	C677T	ALL	Patients with acute leukemia receiving maintenance chemotherapy including MTX. Median age was 32 years (range: 12–72 years). Fifty-six percent were males.	MTX was administered weekly intramuscularly at a variable dosage of 15–30 mg/m ² in ALL, or 15 mg/m ² in APL during maintenance for almost 2 years.	Hepatic, hematological	None
Berkun <i>et al.</i> , [57]	Other	33/60	CC: 15 CT: 14 TT: 4 AA: 22 AC: 8 CC: 3	CC: 30 CT: 24 TT: 6 AA: 28 AC: 12 CC: 20	C677T, A1298C	RA	Ninety-three consecutive patients with rheumatoid arthritis with mean age of 58.74 ± 13.69 years. Females constituted 82.8%.	Duration of MTX treatment 3.70 ± 3.43 years. MTX dose was 11.93 ± 3.81 mg/week.	Overall	Not all

Gemmati <i>et al.</i> , [58]	Caucasian	18/50 for hepatic toxicity	CC: 5 CT: 7 TT: 6 for hepatic toxicity	CC: 17 CT: 29 TT: 4 for hepatic toxicity	C677T, A1298C	NHL	The whole group consisted of 110 patients with a high-grade NHL. Among these 110 patients, 68 received MACOP-B treatment and 42 received CHOP treatment. Mean age was 56.8 ± 17.1 years. Sixty-one percent were males.	No specific information provided.	Hepatic, mucositis hemato-logical	None
Ruiz-Arguelles <i>et al.</i> , [59]	Other	16/12	CC: 0 CT: 10 TT: 6	CC: 2 CT: 4 TT: 6	C677T	ALL	There were 13 patients who were less than 15 years; median age was 16 years, average was 16.5 years, range was from 0 to 40 years. The group included 21 males and seven females.	Patients were treated with a modification of the total XI schedule, which includes high-dose MTX rescued with folinic acid.	Mucositis	All
Warren <i>et al.</i> , [60]	Other	160/180	CC: 75 CT: 67 TT: 18	CC: 84 CT: 77 TT: 19	C677T, A1298C	Psoriasis	Patients with chronic plaque psoriasis treated with MTX were included. Males constituted 61% and mean age of onset was 24 years.	No specific information provided.	Overall	Not all
Kumagai <i>et al.</i> , [61]	Asian	52/63	CC: 24 CT: 18 TT: 10 AA: 36 AC: 14 CC: 2	CC: 22 CT: 29 TT: 12 AA: 44 AC: 15 CC: 4	C677T, A1298C	RA	Ninety-five percent of the patients were females. The mean age was 60 years and the mean disease duration was 11 years.	Patients had been prescribed various doses of MTX (2–12 mg/week) starting with the initial dose of 2–4 mg/week.	Overall	Not all
Wessels <i>et al.</i> , [62]	Caucasian	68/130	AA: 19 AC: 42 CC: 7	AA: 63 AC: 52 CC: 15	A1298C	RA	The group of 247 patients enrolled in this study was a subcohort of 508 patients who participated in the BeSt trial. Mean age was 54.6 ± 13.3 years. Sixty-nine percent were females.	Patients received MTX at an initial dosage of 7.5 mg/week, which was increased to 15 mg/week and combined with folic acid (1 mg/day) after 4 weeks. If the Disease Activity Score in 44 joints (DAS44) was > 2.4 at 3 months, MTX was increased to 25 mg/week.	GI	All
Murphy <i>et al.</i> , [63]	Other	158/29	CC: 78 CT+TT: 80	CC: 12 CT+TT: 17	C677T	Hem. cell transplantation	Had a variety of hematological conditions: AML, CML, lymphomas, myelodysplasia, etc.	15 mg/m ² at day 1 and 10 mg/m ² on days 3, 6 and 11 after transplant.	GVHD	Not all
Shimsaki <i>et al.</i> , [64]	Asian	–	–	–	C677T	ALL, lymphoma	Forty-three courses of high-dose MTX chemotherapy were evaluated in the 15 patients. The median patient age was 6 years (range: 1–14 years). Six patients were males and nine were females.	Patients received high-dose MTX over 12 h in the first course and over 24 h in the remaining courses intravenously with uniform supportive treatments followed by each protocol.	Hepatic, hemato-logical, GI, mucositis	None
Aplenc <i>et al.</i> , [65]	Other	–	–	–	C677T	ALL	We undertook a case-control study of patients treated on CCG-1891, a national intermediate risk ALL study run by CCG from 1989 to 1992 in 137 participating institutions. Intermediate risk was defined as an age between 1 and 10 years at diagnosis, and an initial WBC count of $< 5 \times 10^4/\mu\text{L}$.	All treatment arms were prescribed the identical schedule and dose of age-adjusted I.T. and weekly orally 20 mg/m ² MTX during interim maintenance and maintenance therapy.	Overall	None
Tsuchiya <i>et al.</i> , [66]	Asian	–	–	–	C677T	Urothelial cancer	The patients consisted of 34 males and six females with a mean age of 64.4 years (range: 46–80 years, median: 67 years) at the start of chemotherapy. Primary cancer lesions were located in the bladder in 23 patients and the renal pelvis or urethra in 19.	In brief, classic MVAC consisted of 28-day cycles of 30 mg/m ² intravenous MTX on days 1, 15, and 22; 3 mg/m ² intravenous vinblastine on days 2, 15, and 22; 30 mg/m ² intravenous doxorubicin on day 2 and 70 mg/m ² cisplatin on day 2.	Hematological	None

Table 1 (continued)

Author, year	Descent	No. of cases/controls	Toxicity genotypes	Nontoxicity genotypes	Polymorphism	Disease	Patient's characteristics	Treatment	Toxicities ^a	Folate status ^b
van Kooten <i>et al.</i> , [67]	Caucasian	–	–	–	C677T, A1298C	ALL	Ninety-nine children with non-B ALL diagnosed between 1 January 1992 and 31 December 2005 and treated at least once with HDMTX (5 g/m ²) according to the NOPHO ALL-92 or NOPHO ALL-2000 protocols at the Aarhus University Hospital were eligible for the study.	Each patient was scheduled to receive three or four HDMTX courses in the consolidation phase and five courses in the maintenance phase, all at a dose of 5 g/m ² .	Hematological	None
Hughes <i>et al.</i> , [68]	Caucasian	–	–	–	A1298C	RA	No specific information provided.	The mean MTX dose was 19 mg/week.	Overall	Not all
Weisman <i>et al.</i> , [69]	Other	–	–	–	C677T	RA	Two-hundred fourteen patients (<i>n</i> =53 at Cedars Sinai Medical Center, <i>n</i> =44 at UCLA, <i>n</i> =53 at the Florida site, and <i>n</i> =64 at the Albany site) who were undergoing MTX therapy for at least 1 month. Median age was 63 years (range: 52–71 years) and 79% were females.	Dosage of MTX was 12.5 mg/week (range: 10–17.5 mg/week). No. of months for receiving MTX was 39 (range: 11–89 months).	GI, nervous system, alopecia	Not all
Robien <i>et al.</i> , [70]	Other	158/192	CC: 127 CT: 16 TT: 15	CC: 131 CT: 21 TT: 40	C677T, A1298C	Hem. cell transplantation	Participants in this retrospective cohort study were adults undergoing allogeneic HCT at the Fred Hutchinson Cancer Research Center. Age was 40.4 years (range: 18–67 years). Fifty-six percent were males.	Prophylactic MTX.	GVHD	None
Herrlinger <i>et al.</i> , [71]	Caucasian	22/63	AA: 2 AC: 18 CC: 2	AA: 37 AC: 21 CC: 5	A1298C	IBD	Cohort of unrelated Caucasian IBD patients were recruited and treated with MTX.	The MTX dose for IBD patients varied between 15–25 mg/week with a median weekly dose of 20 mg, patients treated for concomitant RA received lower doses.	Overall	Not all

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; AON, aseptic osteonecrosis; APL, acute promyelocytic leukemia; BME, bone marrow edema; CCG, Children's Cancer Group; CHOP, a chemotherapy without methotrexate; CML, chronic myeloid leukemia; GVHD, graft-versus-host disease; HCT, hematopoietic cell transplantation; HDMTX, high-dose methotrexate; Hem, hematopoietic; HSCT, hematopoietic stem cell transplantation; IBD, inflammatory bowel disease; I.T, intrathecal; LBL, lymphoblastic lymphoma; MACOP-B, a chemotherapy combination containing methotrexate; ML, malignant lymphoma; MTX, methotrexate; MRI, magnetic resonance imaging; MVAC, methotrexate, vinblastine, doxorubicin and cisplatin; NHL, non-Hodgkin's lymphoma; NOPHO, Nordic Society of Pediatric Haematology and Oncology; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; WBC, white blood cell.

^aThe types of MTX adverse effects of the patients.

^bThe folate supplementation of the patients included in each study.

^cThe information is not provided. Odds ratios are given instead.

consistent with the HWE [41,46–61]. The logistic regression methodology (20 articles included [41,46–61]) provided a marginally insignificant association (TT vs. CC, OR: 1.560, 95% CI: 0.950, 2.561). The multivariate model (included 20 studies [41,46–61]) indicated similar results. The λ variable was found to be 0.49, indicating a codominant genetic model.

The univariate method was applied in 41 studies [30–68]. In particular, the contrast of the T allele versus C failed to reveal any association (OR: 1.158, 95% CI: 0.919, 1.459). However, the subgroup analysis provided a significant effect in the TT versus CT + CC contrast in studies consistent with the HWE (OR: 1.575, 95% CI: 1.100, 2.254). Finally, no publication bias was detected, but between-study heterogeneity was observed in some cases. At last, the cumulative meta-analysis revealed a small reduction of the effect over years, which became stable after 2006 (data not shown).

Hepatotoxicity and hematological toxicity

The analysis concerning hepatotoxicity revealed a statistically significant relation between the C677T polymorphism and this type of toxicity. Five studies were used for the logistic regression and the multivariate method [41,51,52,56,58]. The application of the logistic regression-based model noted an OR of 4.191 (95% CI: 1.642, 10.698) for the TT versus CC genotype. Similar significant results were suggested by the multivariate meta-analysis. The λ variable was found to be 0.16, providing evidence for a recessive model of inheritance. The conventional model was subsequently applied in 13 studies [30,32,37,41,42,44,49,51,52,56,58,64,65]. In particular, the OR of the T versus C allele contrast was found to be equal to 1.588 (95% CI: 1.146, 2.202). The subgroup analysis for Caucasians and the analysis for studies in HWE yielded a significant effect (Caucasians: OR: 2.472, 95% CI: 1.341, 4.559; HWE: OR: 2.004, 95% CI: 1.074, 3.738). The TT versus CT + CC genotype contrast demonstrated significant results for the overall population, Caucasians, and populations in HWE. The comparison of TT + CT genotypes versus CC yielded a significant measure of effect for all patients, Caucasians and also Asians. No evidence for publication bias or between-study heterogeneity was found.

Only three studies [51,56,58], which addressed the causal association of hematological disorders and MTX therapy, could be used in a multivariate model. The analysis indicated a marginally insignificant association of the mutant genotype with hematological toxicity (TT vs. CC: OR: 2.590, 95% CI: 0.950, 7.080). The univariate model (included 10 studies [32,44,49,51,56,58,64–67]) indicated a significant OR only in studies consistent with the HWE in the TT versus CT + CC genotype contrast (OR: 2.551, 95% CI: 1.055, 6.169). Publication bias was absent, but heterogeneity was present in some contrasts.

Gastrointestinal toxicity, neurotoxicity, GVHD, and mucositis

The association of the polymorphism with digestive system toxicity, neurotoxicity, GVHD, and mucositis was also studied. The subanalysis concerning gastrointestinal adverse effects found no association with the C677T polymorphism. As far as the nervous system adverse effects are concerned, logistic regression and multivariate models could not be used because of a lack of the necessary data. The univariate random-effects models detected a significant effect on neurotoxicity in the TT versus CC + CT contrast (OR: 3.398, 95% CI: 1.808, 6.389).

Our analysis concerning GVHD presents a protective role of the C677T polymorphism in the incidence of this type of toxicity. The logistic regression methodology produced an OR of 0.446 (TT vs. CC; 95% CI: 0.242, 0.821). The multivariate method provided similar results. Using the univariate method (included six studies [39,40,45,52, 63,70]), we calculated an OR of 0.441 (95% CI: 0.253, 0.771) for the TT versus CT + CC genotype comparison. No publication bias was found, but heterogeneity was detected in the allele contrast (I^2 : 84.6%). Finally, our analysis found no evidence for the association of mucositis with the C677T polymorphism of MTHFR. A recessive genetic model was suggested by the multivariate model in all the above subanalyses.

Rheumatoid arthritis, hematological disorders, and hematopoietic cell transplantation

Rheumatoid arthritis, hematological disorders, and hematopoietic cell transplantation subgroups concern patients who presented any form of toxicity (mild or severe; hematological, or nonhematological) during MTX treatment for the respective disease. Patients who came up with only one side effect by the drug use were included in the analysis.

As far as rheumatoid arthritis is concerned, we came up with a total of 15 studies. The OR of the TT versus CC comparison provided by the logistic regression was 1.991 (95% CI: 0.982, 4.036). The multivariate method provided similar results (included six studies, [49,53,57,61]). A codominant genetic model was proposed by the analysis (λ : 0.7). No association was detected in the TT + CT versus CC genotypes (OR: 1.546, 95% CI: 0.964, 2.479). Four studies [49,53,57,61] were found to be consistent with the HWE. The analysis of this subgroup noted a significant effect of the toxicity on rheumatoid arthritis patients (TT vs. CT + CC; OR: 1.735, 95% CI: 1.129, 2.666). There was no publication bias, but heterogeneity was found in some cases.

The multivariate model concerning patients with hematological disorders (included seven studies [46,51,54–56, 58,59]) provided an OR of 1.752 for the TT versus CC genotype comparison (95% CI: 0.847, 3.626). The classical methodology (included 10 studies [32,44,49,51,56,58,64–67])

presented a marginally insignificant association in the T versus C allele contrast (OR: 1.237, 95% CI: 0.947, 1.616) and in the mutant vs. CT + CC genotypes (CT + CC genotypes contrast (OR: 1.957, 95% CI: 0.935, 4.099). The analysis concerning patients treated with hematopoietic cell transplantation highlighted a protective effect of the polymorphism. In detail, we found an OR of 0.413 (TT vs. CC contrast; 95% CI: 0.226, 0.753) using the logistic regression and multivariate models (included three articles [48,52,70]). A recessive inheritance model emerged from the multivariate analysis (λ : 0.22). Moreover, the protective action of the polymorphism was confirmed by the univariate meta-analysis applied in five studies. The OR of the TT versus CC + CT genotypes contrast was 0.435 (95% CI: 0.252, 0.754). There was no publication bias, but between-studies heterogeneity was shown in some calculations.

Role of folic acid supplementation

The possible genetic effect in the appearance of MTX pharmacotoxicity due to the folate supplementation of patients was aimed to be elucidated. The multivariate methodology recruited 11 studies in the category of no folic acid supplementation, three studies in the category of partial folic acid supplementation, and six articles in the category of full folic acid supplementation. It was shown that the C677T polymorphism is associated with higher levels of toxicity in patients who were treated with MTX and folic acid as well. The OR derived under the random-effects logistic regression for the TT versus CC contrast was equal to 2.871 (95% CI: 1.107, 7.450) and that of CT versus CC was equal to 2.736 (95% CI: 1.347, 5.559) for patients taking folic acid. The same results occurred when the multivariate model was applied. In contrast, no relation between MTX adverse events and the C677T polymorphism was found concerning the other categories (no folic acid supplementation and partial folic acid supplementation).

The classical meta-analysis methodology found no association between the MTHFR polymorphism and MTX toxicity in all cases of folate treatment. Twenty-three studies for the no folate supplementation group, ten studies for the partial supplementation subgroup, and eight articles for full folate supplementation of patients were included. No association between the mutant allele and the MTX toxicity for patients treated with folic acid supplementation (OR: 1.511, 95% CI: 0.897, 2.546) was indicated. The results of the C677T polymorphism analysis are presented in Tables 2 and 3.

MTHFR A1298C polymorphism (rs1801131)

Overall toxicity: more than two adverse effects

Initially, we came up with a total of 12 studies examining the relation of the polymorphism with the presence of at least two adverse effects of MTX treatment. Four papers [60–62,71] were consistent with the HWE in participants without MTX toxicity. The logistic regres-

sion and multivariate methods failed to yield significant results. The univariate meta-analysis provided similar insignificant results. Heterogeneity was present in some contrasts. The cumulative analysis did not provide any trend in the effect estimates over time in any of the contrasts examined (data not shown).

Overall toxicity: more than one adverse effect

In total, 18 studies examined the association of MTHFR A1298C polymorphism with at least one kind of MTX toxicity. In particular, seven studies included Caucasians, four studies included Asians, and four studies included participants of other ethnicities. Six studies [52,55, 60–62,71] were found to be consistent with the HWE. The application of the logistic regression-based model for meta-analysis (included nine studies, [52,55,57,58, 60–62,70,71]) demonstrated a significant effect for the C allele carriers versus AA genotype (OR: 1.889, 95% CI: 1.232, 2.896) and for AC versus AA genotype (OR: 1.690, 95% CI: 1.011, 2.825). The multivariate analysis (9 studies) yielded no statistically significant results (CC vs. AA OR: 1.500, 95% CI: 0.720, 3.140; AC vs. AA OR: 1.520, 95% CI: 0.880, 2.610). These results are consistent with a dominant genetic model of inheritance. The univariate meta-analysis included 18 studies. The C versus A allele comparison of studies in HWE yielded a marginally insignificant association of the polymorphism with the appearance of at least one adverse event (OR: 1.346, 95% CI: 0.951, 1.905). Similarly, the C allele carriers versus AA genotype contrast provided an OR of 1.415 (95% CI: 0.913, 2.193). Between-studies heterogeneity appeared in some cases. No publication bias was detected.

Rheumatoid arthritis, hematological disorders, and hematopoietic cell transplantation

The analysis concerning rheumatoid arthritis patients retrieved a total of eight publications. The logistic regression-based model and the multivariate analysis yielded no statistically significant results. In contrast, the univariate model yielded a significant effect when contrasting CC versus AC + AA genotypes (OR: 0.506, 95% CI: 0.273, 0.936), indicating a protective role of the CC homozygous rheumatoid arthritis patients against the appearance of MTX adverse events. Between-studies heterogeneity was present in some cases. None of the remaining contrasts yielded a significant finding. Studies from patients with hematological diseases were used in a meta-analysis with no significant findings in all the contrasts examined here.

Role of folic acid supplementation

The relation between MTX toxicity and the A1298C polymorphism according to the folate supplementation of patients was explored. The logistic regression OR of the CC versus AA contrast was 4.131 (95% CI: 1.864, 9.155) for patients who were not treated with folic acid

Table 2 Results of the meta-analysis concerning the C677T polymorphism-multivariate methodology [18,19] (see Methods)

Category	Method	Contrast	No. of studies	Odds ratio	95% Confidence interval
Overall toxicity more than two types	Logistic regression	CT vs. CC	8	1.208	0.735, 1.984
		TT vs. CC	8	1.672	1.009, 2.769
	Multivariate model	CT vs. CC	8	1.200	0.730, 1.980
		TT vs. CC	8	1.670	1.000, 2.770
Overall toxicity more than one type	Logistic regression	CT vs. CC	20	1.228	0.868, 1.739
		TT vs. CC	20	1.560	0.950, 2.561
	Multivariate model	CT vs. CC	20	1.320	0.920, 1.890
		TT vs. CC	20	1.480	0.920, 2.390
Hepatotoxicity	Logistic regression	CT vs. CC	5	1.463	0.674, 3.176
		TT vs. CC	5	4.191	1.642, 10.698
	Multivariate model	CT vs. CC	5	1.460	0.680, 3.130
		TT vs. CC	5	4.260	1.750, 10.400
Hematological toxicity	Logistic regression	CT vs. CC	3	0.810	0.372, 1.765
		TT vs. CC	3	2.596	0.951, 7.084
	Multivariate model	CT vs. CC	3	0.800	0.370, 1.760
		TT vs. CC	3	2.590	0.950, 7.080
GVHD	Logistic regression	CT vs. CC	2	0.993	0.508, 1.713
		TT vs. CC	2	0.446	0.242, 0.821
	Multivariate model	CT vs. CC	2	0.930	0.500, 1.700
		TT vs. CC	2	0.440	0.240, 0.820
Mucositis	Logistic regression	CT vs. CC	4	0.690	0.351, 1.360
		TT vs. CC	4	1.069	0.541, 2.114
	Multivariate model	CT vs. CC	4	0.670	0.320, 1.360
		TT vs. CC	4	0.980	0.500, 1.890
Rheumatoid arthritis	Logistic regression	CT vs. CC	6	1.565	0.755, 3.243
		TT vs. CC	6	1.991	0.982, 4.036
	Multivariate model	CT vs. CC	6	1.590	0.750, 3.390
		TT vs. CC	6	2.010	0.960, 4.200
Hematological disorders	Logistic regression	CT vs. CC	7	0.783	0.486, 1.259
		TT vs. CC	7	1.752	0.847, 3.626
	Multivariate model	CT vs. CC	7	0.780	0.480, 1.250
		TT vs. CC	7	1.750	0.840, 3.620
Hematopoietic cell transplantation	Logistic regression	CT vs. CC	3	0.798	0.457, 1.392
		TT vs. CC	3	0.413	0.226, 0.753
	Multivariate model	CT vs. CC	3	0.790	0.450, 1.390
		TT vs. CC	3	0.410	0.220, 0.750
Folate supplement: none	Logistic regression	CT vs. CC	11	0.800	0.539, 1.187
		TT vs. CC	11	1.624	0.728, 3.623
	Multivariate model	CT vs. CC	11	0.790	0.510, 1.200
		TT vs. CC	11	1.390	0.620, 3.110
Folate supplement: not all	Logistic regression	CT vs. CC	3	0.903	0.629, 1.298
		TT vs. CC	3	1.026	0.598, 1.759
	Multivariate model	CT vs. CC	3	0.900	0.620, 1.300
		TT vs. CC	3	1.020	0.590, 1.750
Folate supplement: all	Logistic regression	CT vs. CC	6	2.736	1.347, 5.559
		TT vs. CC	6	2.871	1.107, 7.450
	Multivariate model	CT vs. CC	6	2.820	1.320, 6.000
		TT vs. CC	6	2.920	1.110, 7.770

GVHD, graft-versus-host disease.

supplementation. This was consistent with the multivariate methodology findings with an OR of 3.83 (95% CI: 1.95, 7.49). The univariate method was applied in eight studies, and also provided evidence of a significant relation. The AC + CC versus AA genotype comparison concluded to a significant association between the polymorphism and MTX toxicity in patients who were not treated with folate (OR: 1.926, 95% CI: 1.098, 3.379). Heterogeneity was observed in some contrasts. An association was suggested for the full folate supplementation category when considering the AC + CC versus AA genotype comparison (OR: 2.425, 95% CI: 1.289 4.560), but only one study was eligible. The results of the A1298C polymorphism analyses are presented in Tables 4 and 5.

Discussion

The present meta-analysis aims to clarify the association of MTHFR C677T and A1298C gene polymorphisms with MTX-related toxicity. An association between the C677T MTHFR variant and toxicity as an outcome of MTX treatment is indicated. In detail, logistic regression and multivariate models revealed a significant effect of the TT genotype for patients who presented at least two adverse events. The effect of the polymorphism on hepatotoxicity has been clearer. In contrast, a protective role of C677T polymorphism on the occurrence of GVHD was determined. Our analysis has managed to demonstrate a significant association between 677C > T polymorphism and a lower risk of GVHD in patients

Table 3 Results of the C677T polymorphism meta-analysis: univariate methodology (random-effects) [22] (see Methods)

Category	Contrast	Number of studies	Odds ratio (random effects)	95% Confidence interval	I^2 (%)
Subgroup: overall toxicity: more than two types of toxicities ^a					
All	T vs. C allele	10	1.217	0.889, 1.665	63.2
	TT vs. CT + CC	10	1.566	0.949, 2.582	43.0
	TT + CT vs. CC	14	1.335	0.874, 2.039	72.9
Caucasians	T vs. C allele	3	1.636	0.903, 2.964	29.6
	TT vs. CT + CC	4	3.191	0.514, 19.804	73.2
	TT + CT vs. CC	5	1.342	0.677, 2.663	50.8
Asians	T vs. C allele	4	1.065	0.495, 2.289	78.5
	TT vs. CT + CC	3	1.655	1.029, 2.660	0.0
	TT + CT vs. CC	6	1.540	0.670, 3.543	84.4
Other	T vs. C allele	3	1.021	0.802, 1.299	0.0
	TT vs. CT + CC	3	1.131	0.665, 1.921	0.0
	TT + CT vs. CC	3	0.976	0.690, 1.380	0.0
HWE	T vs. C allele	7	1.335	0.894, 1.996	71.4
	TT vs. CT + CC	7	1.604	1.063, 2.419	14.1
	TT + CT vs. CC	7	1.225	0.640, 2.344	77.9
Subgroup: overall toxicity: more than one type of toxicity ^b					
All	T vs. C allele	24	1.158	0.919, 1.459	65.4
	TT vs. CT + CC	25	1.246	0.840, 1.848	54.2
	TT + CT vs. CC	32	1.239	0.946, 1.624	63.0
Caucasians	T vs. C allele	8	1.162	0.810, 1.666	45.8
	TT vs. CT + CC	11	1.546	0.723, 3.307	66.1
	TT + CT vs. CC	12	1.157	0.803, 1.668	44.4
Asians	T vs. C allele	7	1.394	0.810, 2.401	63.2
	TT vs. CT + CC	6	1.377	0.719, 2.636	29.3
	TT + CT vs. CC	9	1.821	0.938, 3.537	76.1
Other	T vs. C allele	9	1.019	0.737, 1.408	64.0
	TT vs. CT + CC	8	0.917	0.529, 1.590	35.3
	TT + CT vs. CC	11	0.909	0.658, 1.256	28.0
HWE	T vs. C allele	18	1.284	0.985, 1.674	61.1
	TT vs. CT + CC	18	1.575	1.100, 2.254	26.1
	TT + CT vs. CC	18	1.189	0.808, 1.752	62.8
Subgroup: hepatotoxicity ^c					
All	T vs. C allele	8	1.588	1.146, 2.202	23.4
	TT vs. CT + CC	6	4.601	2.139, 9.896	12.6
	TT + CT vs. CC	9	2.170	1.305, 3.608	33.7
Caucasians	T vs. C allele	2	2.472	1.341, 4.559	5.8
	TT vs. CT + CC	3	6.967	3.101, 15.651	0.0
	TT + CT vs. CC	4	2.310	1.470, 3.629	0.0
Asians	T vs. C allele	4	1.560	0.667, 3.647	30.6
	TT vs. CT + CC	3	1.521	0.372, 6.213	0.0
	TT + CT vs. CC	4	2.716	1.034, 7.134	29.6
Other	T vs. C allele	2	1.307	0.963, 1.776	0.0
	TT vs. CT + CC	0	0.494	0.126, 1.941	38.4
	TT + CT vs. CC	1	2.004	1.074, 3.738	5.7
HWE	T vs. C allele	5	3.690	1.605, 8.483	0.0
	TT vs. CT + CC	5	1.841	0.880, 3.853	
	TT + CT vs. CC	5			
Subgroup: hematological toxicity ^d					
All	T vs. C allele	5	1.071	0.792, 1.449	0.0
	TT vs. CT + CC	5	1.861	0.627, 5.519	66.3
	TT + CT vs. CC	6	1.459	0.845, 2.520	0.0
Caucasians	T vs. C allele	2	1.584	0.934, 2.688	0.0
	TT vs. CT + CC	4	1.780	0.539, 5.886	74.2
	TT + CT vs. CC	2	1.401	0.532, 3.688	30.1
Asians	T vs. C allele	2	1.272	0.442, 3.662	0.0
	TT vs. CT + CC	1	3.316	0.120, 91.601	0.0
	TT + CT vs. CC	3	1.387	0.573, 3.360	0.0
Other	T vs. C allele	1	0.844	0.570, 1.249	
	TT vs. CT + CC	0	1.789	0.461, 6.948	
	TT + CT vs. CC	1	1.499	0.913, 2.462	0.0
HWE	T vs. C allele	3	2.551	1.055, 6.169	0.0
	TT vs. CT + CC	3	1.244	0.595, 2.597	0.3
	TT + CT vs. CC	3			
Subgroup: neurotoxicity ^e					
All	T vs. C allele	2	1.308	0.525, 3.257	0.0
	TT vs. CT + CC	2	3.398	1.808, 6.387	0.0
	TT + CT vs. CC	4	0.612	0.329, 1.137	0.0
Asians	TT + CT vs. CC	2	0.474	0.216, 1.041	0.0
	T vs. C allele	2	1.308	0.525, 3.257	0.0
	TT vs. CT + CC	2	3.398	1.808, 6.387	0.0
Other	TT + CT vs. CC	2	0.929	0.340, 2.544	0.0
Subgroup: GVHD ^f					
All	T vs. C allele	2	0.836	0.247, 2.826	84.6
	TT vs. CT + CC	4	0.441	0.253, 0.771	0.0
	TT + CT vs. CC	4	0.773	0.426, 1.401	36.0
Caucasians	TT vs. CT + CC	1	0.636	0.050, 8.123	
	T vs. C allele	1	1.667	0.690, 4.024	
	TT vs. CT + CC	2	0.549	0.065, 4.630	61.0
Asians	TT + CT vs. CC	1	2.895	0.581, 14.430	
	T vs. C allele	1	0.477	0.324, 0.703	
	TT vs. CT + CC	1	0.399	0.211, 0.753	
Other	TT + CT vs. CC	3	0.600	0.399, 0.904	0.0

Table 3 (continued)

Category	Contrast	Number of studies	Odds ratio (random effects)	95% Confidence interval	<i>I</i> ² (%)
Subgroup: mucositis ^g					
All	T vs. C allele	6	1.015	0.706, 1.459	20.4
	TT vs. CT + CC	5	1.115	0.463, 2.687	43.5
	TT + CT vs. CC	5	1.770	0.416, 1.425	27.5
Caucasians	T vs. C allele	3	0.849	0.517, 1.394	41.2
	TT vs. CT + CC	4	0.979	0.306, 3.135	54.2
	TT + CT vs. CC	3	0.675	0.337, 1.355	44.8
Other	T vs. C allele	2	1.601	0.737, 3.476	0.0
HWE	T vs. C allele	3	0.873	0.377, 2.020	60.3
	TT vs. CT + CC	3	0.840	0.505, 1.396	0.0
	TT + CT vs. CC	3	0.829	0.236, 2.914	59.1
Subgroup: rheumatoid arthritis ^h					
All	T vs. C allele	7	1.297	0.807, 2.085	67.7
	TT vs. CT + CC	7	1.422	0.852, 2.375	27.5
	TT + CT vs. CC	13	1.546	0.964, 2.479	76.0
Caucasians	T vs. C allele	2	1.069	0.602, 1.898	0.0
	TT vs. CT + CC	2	0.986	0.071, 13.756	82.2
	TT + CT vs. CC	5	1.335	0.667, 2.673	67.5
Asians	T vs. C allele	3	1.202	0.523, 2.762	83.2
	TT vs. CT + CC	2	1.578	0.846, 2.944	31.8
	TT + CT vs. CC	5	1.843	0.783, 4.337	86.0
Other	T vs. C allele	2	3.268	0.288, 37.046	78.3
	TT vs. CT + CC	3	1.324	0.581, 3.016	0.0
	TT + CT vs. CC	3	1.496	0.463, 4.837	66.5
HWE	T vs. C allele	4	1.285	0.748, 2.209	75.6
	TT vs. CT + CC	4	1.735	1.129, 2.666	0.0
	TT + CT vs. CC	4	1.247	0.448, 3.472	86.1
Subgroup: hematological disorders ⁱ					
All	T vs. C allele	10	1.237	0.947, 1.616	23.5
	TT vs. CT + CC	10	1.957	0.935, 4.099	60.8
	TT + CT vs. CC	9	1.028	0.758, 1.396	0.0
Caucasians	T vs. C allele	4	1.201	0.793, 1.820	43.2
	TT vs. CT + CC	6	1.807	0.690, 4.733	75.1
	TT + CT vs. CC	5	1.052	0.759, 1.458	0.0
Asians	T vs. C allele	3	1.713	0.434, 6.762	54.7
	TT vs. CT + CC	2	3.426	0.544, 21.568	0.0
Other	T vs. C allele	3	1.271	0.903, 1.790	0.0
	TT vs. CT + CC	2	2.152	0.562, 8.238	0.0
	TT + CT vs. CC	3	1.135	0.435, 2.960	0.0
HWE	T vs. C allele	7	1.154	0.845, 1.577	18.1
	TT vs. CT + CC	7	1.904	0.888, 4.081	47.5
	TT + CT vs. CC	7	0.979	0.708, 1.354	0.0
Subgroup: hematopoietic cell transplantation ^j					
All	T vs. C allele	3	0.634	0.254, 1.583	73.6
	TT vs. CT + CC	5	0.435	0.252, 0.754	0.0
	TT + CT vs. CC	5	0.681	0.385, 1.205	35.5
Caucasians	T vs. C allele	2	0.460	0.067, 3.182	0.0
Asians	TT vs. CT + CC	2	0.549	0.065, 4.630	61.0
Other	TT + CT vs. CC	3	0.600	0.399, 0.904	0.0
HWE	T vs. C allele	2	0.736	0.131, 4.143	79.5
	TT vs. CT + CC	2	1.009	0.229, 4.441	0.0
	TT + CT vs. CC	2	0.839	0.078, 9.070	79.2
Subgroup: folate supplementation ^k					
No folate supplementation	T vs. C allele	14	1.123	0.774, 1.631	63.0
	TT vs. CT + CC	14	1.437	0.662, 3.119	66.4
	TT + CT vs. CC	17	1.067	0.756, 1.506	34.8
Partial folate supplementation	T vs. C allele	4	0.978	0.785, 1.218	0.0
	TT vs. CT + CC	4	1.100	0.694, 1.745	0.0
	TT + CT vs. CC	8	1.225	0.773, 1.940	69.0
Full folate supplementation	T vs. C allele	6	1.511	0.897, 2.546	77.2
	TT vs. CT + CC	7	1.282	0.674, 2.442	50.4
	TT + CT vs. CC	7	1.714	0.809, 3.632	79.8

HWE, Hardy–Weinberg equilibrium; GVHD, graft-versus-host disease.

^aThe subanalysis concerning studies with patients who presented more than two types of side effects.

^bThe subanalysis concerning studies with patients who presented more than one type of side effect.

^cThe subanalysis concerning studies with patients who presented liver dysfunction.

^dThe subanalysis concerning studies with patients who presented one or more bone marrow toxicities.

^eThe subanalysis concerning studies with patients who presented one or more neurological disorders.

^fThe subanalysis concerning studies with patients who presented GVHD after hematopoietic cell transplantation.

^gThe subanalysis concerning studies with patients who presented oral mucositis.

^hThe subanalysis concerning studies with rheumatoid arthritis patients.

ⁱThe subanalysis concerning studies with patients suffering from blood diseases.

^jThe subanalysis concerning studies with patients treated with hematopoietic cell transplantation.

^kThe subanalysis based on folate supplementation status of the patients.

undergoing hematopoietic stem cell transplantation. Our findings are consistent with the idea that the lower MTHFR activity due to the 677C > T polymorphism offers patients treated with MTX greater immunosuppres-

sion. However, there are studies that failed to show any relation between MTHFR polymorphisms in patients receiving MTX and the incidence of GVHD. Eventually, evidence was provided by logistic regression and multivariate models

Table 4 Results of the A1298C polymorphism meta-analysis: multivariate methodology [18,19] (see Methods)

Category	Method	Contrast	No. of studies	Odds ratio	95% Confidence interval
Overall toxicity: more than two types	Logistic regression	AC vs. AA	5	1.697	0.762, 3.780
		CC vs. AA	5	1.027	0.419, 2.515
	Multivariate model	AC vs. AA	5	1.690	0.680, 4.150
		CC vs. AA	5	1.150	0.430, 3.060
Overall toxicity: more than one type	Logistic regression	AC vs. AA	9	1.690	1.011, 2.825
		CC vs. AA	9	1.889	1.232, 2.896
	Multivariate model	AC vs. AA	9	1.490	0.860, 2.580
		CC vs. AA	9	1.420	0.730, 2.780
Rheumatoid arthritis	Logistic regression	AC vs. AA	3	1.490	0.812, 2.735
		CC vs. AA	3	0.539	0.185, 1.570
	Multivariate model	AC vs. AA	3	1.470	0.780, 2.760
		CC vs. AA	3	0.540	0.180, 1.570
Hematological disorders	Logistic regression	AC vs. AA	2	0.486	0.137, 1.722
		CC vs. AA	2	4.038	0.771, 21.144
	Multivariate model	AC vs. AA	2	0.520	0.170, 1.530
		CC vs. AA	2	4.220	0.890, 19.900
Folate supplement: none	Logistic regression	AC vs. AA	4	1.278	0.582, 2.803
		CC vs. AA	4	4.131	1.864, 9.155
	Multivariate model	AC vs. AA	4	1.420	0.840, 2.390
		CC vs. AA	4	3.830	1.950, 7.490
Folate supplement: not all	Logistic regression	AC vs. AA	4	2.071	0.859, 4.990
		CC vs. AA	4	1.494	0.733, 3.047
	Multivariate model	AC vs. AA	4	2.140	0.810, 5.600
		CC vs. AA	4	1.590	0.730, 3.450

that the 677T homozygous supplemented with folate have a significantly elevated risk to present MTX toxicity. Finally, the observations of our analysis concerning the C677T polymorphism and toxicity in rheumatoid arthritis patients are consistent with those provided by a previous meta-analysis, [72] concluding in a mild association using random-effects models.

The logistic regression analysis of the A1298C polymorphism highlighted a significant relation between C allele carriers and MTX toxicity in patients who presented at least one adverse effect. Furthermore, evidence for a significantly elevated risk of toxicity is provided for C allele carriers who were not supplemented with folic acid during MTX treatment. Our analysis concluded to a protective role of the CC homozygous rheumatoid arthritis patients against the appearance of MTX adverse events, although this is conflicting with a previous meta-analysis [72].

MTX is a drug with a wide usage in a variety of disorders and in many regimens. Tissues undergoing rapid cell division and growth are the most susceptible to its effects. This mechanism also explains the possible severe side effects of MTX. MTX toxicity includes a significant number of different forms for almost all human organs. Therefore, it is probably a complicated biological issue where many factors are involved. Many researchers have tried to exemplify the metabolic pathways and the genetic impact of MTX toxicity [73–75]. MTX side effects can only partly be explained by folate antagonism and may also depend on its action on other related metabolic pathways. Recent data on the issue include the homocysteine–methionine–polyamine pathway and purine metabolism [75]. Many studies have investigated the role of the adenosine pathway in MTX

toxicities [76,77]. Further investigation is needed to clarify all the aspects of the patient.

In contrast, many investigators have tried to relate the higher risk for MTX adverse effects with gene polymorphisms. There are a large number of case–control studies aimed to investigate the two most common polymorphisms of the MTHFR gene. The C677T and A1298C SNPs of MTHFR have been identified as a risk factor for elevated MTX toxicity by many primary studies, whereas others concluded that there was no association. Other genetic variants are also analyzed. A recent study yielded evidence for the association with ABCB1 3435C > T (ATP-binding cassette) and TLR4 +896A > G (toll-like receptor 4) [78]. SNPs in SLC01B1 (solute carrier organic anion transporter family) and ADORA2a (adenosine receptor 2a) were significantly associated with gastrointestinal toxicity [79,80]. Thus, MTX toxicity may be influenced by additional polymorphisms and these should be examined in further studies [81,82].

The MTHFR gene and its polymorphisms hold an essential position in scientific research concerning the folic acid cycle. The two most common polymorphisms lead to decreased enzyme activity followed by hyperhomocysteinemia. Previous studies have tried to investigate the role of MTHFR SNPs with different diseases and disorders. The C677T polymorphism is considered to be a risk factor for thrombophilia and pregnancy complications [83,84]. Moreover, there is serious evidence that polymorphisms in the MTHFR gene predispose one to neural tube defects and schizophrenia [85,86]. The C677T polymorphism also seems to be associated with a reduced risk of colorectal cancer, but this may not hold true for all populations [87]. Finally, in the last decade, there was

Table 5 Results of the A1298C polymorphism meta-analysis: univariate methodology [22] (see Methods)

Category	Contrast	Number of studies	Odds ratio (random effects)	95% Confidence interval	I ² (%)
Subgroup: overall toxicity: more than two types of toxicities ^a					
All	C vs. A allele	8	0.891	0.568, 1.399	76.2
	CC vs. AC+AA	7	0.631	0.350, 1.135	36.6
	AC+CC vs. AA	7	1.281	0.718, 2.283	75.3
Caucasians	C vs. A allele	3	1.260	0.415, 3.830	80.0
	CC vs. AC+AA	4	0.633	0.318, 1.261	10.2
	AC+CC vs. AA	2	5.230	0.937, 29.183	80.4
Asians	C vs. A allele	2	0.824	0.470, 1.444	0.0
	AC+CC vs. AA	3	0.994	0.618, 1.598	0.0
Other	C vs. A allele	3	0.678	0.403, 1.142	71.0
	CC vs. AC+AA	2	0.540	0.092, 3.177	82.7
	AC+CC vs. AA	2	0.693	0.356, 1.348	51.3
HWE	C vs. A allele	4	1.373	0.871, 2.164	68.7
	CC vs. AC+AA	4	1.041	0.624, 1.737	0.0
	AC+CC vs. AA	4	1.981	0.821, 4.781	83.5
Subgroup: overall toxicity: more than one type of toxicity ^b					
All	C vs. A allele	12	1.099	0.737, 1.640	77.6
	CC vs. AC+AA	12	0.915	0.476, 1.759	67.2
	AC+CC vs. AA	12	1.415	0.913, 2.193	67.3
Caucasians	C vs. A allele	5	1.433	0.757, 2.711	62.1
	CC vs. AC+AA	7	0.843	0.357, 1.992	59.6
	AC+CC vs. AA	4	2.111	0.722, 6.169	70.0
Asians	C vs. A allele	3	0.948	0.578, 1.553	0.0
	CC vs. AC+AA	2	0.751	0.162, 3.482	0.0
	AC+CC vs. AA	4	1.082	0.689, 1.699	0.0
Other	C vs. A allele	4	0.945	0.444, 2.009	90.8
	CC vs. AC+AA	3	1.035	0.275, 3.893	85.6
	AC+CC vs. AA	4	1.319	0.557, 3.120	83.3
HWE	C vs. A allele	6	1.346	0.951, 1.905	48.6
	CC vs. AC+AA	6	1.097	0.671, 1.794	0.0
	AC+CC vs. AA	6	1.700	0.895, 3.229	70.4
Subgroup: rheumatoid arthritis ^c					
All	C vs. A allele	5	0.645	0.304, 1.365	77.7
	CC vs. AC+AA	4	0.506	0.273, 0.936	9.5
	AC+CC vs. AA	5	1.049	0.588, 1.871	61.4
Caucasians	C vs. A allele	2	0.390	0.018, 8.512	84.5
	CC vs. AC+AA	2	0.655	0.326, 1.317	0.0
Asians	C vs. A allele	2	0.824	0.470, 1.444	0.0
	AC+CC vs. AA	3	0.995	0.614, 1.611	0.0
HWE	C vs. A allele	2	1.286	0.820, 2.017	27.8
	CC vs. AC+AA	2	0.803	0.349, 1.847	0.0
	AC+CC vs. AA	2	1.638	0.709, 3.782	63.3
Subgroup: folate supplementation ^d					
No folate supplementation	C vs. A allele	5	1.504	0.841, 2.689	57.3
	CC vs. AC+AA	6	1.417	0.364, 5.522	77.2
	AC+CC vs. AA	5	1.926	1.098, 3.379	25.6
Partial folate supplementation	C vs. A allele	6	0.821	0.464, 1.452	79.1
	CC vs. AC+AA	5	0.632	0.312, 1.283	42.9
	AC+CC vs. AA	6	1.078	0.605, 1.921	66.7
Full folate supplementation	C vs. A allele	1	1.520	0.988, 2.336	–
	CC vs. AC+AA	1	0.880	0.340, 2.273	–
	AC+CC vs. AA	1	2.425	1.289, 4.560	–

HWE, Hardy–Weinberg equilibrium.

^aThe subanalysis concerning studies with patients who presented more than two types of side effects.^bThe subanalysis concerning studies with patients who presented more than one type of side effect.^cThe subanalysis concerning studies with rheumatoid arthritis patients.^dThe subanalysis based on folate supplementation status of the patients.

evidence that elevated plasma homocysteine is associated with an increased risk of atherosclerosis and cardiovascular disease. Several studies have tried to prove an effect of the C677T polymorphism on the risk of cardiovascular events, but the results remain conflicting [88].

Our analysis aims to exemplify the role of folic acid supplementation in the appearance of MTX toxicity in patients carrying a polymorphism in MTHFR gene. We have shown that the C677T polymorphism is associated with a higher risk of toxicity in patients supplemented with folate.

This verification seems paradoxical as one would expect that folate supplementation plays a protective role. A possible explanation could be the fact that patients who received a higher folate dose were probably in the most severe condition, and thus, are more likely to present MTX side effects. This particular phenomenon is common in observational studies, where the term ‘confounding by indication’ has been suggested [89]. Clearly, a meta-analysis of published results can only provide hints in this direction, and additional and carefully designed studies are necessary to elucidate this very important issue. In contrast, the A1298C MTHFR

polymorphism was associated with increased MTX toxicity in studies that did not include folate supplementation. However, many investigators suggest that folate supplementation reduces the risk of MTX adverse effects and it should be prescribed in every case [90].

In conclusion, our analysis suggests that MTHFR SNPs are associated with overall MTX toxicity. Our results also included a significant association with specific types of adverse effects and a positive role of polymorphisms in some cases. Future individual studies and meta-analyses are necessary to confirm our conclusions and set the foundations of individualized treatment. Finally, further investigation is needed to detect the exact biological pathways and mechanisms explaining MTX toxicity and genetic associations.

Acknowledgements

The authors thank the two anonymous reviewers, whose comments helped in improving the quality of the manuscript.

Conflicts of interest

There are no conflicts of interest.

References

- Kremer JM. Methotrexate and emerging therapies. *Clin Exp Rheumatol* 1999; **17**:S43–S46.
- Kremer JM. Methotrexate and emerging therapies. *Rheum Dis Clin North Am* 1998; **24**:651–658.
- Cronstein BN. The mechanism of action of methotrexate. *Rheum Dis Clin North Am* 1997; **23**:739–755.
- Cronstein BN. Molecular therapeutics. *Methotrexate and its mechanism of action*, *Arthritis Rheum* 1996; **39**:1951–1960.
- Chan ES, Cronstein BN. Molecular action of methotrexate in inflammatory diseases. *Arthritis Res* 2002; **4**:266–273.
- Baggott JE, Vaughn WH, Hudson BB. Inhibition of 5-aminoimidazole-4-carboxamide ribotide transformylase, adenosine deaminase and 5'-adenylate deaminase by polyglutamates of methotrexate and oxidized folates and by 5-aminoimidazole-4-carboxamide riboside and ribotide. *Biochem J* 1986; **236**:193–200.
- Kremer JM. Toward a better understanding of methotrexate. *Arthritis Rheum* 2004; **50**:1370–1382.
- Neves C, Jorge R, Barcelos A. The network of methotrexate toxicity. *Acta Reumatol Port* 2009; **34**:11–34.
- Braun J, Rau R. An update on methotrexate. *Curr Opin Rheumatol* 2009; **21**:216–223.
- Rozen R. Molecular genetics of methylenetetrahydrofolate reductase deficiency. *J Inherit Metab Dis* 1996; **19**:589–594.
- Goyette P, Sumner JS, Milos R, Duncan AM, Rosenblatt DS, Matthews RG, et al. Human methylenetetrahydrofolate reductase: isolation of cDNA mapping and mutation identification. *Nat Genet* 1994; **7**:551.
- Van der Put NM, Gabreels F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, et al. A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? *Am J Hum Genet* 1998; **62**:1044–1051.
- Weisberg I, Tran P, Christensen B, Sibani S, Rozen R. A second genetic polymorphism in methylenetetrahydrofolate reductase (MTHFR) associated with decreased enzyme activity. *Mol Genet Metab* 1998; **64**:169–172.
- Frost P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG, et al. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* 1995; **10**:111–113.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; **283**:2008–2012.
- Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the chinese literature. *PLoS Med* 2005; **2**:e334.
- Conn VS, Valentine JC, Cooper HM, Rantz MJ. Grey literature in meta-analyses. *Nurs Res* 2003; **52**:256–261.
- Bagos PG. A unification of multivariate methods for meta-analysis of genetic association studies. *Stat Appl Genet Mol Biol* 2008; **7**:Article 31.
- Bagos PG, Nikolopoulos GK. A method for meta-analysis of case-control genetic association studies using logistic regression. *Stat Appl Genet Mol Biol* 2007; **6**:Article 17.
- Petiti DB. *Meta-analysis decision analysis and cost-effectiveness analysis*. Oxford: Oxford University Press; 1994.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**:557–560.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**:177–188.
- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; **50**:1088–1101.
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**:629–634.
- Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999; **18**:2693–2708.
- Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; **327**:248–254.
- Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. *J Clin Epidemiol* 1995; **48**:45–57, discussion 59–60.
- Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: the Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005; **58**:543–549.
- Bagos PG, Nikolopoulos GK. Generalized least squares for assessing trends in cumulative meta-analysis with applications in genetic epidemiology. *J Clin Epidemiol* 2009 (In press).
- Van Ede AE, Laan RF, Blom HJ, Huizinga TW, Haagsma CJ, Giesendorf BA, et al. The C677T mutation in the methylenetetrahydrofolate reductase gene: a genetic risk factor for methotrexate-related elevation of liver enzymes in rheumatoid arthritis patients. *Arthritis Rheum* 2001; **44**:2525–2530.
- Bohanec Grabar P, Logar D, Lestan B, Dolzan V. Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism. *Eur J Clin Pharmacol* 2008; **64**:1057–1068.
- Aggarwal P, Naik S, Mishra KP, Aggarwal A, Misra R. Correlation between methotrexate efficacy and toxicity with C677T polymorphism of the methylenetetrahydrofolate gene in rheumatoid arthritis patients on folate supplementation. *Indian J Med Res* 2006; **124**:521–526.
- Zeng QY, Wang YK, Xiao ZY, Chen SB. Pharmacogenetic study of 5,10-methylenetetrahydrofolate reductase C677T and thymidylate synthase 3R/2R gene polymorphisms and methotrexate-related toxicity in Chinese Han patients with inflammatory arthritis. *Ann Rheum Dis* 2008; **67**:1193–1194.
- Urano W, Taniguchi A, Yamanaka H, Tanaka E, Nakajima H, Matsuda Y, et al. Polymorphisms in the methylenetetrahydrofolate reductase gene were associated with both the efficacy and the toxicity of methotrexate used for the treatment of rheumatoid arthritis, as evidenced by single locus and haplotype analyses. *Pharmacogenetics* 2002; **12**:183–190.
- Taniguchi A, Urano W, Tanaka E, Furihata S, Kamitsuiji S, Inoue E, et al. Validation of the associations between single nucleotide polymorphisms or haplotypes and responses to disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a proposal for prospective pharmacogenomic study in clinical practice. *Pharmacogenet Genomics* 2007; **17**:383–390.
- Krull KR, Brouwers P, Jain N, Zhang L, Bomgaars L, Dreyer Z, et al. Folate pathway genetic polymorphisms are related to attention disorders in childhood leukemia survivors. *J Pediatr* 2008; **152**:101–105.
- Schmeling H, Biber D, Heins S, Horneff G. Influence of methylenetetrahydrofolate reductase polymorphisms on efficacy and toxicity of methotrexate in patients with juvenile idiopathic arthritis. *J Rheumatol* 2005; **32**:1832–1836.
- Bernbeck B, Mauz-Korholz C, Zolt RB, Gobel U. Methylenetetrahydrofolate reductase gene polymorphism and glucocorticoid intake in children with ALL and aseptic osteonecrosis. *Klin Padiatr* 2003; **215**:327–331.
- Azarpira N, Ramzi M, Aghdaie MH, Daraie M, Geramizadeh B. Methylenetetrahydrofolate reductase C677T genotypes and clinical outcome following hematopoietic cell transplant. *Exp Clin Transplant* 2007; **5**:693–697.

- 40 Sugimoto K, Murata M, Onizuka M, Inamoto Y, Terakura S, Kuwatsuka Y, et al. Decreased risk of acute graft-versus-host disease following allogeneic hematopoietic stem cell transplantation in patients with the 5,10-methylenetetrahydrofolate reductase 677TT genotype. *Int J Hematol* 2008; **87**:451–458.
- 41 Imanishi H, Okamura N, Yagi M, Noro Y, Moriya Y, Nakamura T, et al. Genetic polymorphisms associated with adverse events and elimination of methotrexate in childhood acute lymphoblastic leukemia and malignant lymphoma. *J Hum Genet* 2007; **52**:166–171.
- 42 Campalani E, Arenas M, Marinaki AM, Lewis CM, Barker JN, Smith CH. Polymorphisms in folate, pyrimidine, and purine metabolism are associated with efficacy and toxicity of methotrexate in psoriasis. *J Invest Dermatol* 2007; **127**:1860–1867.
- 43 Ghodke Y, Chopra A, Joshi K, Patwardhan B. Are thymidylate synthase and methylene tetrahydrofolate reductase genes linked with methotrexate response (efficacy, toxicity) in Indian (Asian) rheumatoid arthritis patients? *Clin Rheumatol* 2008; **27**:787–789.
- 44 Chiusolo P, Reddicono G, Farina G, Mannocci A, Fiorini A, Palladino M, et al. MTHFR polymorphisms' influence on outcome and toxicity in acute lymphoblastic leukemia patients. *Leuk Res* 2007; **31**:1669–1674.
- 45 Pihusch M, Lohse P, Reitberger J, Hiller E, Andreesen R, Kolb HJ, et al. Impact of thrombophilic gene mutations and graft-versus-host disease on thromboembolic complications after allogeneic hematopoietic stem-cell transplantation. *Transplantation* 2004; **78**:911–918.
- 46 Seidemann K, Book M, Zimmermann M, Meyer U, Welte K, Stanulla M, et al. MTHFR 677 (C→T) polymorphism is not relevant for prognosis or therapy-associated toxicity in pediatric NHL: results from 484 patients of multicenter trial NHL-BFM 95. *Ann Hematol* 2006; **85**:291–300.
- 47 Toffoli G, Russo A, Innocenti F, Corona G, Tumolo S, Sartor F, et al. Effect of methylenetetrahydrofolate reductase 677C→T polymorphism on toxicity and homocysteine plasma level after chronic methotrexate treatment of ovarian cancer patients. *Int J Cancer* 2003; **103**:294–299.
- 48 Kalayoglu-Besik S, Caliskan Y, Sargin D, Gurses N, Ozbek U. Methylenetetrahydrofolate reductase C677T polymorphism and toxicity in allogeneic hematopoietic cell transplantation. *Transplantation* 2003; **76**:1775–1777.
- 49 Kim SK, Jun JB, El-Sohemy A, Bae SC. Cost-effectiveness analysis of MTHFR polymorphism screening by polymerase chain reaction in Korean patients with rheumatoid arthritis receiving methotrexate. *J Rheumatol* 2006; **33**:1266–1274.
- 50 Fijnheer R, Roest M, Haas FJ, De Groot PG, Homocysteine Derksen RH. methylenetetrahydrofolate reductase polymorphism, antiphospholipid antibodies, and thromboembolic events in systemic lupus erythematosus: a retrospective cohort study. *J Rheumatol* 1998; **25**:1737–1742.
- 51 Shimasaki N, Mori T, Torii C, Sato R, Shimada H, Tanigawara Y, et al. Influence of MTHFR and RFC1 polymorphisms on toxicities during maintenance chemotherapy for childhood acute lymphoblastic leukemia or lymphoma. *J Pediatr Hematol Oncol* 2008; **30**:347–352.
- 52 Kim I, Lee KH, Kim JH, Ra EK, Yoon SS, Hong YC, et al. Polymorphisms of the methylenetetrahydrofolate reductase gene and clinical outcomes in HLA-matched sibling allogeneic hematopoietic stem cell transplantation. *Ann Hematol* 2007; **86**:41–48.
- 53 Ranganathan P, Culverhouse R, Marsh S, Mody A, Scott-Horton TJ, Brasington R, et al. Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis. *J Rheumatol* 2008; **35**:572–579.
- 54 Kishi S, Griener J, Cheng C, Das S, Cook EH, Pei D, et al. Homocysteine, pharmacogenetics, and neurotoxicity in children with leukemia. *J Clin Oncol* 2003; **21**:3084–3091.
- 55 Jazbec J, Kitanovski L, Aplenc R, Debeljak M, Dolzan V. No evidence of association of methylenetetrahydrofolate reductase polymorphism with occurrence of second neoplasms after treatment of childhood leukemia. *Leuk Lymphoma* 2005; **46**:893–897.
- 56 Chiusolo P, Reddicono G, Casorelli I, Laurenti L, Sora F, Mele L, et al. Preponderance of methylenetetrahydrofolate reductase C677T homozygosity among leukemia patients intolerant to methotrexate. *Ann Oncol* 2002; **13**:1915–1918.
- 57 Berkun Y, Levartovsky D, Rubinow A, Orbach H, Aamar S, Grenader T, et al. Methotrexate related adverse effects in patients with rheumatoid arthritis are associated with the A1298C polymorphism of the MTHFR gene. *Ann Rheum Dis* 2004; **63**:1227–1231.
- 58 Gemmati D, Ongaro A, Tognazzo S, Catozzi L, Federici F, Mauro E, et al. Methylenetetrahydrofolate reductase C677T and A1298C gene variants in adult non-Hodgkin's lymphoma patients: association with toxicity and survival. *Haematologica* 2007; **92**:478–485.
- 59 Ruiz-Arguelles GJ, Coconi-Linares LN, Garcés-Eisele J, Reyes-Núñez V. Methotrexate-induced mucositis in acute leukemia patients is not associated with the MTHFR 677T allele in Mexico. *Hematology* 2007; **12**:387–391.
- 60 Warren RB, Smith RL, Campalani E, Eyre S, Smith CH, Barker JN, et al. Outcomes of methotrexate therapy for psoriasis and relationship to genetic polymorphisms. *Br J Dermatol* 2009; **160**:438–441.
- 61 Kumagai K, Hiyama K, Oyama T, Maeda H, Kohno N. Polymorphisms in the thymidylate synthase and methylenetetrahydrofolate reductase genes and sensitivity to the low-dose methotrexate therapy in patients with rheumatoid arthritis. *Int J Mol Med* 2003; **11**:593–600.
- 62 Wessels JA, De Vries-Bouwstra JK, Heijmans BT, Slagboom PE, Goekoop-Ruiterman YP, Allaart CF, et al. Efficacy and toxicity of methotrexate in early rheumatoid arthritis are associated with single-nucleotide polymorphisms in genes coding for folate pathway enzymes. *Arthritis Rheum* 2006; **54**:1087–1095.
- 63 Murphy N, Diviney M, Szer J, Barty P, Grigg A, Hoyt R, et al. Donor methylenetetrahydrofolate reductase genotype is associated with graft-versus-host disease in hematopoietic stem cell transplant patients treated with methotrexate. *Bone Marrow Transplant* 2006; **37**:773–779.
- 64 Shimasaki N, Mori T, Samejima H, Sato R, Shimada H, Yahagi N, et al. Effects of methylenetetrahydrofolate reductase and reduced folate carrier 1 polymorphisms on high-dose methotrexate-induced toxicities in children with acute lymphoblastic leukemia or lymphoma. *J Pediatr Hematol Oncol* 2006; **28**:64–68.
- 65 Aplenc R, Thompson J, Han P, La M, Zhao H, Lange B, et al. Methylenetetrahydrofolate reductase polymorphisms and therapy response in pediatric acute lymphoblastic leukemia. *Cancer Res* 2005; **65**:2482–2487.
- 66 Tsuchiya N, Inoue T, Narita S, Kumazawa T, Saito M, Obara T, et al. Drug related genetic polymorphisms affecting adverse reactions to methotrexate, vinblastine, doxorubicin and cisplatin in patients with urothelial cancer. *J Urol* 2008; **180**:2389–2395.
- 67 Van Kooten Niekerk PB, Schmiegelow K, Schroeder H. Influence of methylene tetrahydrofolate reductase polymorphisms and coadministration of antimetabolites on toxicity after high dose methotrexate. *Eur J Haematol* 2008; **81**:391–398.
- 68 Hughes LB, Beasley TM, Patel H, Tiwari HK, Morgan SL, Baggott JE, et al. Racial or ethnic differences in allele frequencies of single-nucleotide polymorphisms in the methylenetetrahydrofolate reductase gene and their influence on response to methotrexate in rheumatoid arthritis. *Ann Rheum Dis* 2006; **65**:1213–1218.
- 69 Weisman MH, Furst DE, Park GS, Kremer JM, Smith KM, Wallace DJ, et al. Risk genotypes in folate-dependent enzymes and their association with methotrexate-related side effects in rheumatoid arthritis. *Arthritis Rheum* 2006; **54**:607–612.
- 70 Robien K, Bigler J, Yasui Y, Potter JD, Martin P, Storb R, et al. Methylenetetrahydrofolate reductase and thymidylate synthase genotypes and risk of acute graft-versus-host disease following hematopoietic cell transplantation for chronic myelogenous leukemia. *Biol Blood Marrow Transplant* 2006; **12**:973–980.
- 71 Herrlinger KR, Cummings JR, Barnardo MC, Schwab M, Ahmad T, Jewell DP. The pharmacogenetics of methotrexate in inflammatory bowel disease. *Pharmacogenet Genomics* 2005; **15**:705–711.
- 72 Fisher MC, Cronstein BN. Metaanalysis of methylenetetrahydrofolate reductase (MTHFR) polymorphisms affecting methotrexate toxicity. *J Rheumatol* 2009; **36**:539–545.
- 73 Schmiegelow K. Advances in individual prediction of methotrexate toxicity: a review. *Br J Haematol* 2009; **146**:489–503.
- 74 Gervasini G. Polymorphisms in methotrexate pathways: what is clinically relevant, what is not, and what is promising. *Curr Drug Metab* 2009; **10**:547–566.
- 75 Van Ede AE, Laan RF, Blom HJ, De Abreu RA, Van de Putte LB. Methotrexate in rheumatoid arthritis: an update with focus on mechanisms involved in toxicity. *Semin Arthritis Rheum* 1998; **27**:277–292.
- 76 Tian H, Cronstein BN. Understanding the mechanisms of action of methotrexate: implications for the treatment of rheumatoid arthritis. *Bull NYU Hosp Jt Dis* 2007; **65**:168–173.
- 77 Merrill JT, Shen C, Schreiber D, Coffey D, Zakharenko O, Fisher R, et al. Adenosine A1 receptor promotion of multinucleated giant cell formation by human monocytes: a mechanism for methotrexate-induced nodulosis in rheumatoid arthritis. *Arthritis Rheum* 1997; **40**:1308–1315.

- 78 Kooloos WM, Wessels JA, Van der Straaten T, Allaart CF, Huizinga TW, Guchelaar HJ. Functional polymorphisms and methotrexate treatment outcome in recent-onset rheumatoid arthritis. *Pharmacogenomics* 2010; **11**:163–175.
- 79 Trevino LR, Shimasaki N, Yang W, Panetta JC, Cheng C, Pei D, *et al.* Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol* 2009; **27**:5972–5978.
- 80 Hider SL, Thomson W, Mack LF, Armstrong DJ, Shadforth M, Bruce IN. Polymorphisms within the adenosine receptor 2a gene are associated with adverse events in RA patients treated with MTX. *Rheumatology (Oxford)* 2008; **47**:1156–1159.
- 81 Pierik M, Rutgeerts P, Vlietinck R, Vermeire S. Pharmacogenetics in inflammatory bowel disease. *World J Gastroenterol* 2006; **12**:3657–3667.
- 82 Herrlinger KR, Jewell DP. Review article: interactions between genotype and response to therapy in inflammatory bowel diseases. *Aliment Pharmacol Ther* 2006; **24**:1403–1412.
- 83 Vucic N, Frleta M, Petrovic D, Ostojic V. Thrombophilia, preeclampsia and other pregnancy complications. *Acta Med Croatica* 2009; **63**:297–305.
- 84 Kurzawska G, Seremak-Mrozikiewicz A, Drews K, Barlik M, Mrozikiewicz PM. Genetic conditioned changes in activity of 5,10-methylenetetrahydrofolate reductase (MTHFR) and recurrent miscarriages. *Ginek Pol* 2009; **80**:762–767.
- 85 Greene ND, Stanier P, Copp AJ. Genetics of human neural tube defects. *Hum Mol Genet* 2009; **18**:R113–R129.
- 86 Frankenburg FR. The role of one-carbon metabolism in schizophrenia and depression. *Harv Rev Psychiatry* 2007; **15**:146–160.
- 87 Taioli E, Garza MA, Ahn YO, Bishop DT, Bost J, Budai B, *et al.* Meta- and pooled analyses of the methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism and colorectal cancer: a HuGE-GSEC review. *Am J Epidemiol* 2009; **170**:1207–1221.
- 88 Trabetti E. Homocysteine, MTHFR gene polymorphisms, and cardio-cerebrovascular risk. *J Appl Genet* 2008; **49**:267–282.
- 89 Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol* 1999; **149**:981–983.
- 90 Harten P. Reducing toxicity of methotrexate with folic acid. *Z Rheumatol* 2005; **64**:353–358.

Effect of hepatitis C virus on immunological and virological responses in HIV-infected patients initiating highly active antiretroviral therapy: a meta-analysis

C. G. Tsiara,¹ G. K. Nikolopoulos,^{1,2} N. L. Dimou,³ P. G. Bagos,³ G. Saroglou,⁴ E. Velonakis⁵ and A. Hatzakis⁶

¹Hellenic Centre for Disease Control and Prevention, Athens, Greece; ²National Development and Research Institutes (IAS/NIDA Post-Doctoral Research Fellow), New York, NY, USA; ³Department of Computer Science and Biomedical Informatics, University of Central Greece, Lamia, Greece; ⁴Department of Internal Medicine, Faculty of Nursing, University of Athens, Athens, Greece; ⁵Department of Public Health, Laboratory of Prevention, Faculty of Nursing, University of Athens, Athens, Greece; and ⁶Department of Hygiene, Epidemiology and Medical Statistics, Medical School, University of Athens, Athens, Greece

Received December 2012; accepted for publication February 2013

SUMMARY. Co-infection of human immunodeficiency virus (HIV) with hepatitis C virus (HCV) is rather common. In the era of highly active antiretroviral therapy (HAART), viral hepatitis could result in adverse outcomes in HIV+ patients. The current meta-analysis aims to evaluate the impact of HCV on immunological and virological responses after HAART initiation in HIV/HCV co-infected individuals by synthesizing the existing scientific evidence. A comprehensive search of electronic databases was performed. Eligible studies were analysed using univariate and multivariate meta-analytic methods. Totally, 21 studies involving 22533 individuals were eligible. The estimated summary difference in CD4 cell counts increase between HIV and HIV/HCV co-infected subjects after 3–12 months on HAART was 34.86 cells/mm³ [95% confidence interval (CI): 16.82–52.89]. The difference was more prominent in patients with baseline

CD4 counts below 350 cells/mm³ (38.97, 95% CI: 20.00–57.93) and attenuated 2 years later (13.43, 95% CI: 0.83–26.04). The analysis of ratio measures yielded similar findings. The virological control remained unaffected by the presence of HCV (adjusted Hazard Ratio for co-infected patients vs those with HIV alone: 0.99, 95% CI: 0.91–1.07). The bivariate meta-analytic method confirmed the results of the univariate approaches. This meta-analysis supports the adverse effect of HCV on immune recovery of HIV+ patients initiating HAART, especially of those with initially impaired immunologic status. Although this effect diminishes over time, early administration of HAART in the setting of co-infection seems to be justified.

Keywords: antiretroviral treatment, HCV, HIV, immunological response, virological response.

INTRODUCTION

The introduction of highly active antiretroviral therapy (HAART) led to an impressive decrease in morbidity and

mortality of people infected by human immunodeficiency virus (HIV) [1,2], but their prolonged survival allowed the emergence of co-morbidities, including infection with hepatitis C virus (HCV).

Double infection by HIV and HCV is common due to similar modes of transmission. Globally, nearly, 20 per cent of individuals with HIV have chronic hepatitis C [3]. Among HIV+ injecting drug users (IDUs), the prevalence of HCV infection ranges between 82 and 93 per cent [4–7]. HCV incidence has also increased in HIV+ male homosexuals during the last decade [8].

Primary studies assessing the impact of HCV on HIV disease progression have yielded conflicting results [9–12]. A recent meta-analysis showed that HIV/HCV co-infection did not affect the incidence of acquired immune deficiency syndrome (AIDS) but increased overall mortality in the HAART era [13]. In terms of early HAART effectiveness, a quantitative synthesis in 2005 showed a less robust immune reconstitution in co-infected patients who initiated antiretroviral treatment [14].

Abbreviations: AIDS, Acquired Immune Deficiency Syndrome; ART, Antiretroviral Treatment; c-ART, combined Antiretroviral Treatment; CI, Confidence Interval; HAART, Highly Active Antiretroviral Therapy; HCV, Hepatitis C Virus; HIV, Human Immunodeficiency Virus; HR, Hazard Ratio; I², I-squared index; IDUs, Injecting Drug Users; IR, Immunologic Response; NA, Not Available; OR, Odds Ratio; RCTs, Randomized Clinical Trials; RR, Relative Risk; UVL, Undetected Viral Load; VL, Viral Load; VR, Virologic Response.

Correspondence: Angelos Hatzakis, Professor of Epidemiology and Preventive Medicine, Director, Department of Hygiene, Epidemiology and Medical Statistics, Head, National Retrovirus Reference Center, Medical School, University of Athens M. Asias 75, GR-11527, Athens, Greece. E-mail: ahatzak@med.uoa.gr

In the meantime, more relevant studies have been undertaken. The new pieces of information and the lack of a summary appraisal of virological response to HAART in the HIV/HCV setting warranted the conduct of an updated systematic review. The aim of the current meta-analysis was to combine and evaluate the accumulated scientific evidence concerning the effect of HCV on immunological and virological response in HIV-infected patients receiving HAART or combined antiretroviral treatment (c-ART) including at least 3 drugs.

MATERIALS AND METHODS

Eligibility criteria, data sources and search strategy

This meta-analysis was performed according to the MOOSE guidelines [15] and the PRISMA statement [16]. Studies of HIV+ patients were included in the analysis if: (i) they had examined the effect of HIV/HCV co-infection on immunological and/or virological response of persons receiving HAART or c-ART, (ii) they had included HAART-naïve patients at baseline and (iii) they had provided a value for CD4 T-cell counts or HIV-RNA levels difference from baseline to 3–24 months after ART initiation, or an estimate of Relative Risk (RR) for immunological and/or virological response, or sufficient data to calculate these measures. No language restrictions were imposed.

Pertinent studies were identified in Medline, Scopus and ISI Web of Science using a combination of the following terms: 'HIV', 'HCV', 'HAART', 'antiretroviral treatment', 'CD4 T', 'HIV-RNA' and 'viral load' (last search: February 2012). Titles and abstracts were screened to exclude irrelevant records. Full-text versions of the remaining articles were evaluated for eligibility. References of relevant publications and conference abstracts were also appraised to identify further studies for inclusion.

Data extraction

Data were extracted by C.T. and G.N., and discrepancies were resolved by consensus. For each eligible study, the following information, if available, was retrieved: (i) first author's name, year of publication, geographic setting and study design, (ii) diagnostic method for HCV infection, (iii) number of participants and length of follow up, (iv) patients' baseline characteristics and laboratory data, (v) mean differences (with their standard deviations) in CD4 T-cell counts and HIV-RNA levels between measurements at the time of HAART or c-ART initiation and those 3–24 months later and (vi) RRs of achieving a CD4 T-cell increase and/or a HIV-RNA reduction with the corresponding 95% confidence intervals (CIs).

Statistical analysis

Unstandardized mean differences and their 95% CIs were analysed using fixed- and random-effects models [17]. Combined RRs were derived from the univariate random-effects method [18]. Hazard Ratios (HRs) in individual studies, if not directly provided, were computed as suggested by Perneger *et al.* [19]. Heterogeneity tests included the Cochran's Q statistic [17] and the I-squared (I^2) index [20]. The presence of publication bias was investigated by Begg's [21] and Egger's [22] tests. Potential time trends were detected in cumulative analyses [23]. Sensitivity analyses were performed through subgroup investigations or by excluding studies one at a time. A bivariate approach was also implemented to account for the potential correlation of RRs for immunological and virological response [24].

All analyses were conducted in Stata 12 (Stata Corporation, College Station, Texas, USA). Except for heterogeneity statistics (significance was declared if $P < 0.10$), the results were considered significant if the corresponding P value was less than 0.05. All P values were two tailed.

Unless stated differently, only random-effects estimates are presented.

RESULTS

Description of eligible studies

The literature search yielded 31 relevant studies. Ten of them were excluded because they provided insufficient data or were part of multiple publications (Fig. 1). Finally, 21 studies involving 5278 HIV/HCV co-infected patients and 17255 individuals infected by HIV alone were eligible. Among them, 10 included information on immunologic response [9,12,25–32] and 11 about both outcomes [33–43] (Table 1). Immunological response was measured either as CD4 T-cell count change from baseline to 3–24 months after ART initiation [9,12,25,26,28,29,31,32,35–37,40–43] and/or as a CD4 T-cell count increase of at least 50, 75, 100, or 200 cells/mm³ [9,12,26,27,29,30,32–35,38,39]. Virological response was mostly defined as decreasing plasma HIV-RNA levels below 50, 400, or 500 copies/mL or under the limit of detection [33–42]. In 7 studies, the diagnosis of HCV was made on the basis of a positive anti-HCV test and of HCV-RNA detection in plasma [25,27,29,32,36,42,43]. The majority of eligible studies were prospective cohorts [9,12,25,27,30,32–35,37,39,42].

Immunological and virological response

Immunological response

There was no difference between the baseline CD4 T-cell counts of the two patient groups ($P > 0.05$). As presented in Table 2 and in Fig. 2, the summary difference

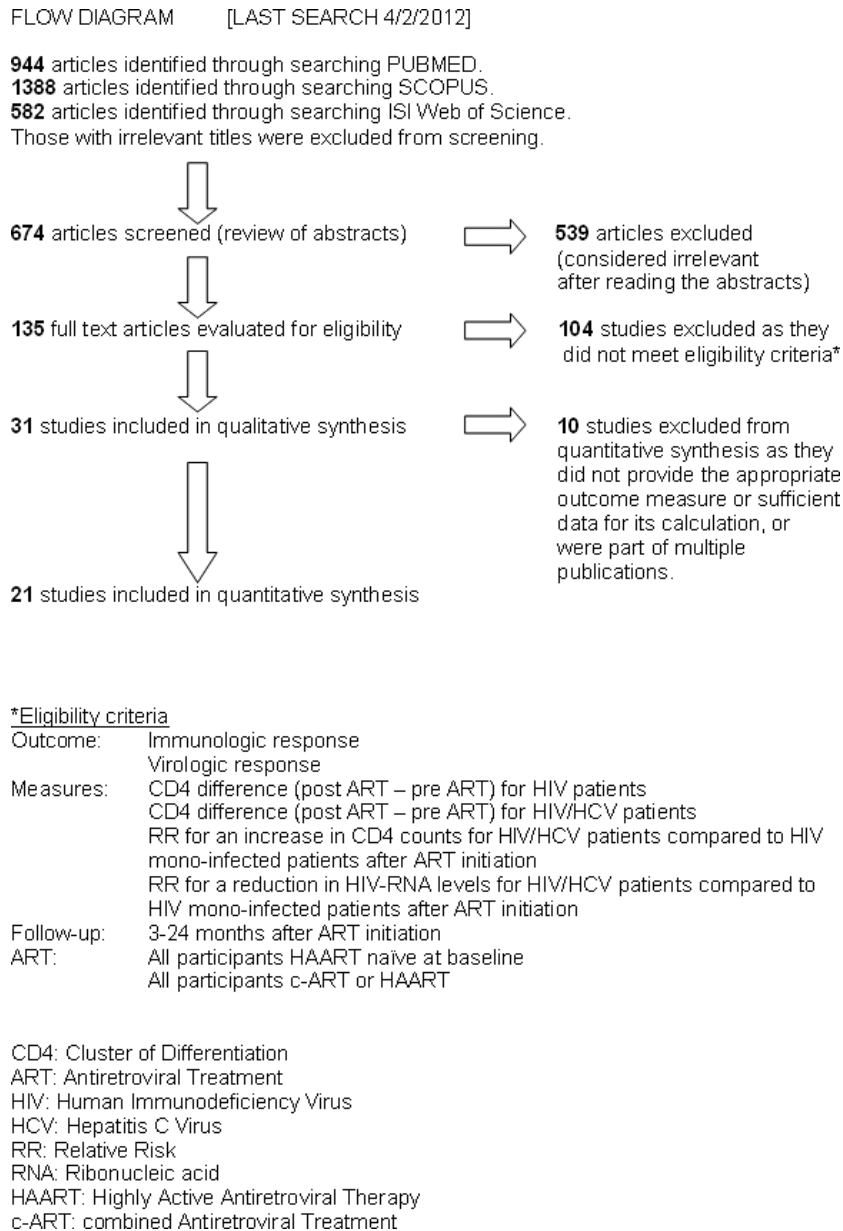


Fig. 1 Identification process of eligible studies.

in CD4 T-cell counts increase between HIV+ and HIV/HCV co-infected subjects in 12 studies after 3–12 months on HAART or c-ART was 34.86 cells/mm³ (95% CI: 16.82–52.89, I² = 92.3%). This difference became bigger when the baseline CD4 T-cell counts were below 350 cells/mm³ (38.97, 95% CI: 20.00–57.93) (Table 2). The results were nonsignificant (27.26, 95% CI: –2.21 to 56.72) (Table 2) in the analysis of studies with dual (serologic and molecular) diagnosis of HCV. The between groups difference attenuated 2 years after ART initiation (13.43, 95% CI: 0.83–26.04, I² = 0.0%) (Table 2). The results of all subgroup analyses are presented in Table 2.

The quantitative synthesis of HRs for increased CD4 T-cell counts (50 or 75 or 100 cells/mm³) in co-infected versus mono-infected patients after the commencement of ART yielded also significant estimates (summary adjusted HR of 8 studies: 0.82, 95% CI: 0.75–0.91) (Table 2).

Virological response

The difference in virological response was examined in 10 studies involving 13394 patients. The summary adjusted HRs for achieving undetectable viral load or HIV-RNA levels less than 50, 400 or 500 copies/mL after HAART or c-ART initiation for HIV/HCV patients compared with HIV

Table 1 Description of primary studies

Study	Type	N		ART-naïve (%)		ART regimen	Endpoint	Outcome measure
		HIV	HIV/HCV	HIV	HIV/HCV			
Santin <i>et al.</i> [25]	Prospective cohort	24	15	100	100	Combined ART, at least 3 drugs	48 weeks after ART initiation	CD4 difference
Macias <i>et al.</i> [32]	Prospective cohort	58	60	100	100	HAART	12 months after HAART initiation	CD4 difference*, HR of CD4 recovery of 50 [†] and 200 cells/mm ³
Braitstein <i>et al.</i> [26]	Nested cohort	580	606	100	100	HAART	48 weeks after HAART initiation, time to IR	CD4 difference, HR of CD4 recovery of ≥ 75 cells/mm ³
Sulkowski <i>et al.</i> [12]	Prospective cohort	122	65	NA	NA	HAART	≥ 1 year ≥ 2 years ≥ 3 years after HAART initiation	CD4 difference, HR of CD4 recovery of ≥ 50 cells/mm ^{3†} , HR of CD4 recovery of ≥ 100 cells/mm ^{3†}
Filippini <i>et al.</i> [42]	Prospective cohort	8	8	NA	NA	HAART	1, 3, 6 months after HAART initiation	CD4 difference [†] , VL pre and after, HR of achieving UVL [†]
Antonucci <i>et al.</i> [27]	Prospective cohort	1219	284	100	100	HAART	Time to IR	HR of CD4 recovery of ≥ 100 cells/mm ³
Castagna <i>et al.</i> [28]	Retrospective multicohort	1053	435	NA	NA	HAART	12, 24 weeks after HAART initiation	CD4 difference
Seminari <i>et al.</i> [29]	Retrospective cohort	673	171	47	35	HAART	At time of UVL (<12 months), 12 months after suppressive HAART (12 < .<24 months)	CD4 difference, OR of CD4 recovery of ≥ 100 cells/mm ³
De Luca <i>et al.</i> [34]	Prospective cohort	674	556	100	100	Combined ART, at least 3 drugs	Time to IR and VR	HR of CD4 recovery of ≥ 100 cells/mm ³ or of achievement of 500 cells/mm ³ , HR of achieving VL<500 copies/mL
Rockstroh <i>et al.</i> [33]	Prospective cohort	1492	768	NA	NA	HAART	Time to IR and VR	HR of CD4 recovery of ≥ 50 cells/mm ³ , HR of achieving VL<500 copies/mL

(continued)

Table 1 (continued)

Study	Type	N		ART-naïve (%)		ART regimen	Endpoint	Outcome measure
		HIV	HIV/HCV	HIV	HIV/HCV			
Klein <i>et al.</i> [35]	Prospective cohort	263	52	NA	NA	HAART	24 months after HAART initiation, time to IR and VR	CD4 difference, HR of CD4 recovery of ≥ 50 cells/mm ³ , HR of achieving VL<500 copies/mL
Chung <i>et al.</i> [36]	Retrospective cohort	129	40	NA	NA	HAART	16, 48 weeks after HAART initiation	CD4 difference, HR of achieving VL<500 copies/mL [†]
Zhou <i>et al.</i> [37]	Prospective cohort	371	43	NA	NA	HAART	180 days after HAART initiation	CD4 difference, HR of achieving VL<400 copies/mL [†]
Carmo <i>et al.</i> [38]	Retrospective cohort	748	76	56	34	HAART	Time to IR and VR	HR of CD4 recovery of ≥ 50 cells/mm ³ , HR of achieving VL<400 copies/mL, HR of CD4 recovery of >50 cells/mm ³ , HR of achieving VL<50 copies/mL
Turner <i>et al.</i> [39]	Prospective cohort	5500	396	100	100	HAART	Time to IR and VR	CD4 difference, HR of achieving VL<50 copies/mL
Stapleton <i>et al.</i> [40]	Longitudinal cohort	1238	148	100	100	Combined ART (at least 3 drugs) and HAART	48 weeks after ART initiation	HR of achieving VL<50 copies/mL [†]
Lincoln <i>et al.</i> [41]	Retrospective cohort	772	112	NA	NA	HAART	12, 24 months after HAART initiation	CD4 difference, HR of achieving UVL [†]
Zhao <i>et al.</i> [43]	Retrospective cohort	178	168	100	100	HAART	1 year after HAART initiation	CD4 difference [†] , VL pre and after
Isa <i>et al.</i> [30]	Prospective cohort	149	68	100	100	HAART	3, 6 months after HAART initiation	HR of CD4 recovery of ≥ 50 cells/mm ³ [†]
Emokpae <i>et al.</i> [31]	Retrospective cohort	50	50	100	100	HAART	6 months after HAART initiation	CD4 difference [†]
Greub <i>et al.</i> [9]	Prospective cohort	1954	1157	52	59	HAART	12 months after HAART initiation	CD4 difference [*] , HR of CD4 recovery of ≥ 50 cells/mm ³

*Data provided by the meta-analysis of Miller *et al.* [14][†]Calculated by provided data.

N, number of subjects; HIV, Human immunodeficiency virus; HCV, Hepatitis C virus; ART, Antiretroviral treatment; CD4, Cluster of differentiation; HAART, Highly active antiretroviral therapy; HR, Hazard ratio; IR, Immunological response; NA, Not available; VL, Viral load; UVL, Undetected VL; OR, Odds ratio; VR, Virological response.

Table 2 Results of meta-analyses of studies evaluating the difference in immunological and virological responses between HIV and HIV/HCV patients receiving antiretroviral treatment

Outcome	Outcome measure	Endpoint	Number of HIV - HIV/HCV patients	Number of studies	Summary Estimate	95% Confidence interval	I ² (%)
Immunological response	CD4 difference (cells/mm ³)	3–12 months	5834, 2604	12	34.86	16.82–52.89	92.3
Subgroup analysis							
	HCV diagnosis based only on positive HCV serology		4764, 2142	6	38.91	15.51–62.30	95.5
	HCV diagnosis based on positive HCV serology and detection of HCV-RNA		1070, 462	6	27.26	–2.21–56.72	73.4
	ART-naïve at baseline		2127, 1047	6	51.87	21.66–82.08	95.4
	ART experienced (HAART-naïve) at baseline		3707, 1557	6	21.77	9.19–34.36	53.1
	Baseline CD4 < 350		5288, 2497	9	38.97	20.00–57.93	91.7
	Baseline CD4 ≥ 350		175, 64	3	1.61	–88.68–91.89	89.8
Immunological response	CD4 difference (cells/mm ³)	13–24 months	1830, 396	4	13.43	0.83–26.04	0.0
Immunological response	Adjusted HR for HIV/HCV vs HIV	Time to IR	12402, 3492	8	0.82	0.75–0.91	53.9
Immunological response	Unadjusted HR for HIV/HCV vs HIV	Time to IR	5523, 2550	8	0.72	0.64–0.81	57.6
Virological response	Adjusted HR for HIV/HCV vs HIV	Time to VR	8677, 1848	5	0.99	0.91–1.07	0.0
Virological response	Unadjusted HR for HIV/HCV vs HIV	Time to VR	4949, 1244	8	0.89	0.76–1.04	50.5

HIV, Human immunodeficiency virus; HCV, Hepatitis C virus; CD4, Cluster of differentiation; RNA, Ribonucleic acid; ART, Antiretroviral treatment; HAART, Highly active antiretroviral therapy; HR, Hazard ratio; IR, Immunological response; VR, Virological response.

mono-infected individuals was 0.99 (95% CI: 0.91–1.07, I²: 0.0%) (Table 2, Fig. 3).

Bivariate meta-analysis

The bivariate synthesis produced similar findings. The adjusted HRs for immunological and virological responses in the HIV/HCV group compared with those patients infected with HIV alone were 0.82 (95% CI: 0.74–0.91) and 0.96 (95% CI: 0.89–1.04), respectively.

Publication bias

In all analyses, the statistical tests of Begg and Egger supported the absence of publication bias.

DISCUSSION

This meta-analysis of 21 studies demonstrated that HCV co-infection adversely affects early immune responses in HIV+ patients who start HAART or c-ART, especially those with baseline CD4 T-cell counts below 350 cells/mm³. This impact was less pronounced 2 years after ART initiation. The virological suppression in persons receiving potent antiretroviral therapy seems to remain unaffected by the simultaneous existence of HCV.

HIV/HCV co-infection hastens HCV-associated hepatic disease in the HAART era [44–47]. The reciprocal effect of HCV on the natural history of HIV disease is not clarified.

Our meta-analysis showed the negative impact of HCV on early CD4 T-cell counts recovery in HIV+ subjects. This phenomenon has biological basis. CD4 T-cells can be infected by HCV and interactions of HIV and HCV at the cellular level could affect the immune efficacy of HAART [48]. Moreover, chronic HCV infection, in the presence of HIV, increases T-cell immune activation, which is known to limit CD4 T-cell gains [47,49–51]. Finally, HCV has also been found to enhance CD4 T-cell apoptosis [52]. As a matter of fact, if apoptosis is partly implicated in the poorer CD4 T-cell recovery of co-infected patients, the restoration of their CD4 T-cells over time that was found in this meta-analysis dovetails with the observation that HIV replication control by HAART gradually attenuates the CD4 T-cell apoptosis rates in HIV+ patients with concomitant HCV infection [52].

This meta-analysis showed that there was a comparable virological response to HAART in HIV+ persons regardless of HCV status. By contrast, a recent analysis of randomized clinical trials (RCTs) reported that HCV carriage was associated with altered HIV-RNA suppression in co-infected individuals [53]. This report, however, considered studies that were not designed to assess the impact of co-infection as the primary outcome of their analysis, recruited HAART-experienced patients and lacked clear description of meta-analytic techniques. Because most of the results of these trials were presented as abstracts, a new analysis of

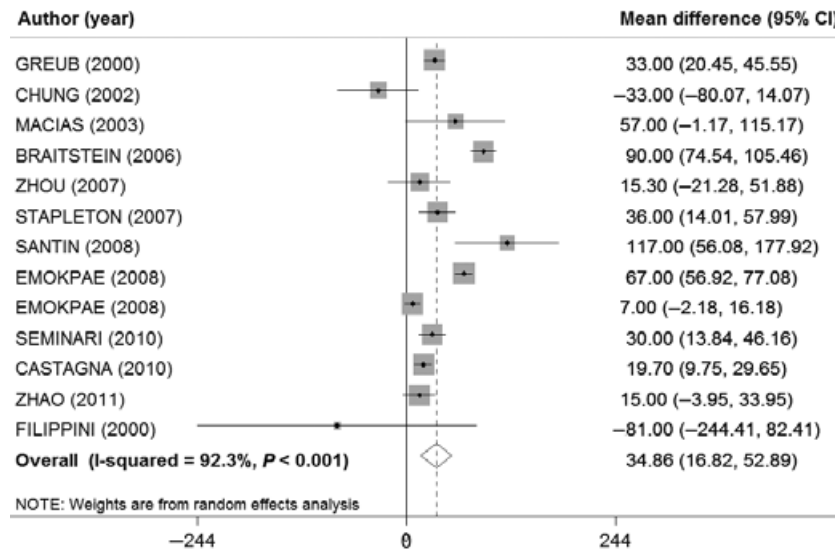


Fig. 2 Univariate random-effects meta-analysis concerning the difference in CD4 T-cell increase between HIV and HIV/HCV subjects after initiation of antiretroviral treatment. Note: CI: Confidence interval.

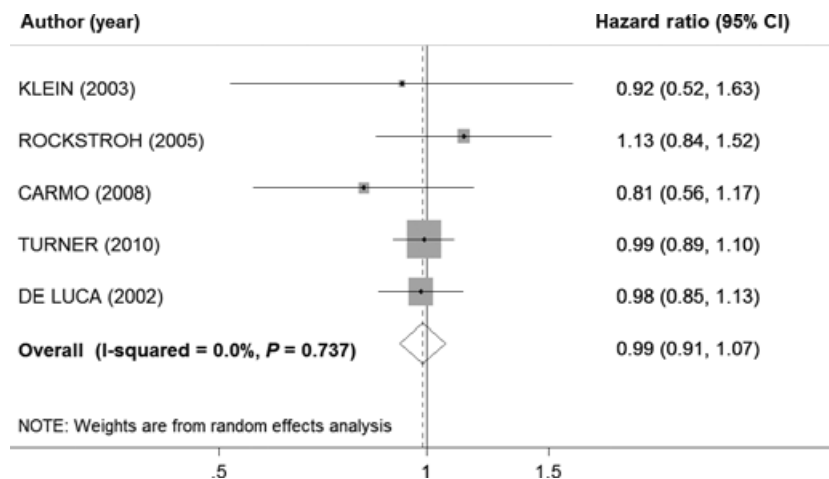


Fig. 3 Univariate random-effects meta-analysis of adjusted Hazard Ratios for virological response in the HIV/HCV group compared with the HIV mono-infection group after initiation of antiretroviral treatment. Note: CI: Confidence interval.

forthcoming full-text publications or of individual data would help clarify potential discrepancies between RCTs and observational research.

The meta-analysis of CD4 T-cell count change was characterized by substantial between-studies heterogeneity that predicated the conduct of subgroup analyses to explore its sources and potential impact. Previous receipt of antiretrovirals could result in heterogeneous effect estimates. Although it would be ideal to consider not only HAART (as we did) but also ART-naïve individuals in this meta-analysis, some of the eligible studies had recruited patients with ART experience. Previous ART therapy could compromise HAART adherence, HIV load control and immunological response [54]. However, as shown in Table 2, in spite

of unexpected differences in magnitude, the effect estimates were statistically significant in both cases indicating a diminished CD4 T-cell response in HIV/HCV individuals irrespective of prior ART exposure.

It is important to diagnose HCV infection based on HCV-RNA detection to avoid misclassification of HCV antibody positive patients who are HCV aviremic [55]. To explore the effect of potential misclassification, the meta-analysis of CD4 T-cell increase was performed separately by method of HCV diagnosis. The subanalyses showed that the results ceased being statistically significant in studies that adopted both serological and molecular testing. Although this could be a chance finding, it certainly calls for more accurate and consistent across studies definition of HCV infection.

The negative effect of co-infection on immunological response to HAART was observed only in patients with baseline CD4 T-cell counts below 350 cells/mm³. This finding supports the current European Guidelines for Antiretroviral Treatment [56], which recommend immediate ART administration in HIV/HCV co-infected patients when CD4 T-cell counts drop below 500 cells/mm³. It should be noted, however, that this meta-analysis included a limited number of studies that had involved patients with initial CD4 counts above 350 cells/mm³. Future research needs to address this stratifying always immunological response by baseline CD4 T-cell counts.

The differences in CD4 T-cell recovery could be attributed to unmeasured confounding effects. For instance, the rates of HCV infection are extremely high among IDUs and previous research has shown that CD4 T-cell recovery is compromised in drug injectors compared with other groups [57], although this finding was not corroborated in other studies [58,59]. It seems, however, that the immunological response of IDUs to ART could be influenced by many factors that result in poor adherence to the prescribed regimen including their potential incarceration, their participation in substitution programmes, their current injection status, the presence of psychiatric conditions or the provision of psychological support [60–63]. The optimal way to elucidate the confounding effect of HIV risk groups in CD4 T-cell increase is to compare HIV mono-infected and HIV/HCV co-infected individuals by HIV risk group. Unfortunately, many of the eligible studies in this meta-analysis lacked the necessary information to perform this type of subgroup investigations.

The current work updates a previous synthesis [14] using thirteen more reports, performing many subgroup analyses, and implementing some new meta-analytic methods.

It indicates a delayed immunological response in HIV/HCV co-infected patients receiving potent combinations of anti-retrovirals, especially among those with CD4 T-cell counts below 350 cells/mm³. The delayed recovery of CD4 T-cells during the first year of HAART could increase the risk of toxicities or non-AIDS events. However, the clinical significance of the blunted CD4 T-cell response in the first months of HAART therapy is not fully elucidated yet and future prospective studies should explore its short-term and long-term consequences.

AUTHORS' CONTRIBUTION

Chrissa G. Tsiara: Acquisition, analysis and interpretation of data; drafting the manuscript; approval of the final version. Georgios K. Nikolopoulos: Analysis and interpretation of data; drafting the manuscript; approval of the final version. Niki L. Dimou: Analysis and interpretation of data; critical revision of the manuscript; approval of the final version. Pantelis G. Bagos: Analysis and interpretation of data; critical revision of the manuscript; approval of the final version. Georgios Saroglou: Conception and design; critical revision of the manuscript; approval of the final version. Emmanuel Velonakis: Conception and design; critical revision of the manuscript; approval of the final version. Angelos Hatzakis: Conception and design; critical revision of the manuscript; approval of the final version.

CONFLICTS OF INTEREST AND SOURCE OF FUNDING

None were declared.

REFERENCES

- 1 UNAIDS. Global report: UNAIDS report on the global AIDS epidemic 2010. Available at: http://www.unaids.org/globalreport/global_report.htm (accessed 3 September 2012).
- 2 Puhan MA, Van Natta ML, Palella FJ, Addessi A, Meinert C, Ocular Complications of AIDS Research Group. Excess mortality in patients with AIDS in the era of highly active antiretroviral therapy: temporal changes and risk factors. *Clin Infect Dis* 2010; 51: 947–956.
- 3 Soriano V, Vispo E, Labarga P, Medrano J, Barreiro P. Viral hepatitis and HIV co-infection. *Antiviral Res* 2010; 85: 303–315.
- 4 Nelson PK, Mathers BM, Cowie B *et al.* Global epidemiology of hepatitis B and hepatitis C in people who inject drugs: results of systematic reviews. *Lancet* 2011; 378: 571–583.
- 5 Larsen C, Pialoux G, Salmon D *et al.* Prevalence of hepatitis C and hepatitis B infection in the HIV-infected population of France, 2004. *Euro Surveill* 2008; 13: 1–4. pii=18888. Available at: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=18888> (accessed 3 September 2012).
- 6 Mohsen AH, Murad S, Easterbrook PJ. Prevalence of hepatitis C in an ethnically diverse HIV-1-infected cohort in south London. *HIV Med* 2005; 6: 206–215.
- 7 Solomon SS, Srikrishnan AK, Mehta SH *et al.* High prevalence of HIV, HIV/hepatitis C virus co-infection, and risk behaviors among injection drug users in Chennai, India: a cause for concern. *J Acquir Immune Defic Syndr* 2008; 49: 327–332.
- 8 Danta M, Rodger AJ. Transmission of HCV in HIV-positive populations. *Curr Opin HIV AIDS* 2011; 6: 451–458.
- 9 Greub G, Ledergerber B, Battegay M *et al.* Clinical progression, survival, and immune recovery during antiretroviral therapy in patients with HIV-1 and hepatitis C virus co-infection.

- tion: the Swiss HIV Cohort Study. *Lancet* 2000; 356: 1800–1805.
- 10 Anderson KB, Guest JL, Rimland D. Hepatitis C virus co-infection increases mortality in HIV-infected patients in the highly active antiretroviral therapy era: data from the HIV Atlanta VA Cohort Study. *Clin Infect Dis* 2004; 39: 1507–1513.
 - 11 Macías J, Pineda JA, Leal M *et al.* Influence of hepatitis C virus infection on the mortality of antiretroviral-treated patients with HIV disease. *Eur J Clin Microbiol Infect Dis* 1998; 17: 167–170.
 - 12 Sulkowski MS, Moore RD, Mehta SH, Chaisson RE, Thomas DL. Hepatitis C and progression of HIV disease. *JAMA* 2002; 288: 199–206.
 - 13 Chen TY, Ding EL, Seage GR III, Kim AY. Meta-analysis: increased mortality associated with hepatitis C in HIV-infected persons is unrelated to HIV disease progression. *Clin Infect Dis* 2009; 49: 1605–1615.
 - 14 Miller MF, Haley C, Koziel MJ, Rowley CF. Impact of hepatitis C virus on immune restoration in HIV-infected patients who start highly active antiretroviral therapy: a meta-analysis. *Clin Infect Dis* 2005; 41: 713–720.
 - 15 Stroup DF, Berlin JA, Morton SC *et al.* Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA* 2000; 283: 2008–2012.
 - 16 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010; 8: 336–341.
 - 17 Petitti DB. Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis, 2nd ed. New York: Oxford University Press, 2000.
 - 18 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–188.
 - 19 Pernerger T. Estimating the relative hazard by the ratio of logarithms of event-free proportions. *Contemp Clin Trials* 2008; 29: 762–766.
 - 20 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–560.
 - 21 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; 50: 1088–1101.
 - 22 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; 315: 629–634.
 - 23 Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med* 1992; 327: 248–254.
 - 24 Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics* 2008; 9: 172–186.
 - 25 Santin M, Mestre M, Shaw E *et al.* Impact of hepatitis C virus co-infection on immune restoration during successful antiretroviral therapy in chronic human immunodeficiency virus type 1 disease. *Eur J Clin Microbiol Infect Dis* 2008; 27: 65–73.
 - 26 Braitstein P, Zala C, Yip B *et al.* Immunologic response to antiretroviral therapy in hepatitis C virus-co-infected adults in a population-based HIV/AIDS treatment program. *J Infect Dis* 2006; 193: 259–268.
 - 27 Antonucci G, Girardi E, Cozzi-Lepri A *et al.* HepaI.Co.N.A. Study Group, Italian Cohort Naive for Antiretrovirals Study Group. Role of hepatitis C virus (HCV) viremia and HCV genotype in the immune recovery from highly active antiretroviral therapy in a cohort of antiretroviral-naïve HIV-infected individuals. *Clin Infect Dis* 2005; 40: e101–e109.
 - 28 Castagna A, Galli L, Torti C *et al.* Predicting the magnitude of short-term CD4 (+) T-cell recovery in HIV-infected patients during first-line highly active antiretroviral therapy. *Antivir Ther* 2010; 15: 165–175.
 - 29 Seminari E, Tinelli C, Ravasi G *et al.* MASTER Cohort. Hepatitis C infection on immune recovery in HIV-positive patients on successful HAART: the role of genotype 3. *Curr HIV Res* 2010; 8: 186–193.
 - 30 Isa SE, Gwamzhi LN, Akolo C, Ginyan J. A prospective cohort study of immunologic and virologic outcomes in patients with HIV/AIDS and hepatitis virus co-infection in Jos, Nigeria. *Niger J Med* 2010; 19: 279–285.
 - 31 Emokpae MA, Nwokedi EE, Jegede EE. Impact of hepatitis C co-infection on CD4 T cell count in HIV infected subjects. *OJHAS* 2008; 7: 1–4.
 - 32 Macías J, Pineda JA, Lozano F *et al.* Impaired recovery of CD4 + cell counts following highly active antiretroviral therapy in drug-naïve patients coinfecting with human immunodeficiency virus and hepatitis C virus. *Eur J Clin Microbiol Infect Dis* 2003; 22: 675–680.
 - 33 Rockstroh JK, Mocroft A, Soriano V. *et al.* EuroSIDA Study Group. Influence of hepatitis C virus infection on HIV-1 disease progression and response to highly active antiretroviral therapy. *J Infect Dis* 2005; 192: 992–1002.
 - 34 De Luca A, Bugarini R, Lepri AC *et al.* Italian Cohort Naive Antiretrovirals Study Group. Co-infection with hepatitis viruses and outcome of initial antiretroviral regimens in previously naive HIV-infected subjects. *Arch Intern Med* 2002; 162: 2125–2132.
 - 35 Klein MB, Lalonde RG, Suissa S. The impact of hepatitis C virus co-infection on HIV progression before and after highly active antiretroviral therapy. *J Acquir Immune Defic Syndr* 2003; 33: 365–372.
 - 36 Chung RT, Evans SR, Yang Y *et al.* AIDS Clinical Trials Group 383 Study Team. Immune recovery is associated with persistent rise in hepatitis C virus RNA, infrequent liver test flares, and is not impaired by hepatitis C virus in co-infected subjects. *AIDS* 2002; 16: 1915–1923.
 - 37 Zhou J, Dore GJ, Zhang F, Lim PL, Chen YM, TREAT Asia HIV Observational Database. Hepatitis B and C virus co-infection in The TREAT Asia HIV Observational Database. *J Gastroenterol Hepatol* 2007; 22: 1510–1518.

- 38 Carmo RA, Guimarães MD, Moura AS *et al.* The influence of HCV co-infection on clinical, immunological and virological responses to HAART in HIV-patients. *Braz J Infect Dis* 2008; 12: 173–179.
- 39 Turner J, Bansi L, Gilson R *et al.* UK Collaborative HIV Cohort (UK CHIC) Study. The prevalence of hepatitis C virus (HCV) infection in HIV-positive individuals in the UK - trends in HCV testing and the impact of HCV on HIV treatment outcomes. *J Viral Hepat* 2010; 17: 569–577.
- 40 Stapleton JT, Bennett K, Bosch RJ, Polgreen PM, Swindells S. Effect of antiretroviral therapy and hepatitis C co-infection on changes in lipid levels in HIV-Infected patients 48 weeks after initiation of therapy. *HIV Clin Trials* 2007; 8: 429–436.
- 41 Lincoln D, Petoumenos K, Dore GJ, Australian HIV Observational Database. HIV/HBV and HIV/HCV co-infection, and outcomes following highly active antiretroviral therapy. *HIV Med* 2003; 4: 241–249.
- 42 Filippini P, Coppola N, Scolastico C *et al.* Can HCV affect the efficacy of anti-HIV treatment? *Arch Virol* 2000; 145: 937–944.
- 43 Zhao J, Yu ZJ, Kan QC, Li XF, Li ZQ, Liang HX. A retrospective study: the influence of human immunodeficiency virus co-infection with hepatitis C virus or hepatitis B virus on the efficacy with HAART in China AIDS area. *Life Science J* 2011; 8: 27–31.
- 44 Thein HH, Yi Q, Dore GJ, Krahn MD. Natural history of hepatitis C virus infection in HIV-infected individuals and the impact of HIV in the era of highly active antiretroviral therapy: a meta-analysis. *AIDS* 2008; 22: 1979–1991.
- 45 Martin-Carbonero L, de Ledinghen V, Moreno A *et al.* Liver fibrosis in patients with chronic hepatitis C and persistently normal liver enzymes: influence of HIV infection. *J Viral Hepat* 2009; 16: 790–795.
- 46 Smit C, van den Berg C, Geskus R, Berkhout B, Coutinho R, Prins M. Risk of hepatitis-related mortality increased among hepatitis C virus/HIV-co-infected drug users compared with drug users infected only with hepatitis C virus: a 20-year prospective study. *J Acquir Immune Defic Syndr* 2008; 47: 221–225.
- 47 Operskalski EA, Kovacs A. HIV/HCV co-infection: pathogenesis, clinical complications, treatment, and new therapeutic technologies. *Curr HIV/AIDS Rep* 2011; 8: 12–22.
- 48 Laskus T, Radkowski M, Jablonska J *et al.* Human immunodeficiency virus facilitates infection/replication of hepatitis C virus in native human macrophages. *Blood* 2004; 103: 3854–3859.
- 49 Hunt PW, Martin JN, Sinclair E *et al.* T cell activation is associated with lower CD4 T+ T cell gains in human immunodeficiency virus-infected patients with sustained viral suppression during antiretroviral therapy. *J Infect Dis* 2003; 187: 1534–1543.
- 50 Sandberg JK, Falconer K, Gonzalez VD. Chronic immune activation in the T cell compartment of HCV/HIV-1 co-infected patients. *Virulence* 2010; 1: 177–179.
- 51 Gonzalez VD, Falconer K, Blom KG *et al.* High levels of chronic immune activation in the T-cell compartments of patients coinfecting with hepatitis C virus and human immunodeficiency virus type 1 and on highly active antiretroviral therapy are reverted by alpha interferon and ribavirin treatment. *J Virol* 2009; 83: 11407–11411.
- 52 Korner C, Kramer B, Schulte D *et al.* Effects of HCV co-infection on apoptosis of CD4 T-cells in HIV positive patients. *Clin Sci* 2009; 116: 861–870.
- 53 Pulido F, Hill A, van Delft Y, Moeclinghoff C. Impact of hepatitis C co-infection on response to antiretroviral treatment. *AIDS Rev* 2012; 14: 124–131.
- 54 Horberg M, Silverberg M, Hurley L, Delorenze G, Quesenberry C. Influence of prior antiretroviral experience on adherence and responses to new highly active antiretroviral therapy regimens. *AIDS Patient Care STDS* 2008; 22: 301–312.
- 55 Thomas DL, Astemborski J, Rai RM *et al.* The natural history of hepatitis C virus infection: host, viral, and environmental factors. *JAMA* 2000; 284: 450–456.
- 56 European AIDS Clinical Society. European Guidelines for treatment of HIV infected adults in Europe - Version 6. Available at: <http://www.europeanaidsclinicalsociety.org/images/stories/EACS-Pdf/EACSGuidelines-v6.0-English.pdf> (accessed 3 September 2012).
- 57 Collazos J, Asensi V, Carton JA, Grupo Español para el Estudio Multifactorial de la Adherencia (GE-EMA). Association of HIV transmission categories with sociodemographic, viroimmunological and clinical parameters of HIV-infected patients. *Epidemiol Infect* 2010; 138: 1016–1024.
- 58 Mocroft A, Madge S, Johnson AM *et al.* A comparison of exposure groups in the EuroSIDA study: starting highly active antiretroviral therapy (HAART), response to HAART, and survival. *J Acquir Immune Defic Syndr* 1999; 22: 369–378.
- 59 Wisaksana R, Indrati AK, Fibriani A *et al.* Response to first-line antiretroviral treatment among human immunodeficiency virus-infected patients with and without a history of injecting drug use in Indonesia. *Addiction* 2010; 105: 1055–1061.
- 60 Mehta SH, Lucas G, Astemborski J, Kirk GD, Vlahov D, Galai N. Early immunologic and virologic responses to highly active antiretroviral therapy and subsequent disease progression among HIV-infected injection drug users. *AIDS Care* 2007; 19: 637–645.
- 61 Giordano TP, Hartman C, Gifford AL, Backus LI, Morgan RO. Predictors of retention in HIV care among a national cohort of US veterans. *HIV Clin Trials* 2009; 10: 299–305.
- 62 Ortego C, Huedo-Medina TB, Llorca J *et al.* Adherence to highly active antiretroviral therapy (HAART): a meta-analysis. *AIDS Behav* 2011; 15: 1381–1396.
- 63 Malta M, Strathdee SA, Magnanini MM, Bastos FI. Adherence to antiretroviral therapy for human immunodeficiency virus/acquired immune deficiency syndrome among drug users: a systematic review. *Addiction* 2008; 103: 1242–1257.