



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΙΑΤΡΙΚΗ

**Μέτα-ανάλυση γενετικής
επιδημιολογίας στο STATA**

ΑΠΟΣΤΟΛΟΥ ΚΑΛΛΙΟΠΗ

ΕΠΙΒΛΕΠΩΝ

ΜΠΑΓΚΟΣ ΠΑΝΤΕΛΕΗΜΩΝ

Λαμία, 19/10/23

Περίληψη

Μέτα-ανάλυση είναι η ποσοτική ανασκόπηση και σύνθεση των αποτελεσμάτων που προκύπτουν από σχετιζόμενες αλλά και ανεξάρτητες μεταξύ τους μελέτες. Αποτελεί ένα βασικό εργαλείο στον κλάδο της επιδημιολογίας λόγω των σύντομων και χρηστικών συμπερασμάτων που μπορεί να προσφέρει σε ερωτήματα τα οποία πρωτογενείς έρευνες δεν μελετούν. Στόχος της συγκεκριμένης εργασίας ήταν η δημιουργία μιας νέας έκδοσης της ήδη δημοσιευμένης εντολής του STATA metagen, η οποία θα πραγματεύεται την ενοποίηση των τεχνικών για τη διεξαγωγή πολυπαραγοντικής μέτα-ανάλυσης σε μελέτες γενετικής συσχέτισης. Ανάμεσα στα μοντέλα που χρησιμοποιήθηκαν, βασικό είναι το γενικό μοντέλο που βασίζεται σε συνοπτικά δεδομένα (summary data), για διακριτά (discrete) και για συνεχή (continuous) αποτελέσματα, αντίστοιχα, με τις αναλυτικές εκφράσεις για τη συνδιακύμανση τους. Τα μοντέλα που βασίζονται στη δυαδική (binary) μορφή των δεδομένων προσαρμόζονται χρησιμοποιώντας τόσο προοπτική (prospective) όσο και αναδρομική (retrospective) πιθανοφάνεια. Επιπλέον, σχεδόν πανομοιότυπες εκτιμήσεις με εκείνες που προκύπτουν από το γενικό διμεταβλητό (bivariate) μοντέλο εμφανίζει η μέθοδος bivariate model-free με την χρήση είτε αθροιστικών είτε δυαδικών δεδομένων. Τέλος, προσθέτοντας στο γενικό μοντέλο για συνεχή αποτελέσματα ένα τυχαίο παράγοντα προκύπτει το γραμμικό μικτό μοντέλο (LMM) με σχεδόν πανομοιότυπα αποτελέσματα. Αυτές οι μεθοδολογίες έχουν εφαρμοστεί σε αρκετές προηγουμένως δημοσιευμένες μετα-αναλύσεις γενετικών μελετών συσχέτισης, καλύπτοντας τόσο διακριτά όσο και συνεχή αποτελέσματα. Προτέρημα της ενοποίησης των μεθόδων μαζί με τους αντίστοιχους ελέγχους σε μια εντολή αποτελεί η εύκολη εφαρμογή τους σε όλα τα κύρια στατιστικά πακέτα, εξαλείφοντας την ανάγκη για εξατομικευμένο λογισμικό.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: μέτα-ανάλυση, γενετική επιδημιολογία, πολυμεταβλητή μέθοδος, STATA

«Υπεύθυνη Δήλωση μη λογοκλοπής και ανάληψης προσωπικής ευθύνης»

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, και γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα και ενυπογράφως ότι η παρούσα εργασία με τίτλο «Μέτα-ανάλυση γενετικής επιδημιολογίας στο STATA» αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές από τις οποίες χρησιμοποίησα δεδομένα, ιδέες, φράσεις, προτάσεις ή λέξεις, είτε επακριβώς (όπως υπάρχουν στο πρωτότυπο ή μεταφρασμένες) είτε με παράφραση, έχουν δηλωθεί κατάλληλα και ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Η ΔΗΛΟΥΣΑ

Αποστόλου Καλλιόπη

Ημερομηνία

19-10-2023

Ευχαριστίες

Με την ολοκλήρωση της πτυχιακής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσοι συνέβαλαν στην εκπόνησή της.

Ιδιαίτερα επιθυμώ να ευχαριστήσω τον καθηγητή μου και επιβλέποντα της παρούσας πτυχικής εργασίας, κ. Παντελεήμονα Μπάγκο, Κοσμήτορα του τμήματος Θετικών Επιστημών του Πανεπιστημίου Θεσσαλίας και Διευθυντή του Εργαστηρίου Μοριακής και Υπολογιστικής Βιολογίας και Γενετικής του Τμήματος Πληροφορικής με Εφαρμογές στην Βϊοιατρική του Πανεπιστημίου Θεσσαλίας, για την εμπιστοσύνη που μου έδειξε εξ' αρχής, αναθέτοντάς μου το συγκεκριμένο θέμα, την επιστημονική του καθοδήγηση, τις υποδείξεις και το αμείωτο ενδιαφέρον σε όλα τα στάδια εκπόνησης της εργασίας.

Επιπλέον, θα ήθελα να ευχαριστήσω την Καλλιόπη Εξάρχου-Κουβέλη, υποψήφια διδάκτορα του Τμήματος Πληροφορικής με Εφαρμογές στην Βϊοιατρική και Ακαδημαϊκή Υπότροφο του Τμήματος Μαθηματικών του Πανεπιστημίου Θεσσαλίας, για την προθυμία της, τη συμπαράσταση, τις συμβουλές και τα σχόλιά της, για την ολοκλήρωση της παρούσας εργασίας.

Τέλος, θέλω να ευχαριστώ την οικογένειά μου για την κατανόηση και τη συμπαράστασή τους.

Περιεχόμενα

Περίληψη	i
Ευχαριστίες.....	iii
Περιεχόμενα	iv
Λίστα με Πίνακες.....	vi
Λίστα με Εικόνες	vii
1. Εισαγωγή	1
1.1 Μέτα-ανάλυση σε μελέτες γενετικής συσχέτισης	2
2. Μέθοδοι που χρησιμοποιήθηκαν.....	5
2.1 Μοντέλα πολυδιάστατης μέτα-ανάλυσης χρησιμοποιώντας συνοπτικά δεδομένα	5
2.1.1 Το γενικό μοντέλο πολυδιάστατης μέτα-ανάλυσης για διακριτά αποτελέσματα	5
2.1.2 Το μοντέλο της γενετικής model-free προσέγγισης.....	7
2.2 Το μοντέλο πολυδιάστατης μέτα-ανάλυσης για συνεχή αποτελέσματα	8
2.3 Πολυμεταβλητή μέτα-ανάλυση με δυαδικά δεδομένα	10
2.3.1 Το γενικό μοντέλο μέτα-ανάλυσης χρησιμοποιώντας προοπτική πιθανοφάνεια.....	10
2.3.2 Το γενικό μοντέλο μέτα-ανάλυσης χρησιμοποιώντας αναδρομική πιθανοφάνεια	11
2.4 Αξιολόγηση αποτελεσμάτων στα γενετικά μοντέλα κληρονομικότητας	13
3. Αποτελέσματα.....	16
3.1 Επεξήγηση προγράμματος	16
3.1.1 Επεξήγηση του προγράμματος της εντολής metagen	16
3.1.2 Επεξήγηση προγραμμάτων των μεθόδων model-free και bivariate	17

3.2 Υλοποίηση προγράμματος	18
3.2.1 Υλοποίηση της εντολής <code>metagen</code>	18
3.2.2 Υλοποίηση προγραμμάτων για τις μεθόδους <code>model-free</code> και <code>bivariate</code>	36
3.3 Παραδείγματα εκτέλεσης του προγράμματος	39
3.3.1 Μελέτες με διακριτά αποτελέσματα	40
3.3.2 Μελέτες με συνεχή αποτελέσματα	41
3.3.3 Απεικονίσεις αποτελεσμάτων στο Stata	44
4. Συζήτηση – Συμπεράσματα.....	53
Βιβλιογραφικές Αναφορές.....	56

Λίστα με Πίνακες

Πίνακας 1: Τυπική διάταξη των δεδομένων ασθενών/μαρτύρων	2
Πίνακας 2: Τυπική διάταξη δεδομένων μετά-ανάλυσης μελετών γενετικής συσχέτισης.....	3
Πίνακας 3: Αποτελέσματα παραδοσιακών και πολυμεταβλητών μεθόδων με διακριτά αποτελέσματα.....	39
Πίνακας 4: Αποτελέσματα παραδοσιακών και πολυμεταβλητών μεθόδων με συνεχή αποτελέσματα.....	41

Λίστα με Εικόνες

Εικόνα 1: Αποτελέσματα εκτέλεσης της metagen με model-free προσέγγιση.....	44
Εικόνα 2: Πρώτη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της model-free μεθόδου με Μέγιστη Πιθανοφάνεια (ML).....	45
Εικόνα 3: Δεύτερη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου προοπτικής πιθανοφάνειας.....	46
Εικόνα 4: Αποτελέσματα εκτέλεσης της metagen με προοπτική πιθανοφάνεια.....	47
Εικόνα 5: Πρώτη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου linear mixed model.....	48
Εικόνα 6: Δεύτερη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της model-free μεθόδου με Μέγιστη Πιθανοφάνεια (ML).....	49
Εικόνα 7: Αποτελέσματα εκτέλεσης της metagen με γραμμικό μεικτό μοντέλο.....	50
Εικόνα 8: Πρώτη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου προοπτικής πιθανοφάνειας.....	51
Εικόνα 9: Δεύτερη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου linear mixed model.....	52

1. Εισαγωγή

Η συλλογή και η σύνθεση των διαθέσιμων πληροφοριών για έναν συγκεκριμένο συσχετισμό γονιδίου-ασθένειας παρέχοντας μια ποσοτική συνολική εκτίμηση, μια διαδικασία γνωστή ως μέτα-ανάλυση, καθίσταται όλο και περισσότερο αναγκαία, δεδομένου ότι ο αριθμός των δημοσιευμένων μελετών γενετικής επιδημιολογίας αυξάνεται συνεχώς. Με το πέρασμα του χρόνου, έχουν προταθεί διάφορες μέθοδοι για τον συνδυασμό των αποτελεσμάτων μελετών βασισμένων στην οικογένεια (family-based), καθώς και μέθοδοι για τον συνδυασμό των αποτελεσμάτων από μελέτες βασισμένες στον πληθυσμό και στην οικογένεια, σε μια μέτα-ανάλυση (Evangelou et al, 2006; Kazeem & Farrall, 2005).

Ιδιαίτερη προσοχή απαιτείται κατά την μέτα-ανάλυση γενετικών παραγόντων που σχετίζονται με την αιτιολογία μιας συγκεκριμένης νόσου. Για μελέτες βασισμένες στον πληθυσμό έχουν προταθεί διάφορες προσεγγίσεις, οι οποίες συνήθως βασίζονται στο μοντέλο ασθενών/μαρτύρων (case/control), αρκετά γνωστό στο χώρο της Επιδημιολογίας. Κάποιες από αυτές βασίζονται σε παραδοσιακές προσεγγίσεις για τη μέτα-ανάλυση επιδημιολογικών μελετών με μικρές τροποποιήσεις για τη διαχείριση των επιδράσεων του γονότυπου (Thakkinstian et al, 2005), ενώ άλλες, πιο εξειδικευμένες μέθοδοι, χρησιμοποιούν πολυμεταβλητές μεθόδους μέτα-ανάλυσης (Minelli et al, 2005a; Minelli et al, 2005b). Δυστυχώς όμως, στη συντριπτική πλειοψηφία των δημοσιευμένων μέτα-αναλύσεων γενετικών συσχετίσεων, τα δεδομένα αναλύονται σε μονομεταβλητό πλαίσιο χωρίς να λαμβάνονται υπόψη οι αντιθέσεις των γονοτύπων εντός των μελετών. Αυτό συμβαίνει διότι, καμία από τις μεθόδους που έχουν προταθεί δεν περιλαμβάνει εξειδικευμένο user-written λογισμικό.

Σε αυτή την εργασία συνοψίζονται οι διάφορες μέθοδοι που προτάθηκαν στο παρελθόν για την εκτέλεση μέτα-αναλύσεων σε μελέτες γενετικής συσχέτισης και παρουσιάζεται ο τρόπος ενοποίησής τους κάτω από την εντολή metagen, καθώς και τα αποτελέσματα που προκύπτουν από την εφαρμογή των μεθόδων σε δεδομένα ήδη δημοσιευμένων μελετών. Οι μέθοδοι εφαρμόζονται τόσο σε διακριτά όσο και σε συνεχή αποτελέσματα και μπορούν εύκολα να εφαρμοστούν σε οποιοδήποτε κύριο στατιστικό πακέτο που μπορεί να προσαρμόσει πολυμεταβλητά γενικευμένα μοντέλα (multivariate generalized linear).

1.1 Μέτα-ανάλυση σε μελέτες γενετικής συσχέτισης

Οι πιθανοί γονότυποι ενός ατόμου για μια ασθένεια η οποία εξαρτάται από ένα γονίδιο με δύο αλληλόμορφα (A, B), όπου το δεύτερο (B) αποτελεί τον παράγοντα κινδύνου, μπορεί να είναι AA, AB και BB. Ο Πίνακας 1 παρουσιάζει τα δεδομένα που χρησιμοποιούνται σε μια μέτα-ανάλυση από k μελέτες, στις οποίες δείγματα περιλαμβανομένων ασθενών (cases) και μαρτύρων (controls) κατηγοριοποιήθηκαν ανάλογα με τους γονοτύπους τους. Το μεγαλύτερο μέρος των δημοσιευμένων μελετών γενετικής συσχέτισης περιλαμβάνει δυαδικά αποτελέσματα που μπορούν να αναπαρασταθούν σε έναν τέτοιο πίνακα. Υπάρχουν όμως και περιπτώσεις που τα αποτελέσματα των μεθόδων είναι συνεχή (Πίνακας 2), όπως συμβαίνει στην Συστολική Αρτηριακή Πίεση (SBP) ή στο επίπεδο χοληστερόλης. Σε αυτήν την περίπτωση, αφού συγκεντρωθεί ένα αντιπροσωπευτικό δείγμα από τον γενικό πληθυσμό ή από έναν υψηλού κινδύνου πληθυσμό, τα άτομα υποβάλλονται σε γονοτυπική ανάλυση και συγκρίνονται οι μέσες τιμές των συνεχών αποτελεσμάτων με διάφορους γονότυπους.

Study	Cases			Controls		
1	AA_{11}	AB_{11}	BB_{11}	AA_{01}	AB_{01}	BB_{01}
2	AA_{12}	AB_{12}	BB_{12}	AA_{02}	AB_{02}	BB_{02}
...	...	...	...	...	...	...
k	AA_{1k}	AB_{1k}	BB_{1k}	AA_{0k}	AB_{0k}	BB_{0k}

Πίνακας 1. Τυπική διάταξη των δεδομένων ασθενών/μαρτύρων.

Ο Πίνακας 1 αποτελεί μια τυπική διάταξη των δεδομένων που χρησιμοποιούνται σε μια μέτα-ανάλυση γενετικής συσχέτισης ασθενών/μαρτύρων που περιλαμβάνουν έναν τόπο με δύο αλληλόμορφα. Η διανομή των διαφόρων γονοτύπων (AA, AB και BB) παρατίθενται για ασθενείς και μάρτυρες, ενώ οι διαφορετικές μελέτες ($i = 1, 2 \dots k$) που συμπεριλαμβάνονται στη μέτα-ανάλυση παρατίθενται σε διαφορετικές σειρές.

Η πλειονότητα των μέτα-αναλύσεων για μελέτες γενετικής συσχέτισης που περιλαμβάνουν ένα τόπο με δύο διακριτικά αλληλόμορφα (A vs B), πραγματοποιούνται με τον συνδυασμό των γονοτύπων σε δύο κατηγορίες, υποθέτοντας ασαφώς ένα

συγκεκριμένο γενετικό μοντέλο. Στη συνέχεια πραγματοποιούν συγκρίσεις μέσω του OddsRatio (OR) για δυαδικά αποτελέσματα ή χρησιμοποιώντας τη μέση διαφορά, επίσης γνωστή ως σταθμισμένη μέση διαφορά (weighted mean difference - WMD), για συνεχή αποτελέσματα. Δηλαδή, λαμβάνοντας υπόψιν την αντίθεση των γονοτύπων BB+AB έναντι του γονότυπου AA, υποθέτουμε ασαφώς ένα επικρατές μοντέλο κληρονομιάς, ενώ το BB έναντι AB+AA αντιστοιχεί σε ένα υπολειπόμενο μοντέλο και ούτω καθεξής. Μια άλλη προσέγγιση, η οποία στην πλειονότητα των περιπτώσεων πραγματοποιείται παράλληλα με την προηγούμενη προσέγγιση, είναι η σύγκριση των συχνοτήτων των αλληλομόρφων μεταξύ των ασθενών και των μαρτύρων, υπονοώντας με αυτόν τον τρόπο ένα συν-επικρατές μοντέλο κληρονομιάς. Τέλος, μια άλλη προσέγγιση είναι η πραγματοποίηση πολλαπλών συγκρίσεων (δηλαδή BB έναντι AB, AB έναντι AA κ.λπ.) με τον κίνδυνο ενός διογκωμένου ποσοστού σφάλματος Τύπου I.

	Genotypes					
Study	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

Πίνακας 2. Τυπική διάταξη δεδομένων μετά-ανάλυσης μελετών γενετικής συσχέτισης

Στον Πίνακα 2 η κατανομή των διαφόρων γονοτύπων (AA, AB και BB) παρατίθενται για όλους τους συμμετέχοντες (N), συνοδευόμενη από τον μέσο όρο και την τυπική απόκλιση του συνεχούς αποτελέσματος σε κάθε ομάδα. Οι διαφορετικές μελέτες ($i = 1, 2 \dots k$) που συμπεριλαμβάνονται στη μετά-ανάλυση παρατίθενται σε διαφορετικές σειρές.

Ανεξάρτητα από την προσέγγιση που χρησιμοποιείται, οι παραδοσιακές μέθοδοι της μετά-ανάλυσης συνοπτικών δεδομένων μπορούν να εφαρμοστούν είτε βασιζόμενες σε μοντέλα σταθερών (fixed-effect) είτε τυχαίων επιδράσεων (random-effects). Σε μια παραδοσιακή μετά-ανάλυση με σταθερές επιδράσεις, η υπόθεση που συνήθως γίνεται είναι ότι οι εκτιμήσεις y_i του logOR ή της μέσης διαφοράς από κάθε μελέτη

κατανέμονται κανονικά με μέση τιμή ίση με θ και διακύμανση σ^2 , η οποία είναι η εκτιμηθείσα διακύμανση του $\log OR$ ή της σταθμισμένης μέσης διαφοράς. Στην περίπτωση ύπαρξης ετερογένειας μια εναλλακτική και προτιμότερη μέθοδος είναι η μέθοδος των τυχαίων επιδράσεων, η οποία υποθέτει ότι οι πραγματικές επιδράσεις ποικίλουν τυχαία μεταξύ των μελετών, ενώ εισάγει και μια τυχαία συνιστώσα της διακύμανσης (τ^2).

Τα δεδομένα στον Πίνακα 1 μπορούν να μοντελοποιηθούν με την θεώρηση των δύο $\log OR$, τα οποία προέρχονται από το μεταλλαγμένο αλληλόμορφο (AB έναντι AA και BB έναντι AA), ως εξαρτημένη μεταβλητή των σχετιζόμενων, μεταξύ τους, εκτιμήσεών τους. Η λήψη αυτού του παράγοντα είναι αναγκαία όταν κάποιος προσπαθεί να εξάγει ταυτόχρονα συμπεράσματα για τη στατιστική σημαντικότητα των δύο $\log OR$, καθώς και όταν προσπαθεί να συγκρίνει τα τελευταία για να αποκρυπτογραφήσει το γενετικό μοντέλο κληρονομικότητας.

2. Μέθοδοι που χρησιμοποιήθηκαν

Μολονότι γενικά μοντέλα για πολυδιάστατη μέτα-ανάλυση είναι διαθέσιμα εδώ και χρόνια, είναι αξιοσημείωτο το γεγονός ότι δεν έχουν γίνει προσπάθειες να προσαρμοστούν στις γενετικές μελέτες συσχέτισης. Αντιθέτως, έχουν προταθεί πολύ ενδιαφέρουσες πολυδιάστατες μέθοδοι. Οι Minelli και συνάδελφοι (Minelli et al, 2005b) πρότειναν μια πολύ ενδιαφέρουσα γενετική model-free προσέγγιση για τη μέτα-ανάλυση γενετικών μελετών συσχέτισης την οποία και επέκτειναν σε ένα μπεϋζιανό πλαίσιο. Οι Salanti και συνάδελφοι εισήγαγαν μια άλλη μπεϋζιανή μέθοδο που συμπεριλαμβάνει απευθείας στη μέτα-ανάλυση τις αποκλίσεις από την ισορροπία Hardy-Weinberg, χρησιμοποιώντας συντελεστές σταθερότητας. Τελευταία, οι Bagos και Nikolopoulos (Bagos & Nikolopoulos, 2007) πρότειναν μια απλούστερη προσέγγιση βασισμένη σε μοντέλα τυχαίων συντελεστών λογιστικής παλινδρόμησης, η οποία είναι εύκολα υλοποιήσιμη στο Stata. Ωστόσο, η πλειονότητα από αυτές τις μεθόδους δεν χρησιμοποιούνται ευρέως στην πράξη για πολλούς λόγους. Παρ' όλα αυτά, θα ήταν επωφελές να υπάρχουν διαθέσιμες μέθοδοι Μέγιστης Πιθανοφάνειας (Maximum Likelihood) που να μπορούν να εκτελούν τις ίδιες αναλύσεις μέσα σε ένα συνοπτικό πλαίσιο. Όλα τα μοντέλα που θα παρουσιαστούν στις επόμενες ενότητες περιγράφονται στο επιστημονικό άρθρο “A Unification of Multivariate Methods for Meta-Analysis of Genetic Association Studies” του Μπάγκου Π.Γ. (P.G. Bagos 2008).

2.1 Μοντέλα πολυδιάστατης μέτα-ανάλυσης χρησιμοποιώντας συνοπτικά δεδομένα

2.1.1 Το γενικό μοντέλο πολυδιάστατης μέτα-ανάλυσης για διακριτά αποτελέσματα

Στη γενική προσέγγιση της διμεταβλητής (bivariate) μέτα-ανάλυσης, τα δύο logOR που προέρχονται από το μεταλλαγμένο αλληλόμορφο (AB έναντι AA και BB έναντι AA) μπορούν να μοντελοποιηθούν ταυτόχρονα ως μια εξαρτημένη μεταβλητή. Ο λογάριθμος του OR_{AB} δίνεται από:

$$y_{1i} = \log\left(\frac{AB_{1i} AA_{0i}}{AA_{1i} AB_{0i}}\right) \quad (2.1)$$

Με την ίδια λογική, ο λογάριθμος του OR_{BB} δίνεται από:

$$y_{2i} = \log\left(\frac{BB_{1i} AA_{0i}}{AA_{1i} BB_{0i}}\right) \quad (2.2)$$

Ενώ οι διακυμάνσεις υπολογίζονται προσεγγιστικά από:

$$s_{1i}^2 = \frac{1}{AA_{1i}} + \frac{1}{AA_{0i}} + \frac{1}{AB_{1i}} + \frac{1}{AB_{0i}} \quad (2.3)$$

$$s_{2i}^2 = \frac{1}{AA_{1i}} + \frac{1}{AA_{0i}} + \frac{1}{AB_{1i}} + \frac{1}{AB_{0i}} \quad (2.4)$$

Χρησιμοποιώντας το μοντέλο τυχαίων επιδράσεων, υποθέτουμε ότι τα δύο logOR ακολουθούν τη διμεταβλητή κανονική κατανομή

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix}, \begin{bmatrix} s_{1i}^2 & \rho_{w12} s_{1i} s_{2i} \\ \rho_{w12} s_{1i} s_{2i} & s_{2i}^2 \end{bmatrix} \right\} \quad (2.5)$$

όπου οι μέσοι όροι (β_{1i}, β_{2i}) θεωρούνται τυχαίοι όροι, κατανεμημένοι με παρόμοιο τρόπο:

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho_{B12} \tau_1 \tau_2 \\ \rho_{B12} \tau_1 \tau_2 & \tau_2^2 \end{bmatrix} \right\} \quad (2.6)$$

Συνεπώς, το τελικό μοντέλο στο οποίο βασιζόμαστε για τις εκτιμήσεις μας είναι:

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \tau_1^2 & \rho_{w12} s_{1i} s_{2i} + \rho_{B12} \tau_1 \tau_2 \\ \rho_{w12} s_{1i} s_{2i} + \rho_{B12} \tau_1 \tau_2 & s_{2i}^2 + \tau_2^2 \end{bmatrix} \right\} \quad (2.7)$$

Σε αυτήν την εξίσωση, η συνδιακύμανση εντός των μελετών είναι $\text{cov}(y_{1i}, y_{2i}) = \rho_{w12} s_{1i} s_{2i}$, όπου το ρ_{w12} είναι η εντός των μελετών συσχέτιση που πρέπει να είναι γνωστή εκ των προτέρων, ενώ η $\rho_{B12} \tau_1 \tau_2$ είναι η συνδιακύμανση των τυχαίων όρων μεταξύ των μελετών, η οποία πρέπει να εκτιμηθεί από τα δεδομένα. Συνεπώς, το ρ_{B12} είναι η μεταξύ των μελετών συσχέτιση, η οποία στα γενικά μοντέλα της διμεταβλητής μέτα-ανάλυσης συνήθως θεωρείται μηδενική ή πρέπει να παρέχεται από τις ατομικές μελέτες. Ωστόσο, στις γενετικές μελέτες συσχέτισης, τα δύο logOR είναι συσχετισμένα, αφού και τα δύο εκτιμούν μια σύγκριση ως προς την ίδια βάση, δηλαδή τα άτομα που

φέρουν το γονότυπο AA. Επιπλέον, η εντός των μελετών συσχέτιση μπορεί να υπολογιστεί αναλυτικά χρησιμοποιώντας τον αριθμό των γονοτύπων από κάθε μελέτη. Η συνδιακύμανση μεταξύ των δύο logOR μπορεί να προκύψει αντιμετωπίζοντας τις παρατηρούμενες μετρήσεις σε κάθε 2×3 πίνακα, που αντιπροσωπεύει μια μόνο μελέτη, ως ανεξάρτητες κατανομές Poisson με $E[Y_i] = \text{var}[Y_i] = Y_i$ και τα logOR ως αντιθέσεις μεταξύ των λογαρίθμων κάθε μελέτης. Στη συνέχεια, χρησιμοποιώντας τη μέθοδο Δέλτα και απλούς υπολογισμούς, μπορούμε να υπολογίσουμε

τη διακύμανση:

$$\text{cov}(y_{1i}, y_{2i}) = \frac{1}{AA_{1i}} + \frac{1}{AA_{0i}} \quad (2.8)$$

και τη συσχέτιση:

$$\rho_{w12} = \frac{\left(\frac{1}{AA_{1i}} + \frac{1}{AA_{0i}}\right)}{s_{1i}s_{2i}} \quad (2.9)$$

2.1.2 Το μοντέλο της γενετικής model-free προσέγγισης

Η γενετική model-free προσέγγιση, μοντέλο των Minelli και συνεργατών (Minelli et al, 2005) αφορά μια επέκταση του γενικού διμεταβλητού μοντέλου χρησιμοποιώντας συνοπτικά (summary) δεδομένα. Αποτελείται από την κοινή μοντελοποίηση του β_2 και του λ , όπου το λ είναι η λόγος του β_1 και του β_2 . Το μοντέλο είναι το εξής:

$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \lambda\beta_2 \\ \beta_2 \end{bmatrix}, \begin{bmatrix} s_{1i}^2 + \lambda^2\tau^2 & \rho_{w12}s_{1i}s_{2i} + \lambda\tau^2 \\ \rho_{w12}s_{1i}s_{2i} + \lambda\tau^2 & s_{2i}^2 + \tau^2 \end{bmatrix} \right\} \quad (2.10)$$

Είναι προφανές ότι αποτελεί ειδική περίπτωση του μοντέλου της εξίσωσης (2.7) με την παραμετροποίηση των β_1 , β_2 χρησιμοποιώντας το $\lambda = \beta_1/\beta_2$. Ωστόσο, το μοντέλο πραγματοποιεί ορισμένες επιπρόσθετες υποθέσεις, καθώς επιβάλλει μία μόνο διακύμανση μεταξύ των μελετών, το τ^2 , υποδεικνύοντας με αυτό τον τρόπο ότι τα β μοιράζονται μια κοινή τυχαία συνιστώσα διακύμανσης, ενώ το λ θεωρείται ως παράμετρος με σταθερές επιδράσεις.

Οι Minelli και συνεργάτες στην αντίστοιχη δημοσίευσή τους, εξέτασαν επίσης τη γενική περίπτωση του διμεταβλητού μοντέλου χωρίς να ορίζουν ρητά τις συσχετίσεις εντός των μελετών. Αν και οι συγγραφείς δεν εξέτασαν αυτήν την πιθανότητα, το μοντέλο που αρχικά προοριζόταν να χρησιμοποιηθεί σε συνοπτικά δεδομένα μπορεί επίσης, να

εφαρμοσθεί απευθείας σε συνεχή δεδομένα, χρησιμοποιώντας τη μέθοδο που περιγράφεται στην επόμενη ενότητα και περιγράφεται στην εξίσωση (2.12).

2.2 Το μοντέλο πολυδιάστατης μέτα-ανάλυσης για συνεχή αποτελέσματα

Στην περίπτωση που έχουμε συνεχή αποτελέσματα, όπως τα δεδομένα που παρουσιάζονται στον Πίνακα 2, από κάθε μελέτη θα λαμβάνουμε το μέσο όρο και την τυπική απόκλιση ανά γονότυπο. Γενικά, αν συμβολίσουμε με y_{ij} το αποτέλεσμα ενός ατόμου: j στη μελέτη i , το οποίο θεωρούμε ότι κατανέμεται κανονικά, και χρησιμοποιήσουμε ψευδείς μεταβλητές, όπως z_{1ij} και z_{2ij} για τα άτομα που φέρουν το γονότυπο AB και BB, αντίστοιχα, μπορούμε να διατυπώσουμε το γραμμικό μοντέλο

$$y_{ij} = a_0 + a_i + \beta_1 z_{1ij} + \beta_2 z_{2ij} \quad (2.11)$$

Εάν τα δεδομένα είναι στη μορφή του Πίνακα 2, οι παρατηρήσεις είναι οι μελέτες και πρέπει να σταθμιστούν με βάση την αντίστροφη διακύμανσή τους. Ωστόσο, εάν έχουμε πρόσβαση στα ατομικά δεδομένα των ασθενών, μπορούμε να τα χρησιμοποιήσουμε απευθείας στο γραμμικό μοντέλο. Σε αυτό το μοντέλο, το a_0 είναι ο συνολικός μέσος που συσχετίζεται με το γονότυπο AA, που θεωρείται η ομάδα αναφοράς, και το a_i είναι οι σταθερές επιδράσεις που αφορούν κάθε μελέτη, οι οποίες χρειάζονται για να διατηρηθεί η ταξινόμηση ανά μελέτη. Οι συντελεστές β_1 και β_2 αντιπροσωπεύουν τις μέσες διαφορές των ατόμων που φέρουν τους γονότυπους AB και BB από το γονότυπο AA, αντίστοιχα. Αυτό το μοντέλο είχε ληφθεί υπόψη και από τον Thakkestian και τους συνεργάτες του (Thakkestian et al, 2005), χωρίς ωστόσο να το επεκτείνουν για να συμπεριλάβουν και τυχαίες επιδράσεις.

Μια επέκταση του μοντέλου με τυχαίες επιδράσεις μπορεί να διατυπωθεί εισάγοντας τυχαίους συντελεστές β_{1i} και β_{2i} για τους γονότυπους:

$$y_{ij} = \beta_0 + \beta_i + (\beta_{1i} + \beta_1)z_{1ij} + (\beta_{2i} + \beta_2)z_{2ij} \quad (2.12)$$

Οι τυχαίοι συντελεστές θεωρούνται ότι κατανέμονται κανονικά με:

$$\begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_2\tau_1 & \tau_2^2 \end{bmatrix} \right\} \quad (2.13)$$

Προσθέτοντας ένα τυχαίο παράγοντα για τις μελέτες οδηγούμαστε σε γραμμικό μικτό

μοντέλο (Linear Mixed Model). Παρόμοια μοντέλα με αυτό της εξίσωσης (2.12) έχουν προταθεί για τη μέτα-ανάλυση δεδομένων συνεχούς αποτελέσματος από κλινικές δοκιμές, αλλά δεν έχουν ποτέ εφαρμοστεί σε γενετικές μελέτες συσχέτισης (Higgins et al, 2001). Το συγκεκριμένο μοντέλο μπορεί να προσαρμοστεί σε οποιοδήποτε στατιστικό πακέτο που είναι ικανό να χρησιμοποιήσει μοντέλα random-coefficient (mixed) weighted regression models, όπως το Stata με την χρήση της εντολής gllamm.

Μια ελαφρώς διαφορετική προσέγγιση που μπορεί επίσης να ακολουθηθεί, περιλαμβάνει την απευθείας μοντελοποίηση των διαφορών μεταξύ των ζευγαριών των αποτελεσμάτων που προκύπτουν από άτομα που φέρουν τους δύο μεταλλαγμένους γονότυπους (AB, BB) από τον γονότυπο αναφοράς (AA):

$$y_{1i} = y_{ABi} - y_{AAi} \quad (2.14)$$

$$y_{2i} = y_{BBi} - y_{AAi} \quad (2.15)$$

Είναι εύκολο να δείξουμε ότι οι δύο διαφορές κατανέμονται κανονικά με μέσο όρο ίσο με τη διαφορά των μέσων όρων και διακύμανση ίση με το άθροισμα των διακυμάνσεων. Επομένως, είναι προφανές ότι το διμεταβλητό μοντέλο που παρουσιάζεται στην εξίσωση (2.7) μπορεί να εφαρμοστεί απευθείας για τη διμεταβλητή μοντελοποίηση των δύο διαφορών. Επίσης, οι δύο ποσότητες είναι συσχετισμένες, αφού και οι δύο εκτιμούν μια διαφορά από μια κοινή βάση (γονότυπος αναφοράς). Με απλούς υπολογισμούς πιθανοτήτων και σημειώνοντας ότι οι τυχαίες μεταβλητές y_{AAi} , y_{ABi} και y_{BBi} είναι ανεξάρτητες μεταξύ τους, προκύπτει η διακύμανση ως εξής:

$$\begin{aligned} \text{cov}(y_{1i}, y_{2i}) &= \text{cov}(y_{ABi} - y_{AAi}, y_{BBi} - y_{AAi}) \\ &= \text{cov}(y_{ABi}, y_{BBi}) - \text{cov}(y_{ABi}, y_{AAi}) - \text{cov}(y_{AAi}, y_{BBi}) + \text{cov}(y_{AAi}, y_{AAi}) \\ &= \text{var}(y_{AAi}) = \frac{sd_{AAi}^2}{AA_i} \end{aligned} \quad (2.16)$$

Το μοντέλο σταθερών επιδράσεων στην εξίσωση (2.11) είναι πλήρως ισοδύναμο με το μοντέλο σταθερών επιδράσεων της εξίσωσης (2.7) για συνεχή αποτελέσματα. Ωστόσο, τα αντίστοιχα μοντέλα με τυχαίες επιδράσεις πιθανόν να παράγουν διαφορετικά αποτελέσματα σε περιπτώσεις όπου υπάρχει μεγάλη ετερογένεια μεταξύ των μελετών, διότι το μοντέλο (2.11) υπολογίζει έναν κοινό παράγοντα (που αντιπροσωπεύει τη βασική γενετική κατηγορία AA), ενώ το μοντέλο (2.7) υπολογίζει απευθείας τις διαφορές μεταξύ των ζευγαριών (AB-AA, BB-AA).

2.3 Πολυμεταβλητή μέτα-ανάλυση με δυαδικά δεδομένα

Ένας άλλος τρόπος για την πραγματοποίηση πολυμεταβλητής μέτα-ανάλυσης, που είναι ισοδύναμος με αυτόν που προτάθηκε στην ενότητα (2.1), είναι να χρησιμοποιήσουμε απευθείας τη δυαδική φύση των δεδομένων. Αντί να υπολογίσουμε τα logOR και να υποθέσουμε ότι ακολουθούν κανονική κατανομή, μπορούμε να χρησιμοποιήσουμε απευθείας τις καταμέτρησεις των γονοτύπων και να πραγματοποιήσουμε την ανάλυση χρησιμοποιώντας τη λογιστική παλινδρόμηση. Παρόμοια μοντέλα έχουν προταθεί εδώ και χρόνια για την πραγματοποίηση μονομεταβλητής (univariate) μέτα-ανάλυσης με τη χρήση λογιστικής παλινδρόμησης σε ένα πλαίσιο πολλαπλών επιπέδων (Thompson & Sharp, 1999; Turner et al, 2000). Τα μοντέλα που παρουσιάζονται παρακάτω μπορούν να θεωρηθούν ως επεκτάσεις των παραπάνω μοντέλων στην περίπτωση πολυμεταβλητής απόκρισης. Στη δυαδική περίπτωση, τα δεδομένα του Πίνακα 1 αναδιατάσσονται και οι 3 γονότυποι κωδικοποιούνται ως j ($j = 0, 1, 2$), ενώ χρησιμοποιούμε το δ_{ij} για να συμβολίσουμε την περίπτωση ή τον έλεγχο (ασθενής=0, μάρτυρας=1) των υποκειμένων που έχουν τον j -οστό γονότυπο στη μελέτη i ($i = 1, 2, \dots, k$). Στη συνέχεια, έχουμε δύο εναλλακτικές για τη μοντελοποίηση, είτε χρησιμοποιώντας την προοπτική είτε την αναδρομική πιθανοφάνεια. Στην πρώτη περίπτωση υποθέτουμε ένα δυαδικό σχήμα δειγματοληψίας, όπου σταθεροί αριθμοί ασθενών και μαρτύρων δειγματοληπτούνται ανεξάρτητα. Στη δεύτερη περίπτωση, αντίστοιχα, υποθέτουμε ένα πολυδιάστατο σχήμα δειγματοληψίας, στο οποίο επιλέγονται υποκείμενα εξαρτημένα από την παθολογική τους κατάσταση και μετέπειτα διαπιστώνεται η κατάσταση της έκθεσής τους.

2.3.1 Το γενικό μοντέλο μέτα-ανάλυσης χρησιμοποιώντας προοπτική πιθανοφάνεια

Χρησιμοποιώντας την προοπτική πιθανοφάνεια, που βασίζεται στην πιθανότητα της νόσου δεδομένης της έκθεσης, η κατάσταση ασθενών/μαρτύρων είναι η εξαρτημένη μεταβλητή και οι γονότυποι θεωρούνται ως συμμεταβλητές. Στη συνέχεια,

συμβολίζουμε $\pi_{ij} = P(\delta_{ij}=1)$ την πιθανότητα να είναι ασθενείς (case) ενός ατόμου που έχει τον j -οστό γονότυπο στη μελέτη i , αντίστοιχα. Δεδομένου ότι το αλληλόμορφο B θεωρείται ως παράγοντας κινδύνου (risk factor), μια λογική επιλογή θα ήταν να θεωρήσουμε τον γονότυπο AA ως γονότυπο αναφοράς και να δημιουργήσουμε ψευδομεταβλητές όπως $z_{1i}=1$ εάν ο γονότυπος είναι AB και $z_{2i}=1$ εάν ο γονότυπος είναι BB.

Αυτό είναι το μοντέλο που περιγράφηκε από τους Bagos και Nikolopoulos (Bagos & Nikolopoulos, 2007) και το οποίο διατυπώνεται ως εξής:

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \text{logit}[P(\delta_{ij} = 1|j)] \\ &= a_0 + a_i + (\beta_1 + \beta_{1i})z_{1i} + (\beta_2 + \beta_{2i})z_{2i} \end{aligned} \quad (2.17)$$

Σε αυτό το μοντέλο, οι τυχαίοι όροι κατανέμονται σύμφωνα με την εξίσωση (2.13)

Οι ψευδομεταβλητές a_i αποτελούν δείκτες, ειδικούς για κάθε μελέτη, των σταθερών επιδράσεων, ενώ οι β_1 και β_2 που προκύπτουν από την εφαρμογή του μοντέλου αποτελούν εκτιμήσεις των logOR (AB έναντι AA και BB έναντι AA, αντίστοιχα). Μια άλλη παρόμοια προσέγγιση είναι να υποθέσουμε από πριν ένα πρόσθετο μοντέλο για τους γονότυπους. Σύμφωνα με αυτήν τη λογική, το μοντέλο θα περιλαμβάνει μια μεταβλητή z που παίρνει τιμές 0, 1 και 2 (για τους γονότυπους AA, AB και BB, αντίστοιχα):

$$\text{logit}(\pi_{ij}) = \text{logit}[P(\delta_{ij} = 1|j)] = a_0 + a_i + (\beta + \beta_i)z_i \quad (2.18)$$

Αυτό το μοντέλο λογιστικής παλινδρόμησης με τυχαίους συντελεστές είναι το ανάλογο ενός μονομεταβλητού (univariate) μοντέλου μέτα-ανάλυσης και, ως εκ τούτου, είναι πιο απλό. Ωστόσο, θα παρέχει αξιόπιστα αποτελέσματα μόνο σε περίπτωση συν-κυριαρχίας (co-dominant).

2.3.2 Το γενικό μοντέλο μέτα-ανάλυσης χρησιμοποιώντας αναδρομική πιθανοφάνεια

Εναλλακτικά, το μοντέλο που παρουσιάστηκε μπορεί να παραμετροποιηθεί υποθέτοντας ένα πολυωνυμικό σχήμα δειγματοληψίας χρησιμοποιώντας την αναδρομική πιθανοφάνεια, δηλαδή την πιθανοφάνεια που βασίζεται στην πιθανότητα

των γονοτύπων δεδομένης της κατάστασης της νόσου. Σε αυτήν την περίπτωση, οι γονότυποι θεωρούνται ως εξαρτημένες μεταβλητές και η κατάσταση ασθενών/μαρτύρων ως τον παράγοντα πρόβλεψης σε μια πολυωνυμική λογιστική παλινδρόμηση (McCullagh & Nelder, 1989):

$$P_{ij} = P(j|\delta_{ij}) = \frac{\exp(a_0 + a_i + \beta_j \delta_{ij})}{\sum_{r=0}^J \exp(a_0 + a_i + \beta_r \delta_{ir})} \quad (2.19)$$

Διαπιστώνοντας ότι ο γραμμικός παράγοντας πρόβλεψης στο παραπάνω μοντέλο γίνεται:

$$U_{ij} = \log\left(\frac{P_{ij}}{P_{i1}}\right) = a_0 + a_i + \beta_j \delta_{ij}, \quad j = 1, 2 \quad (2.20)$$

είναι εύκολο να καταλάβουμε ότι οι β_1 και β_2 που προκύπτουν από την εφαρμογή του μοντέλου είναι εκτιμήσεις των logOR (δηλαδή AB έναντι AA και BB έναντι AA, αντίστοιχα), όπως και οι αντίστοιχοι συντελεστές του μοντέλου στην εξίσωση (2.17). Φυσικά, ισχύει $\beta_0 = 0$, για να μπορεί να αναγνωριστεί, καθώς ο γονότυπος $j=0$ (δηλαδή AA) χρησιμοποιείται ως γονότυπος αναφοράς. Όπως και στο μοντέλο που βασίζεται στην προοπτική πιθανοφάνεια, οι μεταβλητές a_i αντιπροσωπεύουν δείκτες, ειδικούς για κάθε μελέτη, για το μοντέλο σταθερών επιδράσεων. Εάν στο παραπάνω μοντέλο εισαγάγουμε έναν τυχαίο συντελεστή που εξαρτάται από τον γονότυπο (για γονότυπους $j=1, 2$), ο γραμμικός παράγοντας πρόβλεψης γίνεται (Skrondal & Rabe-Hesketh, 2003):

$$U_{ij} = \log\left(\frac{P_{ij}}{P_{i1}}\right) = a_0 + a_i + \beta_j \delta_{ij} + \beta_{ij} \delta_{ij}, \quad j = 1, 2 \quad (2.21)$$

και το μοντέλο είναι πλήρως καθορισμένο ως μια μέτα-ανάλυση δυαδικών τυχαίων στοιχείων, με τυχαίους όρους που κατανέμονται σύμφωνα με την εξίσωση (2.13)

Έχει δειχθεί, στο πλαίσιο των σταθερών επιδράσεων, ότι οι εκτιμήσεις με μέγιστη πιθανοφάνεια που προκύπτουν από την αναδρομική πιθανοφάνεια, εκτός από τη σταθερή τιμή, είναι ίδιες με αυτές που προκύπτουν από την προοπτική πιθανοφάνεια (Chen, 2003). Αυτή η παρατήρηση οδήγησε στη δημοφιλή προσέγγιση εφαρμογής μοντέλων λογιστικής παλινδρόμησης για την προσαρμογή των παραμετροποιημένων μεταβλητών σε παρατηρησιακές μελέτες.

Οι Minelli και συνεργάτες συγκρίνουν τις αναδρομικές με τις προοπτικές προσεγγίσεις πιθανοφάνειας σε ένα μπεϋζιανό πλαίσιο, καταλήγοντας επίσης ότι τα αποτελέσματα στις περισσότερες περιπτώσεις είναι συγκρίσιμα (Minelli et al, 2005a). Ωστόσο, η συγκεκριμένη προσέγγιση διαφέρει από αυτήν που παρουσιάζεται εδώ σε αρκετές

πτυχές. Μια από τις διαφορές είναι το γεγονός ότι στα μοντέλα που παρουσιάζονται εδώ, οι επιδράσεις, που είναι ειδικές σε κάθε μελέτη, θεωρούνται σταθερές παράμετροι. Θα μπορούσαν επίσης να προστεθούν τυχαίες επιδράσεις μελέτης (random study effect), αλλά σε αυτή την περίπτωση η πολυπλοκότητα των υπολογισμών αυξάνεται, χωρίς σημαντική αλλαγή στην ακρίβεια των εκτιμήσεων.

2.4 Αξιολόγηση αποτελεσμάτων στα γενετικά μοντέλα κληρονομικότητας

Η απουσία γενικών τεστ για την εκτίμηση του γενετικού μοντέλου κληρονομικότητας, χωρίς να είναι απαραίτητη η επαναπαραμετροποίηση των μοντέλων των Minelli και συνεργατών, καθιστά δύσκολη την εύρεση συμπερασμάτων σχετικά με το γενετικό μοντέλο. Παρακάτω περιγράφεται ο τρόπος με τον οποίο οι εκτιμήσεις του γενετικού μοντέλου κληρονομικότητας μπορούν να προκύψουν μαζί με τα διαστήματα εμπιστοσύνης τους και πως, στις περισσότερες περιπτώσεις, συμπίπτουν με τις εκτιμήσεις που παρέχονται από την model-free προσέγγιση. Αφού το μοντέλο προσαρμοστεί καταλλήλως, η σύγκριση των δύο εκτιμήσεων ($\widehat{\beta}_1, \widehat{\beta}_2$) θα μπορούσε να παρέχει αποδείξεις για το γενετικό μοντέλο. Στην περίπτωση που και οι δύο εκτιμήσεις είναι σημαντικά διαφορετικές από το μηδέν και είναι ίσες μεταξύ τους, προτείνεται ένα επικρατές (dominant) μοντέλο, ενώ εάν $\widehat{\beta}_2 > \widehat{\beta}_1$, ένα συν-επικρατές μοντέλο θα είναι πιο πιθανό. Αξίζει να τονιστεί ότι, αν $\widehat{\beta}_1 = 0$ και $\widehat{\beta}_2 > 0$ η μόνη επιλογή είναι το υπολειπόμενο μοντέλο (Sasieni, 1997). Οι Thakkinstian και συνεργάτες, εξέτασαν αυτές τις πιθανότητες, αλλά δεν παρείχαν τελικά τεστ στο μονομεταβλητό (univariate) τους πλαίσιο (Thakkinstian κ.ά., 2005). Στην πραγματικότητα, τα τυπικά στατιστικά τεστ μπορούν να παραχθούν μόνο σε ένα πολυμεταβλητό πλαίσιο, ενώ το πιο πιθανό τεστ θα ήταν αυτό της ισότητας των δύο συντελεστών (δηλαδή $\widehat{\beta}_1 = \widehat{\beta}_2$). Αυτό θα οδηγούσε στην διατύπωση της ακόλουθης μηδενικής και εναλλακτικής υπόθεσης, αντίστοιχα:

$$H_0: d = \beta_1 - \beta_2 = 0, H_a: d \neq 0 \quad (2.22)$$

Η διαφορά δείγματος κατανέμεται κανονικά με μέσο

$$\hat{d} = \widehat{\beta}_1 - \widehat{\beta}_2 \quad (2.23)$$

ενώ η διακύμανσή του μπορεί να υπολογιστεί από:

$$var(\widehat{\beta}_1 - \widehat{\beta}_2) = var(\widehat{\beta}_1) + var(\widehat{\beta}_2) - 2cov(\widehat{\beta}_1, \widehat{\beta}_2) \quad (2.24)$$

Έτσι, υπό την μηδενική υπόθεση H_0 , η ακόλουθη στατιστική ποσότητα θα είναι κανονικά κατανοημένη:

$$\frac{\hat{d}}{\sqrt{var(\hat{d})}} = \frac{\widehat{\beta}_1 - \widehat{\beta}_2}{\sqrt{var(\widehat{\beta}_1 - \widehat{\beta}_2)}} \sim N(0,1) \quad (2.25)$$

ενώ το 95% διάστημα εμπιστοσύνης για τη διαφορά υπολογίζεται σύμφωνα με:

$$\hat{d} - 1.96\sqrt{var(\hat{d})}, \hat{d} + 1.96\sqrt{var(\hat{d})} \quad (2.26)$$

Αντί για τον έλεγχο της ισότητας των δύο συντελεστών, μια άλλη προσέγγιση θα ήταν ο έλεγχος του λόγου τους. Σε αυτήν την περίπτωση, η μηδενική και η εναλλακτική υπόθεση θα μπορούσαν να διατυπωθούν ως εξής:

$$H_0: \lambda = \frac{\beta_1}{\beta_2} = 0, H_a: \lambda \neq 0 \quad (2.27)$$

Αυτό είναι ακριβώς το στατιστικό τεστ που αναφέρεται στη model-free προσέγγιση των Minelli και είναι ίσως πιο εύκολο να κατανοηθεί, αφού μοντελοποιεί ταυτόχρονα τη σημαντικότητα των δύο συντελεστών, καθώς και το σχετικό τους μέγεθος. Ουσιαστικά, αυτό που επιδιώκεται είναι η κατάληξη σε συμπεράσματα σχετικά με το λόγο δύο συσχετισμένων κανονικά κατανοημένων μεταβλητών.

Ο λόγος δύο ανεξάρτητων τυπικών κανονικών μεταβλητών κατανέμεται σύμφωνα με την κατανομή Cauchy, εκτός από την περίπτωση μη μηδενικών μέσων (means) ή την περίπτωση συσχετισμού που ακολουθείται μια πιο περίπλοκη διαδικασία. Μια απλούστερη προσέγγιση για τον υπολογισμό της διακύμανσης του λ υπό τη μηδενική υπόθεση μπορεί να χρησιμοποιηθεί εάν ορίσουμε μια συνάρτηση $f(\beta_1, \beta_2) = \beta_1/\beta_2$ και την αναπτύξουμε χρησιμοποιώντας μια διμεταβλητή προσέγγιση Taylor 1ης τάξης γύρω από τα μέσα:

$$\hat{f}(\beta_1, \beta_2) = f(\widehat{\beta}_1, \widehat{\beta}_2) + \frac{\partial f(\widehat{\beta}_1, \widehat{\beta}_2)}{\partial \beta_1}(\beta_1 - \widehat{\beta}_1) + \frac{\partial f(\widehat{\beta}_1, \widehat{\beta}_2)}{\partial \beta_2}(\beta_2 - \widehat{\beta}_2) \quad (2.28)$$

Στη συνέχεια, χρησιμοποιώντας τη μέθοδο Δέλτα και αντικαθιστώντας τις τιμές του δείγματος πληθυσμού με τις δειγματικές τιμές, η διακύμανση θα είναι:

$$var(\hat{\lambda}) = \frac{var(\widehat{\beta}_1)}{\widehat{\beta}_2^2} + \frac{var(\widehat{\beta}_2)\widehat{\beta}_1^2}{\widehat{\beta}_2^4} - 2cov(\widehat{\beta}_1, \widehat{\beta}_2) \frac{\widehat{\beta}_1}{\widehat{\beta}_2^3} \quad (2.29)$$

Τέλος, υπό την H_0 θα έχουμε $\lambda=0$ και έτσι η ακόλουθη στατιστική ποσότητα θα είναι κανονικά κατανοημένη:

$$z = \frac{\hat{\lambda}}{\sqrt{var(\hat{\lambda})}} \sim N(0,1) \quad (2.30)$$

Συνεπώς, μπορεί να υπολογιστεί ένα 95% διάστημα εμπιστοσύνης, χρησιμοποιώντας:

$$\hat{\lambda} - 1.96\sqrt{var(\hat{\lambda})}, \hat{\lambda} + 1.96\sqrt{var(\hat{\lambda})} \quad (2.31)$$

3. Αποτελέσματα

3.1 Επεξήγηση προγράμματος

3.1.1 Επεξήγηση του προγράμματος της εντολής metagen

Το πρόγραμμα της εντολής metagen, είναι υπεύθυνο για την ενοποίηση των πολυμεταβλητών μοντέλων μέτα-ανάλυσης. Αρχικά δημιουργούνται οι μεταβλητές για τις τιμές που πρέπει να δώσει ο χρήστης και εκτελούνται οι απαραίτητοι έλεγχοι σύμφωνα με την μέθοδο που επιλέξει αλλά και ανάλογα με τις προτιμήσεις εμφάνισης των αποτελεσμάτων. Συγκεκριμένα, ο χρήστης καλείται να δώσει δύο ειδών τιμές ανάλογα με τα δεδομένα του και ανάλογα με τον τύπο των αποτελεσμάτων που θέλει. Μπορεί να δώσει είτε τους γονότυπους για ασθενείς (cases) και μάρτυρες (controls) για διακριτά αποτελέσματα, όπως στον Πίνακα 1, είτε να δώσει τα δεδομένα με την μορφή που έχουν στον Πίνακα 2 για συνεχή αποτελέσματα. Εκτός από τα δεδομένα, ο χρήστης πρέπει να επιλέξει και μια συγκεκριμένη μέθοδο μέτα-ανάλυσης, ενώ του δίνεται η επιλογή στα αποτελέσματα να φαίνονται οι επαναλήψεις, να υπολογίζονται τα αποτελέσματα με Μέγιστη Πιθανοφάνεια (ML) ή Περιορισμένη Μέγιστη Πιθανοφάνεια (REML), όπου είναι εφικτό, να εμφανίζονται με εκθετική μορφή τα αποτελέσματα ή ακόμα και να ονομάσει τις μελέτες ώστε να είναι πιο ευδιάκριτες στα γραφήματα.

Αφού γίνουν οι απαραίτητοι έλεγχοι για να μην υπάρχει σύγχυση ανάμεσα στα δεδομένα του χρήστη και στο πρόγραμμα, ανάλογα με την μέθοδο που έχει επιλέξει το άτομο, γίνονται οι απαραίτητες τροποποιήσεις και υπολογισμοί στα δεδομένα ώστε να μπορέσουν να εφαρμοστούν στις κατάλληλες εντολές. Πιο αναλυτικά, για όλες τις μεθόδους υπολογίζονται τα logOR, οι διακυμάνσεις, οι συνδιακυμάνσεις και τα τυπικά σφάλματα, ανάλογα με το αν η μέθοδος που επιλέχθηκε εμφανίζει διακριτά ή συνεχή αποτελέσματα. Στην συνέχεια για τις γενικές μεθόδους summary και continuous υπολογίζονται τα αποτελέσματα της μέτα-ανάλυσης με την χρήση της συνάρτησης pnmmeta, ενώ στις μεθόδους model-free και bivariate, αφού τροποποιηθούν σε κατάλληλη μορφή, υπολογίζονται με την βοήθεια της εντολής ml και των metagen_mf.ado και metagen_mb.ado, αντίστοιχα για κάθε μέθοδο. Να σημειωθεί εδώ, ότι οι τέσσερις μέθοδοι που αναφέρθηκαν μπορούν να υπολογιστούν με ML και με REML, χρησιμοποιώντας την συνάρτηση type(ml/reml) ανάλογα με την επιλογή του χρήστη. Αν ο χρήστης δεν επιλέξει συγκεκριμένο τρόπο τότε το πρόγραμμα αυτόματα ακολουθεί την Περιορισμένη Μέγιστη Πιθανοφάνεια. Οι μέθοδοι που χρησιμοποιούν δυαδικά δεδομένα, δηλαδή η μέτα-ανάλυση με προοπτική ή αναδρομική πιθανοφάνεια,

υπολογίζονται με την βοήθεια της συνάρτησης `gllamm`, όπως και η μέθοδος LMM, αφού πρώτα τροποποιηθούν κατάλληλα τα δεδομένα τους. Όλα τα μοντέλα που χρησιμοποιούνται έχουν υλοποιηθεί από τον Μπάγκο Π. (P.G. Bagos 2008).

Το πρόγραμμα τελειώνει με τον τρόπο εμφάνισης των αποτελεσμάτων, τον σχεδιασμό των γραφημάτων αλλά και τους υπολογισμούς των δ και λ για την αξιολόγηση κάθε μεθόδου. Για τον υπολογισμό του δ χρησιμοποιείται η συνάρτηση `lincom` με όρισμα τη διαφορά των δύο συντελεστών που προκύπτουν μετά την μέτα-ανάλυση. Με την χρήση της `return list` επιστρέφονται το δ και τον τυπικό σφάλμα με τα οποία υπολογίζεται και το 95% διάστημα εμπιστοσύνης, όπως αναφέρεται και στην εξίσωση (2.26). Με την ίδια λογική υπολογίζονται και οι τιμές του λ όπως και το 95% διάστημα εμπιστοσύνης του, με την μόνη διαφορά ότι αντί για τη συνάρτηση `lincom` χρησιμοποιείται η `nlcom` με όρισμα το πηλίκο των συντελεστών.

Για την εμφάνιση των αποτελεσμάτων δημιουργείται ένα πινακάκι με τους συντελεστές, τα τυπικά σφάλματα, τα p-value αλλά και τα 95% διαστήματα εμπιστοσύνης για κάθε περίπτωση αντίστοιχα (AB έναντι AA και BB έναντι AA). Αυτά τα αποτελέσματα μπορούν να εμφανιστούν και στην εκθετική τους μορφή, αν το επιλέξει ο χρήστης, όπως και οι τιμές στα γραφήματα `forest plot`. Στο ίδιο τμήμα του κώδικα γίνεται και ο σχεδιασμός των γραφημάτων με την χρήση της συνάρτησης `metagraph`, χρησιμοποιώντας την επιλογή `combine()` στην οποία εισάγονται οι συντελεστές και τα διαστήματα εμπιστοσύνης που προκύπτουν από την μέτα-ανάλυση. Τα δύο γραφήματα που προκύπτουν για κάθε περίπτωση συνδυάζονται σε ένα γράφημα καθιστώντας ευκολότερη την σύγκριση των αποτελεσμάτων τους.

3.1.2 Επεξήγηση προγραμμάτων των μεθόδων model-free και bivariate

Τα προγράμματα `metagen_mf` και `metagen_mb` δεν χρησιμοποιούνται από τον χρήστη άμεσα. Καλούνται μέσα από το κύριο πρόγραμμα `metagen` σε συνδυασμό με την εντολή `ml` ώστε να υπολογιστούν οι κατάλληλες συναρτήσεις για την μέτα-ανάλυση με την μέθοδο model-free και bivariate, αντίστοιχα. Επιπλέον, ανάλογα με την επιλογή του χρήστη για τον υπολογισμό με Μέγιστη Πιθανοφάνεια (ML) ή με Περιορισμένη Μέγιστη Πιθανοφάνεια (REML) τα δύο προγράμματα τροποποιούν καταλλήλως τις εξισώσεις ώστε να προκύψει το σωστό αποτέλεσμα. Αξίζει να σημειωθεί επίσης, ότι τα δύο προγράμματα ακολουθούν παρόμοια μεθοδολογία και παρ' όλο που βγάζουν πολύ κοντινά αποτελέσματα, έχουν διαφορετική δομή, καθώς τα αποτελέσματα που προκύπτουν από τη μέθοδο model-free καθώς επιβάλλει μία μόνο διακύμανση μεταξύ των μελετών όπως αναφέρεται στην ενότητα 2.1.2.

3.2 Υλοποίηση προγράμματος

3.2.1 Υλοποίηση της εντολής metagen

```
program define metagen , rclass

syntax varlist(min=6 max=9 numeric) [if] [in], method(string)
[studies(varlist)] [type(string)] [iter] [eform]

preserve

if("`method'" == "summary") |("`method'" == "pros") |("`method'" ==
"mb") |("`method'" == "retr") |("`method'" == "con") |("`method'"
== "lmm") |("`method'" == "mf") {

set more off

parse "`varlist'", parse(" ")

if "`if' "~=" {
preserve
quietly keep `if'
}

local summary 0      /*Meta-Analysis with discrete outcome*/
local pros 0         /*Meta-Analysis with prospective likelihood*/
local mb 0           /*Meta-Analysis with bivariate model*/
local retr 0         /*Meta-Analysis with retrospective likelihood*/
local con 0          /*Meta-Analysis with continuous outcome*/
local lmm 0          /*Meta-Analysis with LMM bivariate model*/
local mf 0           /*Meta-Analysis with model free model*/
}

if "`method'" == "summary" {
    local summary=1
    di in gr "The general multivariate model of meta-analysis for discrete
    outcomes"
    di in gr "-----"
```

```

-----"

di " "

    }

if "`method'" == "pros" {
    local pros=1
di in gr "The multivariate model of meta-analysis using prospective
likelihood"
di in gr "-----"
-----"

di " "

    }

if "`method'" == "mb" {
    local mb=1
di in gr "The bivariate model of meta-analysis "
di in gr "-----"
di " "

    }

if "`method'" == "retr" {
    local retr=1
di in gr "The multivariate model of meta-analysis using retrospective
likelihood"
di in gr "-----"
-----"

di " "

    }

if "`method'" == "con" {
    local con=1
di in gr "The general multivariate model of meta-analysis for continuous
outcomes"
di in gr "-----"
-----"

di " "

```

```

    }

if "`method'" == "lmm" {
    local lmm=1
di in gr"The LMM bivariate model of meta-analysis "
di in gr "-----"
di " "
}

if "`method'" == "mf" {
    local mf=1
di in gr"The genetic model-free of meta-analysis "
di in gr "-----"
di " "
}

**In case of specified study names
if "`studies'"!=""{

foreach x of varlist `studies'{
    qui gen _studies =`x'
}

if (`summary' == 1 | `mb' == 1 | `mf' == 1 | `pros'==1 | `retr'==1){
keep `1' `2' `3' `4' `5' `6' _studies
}
else{
keep `1' `2' `3' `4' `5' `6' `7' `8' `9' _studies
}
}
else{
if (`summary' == 1 | `mb' == 1 | `mf' == 1 | `pros'==1 | `retr'==1){
keep `1' `2' `3' `4' `5' `6'
}
}
}

```

```

else{
keep `1' `2' `3' `4' `5' `6' `7' `8' `9'
}
}

tempvar study nt graph1 graph2 graph3 graph4 graph5
gen `study' = _n
local nt = _N

if (`summary' == 1 | `mb' == 1 | `mf' == 1 | `pros'==1 | `retr'==1 ) {

foreach x of varlist `2' `3' `5' `6'{
    qui replace `x' += 0.5 if(`x' >= . | `x'==0)
}

foreach x of varlist `1' `4'{
if (`x' >= . | `x'==0){
    di "Reference genotype has missing values"
    exit
}
}

** Calculation of the log(Odds Ratios)
gen _b1=log(( `5'/ `4')/( `2'/`1'))
gen _b2=log(( `6'/ `4')/( `3'/`1'))
** Calculation of the variances
gen _V11=1/`1' +1/`2'+1/`4' +1/`5'
gen _V22=1/`1' +1/`4'+1/`3' +1/`6'
** Calculation of the covariance
gen _V12=1/`1' +1/`4'
gen _se1=sqrt(_V11)
gen _se2=sqrt(_V22)
}

if (`con' == 1 | `lmm' ==1 ){

```

```

foreach x of varlist `4' `5' `6' `7' `8' `9'{
    qui replace `x' =+ 0.5 if(`x' >= . | `x'==0)
}

foreach x of varlist `1' `2' `3'{
    if (`x' >= . | `x'==0){
        di "Reference genotype has missing values"
        exit
    }
}

** Calculation of the mean differences
gen _b1 = `5' - `4'
gen _b2 = `6' - `4'

** Calculation of the variances
gen _V11 = (`8')^2/`2' + (`7')^2/`1'
gen _V22 = (`9')^2/`3' + (`7')^2/`1'

** Calculation of the covariance
gen _V12 = (`7')^2 /`1'
gen _se1=sqrt(_V11)
gen _se2=sqrt(_V22)
}

if (`summary' == 1 | `con' == 1){

** Fitting the model of Equation 2.7 using REML and ML
if ("`type'" == "reml" | "`type'" == "") {
    qui mvmeta _b _V,vars(_b1 _b2)
}

if "`type'" == "ml" {
    qui mvmeta _b _V,vars(_b1 _b2) ml
}

** Retrieve the covariance matrix
matrix v=e(V)

```

```

local ror2 = _b[_b1]
local rse2 = _se[_b1]

local ror3 = _b[_b2]
local rse3 = _se[_b2]

**Automatic calculation of confidence intervals for d
qui lincom _b[_b2] - _b[_b1]
qui return list
local _del=r(estimate)
local _del_err =r(se)

**Automatic calculation of confidence intervals for l
qui nlcom ratio:_b[_b1]/_b[_b2], post
local _lam= _b[ratio]
local _lam_err=_se[ratio]
}

if (`mb' == 1 | `mf' == 1){
** Fixed-effects estimates
qui gen _wz1=1/_V11
qui gen _wz2=1/_V22
qui gen _z1wz1=_b1*_wz1
qui gen _z2wz2=_b2*_wz2
qui summarize _z1wz1
qui local muz1=r(sum)
qui summarize _wz1
qui local swz1=r(sum)
qui local muz1=`muz1'/'swz1'
qui local tauz1=1/'swz1'
qui summarize _z2wz2
qui local muz2=r(sum)

```

```

qui summarize _wz2
qui local swz2=r(sum)
qui local muz2=`muz2'/'swz2'
qui local tauz2=1/'swz2'
if( "`type'" == "" ){
gen _type=0
}
if("`type'" == "reml"){
gen _type=1
}
if("`type'" == "ml"){
gen _type=2
}
}

if `mb' == 1 {
qui corr _z1wz1 _z2wz2
qui local r=r(rho)
qui local tauz1=log(`tauz1')
qui local tauz2=log(`tauz2')
qui local r=0.5*log((1+'r')/(1-'r'))
qui drop _wz1 _wz2 _z1wz1 _z2wz2
program drop _all

** Model fitting - Equation 2.7
** The results should be identical to running: mvmeta b V,vars(b1 b2)
ml

ml model lf metagen_mb (b1:) (b2:) (logvar1:) (logvar2:) (z_rho:)
ml init b1:_cons=`muz1' b2:_cons=`muz2' logvar1:_cons=`tauz1'
logvar2:_cons=`tauz2' z_rho:_cons=`r'

qui ml check
qui ml search
if "`iter'" ~= "" {
ml maximize, difficult iter(30) nooutput
}

```

```

else{
qui ml maximize, difficult iter(30) nooutput
}
}

if `mf' == 1 {
local lambda=`muz1'/'muz2'
drop _wz1 _wz2 _z1wz1 _z2wz2
local tauz2=log(`tauz2')
program drop _all

** model fitting- Equation 2.12
qui ml model lf metagen_mf (lambda:) (b2:) (logvar2:)
qui ml init lambda:_cons=`lambda' b2:_cons=`muz2'
logvar2:_cons=`tauz2'
qui ml check
qui ml search, repeat(10)
if "`iter'" ~= "" {
ml maximize, difficult iter(30) nooutput
}
else{
qui ml maximize, difficult iter(30) nooutput
}
}

if (`mb'==1 | `mf'==1){

mat b =e(b)
mat V =e(V)

local ror3=b[1,2]
local rse3=sqrt(V[2,2])

```

```

if `mf'==1{
local ror2=b[1,1]* b[1,2]
local rse2=sqrt(V[1,1]*V[2,2])

local _lam=_b[lambda:_cons]
local _lam_err=_se[lambda:_cons]

qui lincom _b[b2:_cons]- `ror2'
qui return list
local _del=r(estimate)
local _del_err =r(se)

}
else{
local ror2=b[1,1]
local rse2=sqrt(V[1,1])

qui lincom _b[b2:_cons]- _b[b1:_cons]
qui return list
local _del=r(estimate)
local _del_err =r(se)

qui nlcom ratio:_b[b1:_cons]/_b[b2:_cons], post
local _lam=_b[ratio]
local _lam_err=_se[ratio]
}

}

if `lmm' == 1{
**Data re-arrangement in a format suitable for fitting the linear mixed
model
rename `1' _gen1
rename `2' _gen2

```

```

rename `3' _gen3
rename `4' _x1
rename `5' _x2
rename `6' _x3
rename `7' _sd1
rename `8' _sd2
rename `9' _sd3
qui reshape long _gen _x _sd,i(`study') j(z)
qui xi i.`study' i.z
**Calculation of the variance
gen _v=_sd^2/_gen
**gllamm for stability parameterizes the residual variance using the
logarithm of the standard error
gen s=log(sqrt(_v))
**Setting up one equation for each study
eq wgt: s
**Creating constraints needed for gllamm in order to perform inverse-
variance weighting
constraint define 1 [lns1]s=1
** Fitting the model of Equation 2.11
eq s2: _Iz_2
eq s3: _Iz_3
gllamm _x _Iz_2 _Iz_3 _I`study'* , i(`study') nrf(2) eqs(s2 s3) s(wgt)
constraint(1 ) nlp(8) noest
matrix A=M_initf
if "`iter'" ~= "" {
gllamm _x _Iz_2 _Iz_3 _I`study'* , i(`study') nrf(2) eqs(s2 s3) s(wgt)
constraint(1 ) nlp(8) from(A) long adapt nodisplay
}
else{
gllamm _x _Iz_2 _Iz_3 _I`study'* , i(`study') nrf(2) eqs(s2 s3) s(wgt)
constraint(1 ) nlp(8) from(A) long adapt nodisplay nolog
}

local ror2=_b[_Iz_2]
local rse2=_se[_Iz_2]

```

```

local ror3=_b[_Iz_3]
local rse3=_se[_Iz_3]

**Automatic calculation of confidence intervals for d
qui lincom _b[_Iz_3] - _b[_Iz_2]
    qui return list
local _del=r(estimate)
local _del_err =r(se)

**Automatic calculation of confidence intervals for lambda
qui nlcom ratio:_b[_Iz_2]/_b[_Iz_3], post
local _lam=_b[ratio]
local _lam_err=_se[ratio]

}

if (`pros' == 1 | `retr' == 1){

rename `1' _aa0
rename `2' _ab0
rename `3' _bb0
rename `4' _aa1
rename `5' _ab1
rename `6' _bb1

tempname group id slope2 slope3 patt a2group a3group gen chosen
tempvar graph1 graph2
**Data re-arrangement in a format suitable for logistic regression
qui reshape long _aa _ab _bb, i(`study') j(`group')

gen `id' = _n
rename _aa _count1

```

```

rename _ab _count2
rename _bb _count3
qui reshape long _count, i(`id') j(`gen')
qui drop `id'
gen wt1=_count
qui xi i.`study' i.`gen'
}
if `pros' == 1{

**Fitting the model based on the prospective likelihood - Equation 2.17
eq `slope2': _I`gen'_2
eq `slope3': _I`gen'_3
if "`iter'" ~= "" {
    gllamm `group' _I`gen'_2 _I`gen'_3 _I`study'_,fam(binom)
    link(logit) i(`study' ) nrf(2) eqs(`slope2' `slope3') w(wt) adapt
    nip(8) nodisplay
}
else {
    gllamm `group' _I`gen'_2 _I`gen'_3 _I`study'_,fam(binom)
    link(logit) i(`study' ) nrf(2) eqs(`slope2' `slope3') w(wt) adapt
    nip(8) nodisplay nolog
}

local ror2=_b[_I`gen'_2]
local rse2=_se[_I`gen'_2]

local ror3=_b[_I`gen'_3]
local rse3=_se[_I`gen'_3]

**Automatic calculation of confidence intervals for d
qui lincom _b[_I`gen'_3] - _b[_I`gen'_2]
qui return list
local _del=r(estimate)
local _del_err =r(se)
**Automatic calculation of confidence intervals for lambda

```

```

qui nlcom ratio: _b[_I`gen'_2]/_b[_I`gen'_3], post
local _lam= _b[ratio]
local _lam_err=_se[ratio]

}

if `retr' == 1 {

**Data re-arrangement in a format suitable for polytomous (multinomial)
logistic regression

qui sort `study' `group' `gen'
gen `patt'=_n
qui expand 3
qui sort `patt'
qui by `patt': gen alt=_n
gen `chosen'=alt==`gen'
qui tab alt,gen(a)
gen `a2group'=a2*`group'
gen `a3group'=a3*`group'

**Fitting the model based on the retrospective likelihood - Equation
2.19

eq `slope2':`a2group'
eq `slope3':`a3group'

if "`iter'" ~= "" {
gllamm alt `group' _I`study'_* , expand(`patt' `chosen' m) i(`study')
link(mlogit) family(binom) nrf(2) eqs(`slope2' `slope3') nip(8)
weight(wt) adapt nodisplay
}
else {
gllamm alt `group' _I`study'_* , expand(`patt' `chosen' m) i(`study')
link(mlogit) family(binom) nrf(2) eqs(`slope2' `slope3') nip(8)
weight(wt) adapt nodisplay nolog
}

mat b =e(b)
mat V =e(V)

```

```

local ror2=b[1,1]
local rse2=sqrt(V[1,1])

local ror3=b[1,9]
local rse3=sqrt(V[1,9])

**Automatic calculation of confidence intervals for d
qui lincom [c3]`group' - [c2]`group'
qui return list
local _del=r(estimate)
local _del_err =r(se)

**Automatic calculation of confidence intervals for lambda
qui nlcom ratio:[c2]`group'/[c3]`group', post
local _lam= _b[ratio]
local _lam_err=_se[ratio]
}

**Outputs**
if (`pros'==1 | `retr'==1 | `lmm'==1){

sort `study'
quietly by `study': gen dup = cond(_N==1,0,_n)
qui drop if dup>1
}

local rp2=2*min((1-normprob(`ror2'/'rse2')),normprob(`ror2'/'rse2'))
local rp3=2*min((1-normprob(`ror3'/'rse3')),normprob(`ror3'/'rse3'))

if "`eform'" ~= "" {
    local ror2=exp(`ror2')
    local ror3=exp(`ror3')

```

```

local rll2=`ror2'-1.96*`rse2'

local rul2=`ror2'+1.96*`rse2'


local rll3=`ror3'-1.96*`rse3'

local rul3=`ror3'+1.96*`rse3'


if "`studies'"~=""{

if (`con' == 1| `lmm' == 1){

qui metagraph _b1 _se1 `study', id(_studies) eform
x(0.95,1,1.15)combined(`ror2' `rll2' `rul2') bgcolor(black)
plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:AB vs. AA} , color(black))
yttitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph3')

qui metagraph _b2 _se2 `study', id(_studies) eform
x(0.95,1,1.15)combined(`ror3' `rll3' `rul3') bgcolor(black)
plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:BB vs. AA} , color(black))
yttitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph4')

}

else{

qui metagraph _b1 _se1 `study', id(_studies) eform
x(0.5,1,2,3)combined(`ror2' `rll2' `rul2') bgcolor(black)
plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:AB vs. AA} , color(black))
yttitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph3')

qui metagraph _b2 _se2 `study', id(_studies) eform
x(0.5,1,2,3)combined(`ror3' `rll3' `rul3') bgcolor(black)
plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:BB vs. AA} , color(black))
yttitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph4')

}

scatter _b1 _b2 [fweight=_V12], bgcolor(black)
plotregion(fcolor(white)) xtitle("OR(AB vs. AA)", color(black))
yttitle("OR(BB vs. AA)", color(black)) title({bf:Meta-analysis})
name(`graph5') mlabel(_studies)

}

else {

if (`con' == 1| `lmm' == 1){

qui metagraph _b1 _se1, id(`study') eform
x(0.95,1,1.15)combined(`ror2' `rll2' `rul2') bgcolor(black)

```

```

plotregion(fcolor(white)                                lcolor(black))
graphregion(color(ltblueishgray))  title({bf:AB vs. AA} , color(black))
yttitle({bf:Study}, color(black))  xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph3`)

qui      metagraph      _b2      _se2,      id(`study`)      eform
x(0.95,1,1.15)combined(`ror3'      `rll3'      `rul3')      bgcolor(black)
plotregion(fcolor(white)                                lcolor(black))
graphregion(color(ltblueishgray))  title({bf:BB vs. AA} , color(black))
yttitle({bf:Study}, color(black))  xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph4`)

}

else{

qui metagraph _b1 _se1, id(`study') eform x(0.5,1,2,3)combined(`ror2'
`rll2' `rul2') bgcolor(black) plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray))  title({bf:AB vs. AA} , color(black))
yttitle({bf:Study}, color(black))  xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph3`)

qui metagraph _b2 _se2, id(`study') eform x(0.5,1,2,3)combined(`ror3'
`rll3' `rul3') bgcolor(black) plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray))  title({bf:BB vs. AA} , color(black))
yttitle({bf:Study}, color(black))  xtitle({bf:Estimate of the Effect},
color(black)) nodraw name(`graph4`)

}

scatter      _b1      _b2      [fweight=_V12]      ,      bgcolor(black)
plotregion(fcolor(white))  xtitle("OR(AB vs. AA)", color(black))
yttitle("OR(BB vs. AA)", color(black))  title({bf:Meta-analysis})
name(`graph5`)

}

graph combine `graph3' `graph4' , cols(2) xcomm title({bf:Meta-
analysis}) altshrink

local rse2=exp(`rse2')
local rse3=exp(`rse3')

di " "

di in gr "          OddsRatio      Stad.Err.      p-value      [95% Conf.
Interval]"

}

else{

```

```

local rll2=`ror2'-1.96*`rse2'
local rul2=`ror2'+1.96*`rse2'

local rll3=`ror3'-1.96*`rse3'
local rul3=`ror3'+1.96*`rse3'

if "`studies'"~=""{
qui metagraph _b1 _se1 `study', id(_studies) combined(`ror2' `rll2'
`rul2') bgcolor(black) plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:AB vs. AA} , color(black))
ytitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) name(`graph1') nodraw

qui metagraph _b2 _se2 `study', id(_studies) combined(`ror3' `rll3'
`rul3') bgcolor(black) plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:BB vs. AA} , color(black))
ytitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) name(`graph2') nodraw

scatter _b1 _b2 [fweight=_V12] , bgcolor(black)
plotregion(fcolor(white)) xtitle("logOR(AB vs. AA)", color(black))
ytitle("logOR(BB vs. AA)", color(black)) title({bf:Meta-analysis})
name(`graph5') mlabel(_studies)
}

else{
qui metagraph _b1 _se1, id(`study') combined(`ror2' `rll2' `rul2')
bgcolor(black) plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:AB vs. AA} , color(black))
ytitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) name(`graph1') nodraw

qui metagraph _b2 _se2, id(`study') combined(`ror3' `rll3' `rul3')
bgcolor(black) plotregion(fcolor(white) lcolor(black))
graphregion(color(ltblueishgray)) title({bf:BB vs. AA} , color(black))
ytitle({bf:Study}, color(black)) xtitle({bf:Estimate of the Effect},
color(black)) name(`graph2') nodraw

scatter _b1 _b2 [fweight=_V12] , bgcolor(black)
plotregion(fcolor(white)) xtitle("logOR(AB vs. AA)", color(black))
ytitle("logOR(BB vs. AA)", color(black)) title({bf:Meta-analysis})
name(`graph5')

}

graph combine `graph1' `graph2' , cols(2) xcomm title({bf:Meta-
analysis}) altshrink

```

```

di " "

di in gr "          logOddsRatio      Stad.Err.      p-value      [95% Conf.
Interval]"

}

di in gr "AB vs. AA" in ye %8.3f `ror2' %17.3f `rse2' %17.3f `rp2'
%16.3f `rll2' %8.3f `rul2'

di in gr "BB vs. AA" in ye %8.3f `ror3' %17.3f `rse3' %17.3f `rp3'
%16.3f `rll3' %8.3f `rul3'

di in gr "-----"
-----"

    local _ulam = `_lam' + invnormal(0.975)*`_lam_err'
    local _llam = `_lam' + invnormal(0.025)*`_lam_err'
    local _udel = `_del' + invnormal(0.975)*`_del_err'
    local _ldel = `_del' + invnormal(0.025)*`_del_err'

di in gr "Confintence intervals for lamda"
di in gr "-----"
di in gr "      Lambda      Stad.Err.      [95% Conf. Interval]"
di in ye %8.3f `_lam' "          " %8.3f `_lam_err' "          "
%8.3f `_llam' "          " %8.3f `_ulam'

di in gr "Confintence intervals for delta"
di in gr "-----"
di in gr "      Delta      Stad.Err.      [95% Conf. Interval]"
di in ye %8.3f `_del' "          " %8.3f `_del_err' "          "
%8.3f `_ldel' "          " %8.3f `_udel'

di in gr "Combined test for gene effect"
di in gr "-----"
di in gr "      Coef.(chi-square)      p-value      "
di in ye %8.3f `_xchi2' "          " %8.3f `_xp'

return scalar b1 = `ror2'
return scalar b2 = `ror3'
return scalar se_b1 = `rse2'
return scalar se_b2 = `rse3'
return scalar p_b1 = `rp2'

```

```

return scalar p_b2 = `rp3'
return local lambda = `_lam'
return local se_lambda = `_lam_err'
return local delta = `_del'
return local se_delta = `_del_err'

drop *
restore

end

```

3.2.2 Υλοποίηση προγραμμάτων για τις μεθόδους model-free και bivariate

Για την μέθοδο της γενικής διμεταβλητής μέτα-ανάλυσης:

```

program define metagen_mb
args ln1 mub1 mub2 logtb1 logtb2 r
quietly {
tempname taub2 rho taub1
tempvar vb1 vb2 corr cov L det tmp
scalar `taub1'=exp(`logtb1')
scalar `taub2'=exp(`logtb2')
scalar `rho'=(exp(2*`r')-1)/(exp(2*`r')+1)
gen double `vb1'=`taub1' + _V11
gen double `vb2'=`taub2' + _V22
gen double `corr'=(`rho'*sqrt(`taub1'*`taub2') +
_V12)/sqrt(`vb1'*`vb2')
gen double `L'=log(`vb1'*`vb2'*(1-`corr'^2))+ /*
*/ ( (_b1-`mub1')^2/`vb1' + /*
*/ (_b2-`mub2')^2/`vb2' - /*
*/ 2*`corr'*(_b1-`mub1')*( _b2-`mub2')/sqrt(`vb1'*`vb2')) /*
*/ / (1-`corr'^2)

```

```

if (_type == 2){
replace `lnl'=-0.5*`L' - 0.5*log(2*_pi)
}

if (_type == 1 | _type == 0) {
gen double `det'=`vb1'*`vb2'-(`corr'*sqrt(`vb1'*`vb2'))^2

count
local n=r(N)
gen double `tmp'=`vb1'/'`det'
sum `tmp'
local xSx11 = r(sum)
replace `tmp'=`vb2'/'`det'
sum `tmp'
local xSx22 = r(sum)
replace `tmp'=(`corr'*sqrt(`vb1'*`vb2'))/'`det'
sum `tmp'
local xSx12 = r(sum)
replace `lnl' = -0.5 * (`L'+log(`xSx11'*`xSx22'-`xSx12'^2)/`n') -
0.5*log(2*_pi)
}
}

end

```

Για την μέθοδο της μέτα-ανάλυσης model-free:

```

program define metagen_mf
args lnl lambda mub2 logtb2
quietly {
tempname taub2 mub1
tempvar vb1 vb2 corr cov L det tmp
scalar `taub2'=exp(`logtb2')

```

```

scalar `mub1'=`mub2'*`lambda'
gen double `vb1'=`lambda'^2*`taub2' + _V11
gen double `vb2'=`taub2' + _V22
gen double `cov'=`lambda'*`taub2' + _V12
gen double `corr'=`cov'/sqrt(`vb2'*`vb1')
gen double `L'=log(`vb1'*`vb2'*(1-`corr'^2)) + /*
*/ ( (_b1 - `mub1' )^2/`vb1' + /*
*/ (_b2 -`mub2')^2/`vb2' - /*
*/ 2*`corr'*( _b1 -`mub1')*( _b2 - `mub2')/sqrt(`vb2'*`vb1') ) /*
*/ / (1-`corr'^2)

if (_type == 2){
replace `lnl'=-0.5*`L' - 0.5*log(2*_pi)
}

if (_type == 1 | _type == 0){
gen double `det'=`vb1'*`vb2'-(`corr'*sqrt(`vb1'*`vb2'))^2

count
local n=r(N)
gen double `tmp'=`vb1'/`det'
sum `tmp'
local xSx11 = r(sum)
replace `tmp'=`vb2'/`det'
sum `tmp'
local xSx22 = r(sum)
replace `tmp'=(`corr'*sqrt(`vb1'*`vb2'))/`det'
sum `tmp'
local xSx12 = r(sum)
replace `lnl' = -0.5 * (`L'+log(`xSx11'*`xSx22'-`xSx12'^2)/`n') -
0.5*log(2*_pi)

}

}

```

end

3.3 Παραδείγματα εκτέλεσης του προγράμματος

Η εντολή metagen εφαρμόστηκε σε αρκετές ήδη δημοσιευμένες μέτα-ανάλυσεις γενετικών μελετών συσχέτισης. Επιλέχθηκαν τρεις δημοσιευμένες μέτα-ανάλυσεις μελετών ασθενών-μαρτύρων (δυαδικό αποτέλεσμα) που χρησιμοποιήθηκαν επίσης ως παραδείγματα εργασίας από τους Minelli και συνεργάτες (Minelli et al, 2005a; Minelli et al, 2005b), προκειμένου τα αποτελέσματα να είναι εύκολα συγκρίσιμα για κάθε μέθοδο. Επιπλέον, χρησιμοποιήθηκαν ακόμα δύο δημοσιεύσεις μέτα-ανάλυσης μελετών για συνεχή αποτελέσματα.

Πιο συγκεκριμένα, για τα διακριτά αποτελέσματα χρησιμοποιήθηκαν οι μελέτες:

- Η μέτα-ανάλυση των Hani και συνεργατών (Hani et al, 1998) που αφορά τον συσχετισμό του Inwardly rectifying K⁺ channel (KIR 6.2) E23K πολυμορφισμού με τον Σακχαρώδη Διαβήτη τύπου II και η οποία περιλαμβάνει 4 μελέτες με 521 ασθενείς και 367 μάρτυρες.
- Η μέτα-ανάλυση των Kato και συνεργατών (Kato et al, 1999) που αφορά τον συσχετισμό του πολυμορφισμού M235T του αγγειοτενσινογόνου (AGT) με την Πρωτοπαθή (ουσιώδη) Υπέρταση και η οποία περιλαμβάνει 7 μελέτες με 1336 ασθενείς και 1225 μάρτυρες.
- Η μέτα-ανάλυση των Wheeler και συνεργατών (Wheeler et al, 2004) που αφορά τον συσχετισμό του πολυμορφισμού Q192R της παραοξονάσης (PON1) με το Έμφραγμα του Μυοκαρδίου και η οποία περιλαμβάνει 19 μελέτες με 5725 ασθενείς και 8072 μάρτυρες.

Για τα συνεχή αποτελέσματα χρησιμοποιήθηκαν οι μελέτες:

- Η μέτα-ανάλυση των Sayed-Tabatabaei και συνεργατών (Sayed-Tabatabaei et al, 2003) που αφορά τον συσχετισμό του πολυμορφισμού I/D του γονιδίου του Μετατρεπτικού Ενζύμου της Αγγειοτενσίνης (ACE) με το πάχος του τοιχώματος της καρωτίδας αρτηρίας και η οποία περιλαμβάνει 23 μελέτες με 7795 συμμετέχοντες.
- Η μέτα-ανάλυση ατομικών δεδομένων ασθενών (individual patients' data) των Boekholdt και συνεργατών (Boekholdt et al, 2005) που αφορά τη σχέση του πολυμορφισμού TaqIB του διαμεσολαβητή για τη μεταφορά των εστέρων

χοληστερόλης (CETP) με τα επίπεδα HDL-C και η οποία περιλαμβάνει 10 μελέτες με 13667 συμμετέχοντες.

3.3.1 Μελέτες με διακριτά αποτελέσματα

Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα

Meta-analysis									
Method	KIR 6.2 E23K and Type II Diabetes Mellitus (Hani et al, 1998)			AGT M235T and Essential Hypertension (Kato et al, 1999)			PON1 Q192R and Myocardial Infarction (Wheeler et al, 2004)		
	OR _{AB} (95% C.I.)	OR _{BB} (95% C.I.)	λ (95% C.I.)	OR _{AB} (95% C.I.)	OR _{BB} (95% C.I.)	λ (95% C.I.)	OR _{AB} (95% C.I.)	OR _{BB} (95% C.I.)	λ (95% C.I.)
Univariate methods									
Summary methods (FE)	1.22 (0.90, 1.64)	2.21 (1.43, 3.40)	-	1.16 (0.77, 1.76)	1.58 (1.06, 2.35)	-	1.09 (1.01, 1.17)	1.15 (1.01, 1.31)	-
Summary methods (RE, MM-DL)	1.24 (0.85, 1.80)	2.21 (1.43, 3.40)	-	1.16 (0.77, 1.76)	1.58 (1.06, 2.35)	-	1.10 (0.99, 1.23)	1.14 (0.99, 1.32)	-
Summary methods (RE, ML)	1.23 (0.88, 1.72)	2.21 (1.43, 3.40)	-	1.16 (0.77, 1.76)	1.58 (1.06, 2.35)	-	1.10 (1.00, 1.21)	1.15 (1.00, 1.31)	-
Summary methods (RE, REML)	1.24 (0.84, 1.81)	2.21 (1.43, 3.40)	-	1.16 (0.77, 1.76)	1.58 (1.06, 2.35)	-	1.10 (0.99, 1.21)	1.15 (1.00, 1.31)	-
Multivariate methods									
Bivariate regression (ML)	1.23 (0.87, 1.73)	2.16 (1.40, 3.33)	0.27 (-0.15, 0.69)	1.09 (0.69, 1.74)	1.77 (1.10, 2.86)	0.16 (-0.56, 0.88)	1.09 (1.00, 1.20)	1.15 (1.00, 1.31)	0.65 (-0.17, 1.47)
Bivariate regression (REML)	1.24 (0.84, 1.83)	2.17 (1.41, 3.35)	0.27 (-0.21, 0.76)	1.09 (0.69, 1.75)	1.78 (1.09, 2.91)	0.15 (-0.57, 0.88)	1.09 (0.99, 1.20)	1.14 (1.00, 1.32)	0.66 (-0.20, 1.53)
Logistic regression (ML-prospective)	1.22 (0.91, 1.63)	2.15 (1.40, 3.30)	0.25 (-0.10, 0.61)	1.06 (0.67, 1.66)	1.72 (1.13, 2.62)	0.11 (-0.67, 0.88)	1.09 (1.01, 1.19)	1.14 (1.00, 1.29)	0.69 (-0.25, 1.62)
Multinomial logistic regression (MLretrospective)	1.21 (0.91, 1.63)	2.15 (1.40, 3.30)	0.25 (-0.10, 0.61)	1.17 (0.76, 1.78)	1.87 (1.16, 3.02)	0.25 (-0.32, 0.81)	1.09 (1.01, 1.17)	1.15 (1.02, 1.30)	0.59 (-0.03, 1.20)
Model-free approach (ML-unbounded λ)	1.21 (0.90, 1.63)	2.14 (1.39, 3.29)	0.25 (-0.11, 0.61)	1.00 (0.66, 1.53)	1.64 (0.99, 2.72)	0.01 (-0.83, 0.85)	1.08 (1.00, 1.17)	1.17 (1.04, 1.33)	0.53 (-0.03, 1.13)
Model-free approach (ML-bounded λ)	1.21 (1.08, 1.63)	2.14 (1.43, 3.29)	0.25 (0.00, 0.69)	1.00 (1.00, 1.62)	1.64 (1.15, 3.05)	0.01 (0.00, 0.52)	1.08 (1.01, 1.17)	1.17 (1.04, 1.33)	0.53 (0.09, 1.00)
Bayesian model-free approach (prospective)	1.16 (0.99, 1.77)	2.01 (0.97, 4.09)	0.24 (0.00, 0.81)	1.08 (1.00, 1.74)	1.81 (1.05, 3.66)	0.16 (0.00, 0.54)	1.08 (1.00, 1.21)	1.15 (1.01, 1.33)	0.63 (0.03, 0.99)
Bayesian model-free approach (retrospective)	1.16 (0.99, 1.80)	2.03 (0.96, 3.96)	0.24 (0.00, 0.79)	1.09 (1.00, 1.71)	1.83 (1.06, 3.60)	0.17 (0.00, 0.54)	1.08 (1.00, 1.21)	1.15 (1.02, 1.34)	0.62 (0.03, 0.99)

Πίνακας 3: Αποτελέσματα παραδοσιακών και πολυμεταβλητών μεθόδων με διακριτά αποτελέσματα.

όπου οι μέθοδοι που προτάθηκαν σε αυτήν την εργασία χρησιμοποιήθηκαν για σύγκριση με τις γενετικές model-free μεθόδους του Minelli και των συνεργατών του, καθώς και με μεθόδους μονομεταβλητών συνοπτικών (summary) δεδομένων.

Στον Πίνακα 3 παρατίθενται τα logOR των ομόζυγων και ετερόζυγων γονοτύπων για το μεταλλαγμένο αλληλόμορφο, η εκτίμηση του γενετικού μοντέλου κληρονομικότητας (λ) και τα αντίστοιχα διαστήματα εμπιστοσύνης 95%. FE: σταθερές επιδράσεις (fixed effects); RE: τυχαίες επιδράσεις (random effects); ML: Μέγιστη Πιθανοφάνεια (Maximum Likelihood); REML: Περιορισμένη Μέγιστη Πιθανοφάνεια (Restricted Maximum Likelihood); MM-DL: Μέθοδος Dersimonian και Laird.

Όλες οι μέθοδοι που εξετάστηκαν παρήγαγαν σχεδόν ταυτόσημες εκτιμήσεις και 95% διαστήματα εμπιστοσύνης. Ο παράγοντας εκτίμησης του γενετικού μοντέλου (λ) είναι σχεδόν ίδιος σε όλες τις περιπτώσεις και κυμαίνεται από 0,24 έως 0,27, υποδεικνύοντας υπολειπόμενο μοντέλο κληρονομικότητας. Αντίθετα, στη μέτα-ανάλυση των Hani και συνεργατών (Hani et al, 1998), στην Bayesian model-free μέθοδο, δηλ σε αυτήν με το μικρότερο αριθμό συμπεριλαμβανόμενων μελετών, προέκυψαν ευρύτερα διαστήματα εμπιστοσύνης. Αυτό υποδηλώνει αβεβαιότητα από την προηγούμενη κατανομή των παραμέτρων και ανακρίβεια στην εκτίμηση της υπολειπόμενης ετερογένειας.

Στη μέτα-ανάλυση που αφορά τη σχέση μεταξύ του πολυμορφισμού M235T AGT και της Πρωτοπαθούς υπέρτασης (Kato et al, 1999), όλες οι μέθοδοι συμφωνούν ότι δεν υπάρχει σημαντικός κίνδυνος που συσχετίζεται με τον ετερόζυγο γονότυπο, υποδεικνύοντας έμμεσα ένα υπολειπόμενο μοντέλο κληρονομικότητας. Οι μέθοδοι που χρησιμοποιούν την αναδρομική πιθανοφάνεια εμφανίζουν ελαφρώς αυξημένες εκτιμήσεις για το OR_{BB} σε σύγκριση με αυτές που χρησιμοποιούν την προοπτική πιθανοφάνεια, καθώς και στις μεθόδους που χρησιμοποιούν συνοπτικά δεδομένα. Επιπλέον, παρατηρείται ότι σε όλες τις περιπτώσεις, οι πολυμεταβλητές μέθοδοι παράγουν ευρύτερα διαστήματα εμπιστοσύνης σε σύγκριση με τις μονομεταβλητές. Τέλος, οι εκτιμήσεις του λ κυμαίνονται από 0,11 έως 0,25 σε όλες τις περιπτώσεις, υποδεικνύοντας ότι είναι πιθανότερο ένα υπολειπόμενο μοντέλο.

Στη μέτα-ανάλυση που αφορά τη συσχέτιση του πολυμορφισμού της παραοξονάσης (PON1) Q192R με το Έμφραγμα του Μυοκαρδίου (Wheeler et al, 2004), που περιλάμβανε 19 μελέτες, παρατηρείται ότι όλες οι μέθοδοι παρήγαγαν σχεδόν πανομοιότυπες εκτιμήσεις. Οι εκτιμήσεις του λ κυμαίνονται από 0,53 έως 0,69, υποδεικνύοντας ότι είναι πιθανότερο ένα συν-επικρατές μοντέλο.

3.3.2 Μελέτες με συνεχή αποτελέσματα

Τα αποτελέσματα που προέκυψαν από την ανάλυση των δεδομένων από δύο μέτα-αναλύσεις μελετών με συνεχές αποτέλεσμα παρουσιάζονται παρακάτω.

Meta-analysis						
Method	CETP TaqIB polymorphism and serum HDL-C levels (mmol/L) (Boekholdt et al, 2005)			ACE I/D polymorphism and carotid artery wall thickness (x10 ⁻¹ mm) (Sayed-Tabatabaei et al, 2003)		
	β_{AB} (95% C.I.)	β_{BB} (95% C.I.)	λ (95% C.I.)	β_{AB} (95% C.I.)	β_{BB} (95% C.I.)	λ (95% C.I.)
Univariate methods						
Summary methods (FE)	0.0397 (0.0297, .0497)	0.1118 (0.0980, 0.1255)	-	0.0672 (0.000, 0.133)	0.1136 (0.0351, 0.1915)	-
Summary methods (RE, MM-DL)	0.0397 (0.0297, .0497)	0.1118 (0.0980, 0.1255)	-	0.1015 (0.0085, 0.1945)	0.2080 (0.0738, 0.3423)	-
Summary methods (RE, ML)	0.0397 (0.0297, 0.0497)	0.1118 (0.0980, 0.1255)	-	0.0918 (0.0072, 0.1763)	0.1905 (0.0696, 0.3114)	-
Summary methods (RE, REML)	0.0399 (0.0297, 0.0499)	0.1118 (0.0980, 0.1255)	-	0.0947 (0.0077, 0.1816)	0.1968 (0.0714, 0.3223)	-
Multivariate methods						
Bivariate regression (ML)	0.0397 (0.0288, 0.0506)	0.1107 (0.0964, 0.1250)	0.36 (0.26, 0.45)	0.1192 (0.0059, 0.2325)	0.2001 (0.0549, 0.3452)	0.60 (0.22, 0.97)
Bivariate regression (REML)	0.0394 (0.0281, 0.0508)	0.1112 (0.0960, 0.1265)	0.35 (0.25, 0.46)	0.1269 (0.0114, 0.2425)	0.2099 (0.0614, 0.3584)	0.60 (0.24, 0.97)
Random coefficient model (ML)	0.0378 (0.0236, 0.0521)	0.1151 (0.1002, 0.1299)	0.33 (0.20, 0.45)	0.2210 (0.0733, 0.3687)	0.4704 (0.2149, 0.7258)	0.47 (0.22, 0.72)
Model-free approach (ML-unbounded λ)	0.0397 (0.0298, 0.0497)	0.1103 (0.0967, 0.1241)	0.36 (0.28, 0.44)	0.1191 (0.0056, 0.2325)	0.1999 (0.0548, 0.3451)	0.60 (0.22, 0.97)

Πίνακας 4: Αποτελέσματα παραδοσιακών και πολυμεταβλητών μεθόδων με συνεχή

αποτελέσματα.

Στον Πίνακα 4 αναπαριστώνται οι μέσες διαφορές (β) για ομόζυγα και ετερόζυγα άτομα για το μεταλλαγμένο αλληλόμορφο παράγοντας εκτίμησης του γενετικού μοντέλου κληρονομικότητας (λ) και τα αντίστοιχα διαστήματα εμπιστοσύνης 95%. FE: σταθερές επιδράσεις (fixed effects); RE: τυχαίες επιδράσεις (random effects); ML: Μέγιστη Πιθανοφάνεια (Maximum Likelihood); REML: Περιορισμένη Μέγιστη Πιθανοφάνεια (Restricted Maximum Likelihood); MM-DL: Μέθοδος των Dersimonian και Laird.

Στη μέτα-ανάλυση ατομικών δεδομένων ασθενών (individual patients' data) των Boekholdt και συνεργατών (Boekholdt et al, 2005) σχετικά με τη σχέση του πολυμορφισμού TaqIB του διαμεσολαβητή για τη μεταφορά των εστέρων χοληστερόλης (CETP) με τα επίπεδα HDL-C, όλες οι μέθοδοι παράγουν σχεδόν ταυτόσημα αποτελέσματα που υποδεικνύουν την απουσία ετερογένειας. Και σε αυτή την ανάλυση, τα διαστήματα εμπιστοσύνης που παράγουν οι πολυμεταβλητές μέθοδοι είναι ελαφρώς ευρύτερα σε σύγκριση με αυτά που παράγονται από τις μονομεταβλητές μεθόδους. Οι εκτιμήσεις του λ από 0,35 έως 0,36 υποδεικνύουν ότι είναι πιο πιθανό ένα συν-επικρατές μοντέλο. Παρόμοια αποτελέσματα προέκυψαν χρησιμοποιώντας το μεικτό μοντέλο (LMM) της εξίσωσης (2.12).

Στη μέτα-ανάλυση των Sayed-Tabatabaei και συνεργατών (Sayed-Tabatabaei et al, 2003) σχετικά με τον συσχετισμό του πολυμορφισμού I/D του γονιδίου του Μετατρεπτικού Ενζύμου της Αγγιοτενσίνης (ACE) με το πάχος του τοιχώματος της καρωτίδας αρτηρίας, η μεγάλη ετερογένεια οδήγησε σε διάφορες εκτιμήσεις που προέκυψαν από τα μοντέλα σταθερών επιδράσεων σε σύγκριση με τα μοντέλα τυχαίων επιδράσεων. Οι πολυμεταβλητές μέθοδοι εμφανίζουν ευρύτερα διαστήματα εμπιστοσύνης για τη μέση διαφορά. Το λ εκτιμάται ίσο με 0,60, υποδεικνύοντας ότι είναι πιθανότερο ένα συν-επικρατές μοντέλο. Σε αυτήν τη μέτα-ανάλυση, το μεικτό μοντέλο που παρουσιάζεται στην εξίσωση (2.12) παράγει διαφορετικά αποτελέσματα σε σύγκριση με το διμεταβλητό μοντέλο της εξίσωσης (2.7). Ο λόγος για αυτήν τη διαφορά πρέπει να αποδοθεί στη μεγάλη μεταβλητότητα μεταξύ των μελετών αφού αυτή η ανάλυση περιλαμβάνει 15 μελέτες που πραγματοποιήθηκαν σε πληθυσμούς χαμηλού κινδύνου (low-risk), καθώς και σε 8 μελέτες με πληθυσμό υψηλού κινδύνου (high-risk). Λαμβάνοντας αυτό υπόψιν, ο μέσος AWT των ατόμων που φέρουν τον QQ γονότυπο (δηλαδή το AA) στους πληθυσμούς χαμηλού κινδύνου είναι 7.2×10^{-1} mm, ενώ η αντίστοιχη τιμή στους πληθυσμούς υψηλού κινδύνου είναι 9.2×10^{-1} mm, υποδεικνύοντας σημαντική ετερογένεια. Επιπλέον, φαίνεται ότι υπάρχει κάποια αλληλεπίδραση μεταξύ του τύπου του πληθυσμού (υψηλού ή χαμηλού κινδύνου) και των επιδράσεων του γονότυπου. Περιλαμβάνοντας στο μεικτό μοντέλο τον τύπο του πληθυσμού (χαμηλός/υψηλός) ως παράγοντα πρόβλεψης, τα αποτελέσματα δεν επηρεάστηκαν, αλλά η εκτίμηση για τη μεταβλητή κινδύνου ήταν σημαντικά υψηλή.

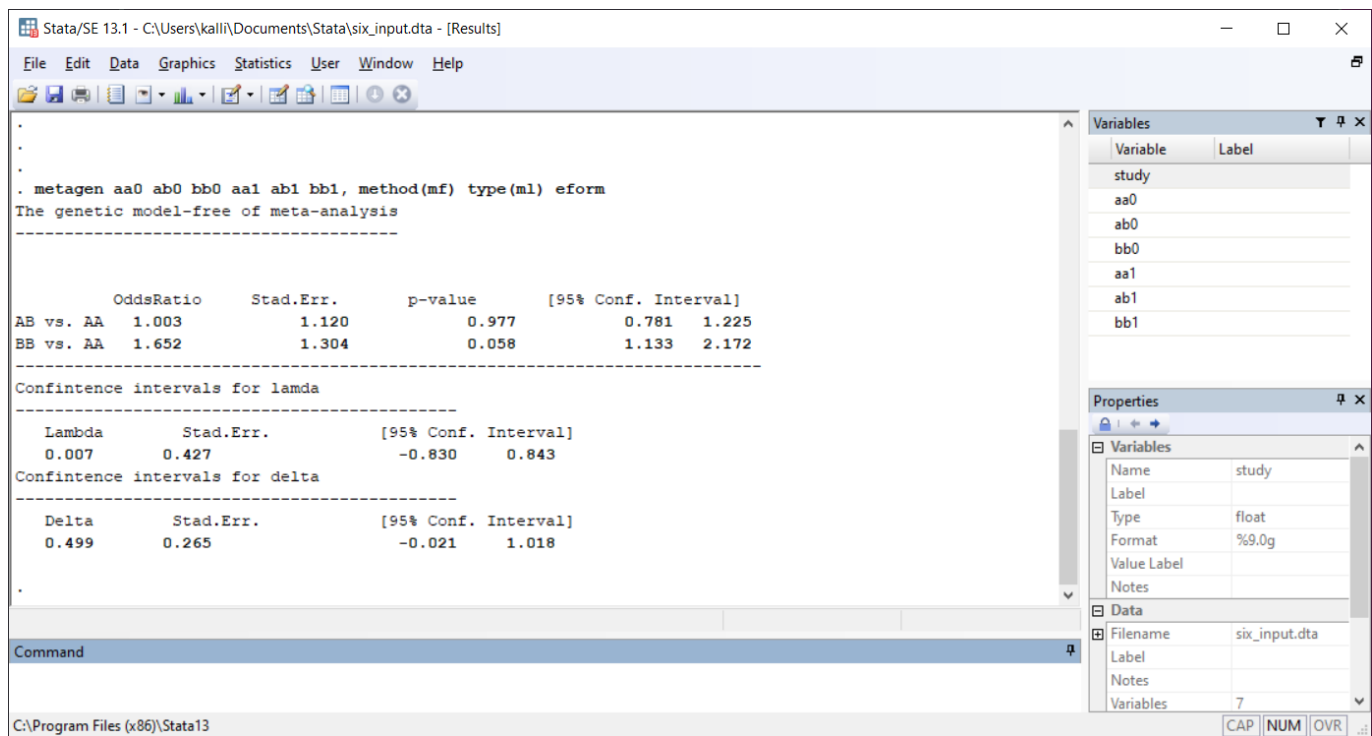
Αναλύοντας τώρα ξεχωριστά τις δύο ομάδες μελετών, το διμεταβλητό ML μοντέλο

παρήγαγε ασήμαντα αποτελέσματα για και τους δύο γονοτύπους όσον αφορά τις 15 μελέτες σε πληθυσμούς χαμηλού κινδύνου (για AB: 0.0494 με 95% CI: [-0.04801, 0.1471] και για BB: 0.0954 με 95% CI: [-0.0244, 0.2152]), ενώ το ίδιο μοντέλο παρήγαγε σημαντικά υψηλά αποτελέσματα για τις 8 μελέτες σε πληθυσμούς υψηλού κινδύνου (για AB: 0.4258 με 95% CI: [0.1122, 0.7393] και για BB: 0.6872 με 95% CI: [0.2841, 1.090]). Το μεικτό μοντέλο για τις ίδιες αναλύσεις παράγει κάπως διαφορετικά αποτελέσματα, καθώς στους πληθυσμούς χαμηλού κινδύνου εμφανίζει σημαντικές εκτιμήσεις (για AB: 0.1952 με 95% CI: [0.0763, 0.3141] και για BB: 0.2485 με 95% CI: [0.1373, 0.3597]), ενώ για τους πληθυσμούς υψηλού κινδύνου η σημαντικότητα των εκτιμήσεων είναι μεγαλύτερη (για AB: 0.5059 με 95% CI: [0.2619, 0.7501] και για BB: 1.020 με 95% CI: [0.5265, 1.5136]). Σε αυτήν τη μέτα-ανάλυση, ο υποκείμενος κίνδυνος (underlying risk) του πληθυσμού φαίνεται να είναι μια σημαντική πηγή ετερογένειας που πρέπει να ληφθεί υπόψη. Επιπλέον, φαίνεται ότι αυτή η μεταβλητότητα επεκτείνεται στο γενετικό μοντέλο, καθώς οι εκτιμήσεις του λ , φαίνεται να κινούνται προς την συν-κυρίαρχη κληρονομικότητα (δηλαδή $\lambda=0,5$) όταν λαμβάνονται υπόψη οι πληθυσμοί χαμηλού κινδύνου.

3.3.3 Απεικονίσεις αποτελεσμάτων στο Stata

Παρακάτω φαίνονται σε πίνακες αλλά και γραφικά τα αποτελέσματα μερικών μεθόδων μέτα-ανάλυσης όπως εμφανίζονται μετά την εκτέλεση του προγράμματος στο Stata με διακριτά και συνεχή αποτελέσματα. Οι μέθοδοι μέτα-ανάλυσης που εκτελέστηκαν είναι η γενική διμεταβλητή μέθοδος μέτα-ανάλυσης, η μέθοδος μέτα-ανάλυσης με την χρήση προοπτικής πιθανοφάνειας και η γενική μέθοδος μέτα-ανάλυσης για συνεχή αποτελέσματα, των οποίων τα αποτελέσματα συζητήθηκαν στις ενότητες 3.3.1 και 3.3.2. Οι μελέτες στις οποίες στηρίχθηκαν είναι η μελέτη των Kato και συνεργατών (Kato et al, 1999) σχετικά με τον συσχετισμό του πολυμορφισμού M235T του αγγειοτενσινογόνου (AGT) με την Πρωτοπαθή (ουσιώδη) Υπέρταση και η μελέτη μέτα-ανάλυσης ατομικών δεδομένων ασθενών (individual patients' data) των Boekholdt και συνεργατών (Boekholdt et al, 2005) σχετικά με τη σχέση του πολυμορφισμού TaqIB του διαμεσολαβητή για τη μεταφορά των εστέρων χοληστερόλης (CETP) με τα επίπεδα HDL-C.

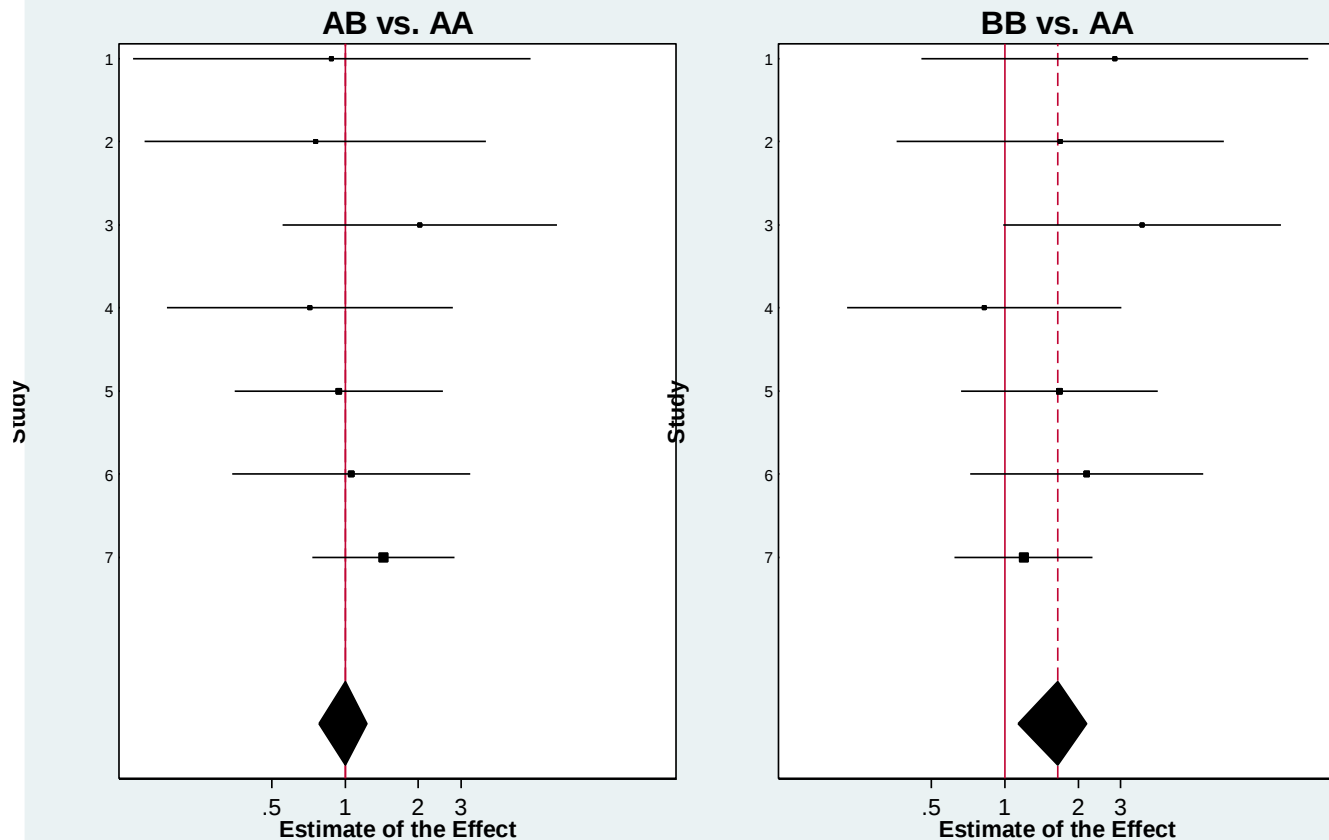
Για τις μεθόδους με διακριτά αποτελέσματα:



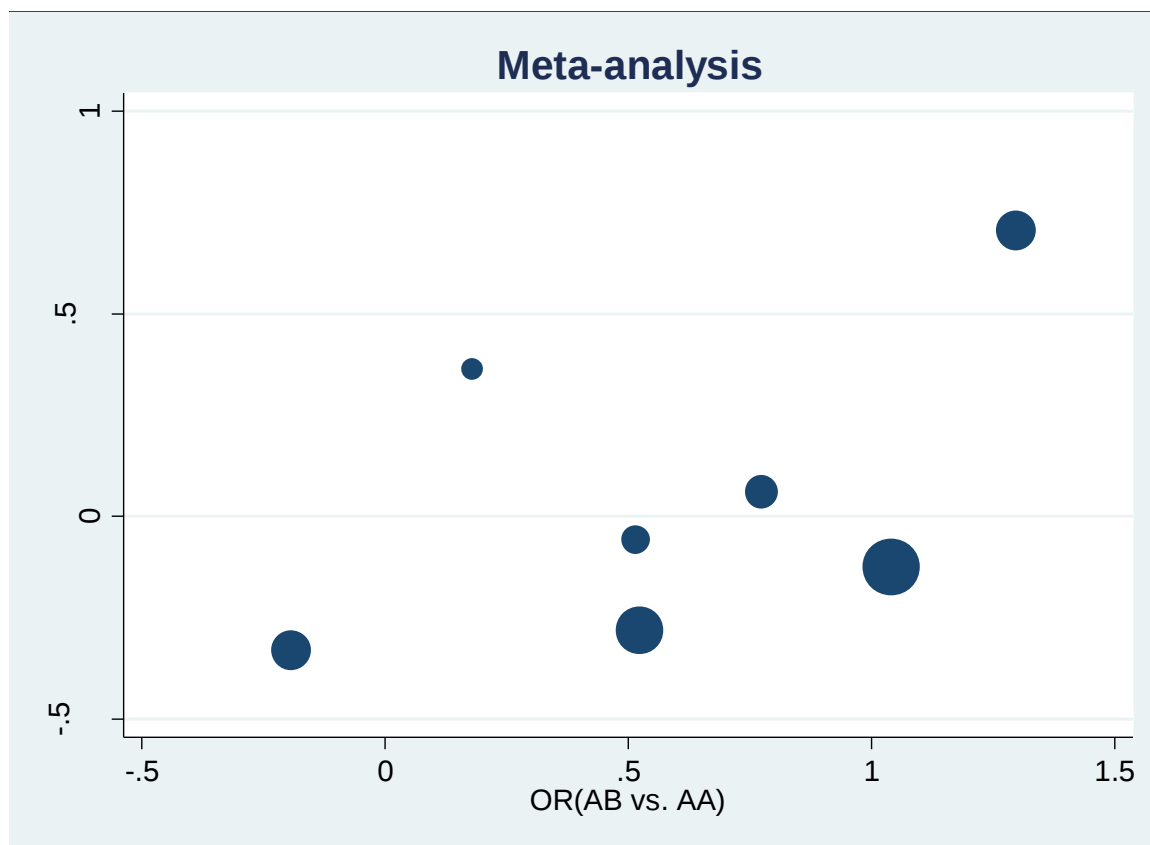
Εικόνα 1: Αποτελέσματα εκτέλεσης της metagen με model-free προσέγγιση

Παραπάνω φαίνονται τα αποτελέσματα μετά την εκτέλεση της εντολής metagen για την μελέτης των Kato και συνεργατών (Kato et al, 1999) με την χρήση της model-free μεθόδου μέτα-ανάλυσης με Μέγιστη Πιθανοφάνεια (ML). Τα αποτελέσματα έχουν εκθετική μορφή.

Meta-analysis



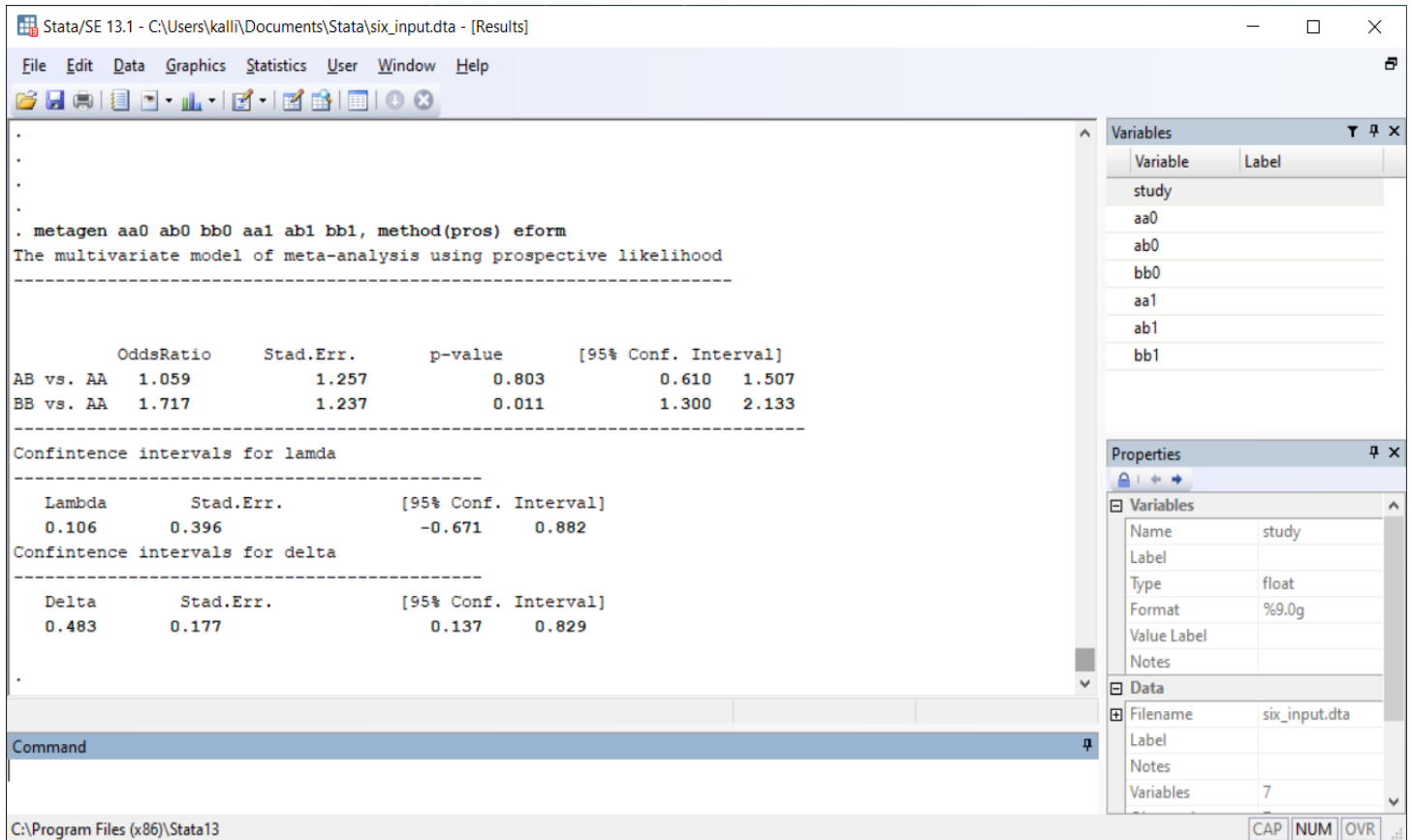
Εικόνα 2: Πρώτη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της model-free μεθόδου με Μέγιστη Πιθανοφάνεια (ML).



Εικόνα 3: Δεύτερη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της model-free μεθόδου με Μέγιστη Πιθανοφάνεια (ML).

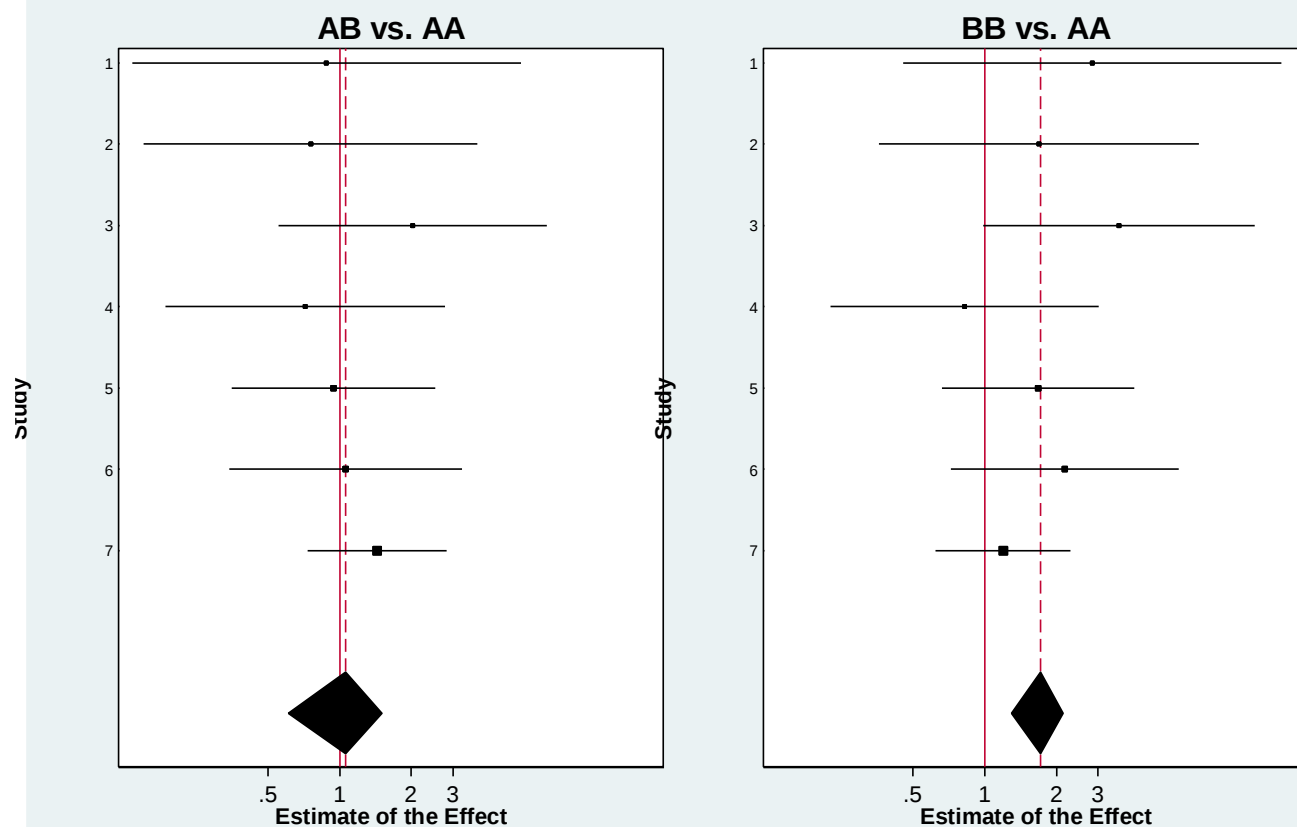
Στην Εικόνα 2 απαναπαρηστώνται τα αποτελέσματα της μελέτης των Kato και συνεργατών (Kato et al, 1999) με την χρήση της model-free μεθόδου μέτα-ανάλυσης με Μέγιστη Πιθανοφάνεια (ML). Φαίνονται τα δύο forest plot για κάθε περίπτωση (AB έναντι AA και BB έναντι AA) σε κοινό άξονα των x. Στην Εικόνα 3 αναπαρηστώνται τα ίδια αποτελέσματα χρησιμοποιώντας scatter plot.

Τα αποτελέσματα της μελέτης των Kato και συνεργατών (Kato et al, 1999) μετά την εκτέλεση της εντολής metagen με την χρήση της μεθόδου μέτα-ανάλυσης με προοπτική πιθανοφάνεια φαίνονται παρακάτω. Τα αποτελέσματα και σε αυτή την μέθοδο έχουν εκθετική μορφή.

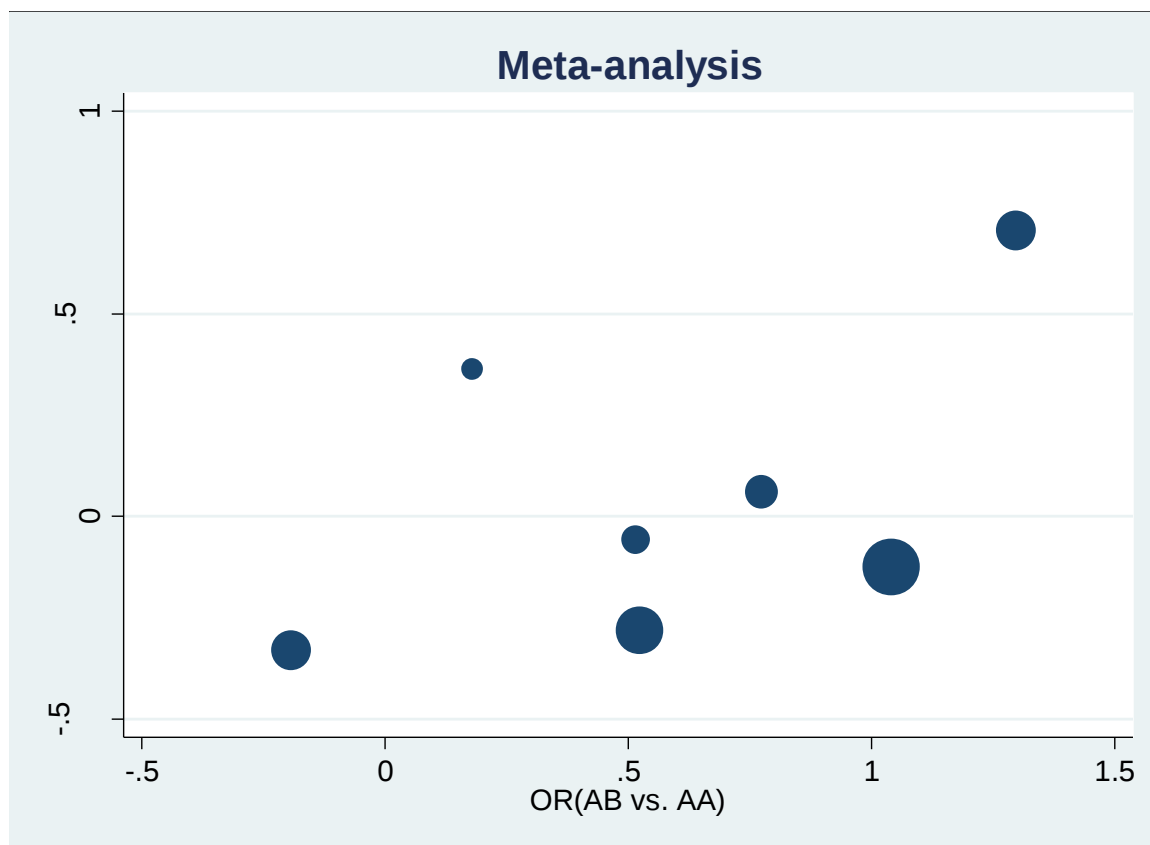


Εικόνα 4: Αποτελέσματα εκτέλεσης της metagen με προοπτική πιθανοφάνεια

Meta-analysis



Εικόνα 5: Πρώτη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου προοπτικής πιθανοφάνειας.



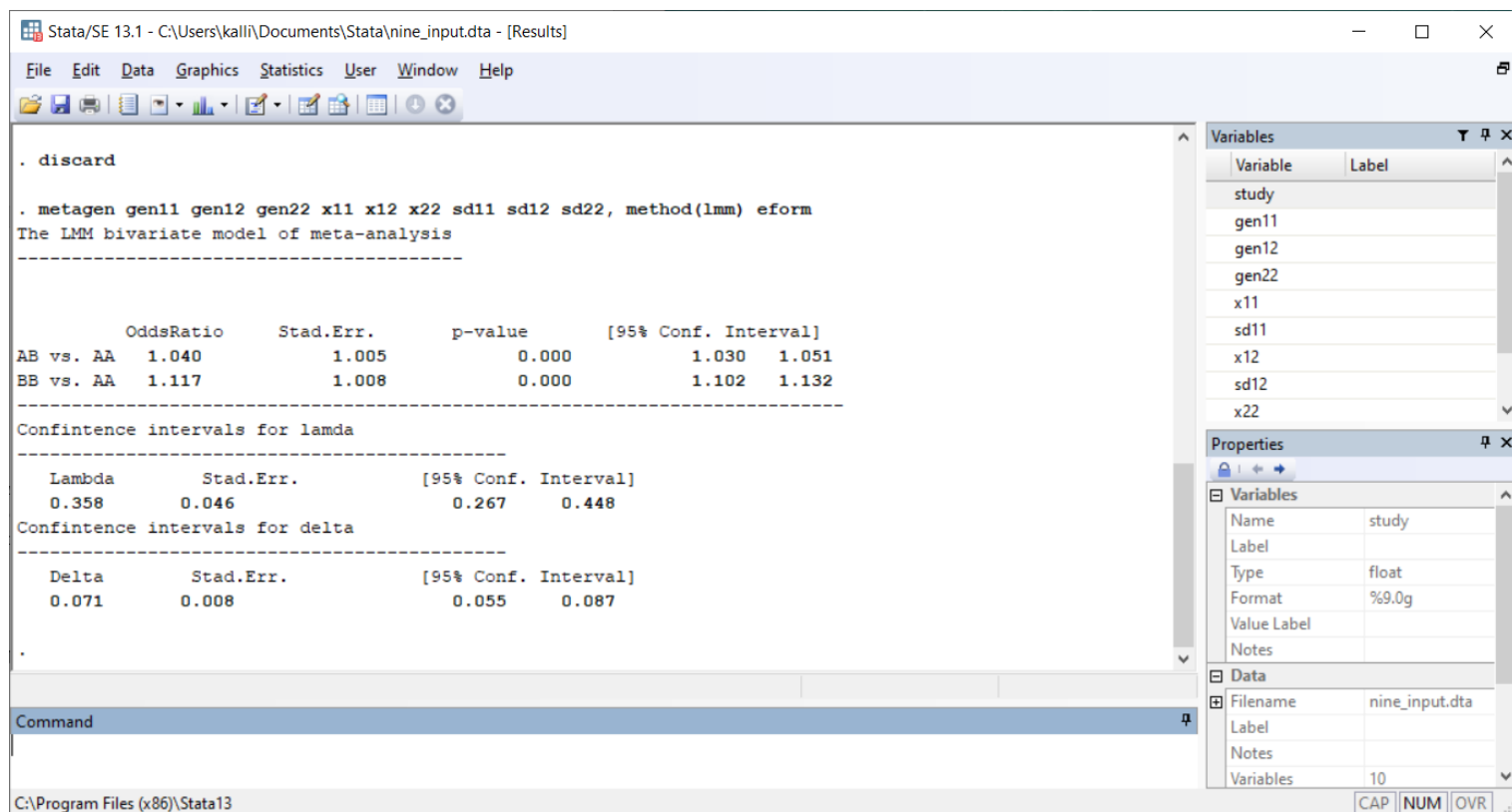
Εικόνα 6: Δεύτερη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου προοπτικής πιθανοφάνειας.

Στην Εικόνα 5 απαναπαρηστώνται τα αποτελέσματα της μελέτης των Kato και συνεργατών (Kato et al, 1999) με την χρήση της μεθόδου μέτα-ανάλυσης με χρήση της προοπτικής πιθανοφάνειας. Φαίνονται τα δύο forest plot για κάθε περίπτωση (AB έναντι AA και BB έναντι AA) σε κοινό άξονα των x. Στην Εικόνα 6 αναπαρηστώνται τα ίδια αποτελέσματα χρησιμοποιώντας scatter plot.

Παρατηρείται ότι στα γραφήματα και των δύο μεθόδων δεν υπάρχει κάποια αξιοσημείωτη διαφορά αφού τα αποτελέσματα που προκύπτουν είναι σχεδόν ίδια ανεξαρτήτως της μεθόδου που χρησιμοποιήθηκε και του τύπου των δεδομένων, συνοπτικά ή δυαδικά.

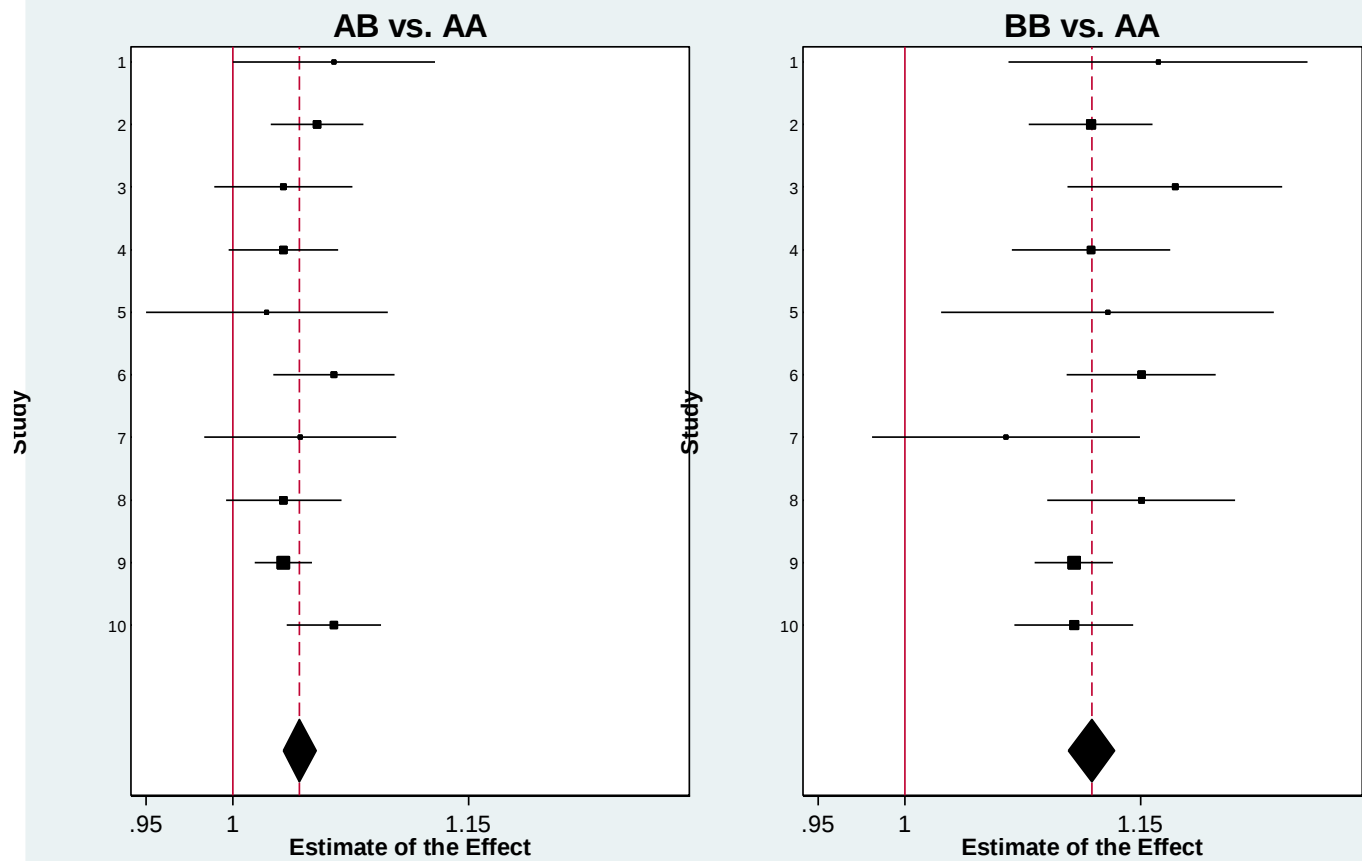
Για τη μέθοδο με συνεχή αποτελέσματα:

Τα ακόλουθα αποτελέσματα προκύπτουν μετά την εκτέλεση της μεθόδου metagen για την μελέτη των Boekholdt και συνεργατών (Boekholdt et al, 2005) σχετικά με τη σχέση του πολυμορφισμού TaqIB του διαμεσολαβητή για τη μεταφορά των εστέρων χοληστερόλης (CETP) με τα επίπεδα HDL-C με την χρήση μέτα-ανάλυσης με την μέθοδο linear mixed model. Τα αποτελέσματα και σε αυτή την εκτέλεση παρουσιάζονται σε εκθετική μορφή ώστε να είναι πιο εύκολη η σύγκριση με τον Πίνακα 4.

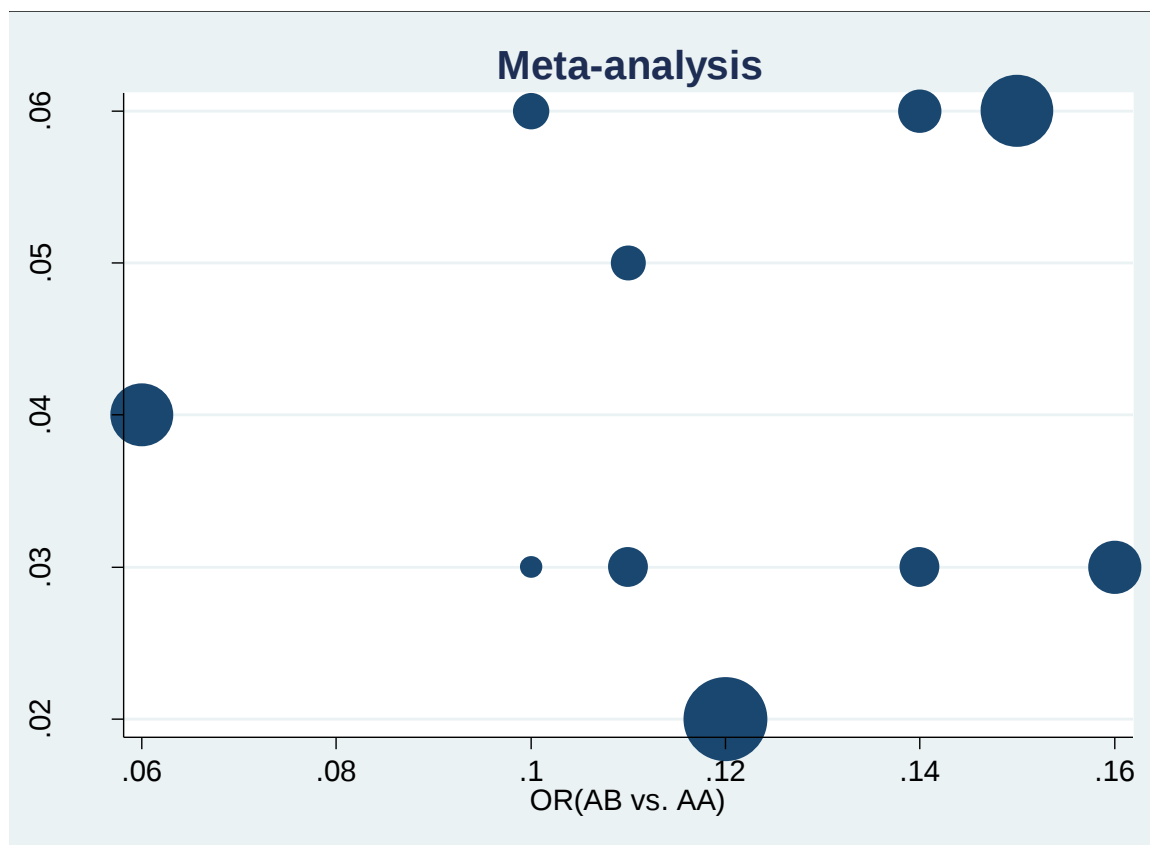


Εικόνα 7: Αποτελέσματα εκτέλεσης της metagen με γραμμικό μεικτό μοντέλο

Meta-analysis



Εικόνα 8: Πρώτη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου linear mixed model.



Εικόνα 9: Δεύτερη αναπαράσταση αποτελεσμάτων μέτα-ανάλυσης με χρήση της μεθόδου linear mixed model.

Στην Εικόνα 8 αναπαριστώνται τα αποτελέσματα της μελέτης των Boekholdt και συνεργατών (Boekholdt et al, 2005) σχετικά με τη σχέση του πολυμορφισμού TaqIB του διαμεσολαβητή για τη μεταφορά των εστέρων χοληστερόλης (CETP) με τα επίπεδα HDL-C με την χρήση linear mixed model μεθόδου μέτα-ανάλυσης για συνεχή αποτελέσματα. Φαίνονται τα δύο forest plot για κάθε περίπτωση (AB έναντι AA και BB έναντι AA) σε κοινό άξονα των x. Στην Εικόνα 9 αναπαριστώνται τα ίδια αποτελέσματα χρησιμοποιώντας scatter plot.

4. Συζήτηση – Συμπεράσματα

Η συντριπτική πλειονότητα των μετα-αναλύσεων γενετικών μελετών συσχετισμού πραγματοποιείται χρησιμοποιώντας μονομεταβλητές τεχνικές (Attia et al, 2003; Ioannidis & Trikalinos, 2005; Ioannidis et al, 2003). Είναι ευρέως γνωστό ότι τα

δεδομένα που προκύπτουν από γενετικές μελέτες σε μορφή ORs μέσα από συγκρίσεις γονοτύπου είναι συσχετισμένα με την κληρονομικότητα, ενώ δεν έχει παρουσιαστεί λεπτομερής προσαρμογή της πολυμεταβλητής μεθοδολογίας στο παρελθόν. Σε αυτό το έργο, το γενικό μοντέλο διμεταβλητής μέτα-ανάλυσης των τυχαίων επιδράσεων τροποποιήθηκε ώστε να εφαρμοστεί στις μελέτες γενετικού συσχετισμού.

Παρουσιάζονται μέθοδοι βασισμένες σε συνοπτικά δεδομένα, καθώς και μέθοδοι βασισμένες στην δυαδική φύση των δεδομένων. Σχετικά με τα συγκεντρωτικά δεδομένα, το μοντέλο καθορίστηκε με την ίδια διατύπωση τόσο για τα διακριτά όσο και για τα συνεχή αποτελέσματα και προκύπτουν αναλυτικές εκφράσεις για την συνδιακύμανση των δύο αποτελεσμάτων που μοντελοποιούνται μαζί. Οι μέθοδοι που παρουσιάστηκαν εδώ, βασισμένες σε συνοπτικά δεδομένα, μοιάζουν πάρα πολύ με τις γενικές πολυμεταβλητές μεθόδους που προτείνονται από τους Berkey και συνεργάτες (Berkey et al, 1998) και τους van Houwelingen και συνεργάτες (van Houwelingen et al, 2002). Η κύρια διαφορά είναι ότι σε αυτές τις μεθόδους, ο συντελεστής συσχέτισης εντός των μελετών θα έπρεπε να δίνεται από τα αρχικά άρθρα, ενώ στις μελέτες γενετικού συσχετισμού μπορεί να προκύψει από τις δημοσιευμένες μετρήσεις γονοτύπων.

Χρησιμοποιώντας τη δυαδική φύση των δεδομένων, προτάθηκαν μέθοδοι βασισμένες τόσο στην προοπτική όσο και στην αναδρομική πιθανοφάνεια. Αυτά τα μοντέλα αποτελούν επεκτάσεις της γενικής μεθοδολογίας της πολυμεταβλητής μέτα-ανάλυσης που προτάθηκε από τους van Houwelingen και συνεργάτες (van Houwelingen et al, 1993). Ωστόσο, με την παρουσιαζόμενη προσέγγιση, τα δύο στοιχεία της διμεταβλητής ανταπόκρισης υπολογίζονται σε σύγκριση με την ίδια ομάδα αναφοράς. Επίσης, οι μέθοδοι μπορούν να θεωρηθούν επεκτάσεις των μονομεταβλητών πολυεπίπεδων μεθόδων μέτα-ανάλυσης με τυχαίες επιδράσεις (Higgins et al, 2001; Thompson & Sharp, 1999; Turner et al, 2000). Αναμένεται ότι αυτά τα μοντέλα θα λειτουργήσουν καλύτερα σε σύγκριση με τα μοντέλα βασισμένα σε συνοπτικά δεδομένα όταν η υπόθεση της κανονικότητας για τα logORs δεν είναι έγκυρη. Επιπλέον, ένα μεγάλο πλεονέκτημα αυτών των μοντέλων είναι ότι μπορούν να χρησιμοποιηθούν άμεσα για συγκεντρωτικές (pooled) μετα-αναλύσεις. Αυτός είναι ο λόγος για τον οποίο αυτά τα μοντέλα συνήθως αποκαλούνται μέθοδοι Individual Patients Data (IPD) (Turner et al, 2000). Από την άλλη πλευρά, ένα μειονέκτημα είναι ότι αυτά τα μοντέλα είναι σημαντικά πιο αργά από τα μοντέλα βασισμένα σε συνοπτικά δεδομένα.

Σε όλα τα μοντέλα που παρουσιάστηκαν, αναπτύχθηκαν επίσημες δοκιμές για τον έλεγχο του γενετικού μοντέλου κληρονομικότητας βασισμένες στην θεωρία της κανονικής κατανομής. Το γενικό μοντέλο συγκρίθηκε με μια πρόσφατα προτεινόμενη γενετική model-free διμεταβλητή προσέγγιση, και διαπιστώθηκε ρητά ότι οι εκτιμήσεις που παρέχονται από αυτές τις δύο προσεγγίσεις είναι παρόμοιες. Υποστηρίζεται ότι οι μικρές διαφορές μεταξύ των εκτιμήσεων που προκύπτουν από τα γενικά μοντέλα και

από τις γενετικές model-free προσεγγίσεις δεν οφείλονται μόνο στη διαφορετική παραμετροποίηση της γενετικής model-free προσέγγισης, αλλά και στις επιπλέον υποθέσεις που γίνονται. Οι επιπλέον υποθέσεις της model-free προσέγγισης γίνονται, καθώς προτείνει μόνο μια διακύμανση τ^2 μεταξύ μελετών (μεταξύ των μελετών μεταβλητότητα), υπονοώντας με αυτόν τον τρόπο ότι τα β_s μοιράζονται μια κοινή τυχαία συνιστώσα διακύμανσης, ενώ το λ θεωρείται ως παράμετρος σταθερών επιδράσεων. Οι Minelli και συνεργάτες στην αντίστοιχη δημοσίευσή τους, εξέτασαν επίσης τη γενική περίπτωση του διμεταβλητού μοντέλου, αλλά κατά την υλοποίησή τους κατάλαβαν ότι σε περιπτώσεις μικρού αριθμού μελετών, η συσχέτιση μεταξύ μελετών δεν εκτιμάται σωστά και συνεπώς χρησιμοποίησαν μια σταθερή τιμή, αυτή του 0,9. Αυτές οι υποθέσεις από τους Minelli και συνεργάτες μπορεί να ευθύνονται για τη μικρή διαφορά στις εκτιμήσεις που προκύπτουν - σε ορισμένες περιπτώσεις - από τα δύο μοντέλα.

Ο μόνος περιορισμός των προτεινόμενων μεθόδων είναι ότι οι μέθοδοι που βασίζονται στη δυαδική φύση των δεδομένων είναι πολύ χρονοβόρες σε μια τέτοια εφαρμογή. Ωστόσο, οι μέθοδοι συνοπτικών δεδομένων εμφανίζουν αποτέλεσμα σε λιγότερο από ένα λεπτό σε έναν κανονικό υπολογιστή προσφέροντας μια ελκυστική προσέγγιση για την ανάλυση τέτοιων δεδομένων. Τέλος, αναμένεται ότι οι μέθοδοι που αναπτύχθηκαν σε αυτήν την εργασία θα εφαρμοστούν ευρέως στη μέτα-ανάλυση γενετικών μελετών συσχέτισης, καθώς είναι εύκολα υλοποιήσιμες σε όλα τα βασικά πακέτα στατιστικής, αποφεύγοντας την ανάγκη για εξατομικευμένο λογισμικό. Οι μέθοδοι εφαρμόστηκαν σε αρκετές ήδη δημοσιευμένες μέτα-αναλύσεις γενετικών μελετών συσχέτισης (και με διακριτά και με συνεχή αποτελέσματα) και τα αποτελέσματα συγκρίθηκαν με τις ευρέως χρησιμοποιούμενες μονομεταβλητές προσεγγίσεις, καθώς και με τις γενετικές model-free προσεγγίσεις.

Βιβλιογραφικές Αναφορές

- Attia, J., Thakkestian, A., & D'Este, C. (2003). Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *J Clin Epidemiol*, 56(4), 297-303.
- Bagos, P.G. (2008). A unification of multivariate methods for meta-analysis of genetic association studies. *Statistical Applications in Genetics and Molecular Biology*, 7(1), 31.
- Bagos, P.G. (2015). Meta-analysis in Stata using gllamm. *Research Synthesis Methods*, 6(4), 310-32.
- Bagos, P.G., & Nikolopoulos, G.K. (2007). A method for meta-analysis of case-control genetic association studies using logistic regression. *Stat Appl Genet Mol Biol*, 6, 17.
- Berkey, C.S., Hoaglin, D.C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G.A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Stat Med*, 17(22), 2537-2550.
- Chen, H.Y. (2003). A note on the prospective analysis of outcome-dependent samples. *J Roy Soc, B* 65(2), 575-584.
- Boekholdt, S.M., Sacks, F.M., Jukema, J.W., Shepherd, J., Freeman, D.J., McMahon, A.D., Cambien, F., Nicaud, V., de Grooth, G.J., Talmud, P.J., Humphries, S.E., Miller, G.J., Eiriksdottir, G., Gudnason, V., Kauma, H., Kakko, S., Savolainen, M.J., Arca, M., Montali, A., Liu, S., Lanz, H.J., Zwinderman, A.H., Kuivenhoven, & J.A., Kastelein, J.J. (2005). Cholesteryl ester transfer protein TaqIB variant, high-density lipoprotein cholesterol levels, cardiovascular risk, and efficacy of pravastatin treatment: individual patient meta-analysis of 13,677 subjects. *Circulation*, 111(3), 278-287.
- Evangelou, E., Trikalinos, T.A., Salanti, G., & Ioannidis, J.P. (2006). Family-based versus unrelated case-control designs for genetic associations. *PLoS Genet*, 2(8), 123.
- Hani, E.H., Boutin, P., Durand, E., Inoue, H., Permutt, M.A., Velho, G., & Froguel, P. (1998). Missense mutations in the pancreatic islet beta cell inwardly rectifying K⁺ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. *Diabetologia*, 41(12), 1511-1515.
- Higgins, J.P., Whitehead, A., Turner, R.M., Omar, R.Z., & Thompson, S.G. (2001). Metaanalysis of continuous outcome data from individual patients. *Stat Med*, 20(15), 2219-2241.

- Ioannidis, J.P., & Trikalinos, T.A. (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), 543-549.
- Ioannidis, J.P., Trikalinos, T.A., Ntzani, E.E., & Contopoulos-Ioannidis, D.G. (2003). Genetic associations in large versus small studies: an empirical assessment. *Lancet*, 361(9357), 567-571.
- Kazeem, G.R., & Farrall, M. (2005). Integrating case-control and TDT studies. *Ann Hum Genet*, 69(Pt 3), 329-335.
- Kato, N., Sugiyama, T., Morita, H., Kurihara, H., Yamori, Y., & Yazaki, Y. (1999). Angiotensinogen gene and essential hypertension in the Japanese: extensive association study and meta-analysis on six reported studies. *J Hypertens*, 17(6), 757-763.
- McCullagh, P., & Nelder, J.A. (1989). *Generalized Linear Models*, London, Chapman & Hall.
- Minelli, C., Thompson, J.R., Abrams, K.R., & Lambert, P.C. (2005). Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies. *Stat Med*, 24(24), 3845-3861.
- Minelli, C., Thompson, J.R., Abrams, K.R., Thakkeinstian, A., & Attia, J. (2005). The choice of a genetic model in the meta-analysis of molecular association studies. *Int J Epidemiol*, 34(6), 1319-1328.
- Minelli, C., Thompson, J.R., Tobin, M.D., & Abrams, K.R. (2004). An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *Am J Epidemiol*, 160(5), 445-452.
- Thakkeinstian, A., McElduff, P., D'Este, C., Duffy, D., & Attia, J. (2005). A method for meta-analysis of molecular association studies. *Stat Med*, 24(9), 1291-130.
- Thompson, S.G., & Sharp, S.J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med*, 18(20), 2693-2708.
- Turner, R.M., Omar, R.Z., Yang, M., Goldstein, H., & Thompson, S.G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Stat Med*, 19(24), 3417-3432.
- Sasieni, P.D. (1997). From genotypes to genes: doubling the sample size. *Biometrics*, 53(4), 1253-1261.
- Sayed-Tabatabaei, F.A., Houwing-Duistermaat, J.J., van Duijn, C.M., & Witteman, J.C. (2003). Angiotensin-converting enzyme gene polymorphism and carotid artery wall thickness: a meta-analysis. *Stroke*, 34(7), 1634-1639.

- Salanti, G., Higgins, J.P., Trikalinos, T.A., & Ioannidis, J.P. (2006). Bayesian meta-analysis and meta-regression for gene-disease associations and deviations from HardyWeinberg equilibrium. *Stat Med*, 26(3), 553-67.
- Skrondal, A., & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, 68(2), 267-287.
- van Houwelingen, H.C., Arends, L.R., & Stijnen, T. (2002). Advanced methods in metaanalysis: multivariate approach and meta-regression. *Stat Med*, 21(4), 589-62.
- van Houwelingen, H.C., Zwinderman, K.H., & Stijnen, T. (1993). A bivariate approach to meta-analysis. *Stat Med*, 12(24), 2273-2284.
- Wheeler, J.G., Keavney, B.D., Watkins, H., Collins, R., & Danesh, J. (2004). Four paraoxonase gene polymorphisms in 11212 cases of coronary heart disease and 12786 controls: meta-analysis of 43 studies. *Lancet*, 363(9410), 689-695.