ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Λογισμικό μέτα-ανάλυσης στο STATA

Σιγάλα Αθηνά

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Υπεύθυνος
Παντελεήμων Γ. Μπάγκος
Καθηγητής

Λαμία, 2024

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Λογισμικό μέτα-ανάλυσης στο STATA

Σιγάλα Αθηνά

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων
Παντελεήμων Γ. Μπάγκος
Καθηγητής

Λαμία, 2024

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις [(1)], που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

*1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάσθηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.*

*2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.*

3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια

4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία: 12/06/2024

Η Δηλούσα

Σιγάλα Αθηνά

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Λογισμικό μέτα-ανάλυσης στο STATA

Σιγάλα Αθηνά

Τριμελής Επιτροπή:

Παντελεήμων Γ. Μπάγκος, Καθηγητής

Γεωργία Μπράλιου, Επίκουρη Καθηγήτρια

Αθανάσιος Σαχλάς, Επίκουρος Καθηγητής

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Αυτή η εργασία είναι αφιερωμένη στην οικογένεια και στους φίλους μου

Η αμέριστη στήριξη και αγάπη τους με ενέπνευσαν και με ενθάρρυναν σε κάθε βήμα αυτής της διαδρομής

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κύριο Παντελή Μπάγκο, Καθηγητή της σχολής Θετικών Επιστημών και εσωτερικό μέλος του Συμβουλίου Διοίκησης του Πανεπιστημίου Θεσσαλίας, για την εμπιστοσύνη που μου έδειξε κατά την ανάθεση του συγκεκριμένου θέματος πτυχιακής εργασίας, την καθοδήγησή του και την πολύτιμη βοήθεια του.

Επιπλέον, θα ήθελα να ευχαριστήσω την Καλλιόπη Εξάρχου-Κουβέλη, υποψήφια διδάκτορα του τμήματος Πληροφορικής με Εφαρμογές στη Βιοϊατρική και Ακαδημαϊκή υπότροφο του τμήματος Μαθηματικών του Πανεπιστημίου Θεσσαλίας, για την προθυμία της να βοηθήσει και την καθοδήγησή της ώστε να ολοκληρωθεί η παρούσα εργασία.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Περιεχόμενα

Ευχαριστίες iv
Περιεχόμενα v
Λίστα με γραφήματα vii
Λίστα με πίνακες viii
Περίληψη ix
Κεφάλαιο 1 1
Εισαγωγή 1
1.1 Σχέσεις γονότυπου-φαινότυπου και γονότυπου-νόσου 2
1.1.1 Γονότυπος – φαινότυπος 2
1.1.2 Γονότυπος-φαινότυπος-νόσος 3
1.2 Μεντελιανή τυχαιοποίηση (MR) 5
1.2.1 Εφαρμογή 5
1.2.2 Βασικές αρχές 6
1.2.3 Παραδείγματα 7
1.3 Επέκταση της MR 8
1.3.1 Χρήση πολλαπλών παραλλαγών 8
1.3.2 MR δύο δειγμάτων 9
1.3.3 MR δύο βημάτων και δύο δειγμάτων 10
Κεφάλαιο 2 12
Μέθοδοι 12
2.1 Μοντέλα για πολυμεταβλητή ανάλυση 14
2.2 Συσχετίσεις εντός-μελέτης και μεταξύ-μελετών 15
2.3 Μοντέλα ML / REML 16
2.3.1 Μοντέλα A ML/REML 16
2.3.2 Μοντέλα B ML/REML 18
Κεφάλαιο 3 21
Αποτελέσματα 21
3.1 Πρόγραμμα 21
3.1.1 Αρχείο mrmeta 21
3.1.2 Αρχείο LL 29

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

3.1.3 Περιγραφή 33
3.2 Εκτέλεση 35
3.2.1 Παράδειγμα 1: Μοντέλο A - Μέθοδος REML 35
3.2.2 Παράδειγμα 2: Μοντέλο B- Μέθοδος ML 42
3.2.3 Παράδειγμα 3: Μοντέλο B- Μέθοδος REML 47
Κεφάλαιο 4 52
Συζήτηση 52
Βιβλιογραφικές αναφορές 54

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Λίστα με γραφήματα

1. Γράφημα 1. Σχηματική αναπαράσταση τριών υποθέσεων για την MR (Zou κ.α., 2022, σελ. 2)....2
2. Γράφημα 2. Ποικιλομορφία φαινοτύπων (Garvan Institute of Medical Research website)....3
3. Γράφημα 3. Υπολογισμός μίας εκτίμησης, χωρίς σύγχυση, της επίδρασης μίας αλλαγής του φαινότυπου σε μια ασθένεια με βάση την έννοια της Μεντελιανής τυχαιοποίησης (Minelli κ.α., 2004, σελ. 446)....5
4. Γράφημα 4. Σχηματική αναπαράσταση της MR (Smith κ.α., 2014, σελ. R90)....6
5. Γράφημα 5. MR δύο βημάτων (Smith κ.α., 2022, σελ. R94)....10
6. Γράφημα 6. MR δύο δειγμάτων (Smith κ.α., 2022, σελ. R94)....11
7. Γράφημα 7. Υπολογισμός μίας εκτίμησης χωρίς σύγχυση, της επίδρασης μίας αύξησης της ομοκυστεΐνης (Hcy) κατά 1 μmol/l στον κίνδυνο CHD, με βάση την τεχνική της Μεντελιανής τυχαιοποίησης (Minelli κ.α., 2004, σελ. 446)....13
8. Γράφημα 8. Διπλό δενδροδιάγραμμα με συγκεντρωτικές εκτιμήσεις για τις σχέσεις γονότυπου-νόσου και γονότυπου-φαινότυπου (Minelli κ.α., 2004, σελ. 448)....14
9. Γράφημα 9. (α) Μοντελοποίηση των ετερογενειών των συσχετισμών στο Μοντέλο Α. (β) Μοντελοποίηση των ετερογενειών των συσχετισμών στο Μοντέλο Β....18
10. Γράφημα 10. Διπλό δενδροδιάγραμμα μελετών που εξετάζει τη σχέση γονότυπου-φαινότυπου, ή γονότυπου-νόσου ή και των δύο παράλληλα με το πρόγραμμα mrmeta και χρήση του μοντέλου A REML....38
11. Γράφημα 11. Γράφημα φυσαλίδων του λογαρίθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta με χρήση του μοντέλου A REML....40
12. Γράφημα 12. Γράφημα διασποράς του λογαρίθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta με χρήση του μοντέλου A REML....40
13. Γράφημα 13. Διπλό δενδροδιάγραμμα του προγράμματος mrmeta και χρήση του μοντέλου B ML....44
14. Γράφημα 14. Γράφημα φυσαλίδων του λογαρίθμου του λόγου σχετικών πιθανοτήτων και της μέσης διαφοράς, του προγράμματος mrmeta και χρήση του μοντέλου B ML....45
15. Γράφημα 15. Γράφημα διασποράς του λογαρίθμου του λόγου σχετικών πιθανοτήτων και της μέσης διαφοράς, του προγράμματος mrmeta με χρήση του μοντέλου B ML....45
16. Γράφημα 16. Διπλό δενδροδιάγραμμα μελετών με το πρόγραμμα mrmeta και χρήση του μοντέλου B REML....49
17. Γράφημα 17. Γράφημα φυσαλίδων του λογάριθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta και χρήση του μοντέλου B REML....50
18. Γράφημα 18. Γράφημα διασποράς του λογάριθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta και χρήση του μοντέλου B REML....50

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Λίστα με πίνακες

1. Πίνακας 1. Παραδείγματα επίδρασης χαρακτηριστικού σε νόσο (Smith κ.α., 2014, σελ. R91)....8
2. Πίνακας 2. Αποτελέσματα των μοντέλων A και B, με χρήση της προσέγγισης Maximum Likelihood....17
3. Πίνακας 3. Αρχικά δεδομένα προγράμματος mrmeta....34
4. Πίνακας 4. Αρχικές βέλτιστες τιμές εκτίμησης του προγράμματος mrmeta με χρήση του μοντέλου A REML....36
5. Πίνακας 5. Μεγιστοποιημένα αποτελέσματα του προγράμματος mrmeta με χρήση του μοντέλου A REML....36
6. Πίνακας 6. Τελικό συγκεντρωτικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων του προγράμματος mrmeta με χρήση του μοντέλου A REML...37
7. Πίνακας 7. Αποτελέσματα του λογαρίθμου του λόγου σχετικών πιθανοτήτων και της μέσης διαφοράς του προγράμματος mrmeta με χρήση του μοντέλου A REML και της εντολής mvmeta.....37
8. Πίνακας 8. Τελικό αποτέλεσμα της αναλογίας του λογαρίθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά από το πρόγραμμα mrmeta με χρήση του μοντέλου A και της mvmeta με τη μέθοδο REML....38
9. Πίνακας 9. Συγκεντρωτικά αποτελέσματα επίδρασης MTHFR στη CHD από το πρόγραμμα mrmeta....41
10. Πίνακας 10. Συγκεντρωτικά αποτελέσματα επίδρασης MTHFR στα επίπεδα ομοκυστεΐνης από το πρόγραμμα mrmeta....41
11. Πίνακας 11. Συγκεντρωτικά αποτελέσματα λόγου logOR προς mean difference από το πρόγραμμα mrmeta....41
12. Πίνακας 12. Δεδομένα παραδείγματος 2 που περιλαμβάνουν τα επίπεδα ομοκυστεΐνης - IHD....42
13. Πίνακας 13. Μεγιστοποιημένα αποτελέσματα του προγράμματος mrmeta με χρήση του μοντέλου B ML....43
14. Πίνακας 14. Τελικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων με το πρόγραμμα mrmeta με χρήση του μοντέλου B και της μεθόδου ML....43
15. Πίνακας 15. Συγκεντρωτικά αποτελέσματα του λογαρίθμου του λόγου σχετικών πιθανοτήτων με το πρόγραμμα mrmeta....46
16. Πίνακας 16. Συγκεντρωτικά αποτελέσματα της μέσης διαφοράς με το πρόγραμμα mrmeta....46
17. Πίνακας 17. Συγκεντρωτικά αποτελέσματα λόγου logOR προς mean difference από το πρόγραμμα mrmeta....46
18. Πίνακας 18. Δεδομένα παραδείγματος 3 που περιλαμβάνουν επίπεδα ομοκυστεΐνης - εγκεφαλικό επεισόδιο....47
19. Πίνακας 19. Μεγιστοποιημένα αποτελέσματα του προγράμματος mrmeta με χρήση του μοντέλου B REML....48
20. Πίνακας 20. Συγκεντρωτικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων με το πρόγραμμα mrmeta και χρήση του μοντέλου B και της μεθόδου REML....48
21. Πίνακας 21. Συγκεντρωτικά αποτελέσματα λόγου σχετικής πιθανότητας από το πρόγραμμα mrmeta....51
22. Πίνακας 22. Συγκεντρωτικά αποτελέσματα μέσης διαφοράς από το πρόγραμμα mrmeta....51
23. Πίνακας 23. Συγκεντρωτικά αποτελέσματα λόγου logOR προς mean difference από το πρόγραμμα mrmeta....51

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Περίληψη

Οι μελέτες γενετικής συσχέτισης εξετάζουν τη σχέση μεταξύ ασθενειών και γενετικών πολυμορφισμών, προσφέροντας εργαλεία για τον εντοπισμό γονιδίων ευαισθησίας. Ωστόσο, παρά την αύξηση του αριθμού των μελετών, τα αποτελέσματα δεν είναι πάντα σταθερά αναπαραγώγιμα, υποδηλώνοντας πιθανές ανακρίβειες. Η μετα-ανάλυση, συνδυάζοντας αποτελέσματα από διάφορες μελέτες, προσφέρει αυξημένη στατιστική ισχύ και ακρίβεια. Η σχέση μεταξύ γονότυπου και φαινότυπου είναι σύνθετη και όχι πάντα ξεκάθαρη, με την ατελή διεισδυτικότητα και τη μεταβλητή εκφραστικότητα να συμβάλλουν στην ποικιλομορφία των φαινοτύπων σε ασθενείς με γενετικές παθήσεις. Η σημασία της Μεντελιανής τυχαιοποίησης (MR), για την κατανόηση της σχέσης γονότυπου-φαινότυπου-νόσου, υπογραμμίζεται από το μεγάλο αριθμό γενετικών μελετών που εξετάζουν τη συσχέτιση αυτή και από μετα-αναλύσεις που αναδεικνύουν ομοιότητες. Η Μεντελιανή τυχαιοποίηση (MR) στην επιδημιολογία χρησιμοποιεί γνωστή (που έχει μετρηθεί) παραλλαγή στα γονίδια για να διερευνήσει αιτιολογικές συνδέσεις μεταξύ έκθεσης και νόσου, εξαλείφοντας τη σύγχυση και ενισχύοντας την ερμηνεία των επιδημιολογικών μελετών. Η κατάλληλη επιλογή γενετικών μελετών για την αξιολόγηση της σχέσης μεταξύ φαινότυπου-ασθένειας, απαιτεί τον συνδυασμό της συσχέτισης γονότυπου-ασθένειας με αυτήν του γονότυπου-φαινότυπου, χρησιμοποιώντας θεωρητικά μοντέλα, όπως η MR, ώστε να προσεγγίσει τα αίτια της νόσου με βάση τις γενετικές παραλλαγές. Η απλή ανάλυση παλινδρόμησης συναντά εμπόδια στην αποσαφήνιση της αιτιώδους κατεύθυνσης μεταξύ συσχετιζόμενων μεταβλητών, ενώ η MR από την άλλη πλευρά, δύναται να αντιμετωπίσει τη σύγχυση στην αξιολόγηση των αιτιακών συνδέσεων μεταξύ γενετικών παραλλαγών και φαινοτύπων. Μία συσχέτιση γονότυπου-φαινότυπου-νόσου που έχει ερευνηθεί ,είναι αυτή του γονιδίου της μεθυλενοτετραϋδροφολικής αναγωγάσης (MTHFR), με τα επίπεδα ομοκυστεΐνης και τη στεφανιαία νόσο (CHD). Δεδομένα από μία τέτοια μετα-ανάλυση των Minelli et all. (2004) έχουν χρησιμοποιηθεί στην παρούσα εργασία, προκειμένου να αναδειχθεί η ύπαρξη συσχέτισης φαινότυπου-νόσου, υλοποιώντας δύο μεθόδους της MR στο περιβάλλον του Stata. Πιο συγκεκριμένα, ως δεδομένα στον κώδικα εισάγονται ο λόγος σχετικών πιθανοτήτων του γονότυπου στη νόσο, η μέση διαφορα στο φαινότυπο και οι διακυμάνσεις τους. Στη συνέχεια, στο πρώτο και βασικό αρχείο συγκεντρώνονται τα μοντέλα A και B μέγιστης πιθανοφάνειας (Maximum Likelihood) και περιορισμένης μέγιστης πιθανοφάνειας (REstricted Maximum Likelihood). Τα αποτελέσματα που προκύπτουν, είναι οι συγκεντρωτικές τιμές της μέσης διαφοράς στο φαινότυπο και του λόγου σχετικών πιθανοτήτων του γονότυπου στη νόσο. Υπολογίζονται επίσης και εμφανίζονται η τελική τυπική απόκλιση και τα διαστήματα εμπιστοσύνης των δύο προηγούμενων αποτελεσμάτων. Σαν επιπλέον επιλογή παρέχεται η εφαρμογή της εντολής mvmeta, για τον υπολογισμό των ίδιων προαναφερόμενων αποτελεσμάτων με διαφορετικό τρόπο. Τέλος, εμφανίζονται δύο γραφήματα στην οθόνη του χρήστη. Το πρώτο είναι ένα διπλό δενδροδιάγραμμα, στο οποίο εμφανίζονται όλες οι μελέτες από τα δεδομένα, κατανεμημένες ανάλογα με τα τελικά διαστήματα εμπιστοσύνης, αλλά και το συγκεντρωτικό αποτέλεσμά τους. Το δεύτερο γράφημα μπορεί να είναι είτε ένα γράφημα διασποράς είτε ένα γράφημα φυσαλίδων με τις αρχικές τιμές των δεδομένων να εμφανίζονται ανάλογα με το βάρος τους.

Λέξεις - κλειδιά: Μεντελιανή τυχαιοποίηση, σχέσεις γονότυπου-φαινότυπου-νόσου, εντός των μελετών συσχέτιση, μεταξύ των μελετών συσχέτιση, μοντέλο μέγιστης πιθανοφάνειας, μοντέλα περιορισμένης μέγιστης πιθανοφάνειας, Stata

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Κεφάλαιο 1

Εισαγωγή

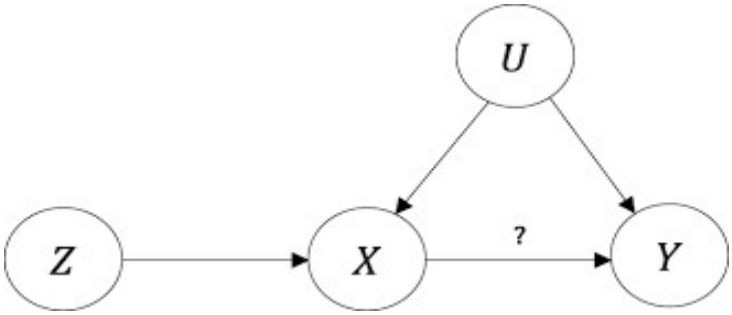
Οι μελέτες γενετικής συσχέτισης γενικά εξετάζουν τη σχέση μεταξύ μιας ασθένειας και γενετικών παραλλαγών ευαισθησίας. Πρόκειται για μια ισχυρή προσέγγιση που επιτρέπει τον εντοπισμό γονιδίων ευαισθησίας για κοινές ασθένειες και παράλληλα παρέχει αναλύσεις συσχέτισης για τη διερεύνηση πολύπλοκων ασθενειών (Tam κ.α., 2019, σελ. 467. Μολονότι διεξάγονται ολοένα και περισσότερες έρευνες, τα ευρήματά τους δεν μπορούν πάντοτε να επαναληφθούν με ακρίβεια. Το γεγονός αυτό υποδηλώνει ότι τα αποτελέσματα μπορεί να ποικίλλουν από μελέτη σε μελέτη, επομένως δεν υπάρχει πάντα η αναμενόμενη συνέπεια και αξιοπιστία στα συμπεράσματα. Μελέτες γονιδίων υπό διερεύνηση έχουν αποδείξει ότι πολλές αρχικές θετικές συσχετίσεις δε μπορούν να αναπαραχθούν σε επόμενες μελέτες. Αυτά τα ευρήματα υποδηλώνουν ψευδώς θετικές αναφορές (σφάλματα τύπου I) ή ψευδώς αρνητικές (σφάλματα τύπου II), δηλαδή μικρές γενετικές επιδράσεις είναι μη ανιχνεύσιμες στην πλειονότητα αυτών των μελετών (Γαλάνης, 2009, σελ. 700).

Η μετα-ανάλυση είναι ένα στατιστικό εργαλείο που αυξάνει τη στατιστική ισχύ και την ακρίβεια των αποτελεσμάτων, μέσω του συνδυασμού τους από διάφορες μελέτες, οι οποίες εξετάζουν τα ίδια χαρακτηριστικά. Επιπροσθέτως, αναλύει τις αποκλίσεις και προσδιορίζει υποομάδες που σχετίζονται με τον παράγοντα ενδιαφέροντος. Η σωστή εφαρμογή της θεωρείται σημαντική απόδειξη και αποτελεί αποτελεσματικό εργαλείο για την κατανόηση πολύπλοκων ασθενειών (Minelli κ.α., 2004, σελ. 446).

Ο πρωταρχικός στόχος της επιδημιολογικής παρατήρησης είναι να εντοπίσει τις βαθύτερες αιτίες της νόσου. Επίκεντρο πολλών τέτοιων αναλύσεων είναι η εξέταση της μεταβολής του βαθμού σοβαρότητας ή της πιθανότητας εμφάνισης μιας ασθένειας σε περίπτωση έκθεσης σε έναν συγκεκριμένο παράγοντα κινδύνου. Τα αιτιώδη συμπεράσματα, όμως, σπάνια δικαιολογούνται με τη χρήση μιας παραδοσιακής ανάλυσης (Belbasis κ.α., 2018, σελ. 1). Ακόμη και όταν μετρώνται ισχυρές στατιστικές συσχετίσεις μεταξύ μιας έκθεσης και του αποτελέσματος, δεν είναι ποτέ βέβαιο ότι όλοι οι συγχυτικοί παράγοντες της συσχέτισης έχουν εντοπιστεί, μετρηθεί και προσαρμοστεί κατάλληλα (Zou κ.α., 2022, σελ. 1). Λαμβάνοντας λοιπόν υπόψιν όλους τους παραπάνω περιορισμούς, στη συγκεκριμένη εργασία εφαρμόζεται η μεθοδολογία της MR, η οποία προσφέρει έναν εναλλακτικό τρόπο διερεύνησης της αιτιότητας στην επιδημιολογική έρευνα, χρησιμοποιώντας πρόσθετες γενετικές παραλλαγές που υποτίθεται ότι ικανοποιούν τις υποθέσεις βοηθητικών μεταβλητών (Instrumental Variables) (Zou κ.α., 2022, σελ. 1).

Εν συνεχεία, για την εξήγηση της λογικής της μελέτης αλλά και για την αποσαφήνιση των IV στα οποία βασίζεται η εγκυρότητά της μελέτης αυτής, χρήσιμο εργαλείο αποτελούν τα κατευθυνόμενα μη κυκλικά γραφήματα (DAGs) (Zou κ.α., 2022, σελ. 1). Παρακάτω παρατίθεται ένα γράφημα DAG, που συσχετίζει την απλούστερη μεμονωμένη μονάδα γενετικής παραλλαγής - έναν πολυμορφισμό μεμονωμένου νουκλεοτιδίου (SNP) Z - με μια έκθεση X, και το αποτέλεσμα Y παρουσία μη μετρημένης σύγχυσης, που αντιπροσωπεύεται από το U (Zou κ.α., 2022, σελ. 1).

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 1. Σχηματική αναπαράσταση τριών υποθέσεων για την MR (Zou κ.α., 2022, σελ. 2).

1.1 Σχέσεις γονότυπου-φαινότυπου και γονότυπου-νόσου

1.1.1 Γονότυπος – φαινότυπος

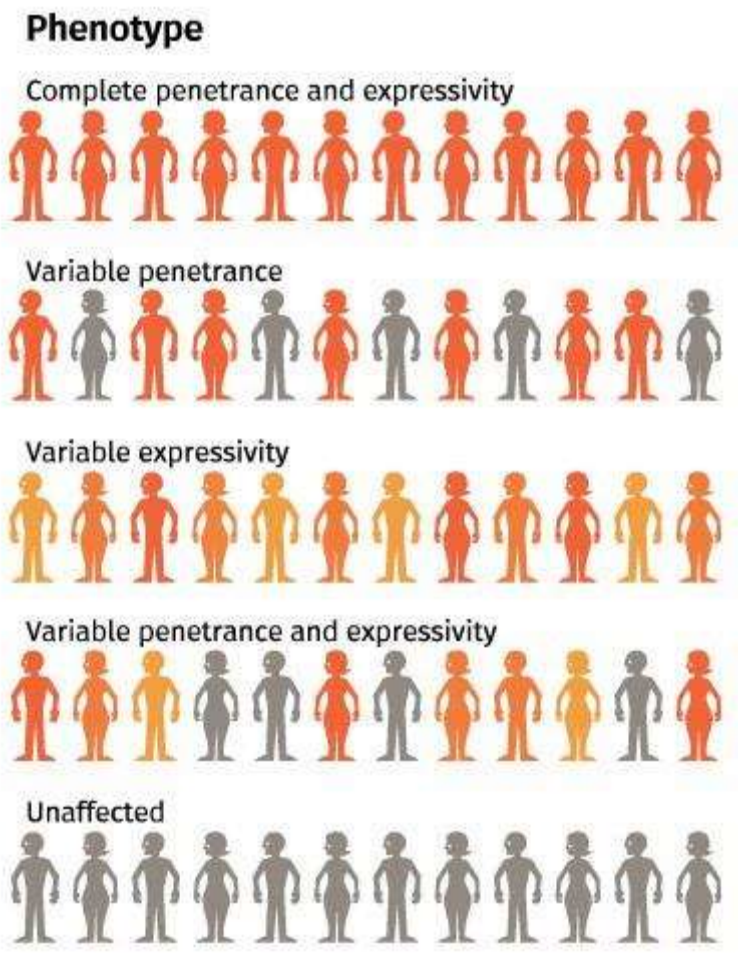
Ο γονότυπος ενός ατόμου είναι η μοναδική αλληλουχία του DNA του. Συγκεκριμένα, ο όρος χρησιμοποιείται για την περιγραφή των δύο αλληλομόρφων που έχει κληρονομήσει για ένα συγκεκριμένο γονίδιο (Elston κ.α., 2012). Ο φαινότυπος είναι η ανιχνεύσιμη έκφραση αυτού του γονότυπου, δηλαδή η κλινική εικόνα ενός ασθενούς. Ο φαινότυπος ενός ατόμου προκύπτει από την αλληλεπίδραση μεταξύ του γονότυπου και του περιβάλλοντός του (Elston κ.α., 2012).

Η σύνδεση μεταξύ γονότυπου και φαινότυπου δεν είναι πάντα ξεκάθαρη. Αυτό σημαίνει ότι μπορεί να υπάρχουν διαφορές μεταξύ του αποτελέσματος της γονιδιωματικής εξέτασης του ασθενούς και της κλινικής του εμφάνισης. Αντίστοιχα, ο φαινότυπος των ασθενών με μια δεδομένη γενετική πάθηση μπορεί να ποικίλλει ακόμη και εντός της οικογένειας (Kingdom κ.α., 2022). Λόγοι για την εμφάνιση αυτού του φαινομένου αποτελούν:

- Η ατελής διεισδυτικότητα: Η διεισδυτικότητα ως έννοια αναφέρεται στο ποσοστό των ατόμων που φέρουν μια συγκεκριμένη γενετική παραλλαγή και εκφράζουν το σχετικό φαινότυπο. Αναφορικά με την ατελή διεισδυτικότητα δεν είναι απαραίτητο κάθε άτομο που έχει την αιτιολογική γενετική παραλλαγή να αναπτύξει συμπτώματα της πάθησης κατά τη διάρκεια της ζωής του. Από την άλλη, ορισμένες καταστάσεις σχετίζονται με 100% διεισδυτικότητα (πλήρης), γεγονός που υποδηλώνει ότι τα συμπτώματα αυτά θα εμφανιστούν υποχρεωτικά (Kingdom κ.α., 2022).

- Η μεταβλητή εκφραστικότητα: Ο όρος εκφραστικότητα αναφέρεται στην ποικιλία των διαφορετικών φαινοτύπων που παρουσιάζονται σε ασθενείς με την ίδια γενετική πάθηση. Η μεταβλητότητά της εξηγείται από το βαθμό εμφάνισης των συμπτωμάτων. Πιο συγκεκριμένα, ένας ασθενής με μια παραλλαγή μπορεί να εμφανίζει ήπια συμπτώματα, ενώ ένας άλλος ασθενής με την ίδια παραλλαγή μπορεί να νοσεί σοβαρά. Όπως και με την ατελή διεισδυτικότητα, οι λόγοι για τη μεταβλητή εκφραστικότητα δεν είναι απόλυτα κατανοητοί. Εάν μια γενετική πάθηση έχει μεταβλητά συμπτώματα, μπορεί να είναι πιο δύσκολη η διάγνωσή της. Τέλος, η μεταβλητή εκφραστικότητα είναι κοινή τόσο σε κυρίαρχους όσο και σε υπολειπόμενους φαινότυπους (Kingdom κ.α., 2022).

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 2. Ποικιλομορφία φαινοτύπων (Garvan Institute of Medical Research website).

1.1.2 Γονότυπος-φαινότυπος-νόσος

Μέσω της πληθώρας γνώσεων που υπάρχει τα τελευταία χρόνια για το ανθρώπινο γονιδίωμα, σημειώνεται και δραματική αύξηση στον αριθμό των γενετικών επιδημιολογικών μελετών που εξετάζουν τη συσχέτιση μεταξύ συγκεκριμένων γονιδίων και ασθενειών, καθώς και μεταξύ των ιδίων αυτών γονιδίων και των παραγόντων κινδύνου ή των φαινοτύπων που θεωρούνται αιτίες εμφάνισης της ασθένειας (Minelli κ.α., 2004, σελ. 445).

Σε πολλές περιπτώσεις, αυτές οι μελέτες προστίθενται σε προϋπάρχουσα έρευνα συσχέτισης του φαινότυπου και της νόσου. Πιο συγκεκριμένα, έχουν αξιολογηθεί οι συσχετίσεις μεταξύ πολυμορφισμών στο γονίδιο της μεθυλενοτετραϋδροφολικής αναγωγάσης (MTHFR) και της στεφανιαίας νόσου (CHD) και μεταξύ του ίδιου γονιδίου MTHFR και του επιπέδου ομοκυστεΐνης (Minelli κ.α., 2004, σελ. 445). Αυτές οι μελέτες παρακινήθηκαν, εν μέρει, από τα προϋπάρχοντα στοιχεία συσχέτισης μεταξύ του επιπέδου ομοκυστεΐνης και της στεφανιαίας νόσου. Ομοίως, έχουν πραγματοποιηθεί πολλές μελέτες πολυμορφισμών στο γονίδιο της απολιποπρωτεΐνης Ε και της στεφανιαίας

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

νόσου ή του εγκεφαλικού επεισοδίου, καθώς και πολλές μελέτες του ίδιου γονιδίου απολιποπρωτεΐνης Ε και των επιπέδων λιπιδίων. Αυτές οι μελέτες βασίστηκαν σε προϋπάρχουσες επιδημιολογικές ενδείξεις απευθείας συσχέτισης των λιπιδίων και της στεφανιαίας νόσου ή του εγκεφαλικού επεισοδίου (Minelli κ.α., 2004, σελ. 445).

Η αύξηση των γενετικών μελετών οδηγεί στην πραγμάτωση μετα-αναλύσεων με στόχο τη σύνθεση ενδείξεων, ξεπερνώντας παράλληλα περιορισμούς ισχύος που εντοπίζονται σε μελέτες, ακόμη και μετρίου μεγέθους (Minelli κ.α., 2004, σελ. 446). Δύο κύριοι παράγοντες προκύπτουν από την ανασκόπηση αυτών των μετα-αναλύσεων:

- Οι μελέτες που εξετάζουν τη σχέση μεταξύ γονιδίου και ενδιάμεσου φαινότυπου είναι λιγότερο συχνές από τις μελέτες που εξετάζουν τη σχέση μεταξύ γονιδίου και νόσου.

- Ενδείξεις για τη συσχέτιση γονότυπου - φαινότυπου λαμβάνονται ως παρεπόμενο μελετών που εστιάζουν κυρίως στη διερεύνηση της σχέσης γονότυπου - νόσου. Έτσι, αυτές οι πληροφορίες λαμβάνονται μόνο για ένα υποσύνολο των υποκειμένων (Minelli κ.α., 2004, σελ. 446).

Σε περιπτώσεις όπου υπάρχει ένδειξη ότι ο φαινότυπος βρίσκεται ενδιάμεσα στην αιτιολογική οδό από το γονότυπο στην ασθένεια, μια μετα-ανάλυση που ενσωματώνει τις τρεις σχέσεις (γονότυπος - φαινότυπος, γονότυπος - νόσος, φαινότυπος - νόσος) θα ήταν λογική. Αυτή η προσέγγιση ενισχύεται από τη Μεντελιανή τυχαιοποίηση, που υποστηρίζει ότι τα γονίδια κληρονομούνται κατά τη σύλληψη μίας φαινομενικά τυχαίας διαδικασίας (Minelli κ.α., 2004, σελ. 446).

Αντίστοιχα, οι επιδημιολογικές μελέτες συσχετίσεων γονότυπου - φαινότυπου και γονότυπου - νόσου αναδεικνύουν ισχυρούς παραλληλισμούς με τυχαιοποιημένες δοκιμές και δεν πρέπει να επηρεάζονται απο σύγχυση ή αντίστροφη αιτότητα με τρόπο, που να καθιστά τις κλασικές επιδημιολογικές μελέτες γονότυπου - νόσου δύσκολο να ερμηνευτούν (Minelli κ.α., 2004, σελ. 445). Πιο συγκεκριμένα, η σύγχυση δυσκολεύει τον προσδιορισμό του παράγοντα που προκαλεί την αλλαγή στον άλλον (ο γονότυπος επηρεάζει τη νόσο ή το αντίστροφο) (Lu CY, 2009, σελ. 695). Η αντίστροφη αιτιότητα, από την άλλη πλευρά, οδηγεί στη συνθήκη ότι η αιτιότητα μεταξύ των σχέσεων είναι αντίθετη από την προβλεπόμενη, προκαλώντας δυσκολίες στην ερμηνεία των δεδομένων και στην κατανόηση των σχέσεων στα πλαίσια ανάλυσης (George Davey Smith & Gibran Hemani, 2014, σελ. R92). Συνδυάζοντας πληροφορίες από τις παραπάνω επιδημιολογικές μελέτες, θα μπορούσε να προκύψει θεωρητικά μια εκτίμηση, χωρίς σύγχυση, της σχέσης φαινότυπου-ασθένειας. Οι ολοκληρωμένες, όμως, μετα-αναλύσεις μπορεί να είναι σε θέση να επωφεληθούν από τη Μεντελιανή τυχαιοποίηση με στόχο να ελέγξουν εάν ο φαινότυπος είναι πραγματικά στην αιτιώδη οδό και να προσφέρουν μια εκτίμηση του μεγέθους της επίδρασης του φαινότυπου στη νόσο (Minelli κ.α., 2004, σελ.445).

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

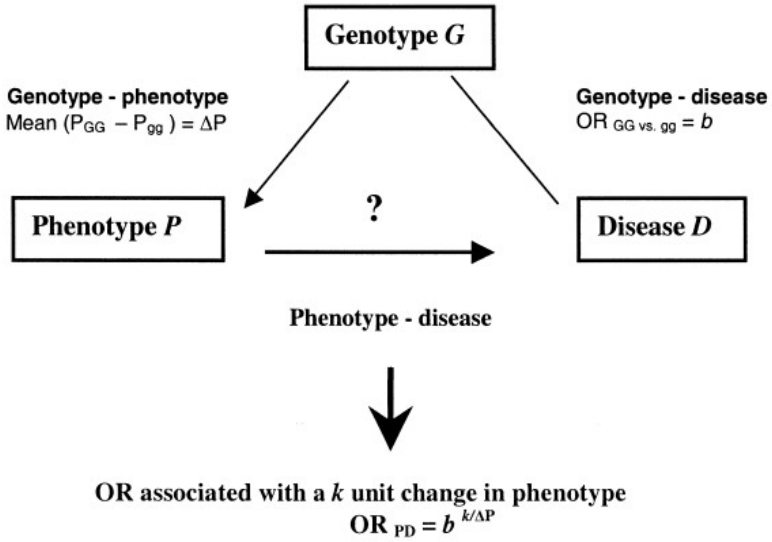
1.2 Μεντελιανή τυχαιοποίηση (MR)

Στην επιδημιολογία, η MR διερευνά την αιτιολογική επίδραση μιας έκθεσης σε έναν παράγοντα, χρησιμοποιώντας γνωστή παραλλαγή στα γονίδια. Σύμφωνα με βασικές παραδοχές, ο σχεδιασμός της συνεισφέρει στη μείωση της αντίστροφης αιτιότητας και της σύγχυσης, προκειμένου να βελτιώσει και να διευκολύνει ουσιαστικά την ερμηνεία των αποτελεσμάτων των επιδημιολογικών μελετών (Minelli κ.α., 2004, σελ. 450).

Η Μεντελιανή τυχαιοποίηση αποτελεί μία μέθοδο χρήσης γνωστής διακύμανσης σε γονίδια γνωστής λειτουργίας, κυρίως για την εξέταση της αιτιώδους επίδρασης μιας τροποποιήσιμης έκθεσης στη νόσο χρησιμοποιώντας μελέτες παρατήρησης (Smith κ.α., 2014, σελ. R89). Ο όρος πρωτοεισήχθη από τους Gray και Wheatley το 1991, περιγράφοντας μια μέθοδο που επιτρέπει την αντικειμενική εκτίμηση των αποτελεσμάτων της θεραπείας του καρκίνου μέσα σε ένα οικογενειακό πλαίσιο (Smith, 2006). Έκτοτε, ο όρος εφαρμόζεται σε μελέτες που χρησιμοποιούν γενετικές παραλλαγές για την απόκτηση πληροφοριών σχετικά με την αιτιολογική επίδραση ενός τροποποιήσιμου παράγοντα κινδύνου (Smith, 2006). Βασικό κριτήριο αποτελεί η αποδεδειγμένη συσχέτιση των γενετικών αυτών παραλλαγών με τον ίδιο παράγοντα κινδύνου. Η απόκτηση των συγκεκριμένων πληροφοριών επιτυγχάνεται μέσω των ιδιοτήτων που παρουσιάζουν οι γενετικές παραλλαγές, καθιστώντας τες ανθεκτικές στην αντίστροφη αιτιότητα και στη σύγχυση. Παράλληλα, με την προσέγγιση αυτή, η MR αντισταθμίζει τα εμπόδια (όπως η αντίστροφη αιτιότητα και η σύγχυση) που συνήθως επηρεάζουν τις επιδημιολογικές μελέτες, επιτρέποντας μια πιο αξιόπιστη ανάλυση επιδράσεων έκθεσης και αποτελέσματος (Zou κ.α., 2022, σελ. 1).

1.2.1 Εφαρμογή

Με σκοπό τη χρήση γενετικών μελετών για την ποσοτικοποίηση της σχέσης μεταξύ φαινότυπου και ασθένειας, απαιτείται ο συνδυασμός της εκτίμησης της συσχέτισης γονότυπου - ασθένειας με την εκτίμηση της συσχέτισης γονότυπουφαινότυπου (Minelli κ.α., 2004, σελ. 446).



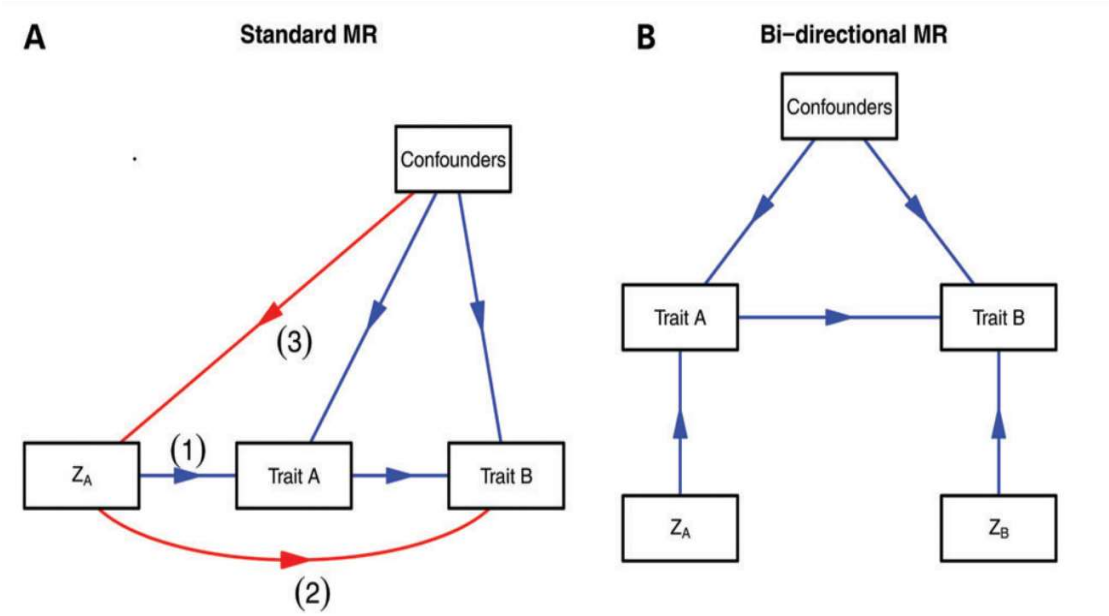
Γράφημα 3. Υπολογισμός μίας εκτίμησης, χωρίς σύγχυση, της επίδρασης μίας αλλαγής του φαινότυπου σε μια ασθένεια με βάση την έννοια της Μεντελιανής τυχαιοποίησης (Minelli κ.α., 2004, σελ. 446).

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Στο Γράφημα 2 φαίνεται ένας μεταλλαγμένος γονότυπος (GG), ο οποίος προκαλεί αυξημένο κίνδυνο εμφάνισης της ασθένειας σε σύγκριση με τον άγριο (κανονικό) τύπο (gg) και μάλιστα αυτό μετριέται από τον λόγο σχετικών πιθανοτήτων, $OR_{GG\ vs\ gg}$. Περαιτέρω θεωρείται ότι το GG σε σύγκριση με το gg προκαλεί μια μέση διαφορά, ΔP, στο επίπεδο του ενδιάμεσου φαινότυπου (Mineli κ.α., 2004, σελ. 446). Στη συνέχεια, θεωρώντας γραμμική τη σχέση μεταξύ διαφοράς φαινότυπου και λογαρίθμου του λόγου σχετικών πιθανοτήτων για τη νόσο, $OR_{GG\ VS\ gg}{}^{1/\Delta P}$, προκύπτει μία εκτίμηση, χωρίς σύγχυση, του λόγου σχετικών πιθανοτήτων της νόσου που είναι απόρροια μιας μοναδιαίας αλλαγής στο φαινότυπο (Minelli κ.α., 2004, σελ. 446).

1.2.2 Βασικές αρχές

Στη βιολογία συναντώνται αρκετοί περιορισμοί στη διαδικασία εύρεσης της αιτιώδους κατεύθυνσης μεταξύ συσχετιζόμενων μεταβλητών, καθώς η απλή ανάλυση παλινδρόμησης δυσκολεύεται να δώσει σαφείς απαντήσεις (Smith κ.α., 2014, σελ. R89). Η συσχέτιση μεταξύ δύο μεταβλητών θα μπορούσε να παρουσιάζει μία αιτιακή σχέση, αλλά η κατεύθυνση της αιτιότητας (λ.χ. το A προκαλεί το B ή το B προκαλεί το A) δεν είναι πάντα ξεκάθαρη (Smith κ.α., 2014, σελ. R89). Επιπλέον, η πιθανή ύπαρξη μη παρατηρούμενων παραγόντων, που επηρεάζουν και τις δύο μεταβλητές και οδηγούν στη συσχέτισή τους, μπορεί να οδηγήσει σε σύγχυση (Smith κ.α., 2014, σελ. R89).



Γράφημα 4. Σχηματική αναπαράσταση της MR (Smith κ.α., 2014, σελ. R90).

Η MR, όπως φαίνεται στο Γράφημα 3(α), μπορεί να χρησιμοποιηθεί για τον έλεγχο της υπόθεσης ότι το χαρακτηριστικό A προκαλεί το χαρακτηριστικό B, υπό την προϋπόθεση ότι πληρούνται επαρκώς οι προϋποθέσεις (1), (2) και (3),που διέπουν ότι το Z_A είναι έγκυρο IV. Πιο συγκεκριμένα, το (1) σχετίζεται με τον ενδιάμεσο φαινότυπο ενδιαφέροντος. Επίσης το (2) δεν έχει καμία σχέση με το αποτέλεσμα παρά μόνο μέσω του ενδιάμεσου φαινότυπου και το (3) δεν σχετίζεται με μετρούμενους ή μη συγχυτικούς παράγοντες (Smith κ.α., 2014, σελ. R90). Στο Γράφημα 3(β), αναπαρίσταται η αμφίδρομη MR (bi-directional), όπου η αιτιολογική κατεύθυνση μεταξύ των χαρακτηριστικών A και B, στην περίπτωση που είναι υπαρκτή, μπορεί να διευκρινιστεί, με την προϋπόθεση χρήσης

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

έγκυρων IV για κάθε χαρακτηριστικό (Smith κ.α., 2014, σελ. R90).

Σε αυτήν την περίπτωση, η επίδραση της ανεξάρτητης μεταβλητής στο αποτέλεσμα μπορεί να είναι ανεπαίσθητη. Ακόμη και αν η υποθετική αιτιώδης κατεύθυνση προσδιορίστηκε σωστά, αν η ανεξάρτητη μεταβλητή σχετίζεται με ορισμένους μη παρατηρούμενους ή ανακριβείς συγχυτικούς παράγοντες, η εκτίμηση της αιτιώδουςεπίδρασής της μπορεί να είναι μεροληπτική (Smith κ.α., 2014, σελ. R89). Η μεθοδολογία MR αποσκοπεί στην αντιμετώπιση αυτού του ζητήματος, επιδιώκοντας την αμερόληπτη ανίχνευση των αιτιωδών αποτελεσμάτων και, όπου είναι εφικτό, την εκτίμηση του μεγέθους τους (Smith κ.α., 2014, σελ. R89).

Με βάση το Γράφημα 3, αν θεωρηθεί ότι το χαρακτηριστικό A και το χαρακτηριστικό B συσχετίζονται, συμπεραίνεται ότι στην περίπτωση που η συσχέτιση αυτή προκύπτει επειδή το A προκαλεί το B, τότε οποιαδήποτε μεταβλητή που επηρεάζει το χαρακτηριστικό A θα πρέπει να επηρεάζει και το χαρακτηριστικό B (Smith κ.α., 2014, σελ. R89). Για να υπάρξει αιτιώδης σχέση μεταξύ A και B, είναι σημαντικό να εντοπιστεί ένα IV που να συσχετίζεται αξιόπιστα με το χαρακτηριστικό A σε γνωστή κατεύθυνση(Smith κ.α., 2014, σελ. R90). Πλεονέκτημα, βέβαια, αποτελεί το γεγονός ότι σχεδόν όλα τα χαρακτηριστικά ενδιαφέροντος επηρεάζονται συχνά από γενετικούς πολυμορφισμούς και μπορούν να λειτουργήσουν ως αξιόπιστα IV για αυτόν τον σκοπό (Smith κ.α., 2014, σελ. R90).

Σε γενετικές συσχετίσεις, η κατεύθυνση της αιτιότητας είναι συνήθως από το γενετικό πολυμορφισμό προς το χαρακτηριστικό ενδιαφέροντος και όχι το αντίστροφο (Smith κ.α., 2014, σελ. R90). Ωστόσο, οι περιβαλλοντικές εκθέσεις συχνά συνδέονται με παράγοντες συμπεριφορικούς, κοινωνικούς και φυσιολογικούς, οι οποίοι προκαλλόυν σύγχυση στο αποτέλεσμα (Smith κ.α., 2014, σελ. R90). Για την επίλυση του συγκεκριμένου ζητήματος οι γενετικές παραλλαγές μπορούν να χρησιμεύσουν ως δείκτες συγκεκριμένων τιμών χαρακτηριστικών (Smith κ.α., 2014, σελ. R90). Η επιλογή τους συνίσταται διότι τα αποτελέσματά τους υπόκεινται σε σχετικά μικρό σφάλμα μέτρησης ή μερική μεροληψία και η πραγματική αιτιολογική παραλλαγή για το χαρακτηριστικό δεν απαιτείται (Smith κ.α., 2014, σελ. R90). Συγκεκριμένα, τις προϋποθέσεις για την MR ικανοποιεί ένας δείκτης σε ανισορροπία σύνδεσης (LD) με την αιτιολογική παραλλαγή. Τέλος, με τη διεξαγωγή μελετών συσχέτισης σε όλο το γονιδίωμα (GWAS) και την ανάπτυξη τεχνολογιών γονιδιώματος υψηλής απόδοσης, τα γενετικά δεδομένα είναι συνήθως διαθέσιμα σε μεγάλες φαινοτυπικές μελέτες (Smith κ.α., 2014, σελ. R90).

1.2.3 Παραδείγματα

Σε αυτήν την ενότητα παρατίθεται ένας πίνακας, που περιλαμβάνει κάποια συμπεράσματα συσχετίσεων έκθεσης σε ένα αποτέλεσμα, με χρήση της MR (Smith κ.α., 2014, σελ. R91). Ορισμένες συσχετίσεις από αυτές που εξετάστηκαν, αποδείχτηκαν υπαρκτές. Άλλες εμφάνισαν στα συμπεράσματα τους σύγχυση ή / και αντίστροφη αιτιότητα, με αποτέλεσμα να απορριφθεί η πιθανότητα ύπαρξης οποιασδήποτε συσχέτισης μεταξύ έκθεσης-αποτελέσματος.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Πίνακας 1. Παραδείγματα επίδρασης χαρακτηριστικού σε νόσο (Smith κ.α., 2014, σελ. R91).

Τύπος της έκθεσης	Έκθεση/Χαρακτηριστικό	Ασθένεια/Αποτέλεσμα	Συμπέρασμα
Βιολογικοί Δείκτες	CRP	Στεφανιαία νόσος	Η συσχέτιση παρατήρησης μεταξύ της CRP και της στεφανιαίας νόσου είναι αποτέλεσμα σύγχυσης και/ή αντίστροφης αιτιότητας
	Σίδηρος (Serum iron)	Parkinson	Τα υψηλά επίπεδα σιδήρου μειώνουν τον κίνδυνο της νόσου του Parkinson
	Ουρικό οξύ	Στεφανιαία νόσος	Η συσχέτιση παρατήρησης μεταξύ ουρικού οξέος και στεφανιαίας νόσου οφείλεται, εν μέρει, σε σύγχυση από τον BMI
	Ανασταλτικός παράγοντας μετανάστευσης μακροφάγων (MIF)	Διαβήτης τύπου 2	Η αυξημένη MIF, μεταξύ άλλων παραγόντων, αυξάνει τον κίνδυνο διαβήτη τύπου 2
	Ιντερλευκίνη 6 (IL6)	Στεφανιαία νόσος	Η IL6 αυξάνει τον κίνδυνο στεφανιαίας νόσου
Συμπεριφορές	Κάπνισμα	Άγχος/κατάθλιψη	Το άγχος και η κατάθλιψη μεταξύ των καπνιστών δε φαίνεται να είναι συνέπεια του καπνίσματος
	Κατανάλωση αλκοόλ	Πίεση αίματος	Η κατανάλωση αλκοόλ αυξάνει την αρτηριακή πίεση
Μετρήσεις	BMI (δείκτης μάζας σώματος)	Συμπτωματική νόσος των χολόλιθων	Ο υψηλότερος BMI αυξάνει τον κίνδυνο συμπτωματικής πέτρας στη χολή
	Κατανάλωση αλκοόλ	Σχολική επίδοση	Το εύρημα παρατήρησης ότι η μέτρια κατανάλωση αλκοόλ από τη μητέρα σχετίζεται με πιο ευνοϊκή σχολική επίδοση, οφείλεται σε σύγχυση και η περιστασιακή συσχέτιση είναι προς την αντίθετη κατεύθυνση.
	Μητρικός BMI	Λίπος μάζας απογόνων	Η μάζα λίπους σε παιδιά ηλικίας 9-11 ετών δεν επηρεάζεται έντονα από το BMI της μητέρας και τη διάρκεια της εγκυμοσύνης

1.3 Επέκταση της MR

1.3.1 Χρήση πολλαπλών παραλλαγών

Η MR ιδανικά εφαρμόζεται με τη χρήση ενός μόνο IV του οποίου η βιολογική επίδραση του χαρακτηριστικού είναι κατανοητή (Smith κ.α., 2022, σελ. R90). Ακόμη και αυτή η περίπτωση, όμως, υπόκειται σε ορισμένους πιθανούς περιορισμούς, οι οποίοι μπορούν να μετριαστούν εν μέρει αυξάνοντας τον αριθμό των παραλλαγών που χρησιμοποιούνται

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

ως IV (Smith κ.α., 2022, σελ. R91).

Η γενετική επίδραση μπορεί να μην είναι ιδιαίτερα μεγάλη, με αποτέλεσμα ένα αδύναμο IV και την απαίτηση πολύ μεγάλου μεγέθους δείγματος. Αυξάνοντας τον αριθμό των παραλλαγών, το ποσοστό διακύμανσης που εξηγείται από το IV αυξάνεται, βελτιώνοντας έτσι την ακρίβεια της μεθόδου των ελαχίστων τετραγώνων σε δύο στάδια (Smith κ.α., 2022, σελ. R91).

Η συμβατική ανάλυση των IV απαιτεί τις μεταβλητές να είναι έγκυρες, δηλαδή να πληρούν τρεις προϋποθέσεις. Ένα όργανο για το χαρακτηριστικό A πρέπει να:

- Συνδέεται αξιόπιστα με το χαρακτηριστικό A.
- Συνδέεται με το αποτέλεσμα (χαρακτηριστικό B) μόνο μέσω του χαρακτηριστικού A.
- Είναι ανεξάρτητο από μη παρατηρούμενους συγχυτικούς παράγοντες που επηρεάζουν τα γνωρίσματα A και B μετά την προετοιμασία σε παρατηρούμενες συγχυτικές ουσίες (Smith κ.α., 2022, σελ. R91).

Συνεχίζοντας, η παραλλαγή μπορεί να είναι πλειοτροπική ή σε ανισορροπία σύνδεσης (LD) με μια παραλλαγή που επηρεάζει το αποτέλεσμα, παραβιάζοντας τις προϋποθέσεις για την έγκυρη χρήση του IV (Smith κ.α., 2022, σελ. R91). Αυτό μπορεί να διερευνηθεί χρησιμοποιώντας πολλαπλά IV. Για παράδειγμα, είναι ολοένα και πιο απίθανο να συμφωνούν δύο, τρία ή και περισσότερα ανεξάρτητα IV σε ένα κοινό συμπέρασμα, λόγω της τέλειας εξισορρόπησης των πλειοτροπικών επιδράσεων και στα δύο χαρακτηριστικά. Εννέα πολυμορφισμοί από έξι γονίδια οδηγούν, με ανεξάρτητο τρόπο, σε πολύ παρόμοιες προβλεπόμενες αιτιολογικές επιδράσεις της LDL-C, αποτελώντας χαρακτηριστικό παράδειγμα που αποδεικνύει την αιτιολογική επίδραση της λιποπρωτεϊνικής χοληστερόλης χαμηλής πυκνότητας (LDL-C) στη CHD (Smith κ.α., 2022, σελ. R91).

Τέλος, πολλές παραλλαγές μπορούν επίσης να παρέχουν κάποια στοιχεία σχετικά με την πολυπλοκότητα των συσχετίσεων στις μελέτες MR (Smith κ.α., 2022, σελ. R91). Εάν πολλές παραλλαγές, που σχετίζονται με έναν συγκεκριμένο ενδιάμεσο φαινότυπο μέσω διαφορετικών μηχανισμών, συσχετίζονται και με την έκβαση της νόσου με τον τρόπο που προβλέπεται από τη συσχέτισή τους με τον ενδιάμεσο φαινότυπο (όπως στην περίπτωση πολλαπλών παραλλαγών που σχετίζονται με την LDL-C και τη CHD), τότε ο συγκεκριμένος τρόπος με τον οποίο μια παραλλαγή σχετίζεται με τον ενδιάμεσο φαινότυπο είναι απίθανο να επηρεάσει τα συγκεντρωτικά αποτελέσματα (Smith κ.α., 2022, σελ. R91).

1.3.2 MR δύο δειγμάτων

Συχνή περίπτωση αποτελεί η ύπαρξη μιας παρατηρούμενης συσχέτισης μεταξύ δύο μεταβλητών, αλλά το υψηλό κόστος μέτρησης ή η έλλειψη κατάλληλων βιοδειγμάτων οδηγεί σε σχετικά μικρά σύνολα δεδομένων με ενδιάμεσους φαινοτύπους (Smith κ.α., 2014, σελ. R93). Έχουν αναπτυχθεί μέθοδοι για την εκτέλεση ανάλυσης τύπου IV όταν ο ενδιάμεσος φαινότυπος και η μεταβλητή αποτελέσματος μετρώνται σε δύο ανεξάρτητα σύνολα δεδομένων, και αυτές μπορούν να εφαρμοστούν στο πλαίσιο της MR (Smith κ.α., 2014, σελ. R93). Αυτή η προσέγγιση μπορεί να είναι ιδιαίτερα πολύτιμη όταν εφαρμόζεται σε πολύ μεγάλα σύνολα δεδομένων που υπάρχουν και συσχετίζουν τα δεδομένα GWAS με τα αποτελέσματα της νόσου, αλλά τα οποία στερούνται ενδιάμεσων δεδομένων φαινότυπου (Tam κ.α., 2019, σελ. 469).

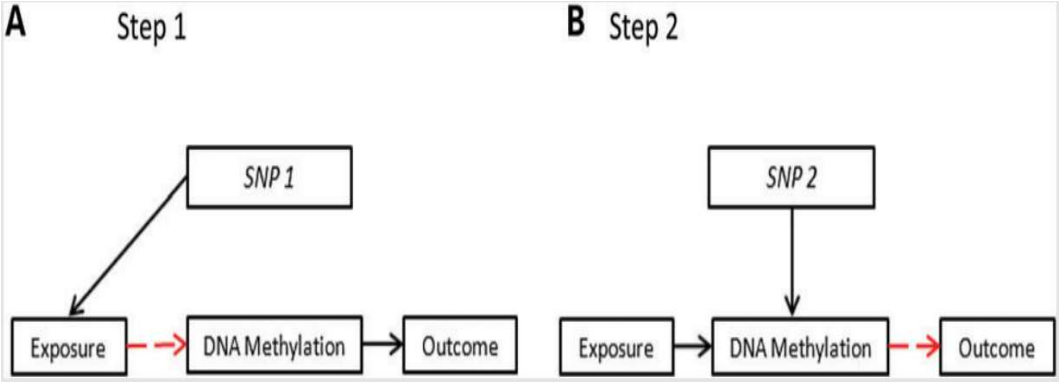
Ένα άλλο σενάριο στο οποίο μπορεί να χρησιμοποιηθεί η MR δύο δειγμάτων είναι εάν το

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

σύνολο δεδομένων στο οποίο εκτελείται η MR είναι το ίδιο με αυτό που χρησιμοποιείται για την αναγνώριση IV (Smith κ.α., 2014, σελ. R93). Τα GWAS είναι γνωστό ότι οδηγούν σε υπερεκτίμηση των μεγεθών γενετικής επίδρασης λόγω της κατάρας του νικητή, πηγή εμφάνισης μεροληψίας στην MR (Mounier κ.α., 2022, σελ. 315). Η διαίρεση του συνόλου δεδομένων σε δύο ή περισσότερα δείγματα για εκτίμηση και δοκιμή μπορεί να μετριάσει αυτό το πρόβλημα. Αυτή η μέθοδος έχει εφαρμοστεί σε μια μελέτη σωματικής δραστηριότητας και παιδικής παχυσαρκίας(Smith κ.α., 2014, σελ. R93).

1.3.3 MR δύο βημάτων και δύο δειγμάτων

Οι γενετικές παραλλαγές μπορούν να χρησιμοποιηθούν ως IV σε ένα πλαίσιο δύο βημάτων για να εξεταστεί εάν η έκθεση και η ασθένεια οφείλονται σε συγκεκριμένο βαθμό μεθυλίωσης DNA (Smith κ.α., 2022, σελ. R94).

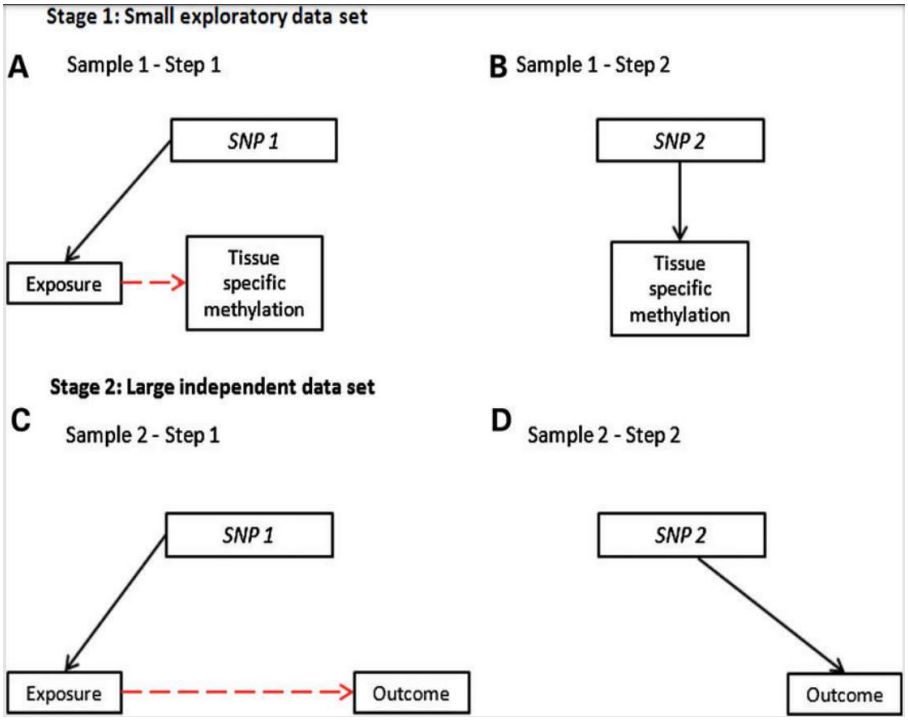


Γράφημα 5. MR δύο βημάτων (Smith κ.α., 2022, σελ. R94).

Στο πρώτο βήμα, χρησιμοποιείται ένας μονονουκλεοτιδικός πολυμορφισμός (SNP) για την αντιπροσώπευση της περιβαλλοντικά τροποποιήσιμης έκθεσης που εξετάζεται (π.χ. κάπνισμα), προκειμένου να διερευνηθεί πώς αυτή η έκθεση επηρεάζει τη μεθυλίωση του DNA. Στο δεύτερο βήμα, χρησιμοποιείται ένα διαφορετικό SNP (που δεν σχετίζεται με την έκθεση), κατά προτίμηση μία παραλλαγή cis, για να αντιπροσωπεύσει τη συγκεκριμένη διαφορά μεθυλίωσης του DNA και να εξεταστεί η συσχέτισή της με την έκβαση της νόσου υπό διερεύνηση (Smith κ.α., 2022, σελ. R94).

Η MR δύο δειγμάτων, δύο σταδίων εφαρμόζεται για την διερεύνηση σχετικά με τη μεθυλίωση του DNA ιστού ως δυνητικού αιτιολογικού ενδιάμεσου φαινότυπου (Smith κ.α., 2022, σελ. R94).

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 6. MR δύο δειγμάτων (Smith κ.α., 2022, σελ. R94).

Στο μικρότερο πρώτο δείγμα, η συσχέτιση της έκθεσης σε ειδική για τον ιστό μεθυλίωση DNA αποδεικνύεται χρησιμοποιώντας μια προσέγγιση MR (με το SNP1 που σχετίζεται με την έκθεση, A) και μια παραλλαγή cis που σχετίζεται με την ίδια διαφορά μεθυλίωσης αλλά δεν σχετίζεται με την έκθεση (Smith κ.α., 2022, σελ. R94). Ταυτοποιήθηκε ότι (SNP2, B) στο μεγαλύτερο δεύτερο δείγμα, η έκθεση φαίνεται ότι επηρεάζει το αποτέλεσμα μέσω της χρήσης του SNP1, είτε μέσω της συσχέτισης του SNP1 τόσο με την έκθεση (εάν υπάρχουν διαθέσιμα δεδομένα για αυτό) όσο και με το αποτέλεσμα, είτε εάν δεν υπάρχουν διαθέσιμα δεδομένα έκθεσης, τότε απλά συσχετίζοντας το SNP1 με το αποτέλεσμα (C). Τέλος, η μεθυλίωση που σχετίζεται με την έκθεση φαίνεται ότι επηρεάζει το αποτέλεσμα μέσω της χρήσης του SNP2, το οποίο σχετίζεται άμεσα με το αποτέλεσμα (D) (Smith κ.α., 2022, σελ. R94).

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Κεφάλαιο 2

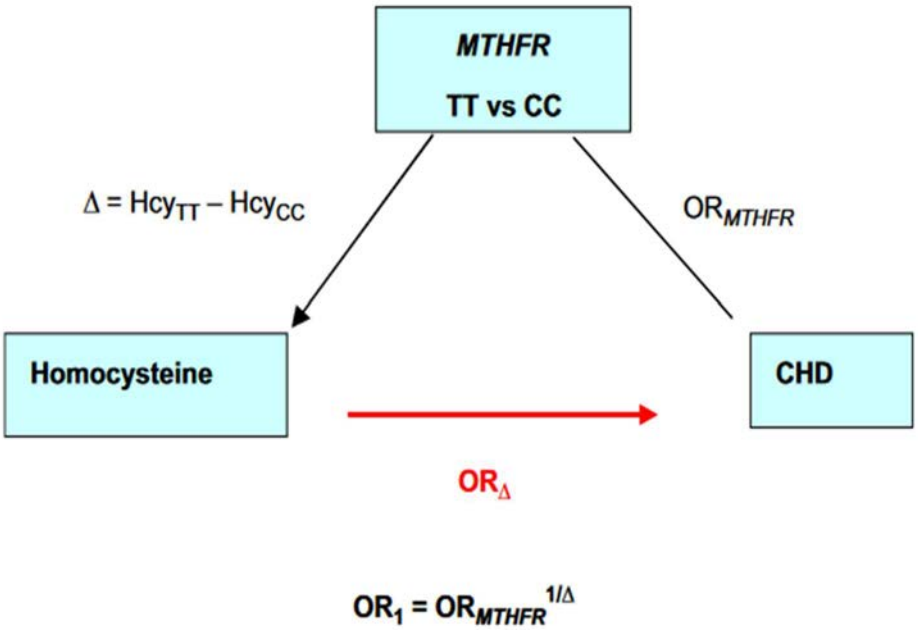
Μέθοδοι

Για την εφαρμογή των μοντέλων της MR, εξετάζεται η χρήση του γονιδίου MTHFR ως IV στην εκτίμηση του βαθμού επίδρασης της ομοκυστεΐνης στη CHD. Πολλές παραδοσιακές μελέτες παρατήρησης, μάλιστα, παρέχουν ενδείξεις ότι η αύξηση του επιπέδου της ομοκυστεΐνης στο αίμα σχετίζεται με αυξημένο κίνδυνο CHD (Wald κ.α., 2002, σελ. 7) κ.α., 2004, σελ. 447). Ωστόσο, σημαντικοί πιθανοί παράγοντες που προκαλούν σύγχυση, όπως το κάπνισμα και η αρτηριακή πίεση, μπορεί να ευθύνονται για μέρος ή για το σύνολο της παρατηρούμενης συσχέτισης. Αυτή η σύγχυση μπορεί να εξηγήσει την ετερογένεια μεταξύ των μελετών που βρέθηκε σε μια πρόσφατη μετα-ανάλυση μεμονωμένων δεδομένων ασθενών (Minelli κ.α., 2004, σελ. 447).

Τη σχέση αυτή εκτός από τη σύγχυση, θεωρείται ότι επηρεάζουν και άγνωστοι (και επομένως πρακτικά μη μετρήσιμοι) παράγοντες αντίστροφης αιτιότητας, όπως για παράδειγμα αυξήσεις στην ομοκυστεΐνη του αίματος μπορεί να προκύψουν από αθηροσκλήρωση και CHD (Minelli κ.α., 2004, σελ. 447). Η πιθανή παρουσία μη προσαρμοσμένης σύγχυσης καθιστά πολύ δύσκολη την εξακρίβωση της αιτιολογικής σχέσης μεταξύ ομοκυστεΐνης και CHD. Το γονίδιο MTHFR έχει έναν κοινό πολυμορφισμό που έχει αποδειχθεί ότι επηρεάζει το επίπεδο της ομοκυστεΐνης στο αίμα, γεγονός που το καθιστά μια πιθανή επιλογή ως IV (Minelli κ.α., 2004, σελ. 447).

Ο πολυμορφισμός περιλαμβάνει μια αντικατάσταση του αλληλόμορφου C σε αλληλόμορφο T στη βάση 677 και ως εκ τούτου ο ομόζυγος γονότυπος άγριου τύπου αναφέρεται ως CC και ο μεταλλαγμένος ομόζυγος γονότυπος ως TT. Ο πολυμορφισμός TT οδηγεί σε μειωμένη ενζυμική δραστηριότητα και κατά συνέπεια σε υψηλότερα επίπεδα ομοκυστεΐνης. Επομένως, εάν η ομοκυστεΐνη σχετίζεται αιτιολογικά με τη CHD, ο πολυμορφισμός TT θα συσχετιστεί με τη CHD (Minelli κ.α., 2004, σελ. 447).

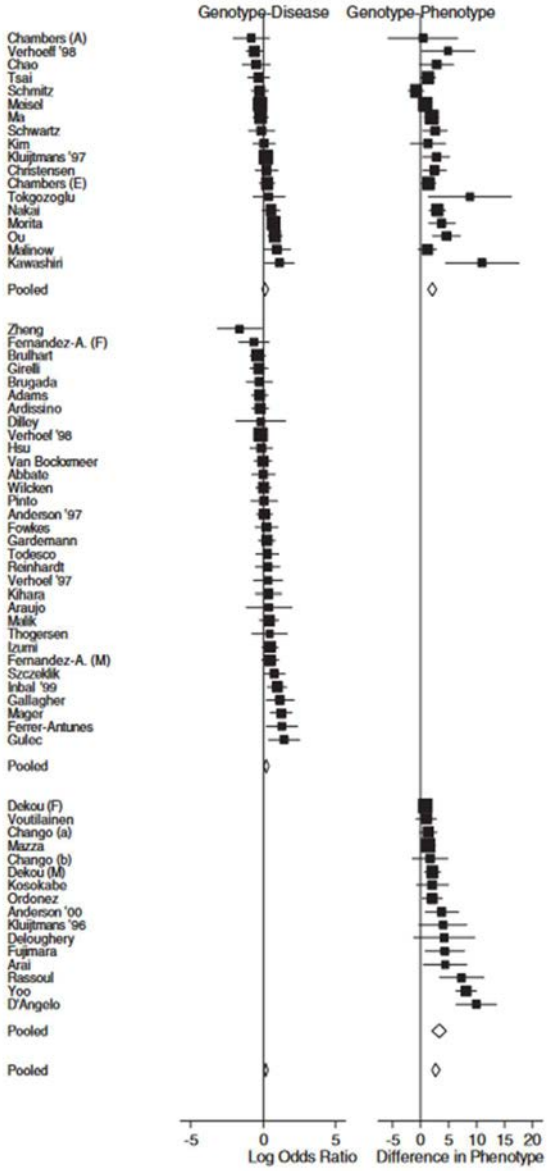
Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 7. Υπολογισμός μίας εκτίμησης χωρίς σύγχυση, της επίδρασης μίας αύξησης της ομοκυστεΐνης (Hcy) κατά 1 μmol/l στον κίνδυνο CHD, με βάση την τεχνική της Μεντελιανής τυχαιοποίησης (Minelli κ.α., 2004, σελ. 446).

Μια πρόσφατη γενετική μετα-ανάλυση δεδομένων από μεμονωμένους ασθενείς έδειξε αυξημένο κίνδυνο CHD κατά περίπου 16%, που σχετίζεται με τον γονότυπο TT σε σύγκριση με τον CC $(\mathrm{OR}: 1.16;\ 95\%\ \mathrm{CI}: 1.05 - 1.28)$ (Minelli κ.α., 2004, σελ. 448). Αυτό το αποτέλεσμα είναι παρόμοιο με αυτό μιας άλλης μετα-ανάλυσης που βασίζεται σε συγκεντρωτικά δεδομένα και το οποίο έδειξε ότι το $\mathrm{OR} = 1.21\ (95\%\ \mathrm{CI}: 1.06 - 1.39)$ (Wald κ.α., 2002, σελ. 2). Η τελευταία μελέτη εξέτασε επίσης στοιχεία σχετικά με τη συσχέτιση μεταξύ γονότυπου και φαινότυπου και ανέφερε μια μέση διαφορά $2.7\ (95\%\ \mathrm{CI}: 2.1 - 3.4)$ (Wald κ.α., 2002, σελ. 2). Συνδυάζοντας τις δύο γενετικές μετα-αναλύσεις, εντοπίστηκαν συνολικά 66 γενετικές μελέτες. Από αυτές, 32 αξιολόγησαν μόνο τη συσχέτιση γονότυπου-νόσου, 16 μόνο τη συσχέτιση γονότυπου-φαινότυπου και 18 και τις δύο (Minelli κ.α., 2004, σελ. 449). Το διπλό δενδροδιάγραμμα παρακάτω (Γράφημα 7) απεικονίζει τα αποτελέσματα των μελετών.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 8. Διπλό δενδροδιάγραμμα με συγκεντρωτικές εκτιμήσεις για τις σχέσεις γονότυπου-νόσου και γονότυπου-φαινότυπου (Minelli κ.α., 2004, σελ. 448).

2.1 Μοντέλα για πολυμεταβλητή ανάλυση

Σε κάθε πολυμεταβλητή μετα-ανάλυση, είναι σημαντικό να επιτρέπεται η συσχέτιση στις εκτιμήσεις σε σύγκριση με τις μελέτες που παρέχουν πληροφορίες για περισσότερα από ένα αποτελέσματα (Ho κ.α., 2022). Εάν η i^{th} μελέτη παρέχει συσχετισμένα αποτελέσματα, x_i και y_i, τα οποία υποτίθεται ότι ακολουθούν την πολυμεταβλητή κανονική (MVN) κατανομή με διακυμάνσεις vx_i και vy_i, τότε,

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_{xi} \\ \mu_{yi} \end{pmatrix}, \begin{pmatrix} v_{xi} & \psi\sqrt{v_{xi}v_{yi}} \\ \psi\sqrt{v_{xi}v_{yi}} & v_{yi} \end{pmatrix}\right) \tag{2.1}$$

Στο επόμενο επίπεδο της ιεραρχίας θεωρείται ότι διαφορετικές μελέτες ποικίλλουν σχετικά

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

με τα κοινά μέσα μ_x και μ_y με συσχετισμένες ετερογένειες τ_x και τ_y.

$$\begin{pmatrix}\mu_{xi}\\ \mu_{yi}\end{pmatrix} \sim MVN\left(\begin{pmatrix}\mu_x\\ \mu_y\end{pmatrix}, \begin{pmatrix}\tau_x & \rho\sqrt{\tau_x\tau_y}\\ \rho\sqrt{\tau_x\tau_y} & \tau_y\end{pmatrix}\right) \tag{2.2}$$

Η παράμετρος ρ αντιπροσωπεύει τη υσχέτιση μεταξύ των μελετών, οι οποίες μελέτες εξετάζουν και τα δύο αποτελέσματα και η παράμετρος ψ αντιπροσωπεύει τη συσχέτιση εντός της μελέτης (Jennrich κ.α., 1986). Για εκείνες τις μελέτες που μετρούν μόνο το πρώτο ή δεύτερο αποτέλεσμα χρησιμοποιούμε τις αντίστοιχες μονομεταβλητές κανονικές κατανομές.

$$x_i \sim N(\mu_{xi}, v_{xi})\ and\ \mu_{xi} \sim N(\mu_x, \tau_x) \tag{2.3}$$

και

$$y_i \sim N(\mu_{yi}, v_{yi})\ and\ \mu_{yi} \sim N(\mu_y, \tau_y) \tag{2.4}$$

Σε μοντέλα μετα-ανάλυσης στη Μεντελιανή τυχαιοποίηση, το x θα αντιπροσωπεύει το logOR του δεδομένου γονότυπου της νόσου και το y θα αντιπροσωπεύει τη μέση διαφορά στον φαινότυπο. Το αντικείμενο υπό εξέταση θα είναι η εκτίμηση της αναλογίας x προς y καθώς αυτό θα δώσει το logOR της επίδρασης του φαινότυπου στην ασθένεια.
Η βιβλιογραφία θα παρέχει τις εκτιμήσεις της κάθε μελέτης και τις διακυμάνσεις τους, vx_i και vy_i, οι οποίες μπορεί να αναφέρονται με τη μορφή τυπικών σφαλμάτων ή διαστημάτων εμπιστοσύνης. Θα υποτεθεί ότι αυτές οι αποκλίσεις είναι γνωστές χωρίς σφάλμα.

2.2 Συσχετίσεις εντός-μελέτης και μεταξύ-μελετών

Η συσχέτιση εντός της μελέτης προσδιορίζει την πιθανότητα εκτίμησης ενός μέτρου, μίας συγκεκριμένης μελέτης, ως μεγαλύτερο (ή μικρότερο) από την πραγματική του τιμή. Σε αυτήν την περίπτωση, υπάρχει τάση για την άλλη εκτίμηση, να είναι επίσης μεγαλύτερη (ή μικρότερη) από την πραγματική της τιμή (Wald κ.α., 2002, σελ. 5). Αυτή η συσχέτιση μπορεί να εκτιμηθεί μόνο από μεμονωμένα δεδομένα ασθενών και δεν αναφέρεται ποτέ από πρωτογενείς μελέτες. Η τυχαιοποιημένη κατανομή των γονιδιακών παραλλαγών και η ανεξαρτησία των εργαλείων από παράγοντες σύγχυσης αποτελούν δύο καλούς λόγους να υποτεθεί ότι η συσχέτιση εντός της μελέτης θα είναι αμελητέα στις περισσότερες μελέτες που χρησιμοποιούνται σε μια Μεντελιανή τυχαιοποίηση. Οι μελέτες ειδικά για το OR του γονότυπου στη νόσο βασίζονται σε συγκεντρωτικά δεδομένα, και στις περισσότερες από αυτές, το επίπεδο φαινότυπου μετράται μόνο σε ένα υπόδειγμα των ατόμων (Wald κ.α., 2002, σελ. 5).

Με σκοπό, λοιπόν, την εκτίμηση του πιθανού μεγέθους της συσχέτισης εντός της μελέτης, πραγματοποιήθηκε μία μικρή μελέτη προσομοίωσης, η οποία. περιλάμβανε τον ίδιο αριθμό περιπτώσεων και ελέγχων $(\mathrm{n} = 100)$. Η πιθανότητα ασθένειας δεδομένου του φαινότυπου θεωρήθηκε ότι ελέγχεται από μια λογαριθμική συνάρτηση του φαινότυπου με παραμέτρους α και β. Η κατανομή του φαινότυπου στον πληθυσμό θεωρήθηκε ότι είναι φυσιολογική με μικτή επίδραση του γονότυπου στον φαινότυπο (Klerk κ.α., 2022).

Επαναλαμβανόμενες μελέτες προσομοιώθηκαν κάτω από τις ίδιες συνθήκες και

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

υπολογίστηκε η συσχέτιση μεταξύ του εκτιμώμενου logOR για τη συσχέτιση γονότυπου-ασθένειας και της μέσης διαφοράς στον φαινότυπο (Klerk κ.α., 2022). Οι παράμετροι επιλέχθηκαν για να αντικατοπτρίζουν τις τιμές σε μελέτες της οδού MTHFR-ομοκυστεΐνη-CHD. Αρχικά, λήφθηκαν υπόψη οι ακόλουθες παράμετροι:

- Μια μεταβολή OR ανά μονάδα στον φαινότυπο της νόσου $[\beta = \log(1{:}1) = 0{:}095]$
- Μια διαφορά στο φαινότυπο λόγω γονότυπου 4 μονάδων, που δίνουν OR γονότυπου σε ασθένεια 1,46,
- συχνότητα μεταλλαγμένου γονότυπου 12% και
- μια διακύμανση μεταξύ υποκειμένων στα επίπεδα του φαινότυπου που κανονικά κατανέμεται με παραμέτρους N(8; 2) ή N(12; 2) ανάλογα με τον γονότυπο.
- Ένας βασικός κίνδυνος ασθένειας (α) με μέσο όρο -4 και τυπική απόκλιση 0,2, που αντικατοπτρίζει μη μετρήσιμους παράγοντες σύγχυσης (Klerk κ.α., 2022).

Υπό αυτές τις συνθήκες, το 3,9% των ατόμων με τον γονότυπο άγριου (wild) τύπου και το 5,6% εκείνων με τον μεταλλαγμένο (mutant) γονότυπο αναπτύσσουν τη νόσο. Όπως συνιστάται, η επίδραση του γονότυπου στον φαινότυπο εκτιμήθηκε μόνο από τα controls (Klerk κ.α., 2022).

Από τις 5000 επαναλαμβανόμενες προσομοιώσεις, η συσχέτιση μεταξύ του εκτιμώμενου logOR του γονότυπου στη νόσο και της μέσης διαφοράς στον φαινότυπο ήταν 0,015. Χρησιμοποιώντας περαιτέρω σετ 5000 προσομοιώσεων, η βασική κατάσταση άλλαξε αλλάζοντας μία παράμετρο (μέγεθος δείγματος 50 ή 100, διαφορά φαινότυπου 2 ή 6, OR 1,05 ή 1,2, βασικός κίνδυνος ασθένειας -3). Σε όλες αυτές τις προσομοιώσεις, η εκτιμώμενη συσχέτιση ήταν εντός του εύρους ±0,05 (Klerk κ.α., 2022).

Η συσχέτιση μεταξύ των μελετών, ρ, αντιπροσωπεύει την τάση για μελέτες που διεξάγονται σε πληθυσμούς όπου η πραγματική επίδραση του γονότυπου στον φαινότυπο είναι μεγάλη και εμφανίζουν ακόμα μεγαλύτερη μέση επίδραση του γονότυπου στη νόσο. Σε αντίθεση με τη συσχέτιση εντός της μελέτης, για μια μετα-ανάλυση μελετών με ένα ευρύ φάσμα πληθυσμών, αυτή η συσχέτιση μπορεί να είναι ουσιαστική (Wald κ.α., 2002, σελ. 5).

2.3 Μοντέλα ML / REML

2.3.1 Μοντέλα A ML/REML

Στο πρώτο μοντέλο, που αναπαρίσταται σχηματικά στο Γράφημα 8(α) παρακάτω, υιοθετούνται οι ιεραρχικές διμεταβλητές κανονικές κατανομές (Jennrich κ.α., 1986). Χωρίς μεμονωμένα δεδομένα δεν είναι δυνατό να εκτιμηθεί η συσχέτιση εντός-της-μελέτης ψ, αλλά για τους λόγους που αναφέρονται παραπάνω το ψ θα είναι πολύ μικρό και θεωρείται ίσο με 0. Δηλώνοντας το logOR του φαινότυπου της νόσου με $\theta = \frac{\mu_x}{\mu_y}$, η οριακή κατανομή γίνεται:

$$\begin{bmatrix} x_i \\ y_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} \theta\mu_y \\ \mu_y \end{bmatrix}, \begin{bmatrix} v_{xi} + \tau_x & \rho\sqrt{\tau_x\tau_y} \\ \rho\sqrt{\tau_x\tau_y} & u_{yi} + \tau_y \end{bmatrix}\right) \tag{2.5}$$

Όταν αναφέρεται μόνο μία από τις δύο εκτιμήσεις, τότε θεωρούνται ως μονομεταβλητές

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

κανονικές κατανομές. Έτσι, το x_i κατανέμεται κανονικά με μέσο $\theta\mu_y$ και διακύμανση $v_{xi} + \tau_x$, ενώ το y_i κατανέμεται κανονικά με μέσο μ_y και διακύμανση $v_{yi} + \tau_y$.

Οι μελέτες θεωρούνται ανεξάρτητες και, ως εκ τούτου, η πιθανότητα μπορεί να ληφθεί και να μεγιστοποιηθεί προκειμένου να παρέχονται εκτιμήσεις μέγιστης πιθανότητας για όλες τις παραμέτρους. Ωστόσο, συνήθως χρησιμοποιείται η υπολογιστική μέθοδος Restricted Maximum Likelihood (REML) για τον υπολογισμό του πίνακα διακύμανσης μεταξύ των μελετών (Jennrich κ.α., 1986). Αυτή η εκτίμηση μπορεί να πραγματοποιηθεί μεγιστοποιώντας μια ειδική περίπτωση της έκφρασης REML που παρέχεται από τους Jennrich και Schluchter, με την προϋπόθεση ότι ο πίνακας διακύμανσης μεταξύ των μελετών είναι θετικά ημι-καθορισμένος. Συγκεκριμένα η έκφραση για μεγιστοποίηση είναι

$$\lambda_{REML} = -\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma + \Delta_i| - \frac{1}{2}\log\left|\sum_{i=1}^{n}(\Sigma + \Delta_i)^{-1}\right| - \frac{1}{2}\sum_{i=1}^{n} r_i^T(\Sigma + \Delta_i)^{-1} r_i \qquad (2.6)$$

όπου Δi είναι ο πίνακας διακύμανσης της i^{th} μελέτης

$$\Delta_i = \begin{pmatrix} \sigma_{x_i}^2 & \rho_i\sigma_{x_i}\sigma_{y_i} \\ \rho_i\sigma_{x_i}\sigma_{y_i} & \sigma_{y_i}^2 \end{pmatrix} \qquad (2.7)$$

Το Σ υποδηλώνει τον πίνακα διακύμανσης μεταξύ των μελετών

$$\sum\begin{pmatrix} \tau_x^2 & \kappa\tau_x\tau_y \\ \kappa\tau_x\tau_y & \tau_y^2 \end{pmatrix} \qquad (2.8)$$

Το μοντέλο έχει παραμέτρους $(\theta, \mu_y, \tau_x, \tau_y, \rho)$ και η αριθμητική μεγιστοποίηση της πιθανότητας καταγραφής χρησιμοποιώντας την εντολή ml στο Stata βασίζεται στους μετασχηματισμούς των παραμέτρων $\left(\theta, \mu_y, \log\tau_x, \log\tau_\upsilon, \log\frac{1+\rho}{1-\rho}\right)$. Αυτό αποφεύγει την ανάγκη ελέγχου των περιορισμών φυσικού εύρους στις παραμέτρους, π.χ. τ_x και τ_y, οι οποίοι πρέπει να είναι μεγαλύτεροι από 0 εάν δεν έχουν μετασχηματιστεί σε λογάριθμο. Οι αρχικές τιμές υπολογίστηκαν από τα πρωτογενή δεδομένα ως

$$\theta = \frac{\bar{x}}{\bar{y}}, \qquad \mu_y = \bar{y}, \qquad \tau_x = Var(x), \qquad \tau_y = Var(y), \qquad \rho = Corr(x, y) \qquad (2.9)$$

Στην περίπτωση του MTHFR, η μεγιστοποίηση των δεδομένων ξεκίνησε με τιμές $\theta = 0.06, \mu_y = 3.33,\ \tau_x = 0.35,\ \tau_y = 7.75, \rho = 0.51$ και συγκλίνει στις τιμές που αναφέρονται στον Πίνακα 2.

Πίνακας 2. Αποτελέσματα των μοντέλων A και B, με χρήση της προσέγγισης Maximum Likelihood.

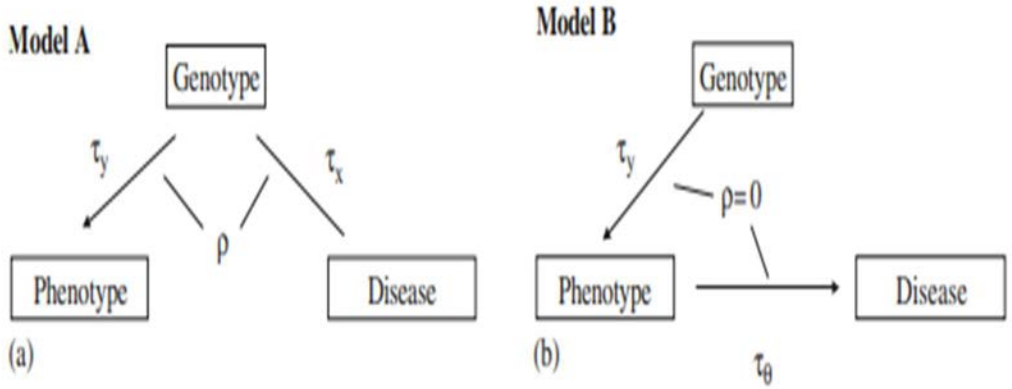
Model	*Parameter*	*Estimate*	*95 per cent CI*
A	θ	0.085	0.038–0.132
	μ_y	2.678	1.989–3.366
	τ_x	0.127	0.056–0.287

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

	τ_y	3.083	1.513–6.281
	ρ^{*}	1.000	—
B	θ	0.087	0.033–0.141
	μ_y	2.714	1.932–3.495
	τ_x	0.009	0.002–0.035
	τ_y	3.370	1.589–7.148

Συμπερασματικά, το αποτέλεσμα της αύξησης της ομοκυστεΐνης κατά 5 μmol/l στην CHD, αντιστοιχεί σε OR 1,53 (95% CI: 1,21–1,93). Η συσχέτιση μεταξύ της μελέτης εκτιμάται ελάχιστα, και η πιθανότητα της συσχέτισης αυξάνεται με μέγιστο στο ρ=1.

Σύμφωνα με το μοντέλο Α, η συνολική εκτίμηση του OR εξαρτάται από τη συσχέτιση μεταξύ της μελέτης, η οποία εκτιμάται ελάχιστα. Αυτό αποδείχθηκε από αναλύσεις ευαισθησίας που προσδιόρισαν το μοντέλο με μια σειρά από διαφορετικές γνωστές τιμές για τη συσχέτιση. Τα αποτελέσματα διέφεραν κατά OR 1,42 (95% CI: 1,09–1,85), που λήφθηκε όταν η συσχέτιση μεταξύ της μελέτης ήταν 0 έως 1,53 (95% CI: 1,21–1,93) όταν η συσχέτιση ήταν 1. Συνήθως, είναι δύσκολο να χρησιμοποιηθεί το μοντέλο Α επειδή τα αποτελέσματα εξαρτώνται από τη συσχέτιση μεταξύ της μελέτης, και είναι απίθανο να υπάρχουν επαρκείς πληροφορίες για την ακριβή εκτίμησή της.



Γράφημα 9. (α) Μοντελοποίηση των ετερογενειών των συσχετισμών στο Μοντέλο Α. (β) Μοντελοποίηση των ετερογενειών των συσχετισμών στο Μοντέλο Β.

2.3.2 Μοντέλα Β ML/REML

Ένα δεύτερο μοντέλο, που αναφέρεται ως μοντέλο Β, ξεπερνά τους περιορισμούς του μοντέλου Α μοντελοποιώντας τις ετερογένειες των συσχετίσεων με έναν εναλλακτικό τρόπο. Τα τρία στάδια στο τρίγωνο που φαίνεται στον παραπάνω Γράφημα 8(β), δηλαδή γονότυπος-φαινότυπος, φαινότυπος-ασθένεια και γονότυπος-ασθένεια, υπόκεινται όλα σε ετερογένεια. Σύμφωνα με το αιτιολογικό μοντέλο που υπονοείται σε αυτό το σκεπτικό, οποιαδήποτε σχέση αντλείται από τις άλλες δύο. Στο μοντέλο Β παραμετροποιούνται οι ετερογένειες στα στάδια γονότυπος-φαινότυπος και φαινότυπος-νόσος και θεωρούνται ανεξάρτητες, αλλά θα εξακολουθήσει να υπάρχει συσχέτιση στις προκύπτουσες ετερογένειες σε γονότυπο-φαινότυπο και γονότυπο-ασθένεια. Η ανεξαρτησία αυτή, που

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

θεωρείται στις ετερογένειες, σημαίνει ότι οι μελέτες αυτές και μία μεγάλη επίδραση του γονότυπου στο φαινότυπο, δε θα έχουν απαραίτητα την τάση να παρουσιάζουν σχετικά μεγαλύτερα ή μικρότερα αποτελέσματα του φαινότυπου στη νόσο (Jennrich κ.α., 1986). Στην πραγματικότητα, είναι απίθανο η επίδραση ενός συγκεκριμένου επιπέδου φαινότυπου στη νόσο να εξαρτάται από την αιτία αυτού του επιπέδου (π.χ. γονότυπος και όχι οποιαδήποτε άλλη αιτία). Αυτό είναι σχεδόν σίγουρα λογικό στην περίπτωση του MTHFR, της ομοκυστεΐνης και της CHD, και είναι πολύ πιθανό να ισχύει στις περισσότερες από τις άλλες περιπτώσεις.

Υποθέτοντας διμεταβλητές κανονικές κατανομές έχουμε ένα ιεραρχικό Μοντέλο, στο χαμηλότερο επίπεδο, με μέσους όρους μ_{xi} και μ_{yi}, διακυμάνσεις v_{xi} και v_{yi} και 0 συσχέτιση εντός της μελέτης. Θεωρείται ότι το θ_i και το μ_{yi} αντιπροσωπεύουν το logOR του φαινότυπου στη νόσο και τη μέση μεταβολή του φαινότυπου στην ι-οστή μελέτη, και ότι οι αντίστοιχοι μέσοι πληθυσμού είναι θ και μ_y (Jennrich κ.α., 1986). Εάν οι διακυμάνσεις μεταξύ των μελετών είναι τ_θ και τ_y, αντίστοιχα, και υπάρχει μηδενική συσχέτιση μεταξύ των μελετών, τότε εφόσον $\mu_{xi} = \theta_i \mu_{yi}$, εφαρμόζεται μια προσέγγιση της σειράς Taylor (μέθοδος δέλτα), για να εξαχθεί ο μέσος όρος και οι διακυμάνσεις των δύο μετρήσεων που παρατηρήθηκαν (γονότυπος-φαινότυπος και γονότυπος-ασθένεια). Αν προσεγγιστεί η κατανομή τους με μια διμεταβλητή κανονική προκύπτει:

$$\begin{bmatrix}\mu_{xi}\\ \mu_{yi}\end{bmatrix} \sim MVN\left(\begin{bmatrix}\theta\mu_y\\ \mu_y\end{bmatrix}, \begin{bmatrix}\mu_y^2\tau_\theta + \theta^2\tau_y & \theta\tau_y\\ \theta\tau_y & \tau_y\end{bmatrix}\right) \tag{2.10}$$

Και ως εκ τούτου

$$\begin{bmatrix}x_i\\ y_i\end{bmatrix} \sim MVN\left(\begin{bmatrix}\theta\mu_y\\ \mu_y\end{bmatrix}, \begin{bmatrix}v_{xi} + \mu_y^2\tau_\theta + \theta^2\tau_y & \theta\tau_y\\ \theta\tau_y & v_{yi} + \tau_y\end{bmatrix}\right) \tag{2.11}$$

Οι μελέτες θεωρούνται ανεξάρτητες, άρα η πιθανότητα μπορεί να ληφθεί και να μεγιστοποιηθεί ώστε να παρέχονται εκτιμήσεις μέγιστης πιθανοφάνειας για όλες τις παραμέτρους. Ωστόσο, είναι πιο σύνηθες να υπολογίζονται οι τιμές του πίνακα διακύμανσης μεταξύ των μελετών χρησιμοποιώντας την μέθοδο REML. Αυτή η εκτίμηση μπορεί εύκολα να πραγματοποιηθεί μεγιστοποιώντας μια ειδική περίπτωση της έκφρασης REML που δίνεται από τους Jennrich και Schluchter (1986), (με την επιφύλαξη των περιορισμών ότι ο πίνακας διακύμανσης μεταξύ των μελετών είναι θετικός ημι-καθορισμένος). Συγκεκριμένα η έκφραση για μεγιστοποίηση είναι:

$$\lambda_{REML} = -\frac{1}{2}\sum_{i=1}^{n}\log|\Sigma + \Delta_i| - \frac{1}{2}\log\left|\sum_{i=1}^{n}(\Sigma + \Delta_i)^{-1}\right| - \frac{1}{2}\sum_{i=1}^{n} r_i^T(\Sigma + \Delta_i)^{-1} r_i \tag{2.12}$$

Όπου Δi είναι ο πίνακας διακύμανσης εντός της μελέτης της i^{th}μελέτης

$$\Delta_i = \begin{pmatrix}\sigma_{x_i}^2 & \rho_i\sigma_{x_i}\sigma_{y_i}\\ \rho_i\sigma_{x_i}\sigma_{y_i} & \sigma_{y_i}^2\end{pmatrix} \tag{2.13}$$

Το Σ υποδηλώνει τον πίνακα διακύμανσης μεταξύ των μελετών

$$\sum\begin{pmatrix}\tau_x^2 & \kappa\tau_x\tau_y\\ \kappa\tau_x\tau_y & \tau_y^2\end{pmatrix} \tag{2.14}$$

Όπως και στο μοντέλο Α, όταν παρατηρείται μόνο ένα από τα ζευγάρια, το θ θεωρείται ως μονομεταβλητό. Ως εκ τούτου το x_i κατανέμεται κανονικά με μέσο $\theta\mu_y$ και η διακύμανση

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

$v_{xi} + y^2\tau_\theta + \theta^2\tau_y$, ή το y_i κατανέμεται κανονικά με το μέσο μ_y και η διακύμανση $v_{yi} + \tau_y$. Αυτό το νέο μοντέλο έχει τέσσερις παραμέτρους $(\theta, \mu_y, \tau_\theta, \tau_y)$ και η αριθμητική μεγιστοποίηση της πιθανοφάνειας βασίζεται σε μετασχηματισμένες τιμές παραμέτρων $(\theta, \mu_y, \log\ \tau_\theta, \log \tau_y)$. Για να ληφθούν οι αρχικές τιμές ισχύει ότι:

$$\theta = \frac{\bar{x}}{\bar{y}}, \qquad \mu_y = \bar{y}, \qquad \tau_x = Var(x), \qquad \tau_y = Var(y) \tag{2.15}$$

Για το παράδειγμα MTHFR–ομοκυστεΐνη–CHD, η μεγιστοποίηση ξεκινά από $\theta = 0.06, \mu_y = 3.33, \tau_\theta = 0.31, \tau_y = 7.75$ και συγκλίνει στις τιμές που αναφέρονται στον παραπάνω πίνακα. Οι ενδείξεις οδηγούν σε μια εκτίμηση της επαγόμενης συσχέτισης μεταξύ των πραγματικών μέσων μελετών που εξετάζουν τις σχέσεις φαινότυπου-γονότυπου και γονότυπου-ασθένειας, δηλ. στο ρ = 0,52. Το αποτέλεσμα του μοντέλου Β που εκφράζεται ως αποτέλεσμα μιας αύξησης της ομοκυστεΐνης κατά 5 μmol/l στη CHD είναι ένα OR 1,54 (95 % CI: 1,18– 2,02).

Μια ανάλυση ευαισθησίας που λαμβάνει υπόψη μια μικρή, γνωστή ποσότητα συσχέτισης εντός της μελέτης ανέδειξε ότι, με ένα $\psi = -0.1$, το OR για μια αύξηση 5 μmol/l στην ομοκυστεΐνη ήταν $1{,}56\ (95\%\ CI: 1{,}18 – 2{,}05)$, ενώ με ένα $\psi = +0{,}1$ το OR ήταν $1{,}53\ (95\%\ CI: 1{,}17 – 2{,}00)$. Συνεπώς, ακόμη και αν υπάρχει μικρή συσχέτιση εντός της μελέτης, αυτή έχει περιορισμένη επίδραση στην τελική εκτίμηση.

Η κύρια διαφορά μεταξύ των μοντέλων Α και Β είναι ότι το μοντέλο Β υποθέτει ότι η ετερογένεια στη μέση διαφορά του φαινότυπου με τον γονότυπο είναι ανεξάρτητη από την ετερογένεια στο logOR του φαινότυπου της νόσου. Σύμφωνα με το μοντέλο Β η γραφική παράσταση της πραγματικής μελέτης ή του γονότυπου σε ασθένεια για κάθε μελέτη, $\mu_{xi} = \theta_i \mu_{yi}$, θα παρήγαγε σημεία που θα βρίσκονταν σε μια ευθεία γραμμή με κλίση θ (που διέρχεται από την αρχή των αξόνων), και η συσχέτιση θα ήταν περίπου $\theta\sqrt{\frac{\tau_y}{\tau_x}}$. Χρησιμοποιώντας τις εκτιμήσεις παραμέτρων από την προσαρμογή του μοντέλου Α, αυτή η συσχέτιση θα πρέπει να είναι περίπου $0.085 * \sqrt{\frac{3.083}{0.127}} = 0.42$, παρόμοια με την εκτιμώμενη συσχέτιση κατά την προσαρμογή του μοντέλου Β. Το γεγονός ότι το μοντέλο Α αναζητά μια απίθανη λύση συσχέτισης οφείλεται στην έλλειψη δεδομένων, επιπλέον υποθέσεων και της ανίχνευσης κάποιας μη γραμμικότητας. Η μεγιστοποιημένη πιθανότητα καταγραφής για το μοντέλο Α είναι -37,3, ενώ για το μοντέλο Β είναι -39,2. Καθώς υπάρχει μία επιπλέον παράμετρος στο μοντέλο Α, η επιλογή του μοντέλου δεν είναι ξεκάθαρη, αλλά προτιμάται το μοντέλο Β, καθώς συμφωνεί περισσότερο με το μοτίβο της παραλλαγής μεταξύ των μελετών.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Κεφάλαιο 3

Αποτελέσματα

Το Stata είναι ένα στατιστικό πακέτο που δημιουργήθηκε στις αρχές του 1985 από την Statacorp. Χρησιμοποιείται από πολλές επιχειρήσεις και ακαδημαϊκά ιδρύματα σε όλο τον κόσμο αλλά κυρίως χρησιμοποιείται σε κλάδους της οικονομικής επιστήμης. Περιλαμβάνει επιπλέον ένα ευρύ φάσμα από εντολές για την παλινδρόμηση, για την ανάλυση δεδομένων, χρονοσειρών, κ.α.. Δίνει τη δυνατότητα εμφάνισης γραφημάτων, όπως των δενδροδιαγραμμάτων και των γραφημάτων διασποράς, τα οποία εμφανίζονται και στην παρούσα εργασία. Ένα δυνατό του πλεονέκτημα, μάλιστα, είναι η ταχύτητα και η ευκολία της χρήσης του. Σύμφωνα με την StataCorp (2016), το Stata είναι «ένα πλήρες, ολοκληρωμένο στατιστικό πακέτο λογισμικού που παρέχει όλα όσα είναι απαραίτητα για ανάλυση δεδομένων, διαχείριση δεδομένων και γραφικών». Συνοψίζοντας, το Stata είναι ένα λογισμικό που επιτρέπει την αποθήκευση και τη διαχείριση δεδομένων (για μεγάλα και μικρά σύνολα δεδομένων), τη στατιστική ανάλυση τους και την σχεδίαση γραφημάτων (Επίσημη ιστοσελίδα StataCorp).

Συγκεκριμένα, στον τομέα της υγείας το Stata χρησιμοποιείται όταν απαιτείται η διαχείριση πολύ μεγάλων σετ δεδομένων. Οι τρείς πιο βασικές εφαρμογές του στον τομέα αυτόν είναι η ανάλυση επιπτώσεων θεραπείας, με χρήση πολλαπλών παλινδρομήσεων με στόχο την αξιολόγηση της επίδρασης της θεραπείας σε διάφορες παραμέτρους υγείας. Επιπλέον, η μελέτη συσχέτισης παραγόντων κινδύνου με χρήση μοντέλων logit/probit για την εκτίμηση του κινδύνου παθήσεων με βάση διάφορους παράγοντες. Τέλος, η ανάλυση επιπτώσεων ενός προγράμματος υγείας με εφαρμογή γραφημάτων και παλινδρομήσεων για να αξιολογηθούν οι αλλαγές που προκαλεί ένα πρόγραμμα υγείας σε έναν πληθυσμό (Επίσημη ιστοσελίδα StataCorp).

3.1 Πρόγραμμα

3.1.1 Αρχείο mrmeta

```
*****************************************************************
************
* Meta-Analysis of Genetic Association Studies using Mendelian
Randomization
*****************************************************************
************

program define mrmeta, rclass
version 10.0

  syntax varlist(min=4 max=4 numeric) [if] [in], model(string)
method(string) [mvmeta] [study(string)] [plot(string)]
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
set more off
program drop _all
parse "`varlist'", parse(" ")


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


if "`model'" == "b" & "`mvmeta'" ~= "" {
     di in gr ""
     di as error "Please select method ML or REML to perform
Mendelian Randomization with Model B"
}

else {

tempvar z varz y vary
qui gen `z'=`1'            //z  =  logOddsRatio  genotype-disease
association
qui gen `varz'=`2'
qui gen `y'=`3'          //y = meanDifference genotype-phenotype
association
qui gen `vary'=`4'

tempvar wz wy
qui gen `wz' = 1/`varz'  //weight of z
qui gen `wy' = 1/`vary'  //weight of y

//Variables for LL.ado
qui gen _y= `y'
qui gen _z= `z'
qui gen _varz=`varz'
qui gen _vary=`vary'


//Models ----
local a 0
local b 0


//Methods ----
qui gen method=0
qui replace method=1 if ("`model'" == "a" & "`method'" == "ml")
qui replace method=2 if ("`model'" == "a" & "`method'" == "reml")
qui replace method=3 if ("`model'" == "b" & "`method'" == "ml")
qui replace method=4 if ("`model'" == "b" & "`method'" == "reml")


tempvar id
qui egen `id' = fill(1 2)
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
qui replace `id'=. if  `1'==. |`2'==.


//MVMETA ----

if "`model'" == "a" & "`mvmeta'" ~= "" {

if "`model'" == "a" & "`method'" == "ml" & "`mvmeta'" ~= "" {
      local a=1
      di in gr ""
      di in gr "Performing Mendelian Randomazition with " in ye
"Model A " in gr " method " in ye "ML " in gr "using " in ye "mvmeta
"
      di in gr ""
      }
if "`model'" == "a" & "`method'" == "reml" & "`mvmeta'" ~= "" {
      local a=1
      di in gr ""
      di in gr "Performing Mendelian Randomazition with " in ye
"Model A " in gr " method " in ye "REML " in gr "using " in ye
"mvmeta "
      di in gr ""
      }


************************************************
* MVMETA FOR MODEL A
************************************************

qui gen b1=`z'
qui gen b2=`y'

qui gen V11=`varz'
qui gen V22=`vary'


if "`model'" == "a" & "`method'" == "ml" & "`mvmeta'" ~= "" {
mvmeta b V, vars(b1 b2) ml wscorr(0)
}

if "`model'" == "a" & "`method'" == "reml" & "`mvmeta'" ~= "" {
mvmeta b V, vars(b1 b2) reml wscorr(0)
}

drop b1 b2 V11 V22


matrix b=e(b)
matrix v=e(V)

qui local oddsratio=b[1,1]  // Pooled log(or)
qui local meandiff=b[1,2]   // Pooled MD

//Mean difference CI 95%
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
qui local se_est = sqrt(v[2,2])
qui local lb_est = `meandiff' - 1.96 * `se_est'
qui local ub_est = `meandiff' + 1.96 * `se_est'

//Log Odds Ratio CI 95%
qui local see = sqrt(v[1,1])
qui local cilow = `oddsratio' - 1.96 * `see'
qui local ciup = `oddsratio' + 1.96 * `see'


//NLCOM ----
nlcom divisionLogORMeanDiff: _b[b1]/_b[b2]


//Results
di ""
di in gr "Pooled log odds ratio= " in ye %7.4f `oddsratio' in gr "
and its standard error se=" in ye %7.4f `see'
di "95% Conf. Interval= [" in ye %7.4f `cilow' , %7.4f `ciup' in
gr "]"


}

// ML and REML methods ----

      else{

if "`model'" == "a" & "`method'" == "ml" {
      local a=1
      di in gr ""
      di in gr "Performing Mendelian Randomazition with " in ye
"Model A " in gr "and method " in ye "ML "
      }
if "`model'" == "a" & "`method'" == "reml" {
      local a=1
      di in gr ""
      di in gr "Performing Mendelian Randomazition with " in ye
"Model A " in gr "and method " in ye "REML "
      }
if "`model'" == "b" & "`method'" == "ml" {
      local b=1
      di in gr ""
      di in gr "Performing Mendelian Randomazition with " in ye
"Model B " in gr "and method " in ye "ML"
      }
if "`model'" == "b" & "`method'" == "reml" {
      local b=1
      di in gr ""
      di in gr "Performing Mendelian Randomazition with " in ye
"Model B " in gr "and method " in ye "REML "
      }
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
*------------------------------
* gpd denotes the available data
qui gen gpd=1
qui replace gpd=2 if `z' ==. & `y' ~=.
qui replace gpd=3 if `y' ==. & `z' ~=.


************************************************
* LIKELIHOOD ANALYSIS FOR MODEL A
************************************************

if  "`model'" == "a" {

*------------------------------
* Initial Guess

tempvar tauz ratio mu tauy

summarize `z'
local tauz=r(Var)
local ratio=r(mean)
summarize `y'
local mu=r(mean)
local tauy=r(Var)
local ratio=`ratio'/`mu'
corr `z' `y'
local r=r(rho)
di "Initial Guess "%7.3f `ratio' %7.3f `mu' %7.3f `tauz' %7.3f
`tauy' %7.3f `r'
local tauz=log(`tauz')
local tauy=log(`tauy')
local r=log((1+`r')/(1-`r'))


//Maximum Likelihood Estimation---------

ml model lf LL (ratio:) (mu:) (tauz:) (tauy:) (r:)
ml   init   ratio:_cons=`ratio'   mu:_cons=`mu'   tauz:_cons=`tauz'
tauy:_cons=`tauy' r:_cons=`r'
if "`method'" == "ml" {
qui ml check
}
qui ml search

}


************************************************
* LIKELIHOOD ANALYSIS FOR MODEL B
************************************************

else if "`model'" == "b" {

*------------------------------
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
* Initial Guess

tempvar _wz1 _wz2 _z1wz1 _z2wz2
qui gen `_wz1'=1/`varz'
qui gen `_wz2'=1/`vary'
qui gen `_z1wz1'=`z'*`_wz1'
qui gen `_z2wz2'=`y'*`_wz2'
summarize `_z1wz1'

tempvar muz1 swz1 muz tauz muz2 swz2 mu tauy ratio

local muz1=r(sum)
summarize `_wz1'
local swz1=r(sum)
local muz=`muz1'/`swz1'
local tauz=1/`swz1'

summarize `_z2wz2'
local muz2=r(sum)

summarize `_wz2'
local swz2=r(sum)
local mu=`muz2'/`swz2'
local tauy=1/`swz2'
local ratio=`muz1'/`muz2'

drop `_wz1' `_wz2' `_z1wz1' `_z2wz2'

local tauz=log(`tauz')

tempvar theta taut
qui gen `theta'=`z'/`y'
summarize `theta'
local taut=r(Var)
di "Initial Guess "%7.3f `ratio' %7.3f `mu' %7.3f `taut' %7.3f
`tauy'
local tauz=log(`tauz')
local tauy=log(`tauy')

drop `theta'

program drop _all


//Maximum Likelihood Estimation---------

ml model lf LL (ratio:) (mu:) (taut:) (tauy:)
ml   init   ratio:_cons=`ratio'   mu:_cons=`mu'   taut:_cons=`taut'
tauy:_cons=`tauy'
qui ml check
qui ml search

 }
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
ml maximize, difficult iter(30)

matrix b=e(b)
matrix v=e(V)

qui local meandiff=b[1,2]   // Pooled MD

//Mean difference CI 95%
qui local se_est = sqrt(v[2,2])
qui local lb_est = `meandiff' - 1.96 * `se_est'
qui local ub_est = `meandiff' + 1.96 * `se_est'


**********************NLCOM************************

nlcom LogOddsRatio: _b[ratio:_cons]*_b[mu:_cons]

matrix bnl=r(b)
qui local oddsratio = bnl[1,1]  // Pooled log(or)


// Results
local mpool=(b[1,1]*b[1,2])
local       see=sqrt(b[1,1]^2*v[2,2]       +       b[1,2]^2*v[1,1]
+2*v[1,2]*b[1,2]*b[1,1])
local     cilow=(b[1,1]*b[1,2]-1.96*     sqrt(b[1,1]^2*v[2,2]     +
b[1,2]^2*v[1,1] +2*v[1,2]*b[1,2]*b[1,1]))
local     ciup=(b[1,1]*b[1,2]+1.96*     sqrt(b[1,1]^2*v[2,2]     +
b[1,2]^2*v[1,1] +2*v[1,2]*b[1,2]*b[1,1]))

di ""
di in gr "Pooled log odds ratio= " in ye %7.4f `mpool' in gr " and
its standard error se=" in ye %7.4f `see'
di "95% Conf. Interval= [" in ye %7.4f `cilow' , %7.4f `ciup' in
gr "]"

drop gpd

}


**********************FOREST PLOTS**********************

tempvar zi yi sezi seyi
qui gen `zi' = `z'
qui gen `yi' = `y'
qui gen `sezi'=sqrt(`varz')
qui gen `seyi'=sqrt(`vary')


quietly {
replace `zi'=0 if `zi' ==.
replace `yi'=0 if `yi' ==.
replace `sezi'=0 if `sezi' ==.
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
replace `seyi'=0 if `seyi' ==.
}

capture confirm variable y
if (_rc == 0) {
   rename y psiy
}


if "`study'" != "" {

qui    metagraph    `zi'    `sezi',    id(`study')    mscale(0.22)
combined(`oddsratio'   `cilow'   `ciup')   title({bf:Genotype   -
Disease})  xtitle({bf:LogOddsRatio})  nodraw  saving(plotGenDis,
replace)
qui    metagraph    `yi'    `seyi',    id(`study')    mscale(0.22)
combined(`meandiff'   `lb_est'   `ub_est')   title({bf:Genotype   -
Phenotype})       xtitle({bf:Mean       difference})       nodraw
saving(plotGenPhen, replace)

qui   graph   combine   plotGenDis.gph   plotGenPhen.gph,   col(2)
title("Two-Column Forest Plot") nodraw saving(twocolumnForestPlot,
replace)

}

else {

qui    metagraph    `zi'    `sezi',    id(`id')    mscale(0.22)
combined(`oddsratio'   `cilow'   `ciup')   title({bf:Genotype   -
Disease})  xtitle({bf:LogOddsRatio})  nodraw  saving(plotGenDis,
replace)
qui    metagraph    `yi'    `seyi',    id(`id')    mscale(0.22)
combined(`meandiff'   `lb_est'   `ub_est')   title({bf:Genotype   -
Phenotype})       xtitle({bf:Mean       difference})       nodraw
saving(plotGenPhen, replace)

qui   graph   combine   plotGenDis.gph   plotGenPhen.gph,   col(2)
title("Two-Column Forest Plot") nodraw saving(TwocolumnForestPlot,
replace)

}

capture confirm variable psiy
if (_rc == 0) {
   rename psiy y
}


******************** SCATTER PLOT**********************

if "`plot'" == "scatter" {

qui twoway scatter `z' `y', title("Genotype - Phenotype - Disease")
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
ytitle("Log   Odds   ratio   (logoddsratio)")   ylabel(,   grid)
xtitle("Mean    difference    (meandiff)")    xlabel(,    grid)
saving(ScatterPlot, replace)

}

********************* BUBBLE PLOT**********************

if "`plot'" == "bubble" {

qui twoway scatter `z' `y' [aweight=`wz'], symbol(Oh) mlstyle(dot)
title("Genotype - Phenotype - Disease") ytitle("Log Odds ratio
(logoddsratio)")   ylabel(,   grid)   xtitle("Mean   difference
(meandiff)") xlabel(, grid) saving(BubblePlot, replace)

}


graph use TwocolumnForestPlot.gph



drop method _y _z _vary _varz

}

end
```

3.1.2 Αρχείο LL

```
********************************
* Program to evaluate LogL
********************************

program define LL


************************************************
* Model A and method ML
************************************************

if method==1 {

      args lnl ratio muy tz ty r

      tempname tauz tauy rho muz cov
    tempvar vz vy L3 L2 L1 fis


      quietly {

            scalar `tauz'=exp(`tz')
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
            scalar `tauy'=exp(`ty')
            scalar `rho'=(exp(`r')-1)/(exp(`r')+1)
            scalar `muz'=`muy'*`ratio'

            gen double `vz'=`tauz'+ _varz
            gen double `vy'=`tauy'+ _vary

            scalar `cov'=`rho'*sqrt(`tauz'*`tauy')

            gen double `fis'=`cov'/sqrt(`vz'*`vy')

      gen double `L3'=log(`vz')+(_z-`muz')^2/`vz'
            gen double `L2'=log(`vy')+(_y-`muy')^2/`vy'

            gen  double  `L1'=log(`vz'*`vy'*(1-`fis'*`fis'))+((_z-
`muz')^2/`vz'- /*
            */   2*`fis'*(_z-`muz')*(_y-`muy')/sqrt(`vz'*`vy')+(_y-
`muy')^2/`vy')/(1-`fis'*`fis')
            replace `lnl'=-0.5*`L1' if gpd==1
            replace `lnl'=-0.5*`L2' if gpd==2
            replace `lnl'=-0.5*`L3' if gpd==3

            drop `vz' `vy' `L1' `L2' `L3' `fis'

            }

      }



**************************************************
* Model A and method REML
**************************************************

else if method==2 {

      args lnl ratio muy tz ty r

      tempvar tauz tauy rho muz vz vy cov det ro

      quietly {

            gen double `tauz'=exp(`tz')
            gen double `tauy'=exp(`ty')
            gen double `rho'=(exp(`r')-1)/(exp(`r')+1)
            gen double `muz'=`muy'*`ratio'
            gen double `vz'=`tauz'+ _varz
            gen double `vy'=`tauy'+ _vary
            gen double `cov'=`rho'*sqrt(`tauz'*`tauy')
         gen double `det'=`vy'*`vz'-`cov'^2
            gen double `ro'=`cov'/sqrt(`vz'*`vy')

            tempvar L3 n3 tmp xSx L2 n2 L1 n1 xSx11 xSx22 xSx12
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
           gen double `L3' = log(`vz') + (_z-`muz')^2 / `vz'
           count if gpd==3
        local n3=r(N)
        gen double `tmp'=1/`vz'
        sum `tmp' if gpd==3
        local xSx = r(sum)
        replace  `lnl'  =  -0.5  *  (`L3'+log(`xSx')/`n3')     -
0.5*log(2*_pi) if gpd == 3

        gen double `L2' = log(`vy') + (_y-`muy')^2 / `vy'
           count if gpd==2
        local n2=r(N)
        replace `tmp'=1/`vy'
        sum `tmp' if gpd==2
        local xSx = r(sum)
        replace  `lnl'  =  -0.5  *  (`L2'+log(`xSx')/`n2')   -
0.5*log(2*_pi)  if gpd == 2

     gen      double      `L1'=log(`vz'*`vy'*(1-`ro'*`ro'))+((_z-
`muz')^2/`vz'- /*
           */   2*`ro'*(_z-`muz')*(_y-`muy')/sqrt(`vz'*`vy')+(_y-
`muy')^2/`vy')/(1-`ro'*`ro')

        count if gpd==1
        local n1=r(N)
        replace `tmp'=`vy'/`det'
        sum `tmp' if gpd==1
        local xSx11 = r(sum)
        replace `tmp'=`vz'/`det'
        sum `tmp' if gpd==1
        local xSx22 = r(sum)
        replace `tmp'=`cov'/`det'
        sum `tmp' if gpd==1
        local xSx12 = r(sum)
        replace  `lnl'  =  -0.5  *  (`L1'+log(`xSx11'*`xSx22'-
`xSx12'^2)/`n1') - 0.5*log(2*_pi) if gpd == 1

           }

     }



**************************************************
* Model B and method ML
**************************************************

else if method==3 {

args lnl ratio muy tt ty

     tempname taut tauy muz
    tempvar vz vy cov r L3 L2 L1
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
      quietly {

            scalar `taut'=exp(`tt')
            scalar `tauy'=exp(`ty')

            scalar `muz'=`muy'*`ratio'

            gen double `vz'=`ratio'^2*`tauy'+`muy'^2*`taut'+ _varz

            gen double `vy'=`tauy'+ _vary
            gen double `cov'=`ratio'*`tauy'

            gen double `r'=`cov'/sqrt(`vz'*`vy')
            gen double `L3'=log(`vz')+(_z-`muz')^2/`vz'
            gen double `L2'=log(`vy')+(_y-`muy')^2/`vy'
            gen   double   `L1'=log(`vz'*`vy'*(1-`r'*`r'))+(   (_z-
`muz')^2/`vz'- /*
                 */                          2*`r'*(_z-`muz')*(_y-
`muy')/sqrt(`vz'*`vy')+(_y-`muy')^2/`vy' )/(1-`r'*`r')
            replace `lnl'=-0.5*`L1' if gpd == 1
            replace `lnl'=-0.5*`L2' if gpd == 2
            replace `lnl'=-0.5*`L3' if gpd == 3
            drop `vz' `vy' `r' `cov' `L1' `L2' `L3'
            }

      }



**************************************************
* Model B and method REML
**************************************************

else if method==4 {

args lnl ratio muy tt ty

    tempvar taut tauy muz vz vy cov det ro

      quietly {

            gen double `taut'=exp(`tt')
            gen double `tauy'=exp(`ty')
            gen double `muz'=`muy'*`ratio'
            gen double `vz'=`ratio'^2*`tauy'+`muy'^2*`taut'+ _varz
            gen double `vy'=`tauy'+ _vary
            gen double `cov'=`ratio'*`tauy'
            gen double `det'=`vy'*`vz'-`cov'^2
            gen double `ro'=`cov'/sqrt(`vz'*`vy')

            tempvar L3 n3 tmp xSx L2 n2 L1 n1 xSs11 xSx22 xSx12

            gen double `L3' = log(`vz') + (_z-`muz')^2 / `vz'
         count if gpd==3
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

```
        local n3=r(N)
        gen double `tmp'=1/`vz'
        sum `tmp' if gpd==3
        local xSx = r(sum)
        replace  `lnl'  =  -0.5  *  (`L3'+log(`xSx')/`n3')     -
0.5*log(2*_pi) if gpd == 3

        gen double `L2' = log(`vy') + (_y-`muy')^2 / `vy'
        count if gpd==2
        local n2=r(N)
        replace `tmp'=1/`vy'
        sum `tmp' if gpd==2
        local xSx = r(sum)
        replace  `lnl'  =  -0.5  *  (`L2'+log(`xSx')/`n2')   -
0.5*log(2*_pi)  if gpd == 2

     gen      double      `L1'=log(`vz'*`vy'*(1-`ro'*`ro'))+((_z-
`muz')^2/`vz'- /*
          */   2*`ro'*(_z-`muz')*(_y-`muy')/sqrt(`vz'*`vy')+(_y-
`muy')^2/`vy')/(1-`ro'*`ro')

        count if gpd==1
        local n1=r(N)
        replace `tmp'=`vy'/`det'
        sum `tmp' if gpd==1
        local xSx11 = r(sum)
        replace `tmp'=`vz'/`det'
        sum `tmp' if gpd==1
        local xSx22 = r(sum)
        replace `tmp'=`cov'/`det'
        sum `tmp' if gpd==1
        local xSx12 = r(sum)
        replace  `lnl'  =  -0.5  *  (`L1'+log(`xSx11'*`xSx22'-
`xSx12'^2)/`n1') - 0.5*log(2*_pi) if gpd == 1
          }

     }

end
```

3.1.3 Περιγραφή

Τα παραπάνω αρχεία υλοποιηθήκαν στο Stata 13.1, με ονομασίες mrmeta και LL. Το κύριο αρχείο είναι το mrmeta, στο οποίο αρχικά ορίζονται οι επιλογές που πρέπει να καθορίσει ο χρήστης, με σκοπό την εκτέλεση του προγράμματος. Η σύνταξή του είναι:

```
syntax varlist(min=4 max=4 numeric) [if] [in], model(string) /*
   *\ method(string) [mvmeta] [study(string)] [plot(string)]
```

Πιο συγκεκριμένα, μετά το όνομα του προγράμματος, ο χρήστης καλείται να καθορίσει από το αρχείο των δεδομένων προς επεξεργασία, το λογάριθμο του λόγου σχετικών πιθανοτήτων του γονότυπου στην ασθένεια (logOR ως πρώτη επιλογή) και τη διακύμανσή

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

του (ως δεύτερη). Τις επόμενες δύο επιλογές αποτελούν η μέση διαφορά στο φαινότυπο (ΔΡ) και η διακύμανσή της, αντίστοιχα. Με το `model(string)` καθορίζεται ποιο από τα δύο μοντέλα a ή b, που περιγράφηκαν στο δεύτερο κεφάλαιο, επιθυμεί ο χρήστης να χρησιμοποιήσει. Το `method(string)` δίνει την επιλογή του τύπου του μοντέλου, ML (ml) ή REML (reml). Στη συνέχεια, δίνονται κάποιες επιπλέον προαιρετικές επιλογές όπως, το `[mvmeta]`, σε περίπτωση που ο χρήστης επιθυμεί να εξάγει τα αποτελέσματά του από την ομώνυμη εντολή του Stata, το οποίο όμως μπορεί να εφαρμοστεί επιλέγοντας μόνο το μοντέλο Α. Στην περίπτωση που ο χρήστης επιλέξει να τρέξει το μοντέλο Β ML ή Β REML με mvmeta, εμφανίζεται μήνυμα σφάλματος, καθώς αυτή η επιλογή δεν είναι δυνατή. Μία άλλη τέτοια επιλογή είναι το `[study(string)]`, το οποίο λαμβάνει ως μεταβλητή εκείνη που περιέχει τον αριθμό ταυτοποίησης ή τις ονομασίες των μελετών, προκειμένου να εμφανιστούν στο διπλό δενδροδιάγραμμα του προγράμματος. Αν ο χρήστης δεν εισάγει την επιλογή αυτή, τότε το πρόγραμμα δημιουργεί τη μεταβλητή 'id', η οποία δημιουργεί τους αριθμούς ταυτοποίησης των μελετών. Τέλος με το `[plot(string)]` ο χρήστης μπορεί να επιλέξει την εμφάνιση ενός διαγράμματος φυσαλίδων `plot(bubble)` ή διασποράς `plot(scatter)`. Και στα δύο αυτά γραφήματα εμφανίζονται μόνο οι μελέτες που έχουν αξιολογήσει και τις δύο σχέσεις γονότυπος-νόσος και γονότυπος-φαινότυπος. Η διαφορά τους είναι ότι το πρώτο (διάγραμμα φυσαλίδων) εμφανίζει τις μελέτες αυτές σε φυσαλίδες, των οποίων το μέγεθος καθορίζει το βάρος της μελέτης. Όλα τα γραφήματα αποθηκεύονται αυτόματα στο φάκελο του χρήστη (path που έχει ορίσει) με τις ονομασίες TwocolumnForestPlot, BubblePlot και ScatterPlot, αντίστοιχα με τη μορφή αρχείου .gph.

Πριν προχωρήσει η εργασία στην ανάλυση του κώδικα, παραθέτονται τα δεδομένα που χρησιμοποιήθηκαν για τη δημιουργία του, καλώντας το χρήστη να παρέχει την αντίστοιχη δομή.

Πίνακας 3. Αρχικά δεδομένα προγράμματος mrmeta.

	id	z	varz	y	vary
1	1	-.82	.42	.4	10
2	2	-.48	.25	2.9	2.44
3	3	-.31	.14	1.3	.46
4	4	-.22	.1	-.8	.48
5	5	-.2	.03	.9	.19
6	6	-.17	.07	2	.36
7	7	-.11	.21	2.6	1.23
8	8	.07	.15	1.3	2.63
9	9	.19	.03	2.8	1.47
10	10	.25	.16	2.5	1.15
11	11	.28	.07	1.4	.47
12	12	.41	.32	8.8	14.29
13	13	.55	.1	3	.57
14	14	.73	.04	3.8	1.47
15	15	.84	.06	4.6	1.61
16	16	.96	.23	1.2	.71
17	17	1.14	.26	11	11.11
18	18	-1.61	.61	.	.
19	19	-.63	.28	.	.
20	20	-.37	.07	.	.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Στη συνέχεια, καταγράφεται η ροή του κώδικα ανάλογα με την κάθε επιλογή του χρήστη. Στην περίπτωση που ο χρήστης επιλέξει να τρέξει το μοντέλο A ML/REML με mvmeta, αποθηκεύονται τα logOR, η ΔP και οι διακυμάνσεις τους στις μεταβλητές b1, b2, V11 και V22, αντίστοιχα, προκειμένου να εισαχθούν στη σύνταξη της εντολής mvmeta. Το mvmeta, πιο συγκεκριμένα είναι μια εντολή στο Stata που χρησιμοποιείται για την εκτέλεση πολυμεταβλητής μετα-ανάλυσης (Επίσημη ιστοσελίδα Stata). Το αποτέλεσμα της εντολής αυτής είναι ο υπολογισμός των συγκεντρωτικών αποτελεσμάτων των logOR και ΔP, στην παρούσα εργασία. Μία άλλη εντολή του Stata που χρησιμοποιείται, είναι το nlcom, η οποία υπολογίζει μη γραμμικές συναρτήσεις των εκτιμήσεων των παραμέτρων από ένα στατιστικό μοντέλο (Επίσημη ιστοσελίδα Stata).

Σε οποιαδήποτε άλλη επιλογή του χρήστη, ο κώδικας συνεχίζει στην επεξεργασία των αρχικών τιμών ('initial guess'), δηλαδή των τιμών που καθορίζονται πριν ξεκινήσει η διαδικασία εκτίμησης και βελτιστοποίησης, από το μοντέλα A ή B, ML ή REML. Αυτή η αρχική υπόθεση βοηθά τον αλγόριθμο να ξεκινήσει την παραπάνω διαδικασία από ένα σημείο που θεωρείται πιθανό να είναι κοντά στην τελική βέλτιστη λύση. Μέσα από τις εντολές ml που εκτελούνται, περνά η λογαριθμική αρχική υπόθεση στο δεύτερο αρχείο της εργασίας, το LL, όπου υπολογίζεται το ML ή το REML. Η τελευταία εντολή ml maximize είναι αυτή η οποία παρέχει εκτιμήσεις μέγιστης πιθανοφάνειας για όλες τις παραμέτρους. Συγκεκριμένα, εμφανίζει σαν 'ratio' το λόγο των logOR ως προς τη ΔP και σαν 'mu' το συγκεντρωτικό αποτέλεσμα της ΔP στο φαινότυπο. Επομένως, το συγκεντρωτικό logOR στην ασθένεια θα υπολογιστεί αργότερα από την εντολή nlcom πολλαπλασιάζοντας το λόγο που περιέχει το 'ratio' με το 'mu'.

Τελικά, εμφανίζονται τα γραφήματα της εργασίας, τα οποία είναι ένα δενδροδιάγραμμα, με την εντολή metagraph, αντίστοιχο με το παραπάνω **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** και προαιρετικά ένα γράφημα φυσαλίδων ή διασποράς με τις αρχικές τιμές των z και y. Τα γραφήματα αυτά αποθηκεύονται αυτόματα στο φάκελο εργασίας του χρήστη και θα εξηγηθούν περαιτέρω στο επόμενο κεφάλαιο 'Εκτέλεση'.

3.2 Εκτέλεση

Για τον έλεγχο της σωστής λειτουργίας του κώδικα διασταυρώθηκαν τα αποτελέσματα του κάθε μοντέλου A και B, ML ή REML με τα αποτελέσματα που προκύπτουν από την εντολή mvmeta. Επίσης χρησιμοποιήθηκαν άλλα δύο σετ δεδομένων τα οποία εξετάζουν τη σχέση μεταξύ των συγκεντρώσεων της ομοκυστεΐνης στο αίμα και του κινδύνου για στεφανιαία καρδιακή νόσο (IHD) ή το εγκεφαλικό επεισόδιο.

Ακολουθεί η εκτέλεση του προγράμματος για κάθε μία από τις πιθανές επιλογές του χρήστη. Παρέχονται στιγμιότυπα των αποτελεσμάτων, που λήφθηκαν από το γραφικό περιβάλλον του Stata, καθώς και εξηγήσεις για κάθε μέρος των αποτελεσμάτων.

3.2.1 Παράδειγμα 1: Μοντέλο A - Μέθοδος REML

Η σύνταξη για την εκτέλεση του προγράμματος με μοντέλο το A, μέθοδο τη REML και με τους αριθμούς ταυτοποίησης των μελετών είναι η ακόλουθη:

```
mrmeta z varz y vary, model(a) method(reml) study(id)
```


Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Με τα δεδομένα που παραθέτονται παραπάνω στον **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**, οι αρχικές υποθέσεις που υπολογίζονται είναι οι εξής:

Πίνακας 4. Αρχικές βέλτιστες τιμές εκτίμησης του προγράμματος mrmeta με χρήση του μοντέλου A REML.

Initial Guess	0.060	3.285	0.338	7.916	0.603

Οι τιμές αυτές κατά σειρά είναι τα 'ratio' και 'mu', που αναφέρθηκαν παραπάνω, τα 'tauz' και 'tauy', τα οποία είναι οι διακυμάνσεις των logOR και ΔP αντίστοιχα. Τελικά το 'r' είναι μεταβλητή από το μετασχηματισμό του Fisher, ο οποίος εφαρμόστηκε για την ομαλή επεξεργασία των δεδομένων από τις εντολές ml.

Χρησιμοποιώντας, λοιπόν, τις τιμές αυτές ως βέλτιστες, το πρόγραμμα συνεχίζει στις εντολές ml, με την τελευταία ml maximize να εμφανίζει τα αποτελέσματα:

Πίνακας 5. Μεγιστοποιημένα αποτελέσματα του προγράμματος mrmeta με χρήση του μοντέλου A REML.

				Number of obs	=	64
				Wald chi2(0)	=	.
Log likelihood = -98.566159				Prob > chi2	=	.

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ratio						
_cons	.0866931	.0243699	3.56	0.000	.0389289	.1344572
mu						
_cons	2.655228	.3666451	7.24	0.000	1.936616	3.373839
tauz						
_cons	-2.099267	.4415943	-4.75	0.000	-2.964776	-1.233758
tauy						
_cons	1.248504	.3681493	3.39	0.001	.5269451	1.970064
r						
_cons	16.50275	1172.073	0.01	0.989	-2280.719	2313.724

Στον **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** εμφανίζεται ένα Log Likelihood ίσο με -98.566159, σε ένα σύνολο 64 παρατηρήσεων από τα δεδομένα. Η εκτίμηση του λόγου logOR προς ΔP (ratio) είναι 0.0866931, της διακύμανσής του tauz -2.099267, της ΔP είναι 2.655228 και της διακύμανσής της tauy είναι 1.248504. Δίπλα από τις εκτιμήσεις αυτές έχουν υπολογιστεί τα αντίστοιχα τυπικά σφάλματά τους και οι τιμές z της ανάλυσης. Όλες οι εκτιμήσεις, εκτός από την r, είναι στατιστικά σημαντικές με τις τιμές p-value κατά πολύ μικρότερες από το 0.05.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Ο υπολογισμός του συγκεντρωτικού logOR, όπως αναφέρθηκε παραπάνω γίνεται με την εντολή nlcom:

Πίνακας 6. Τελικό συγκεντρωτικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων του προγράμματος mrmeta με χρήση του μοντέλου A REML.

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LogOddsRatio	.2301898	.0705194	3.26	0.001	.0919742	.3684053

Pooled log odds ratio= 0.2302 and its standard error se= 0.0705
95% Conf. Interval= [0.0920 0.3684]

Η εκτίμηση του logOR υπολογίζεται στο 0.2301898, με τυπικό σφάλμα σχετικά μικρό 0.705194 και τιμή z = 3.26. Η τιμή αυτή είναι στατιστικά σημαντική με p-value = 0.001, κατά πολύ μικρότερη από το 0.05 και ένα διάστημα εμπιστοσύνης [0.0919742; 0.3684053] σχετικά μικρό.
Για τον έλεγχο της ορθότητας των αποτελεσμάτων θα ακολουθήσει η εκτέλεση της επιπλέον επιλογής της εντολής mvmeta, με τα ίδια δεδομένα. Η σύνταξη του προγράμματος είναι:

```
mrmeta z varz y vary, model(a) method(reml) mvmeta study(id)
```

Το αποτελέσματα που προκύπτουν είναι:

Πίνακας 7. Αποτελέσματα του λογαρίθμου του λόγου σχετικών πιθανοτήτων και της μέσης διαφοράς του προγράμματος mrmeta με χρήση του μοντέλου A REML και της εντολής mvmeta.

Multivariate meta-analysis
Variance-covariance matrix = unstructured
Method = reml — Number of dimensions = 2
Restricted log likelihood = -110.12562 — Number of observations = 64

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Overall_mean						
b1	.2287775	.0695878	3.29	0.001	.0923879	.3651671
b2	2.647303	.3609679	7.33	0.000	1.939819	3.354787

Estimated between-studies SDs and correlation matrix

	SD	b1	b2
b1	.33954598	1	.
b2	1.8144783	1	1

Παρατηρείται ένα Restricted log likelihood μικρότερο από το προηγούμενο με τιμή -110.12562. Η εκτίμηση του logOR, η οποία έχει αποθηκευτεί στη μεταβλητή b1, είναι 0.2287775, με τυπικό σφάλμα 0.0695878. Ομοίως, η ΔΡ, αποθηκευμένη στη b2, εκτιμάται

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

2.647303, με τυπικό σφάλμα 0.3609679. Συμπερασματικά, τα αποτελέσματα αυτά, είναι κατά προσέγγιση ίδια με τα προηγούμενα από τους Πίνακας 5 καιΠίνακας 6 και στατιστικά σημαντικά σύμφωνα με τις τιμές p-value.

Με την εντολή nlcom υπολογίζεται ο λόγος του σχετικού λόγου πιθανοτήτων με τη μέση διαφορά:

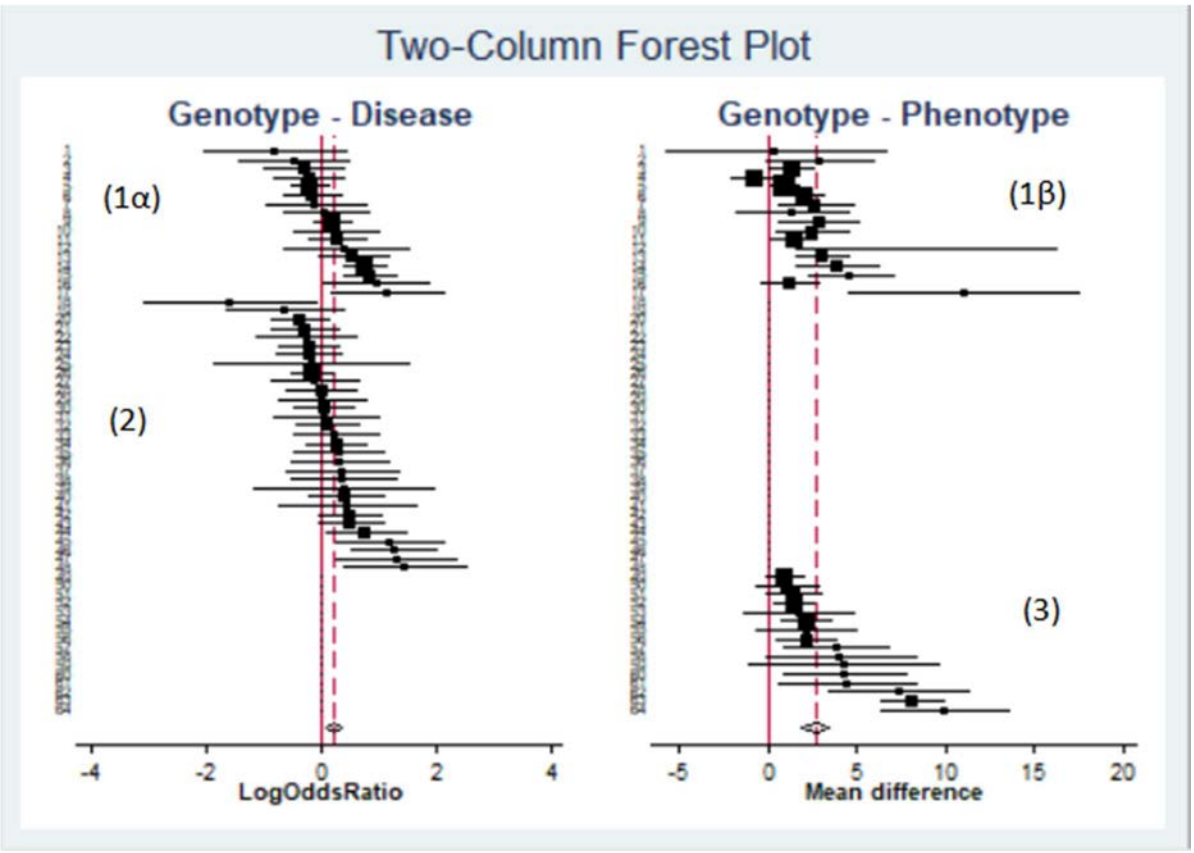
Πίνακας 8. Τελικό αποτέλεσμα της αναλογίας του λογαρίθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά από το πρόγραμμα mrmeta με χρήση του μοντέλου A και της mvmeta με τη μέθοδο REML

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
divisionLogORMeanDiff	.0864191	.0242583	3.56	0.000	.0388737	.1339645

Pooled log odds ratio= 0.2288 and its standard error se= 0.0696
95% Conf. Interval= [0.0924 0.3652]

Τα αποτελέσματα που εμφανίζονται για το λόγο των logOR με τη μέση διαφορά είναι πολύ κοντά με τα προηγούμενα απο το μοντέλο A και τη μέθοδο REML, αποδεικνύοντας μία σχετική ακρίβεια των αποτελεσμάτων.

Στο τέλος του κώδικα περιέχονται οι εντολές για τα γραφήματα της εργασίας, τα οποία να σημειωθεί ότι αποθηκεύονται στο φάκελο εργασίας του χρήστη. Το πρώτο είναι το δενδροδιάγραμμα:



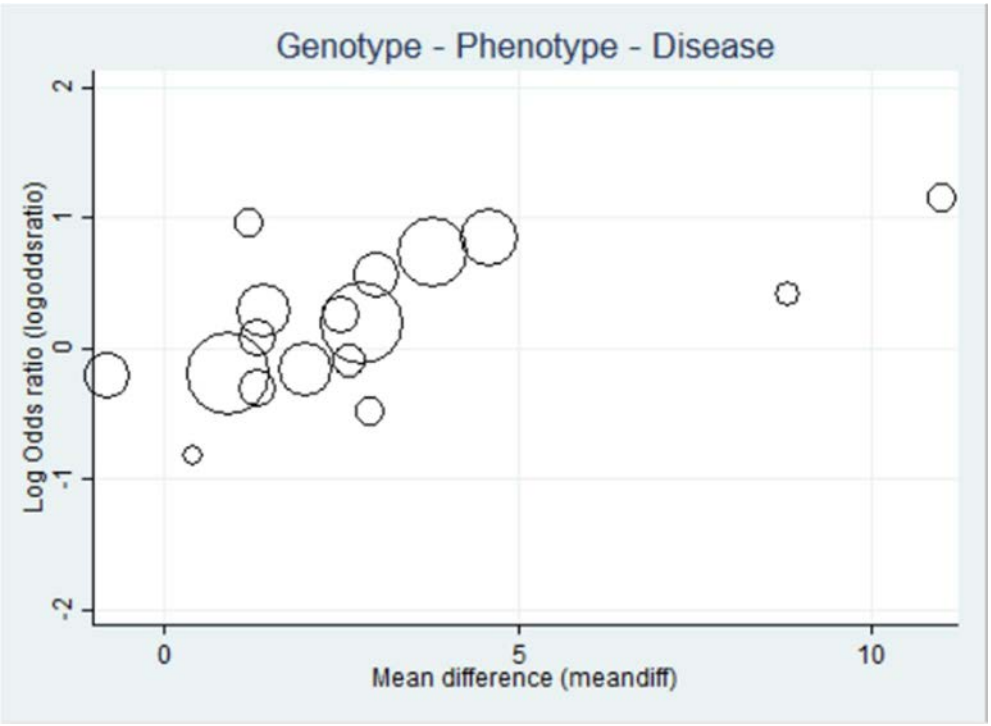
Γράφημα 10. Διπλό δενδροδιάγραμμα μελετών που εξετάζει τη σχέση γονότυπου-φαινότυπου, ή γονότυπου-νόσου ή και των δύο παράλληλα με το πρόγραμμα mrmeta και χρήση του μοντέλου A REML.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

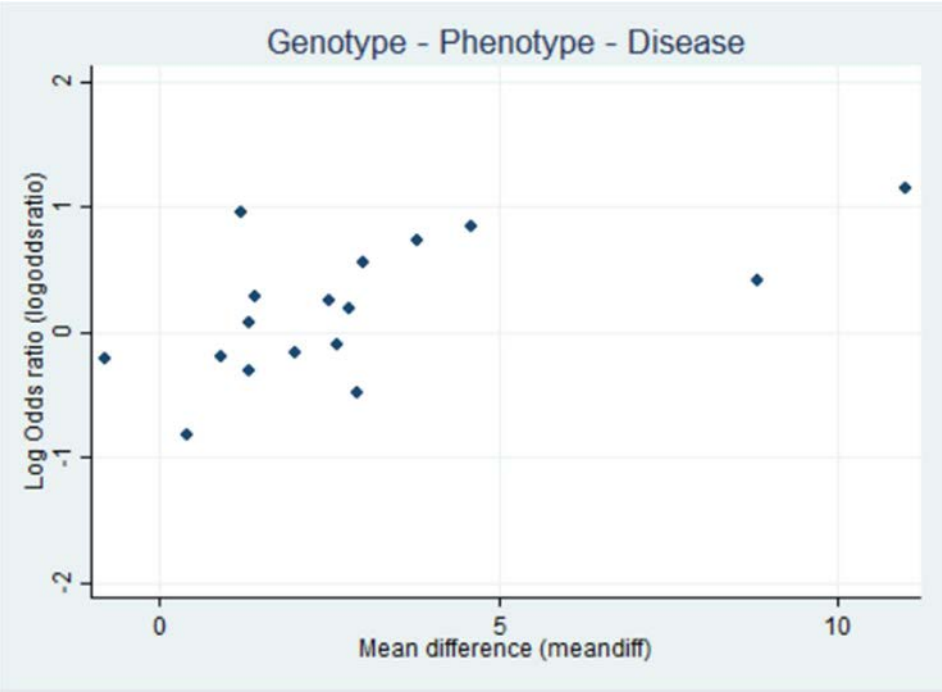
Στην αριστερή πλευρά του κάθε επιμέρους δενδροδιαγράμματος εμφανίζονται οι ονομασίες ή στη συγκεκριμένη περίπτωση, οι αριθμοί ταυτοποίησης των μελετών. Το γράφημα αυτό σαν σύνολο, φαίνεται να χωρίζεται νοητά σε 3 μέρη. Το πάνω (1) μέρος του περιλαμβάνει τις 17 μελέτες που εξέτασαν και τις δύο σχέσεις γονότυπος-νόσος και γονότυπος-φαινότυπος. Το κάτω αριστερά (2), αναπαριστά τις 31 μελέτες που εξέτασαν μόνο τη σχέση γονότυπου-νόσου και κάτω δεξιά (3) τις 16 που εξέτασαν μόνο τη σχέση γονότυπου-φαινότυπου. Στο (1α), λοιπόν, οι 7 μελέτες φαίνεται να έχουν μία ικανοποιητική επίδραση του γονότυπου στη νόσο, με σχετικά μεσαίου μεγέθους διαστήματα εμπιστοσύνης, ενώ το logOR τους είναι πριν από το 0, υποδηλώνοντας θετική επίδραση. Στις υπόλοιπες 10 το logOR είναι μεγαλύτερο από το 0, με τις περισσότερες από αυτές να εμφανίζουν μία μεγάλη αρνητική επίδραση και να έχουν σχετικά μεσαίου μεγέθους διαστήματα εμπιστοσύνης. Παρ' όλα αυτά όμως, οι περισσότερες μελέτες φαίνεται να περιέχουν το 0 στο διάστημα εμπιστοσύνης τους, έχοντας ως συνέπεια το αποτέλεσμά τους (το μέγεθος επίδρασής τους) να μην θεωρείται στατιστικά σημαντικό. Στο (2) γίνονται αντίστοιχες παρατηρήσεις με το (1α), καθώς παρατηρείται ότι οι λιγότερες μελέτες έχουν ικανοποιητικό μέγεθος επίδρασης και μικρά διαστήματα εμπιστοσύνης, ενώ τα περισσότερα logOR βρίσκονται πριν από το 0. Και εδώ οι περισσότερες μεμονωμένες μελέτες θεωρούνται μη στατιστικά σημαντικές. Το διαμάντι, ωστόσο, στο τέλος του γραφήματος, αντιπροσωπεύει το συγκεντρωτικό αποτέλεσμα των μελετών, το οποίο αποδεικνύει μία αρνητική τελική επίδραση του γονότυπου στη νόσο, με το μέγεθός του να υποδηλώνει μικρό διάστημα εμπιστοσύνης. Στο (1β) μόνο 1 από τις 17 μελέτες εμφανίζει θετικό και μάλιστα αρκετά μεγάλο βαθμό επίδρασης και πολύ μικρό διάστημα εμπιστοσύνης. Οι υπόλοιπες βρίσκονται δεξιά του μηδενός, φανερώνοντας την αρνητική επίδραση του γονότυπου στο φαινότυπο, ενώ εμφανίζονται και κάποιες μη-στατιστικά σημαντκές. Στο (3) όλες οι μελέτες, ανεξαιρέτως, βρίσκονται δεξιά από το 0, με κάποιες πάλι να είναι μη-στατιστικά σημαντικές. Το συγκεντρωτικό αποτέλεσμα, λοιπόν, είναι δεξιά από το 0, με πολύ μικρό διάστημα εμπιστοσύνης, αποδεικνύοντας την αρνητική τελική επίδραση του γονότυπου στο φαινότυπο. Συμπερασματικά, παρατηρούνται αρνητική επίδραση του MTHFR στη CHD, όπως και του επιπέδου ομοκυστεΐνης στη CHD.

Στη συνέχεια δίνεται η επιλογή στο χρήστη ανάμεσα σε ένα διάγραμμα φυσαλίδων `plot(bubble)` και διασποράς `plot(scatter)`.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 11. Γράφημα φυσαλίδων του λογαρίθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta με χρήση του μοντέλου A REML



Γράφημα 12. Γράφημα διασποράς του λογαρίθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta με χρήση του μοντέλου A REML.

Οι μελέτες – κουκίδες/κύκλοι και στα 2 γραφήματα είναι κατανεμημένες στον άξονα x, που αντιπροσωπεύει τη μέση διαφορά, ενώ ο άξονας y αναφέρεται στο λογάριθμο του λόγου σχετικών πιθανοτήτων. Οι τιμές τους είναι από τα δεδομένα που χρησιμοποιήθηκαν. Στο γράφημα φυσαλίδων το βάρος των μελετών ορίζει το μέγεθος των κύκλων που εμφανίζονται. Να σημειωθεί ότι ο χρήστης δε χρειάζεται να εισάγει και τα βάρη των μελετών, καθώς υπολογίζονται από τον κώδικα.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Με σκοπό τη σύγκριση των αποτελεσμάτων όλων των μοντέλων της εργασίας, παρατίθεται ένας συγκεντρωτικός πίνακας τιμών.

Πίνακας 9. Συγκεντρωτικά αποτελέσματα επίδρασης MTHFR στη CHD από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Log (Odds Ratio)	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	0.2279932	0.0690846	0.0925899	0.3633966	0.001
A REML	0.2301898	0.0751944	0.0919742	0.3684053	0.001
B ML	0.2482481	0.0670263	0.1168789	0.3796173	0.000
B REML	0.249348	0.0688111	0.1144808	0.3842153	0.000
mvmeta ML	0.2279932	0.0690846	0.0925899	0.3633966	0.001
mvmeta REML	0.2287775	0.0695878	0.0923879	0.3651671	0.001

Πίνακας 10. Συγκεντρωτικά αποτελέσματα επίδρασης MTHFR στα επίπεδα ομοκυστεΐνης από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Mean Difference	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	2.639602	0.3561782	1.941505	3.337698	0.000
A REML	2.655228	0.3666451	1.936616	3.373839	0.000
B ML	2.644506	0.4088018	1.823269	3.425742	0.000
B REML	2.647287	0.4243705	1.815537	3.479038	0.000
mvmeta ML	2.639602	0.3561782	1.941505	3.337698	0.000
mvmeta REML	2.647303	0.3609679	1.939819	3.354787	0.000

Πίνακας 11. Συγκεντρωτικά αποτελέσματα λόγου logOR προς mean difference από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Ratio	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	0.0863741	0.0250286	0.037319	0.1354292	0.001
A REML	0.0866931	0.0243699	0.0389289	0.13445720	0.000
B ML	0.0945885	0.0282464	0.0392265	0.1499505	0.001
B REML	0.09419	0.0287535	0.0378342	0.1505459	0.001
mvmeta ML	0.0863741	0.0242363	0.0388719	0.1338763	0.000
mvmeta REML	0.0864191	0.0242583	0.0388737	0.1339645	0.000

Στα παραπάνω πινακάκια παρατηρείται ότι όλα τα αποτελέσματα είναι στατιστιστικά σημαντικά με p-value μικρότερα του 0.05. Το τελικό log(Odds Ratio) από τα μοντέλα A ML/REML και από την εντολή mvmeta είναι αρκετά κοντά στην τιμή 0.23, ενώ από τα μοντέλα B υπάρχει μία μεγαλύτερη διαφοροποίηση στο 0.25. Το Mean Difference, με τη σειρά του, έχει ένα μεγαλύτερο εύρος τιμών από 2.64 μέχρι και 2.65, συγκεκριμένα στο μοντέλο A με τη μέθοδο REML. Το ratio, από την άλλη υπολογίζεται 0.086 από τα μοντέλα A και την εντολή mvmeta, αλλά από τα μοντέλα B εκτιμάται 0.09. Τελικά, συμπεραίνεται ότι ειδικά το μοντέλο A ML με την εντολή mvmeta ml εμφανίζουν ολόιδιες τιμές με το συγκεκριμένο σετ δεδομένων.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

3.2.2 Παράδειγμα 2: Μοντέλο Β- Μέθοδος ML

Το συγκεκριμένο παράδειγμα βασίζεται στη μετα-ανάλυση 'Homocysteine and Risk of Ischemic Heart Disease and Stroke' (Clarke κ.α., 2002), η οποία αποτελείται από 25 μελέτες, οι οποίες εξετάζουν την πιθανή συσχέτιση μεταξύ αυξημένων επιπέδων ομοκυστεΐνης στο αίμα και του κινδύνου ισχαιμικής καρδιοπάθειας (IHD) και εγκεφαλικού επεισοδίου σε υγιείς πληθυσμούς.

Τα δεδομένα της μετα-ανάλυσης αυτής για την εξέταση της συσχέτισης των επιπέδων ομοκυστεΐνης με την IHD, μετά τον υπολογισμό του log(Odds Ratio), έχουν τη μορφή:

Πίνακας 12. Δεδομένα παραδείγματος 2 που περιλαμβάνουν τα επίπεδα ομοκυστεΐνης - IHD

	id	author	logor	vor	md	vmd
1	1	Stehouver et al	-.1625189	.0216897	14.5	.0343546
2	2	Bots et al	-.4942963	.0312665	14.9	.0193566
3	3	Arnesen et al	-.4307829	.0175153	11.4	.0259772
4	4	Alfthan et al	-.0833816	.0319616	9.5	.022113
5	5	Omenn et al	-.2744369	.0241658	11.5	.0342524
6	6	Ubbink et al	-.1165338	.0104586	12.1	.0094544
7	7	Evans et al	.1739534	.0172711	12.9	.034195
8	8	Whincup et al	-.0833816	.0126486	13.4	.0342407
9	9	Wald et al	-.4155154	.011708	11.8	.0102208
10	10	Folsom et al	-.1743534	.0065996	9.6	.020066
11	11	Verhoef et al291	-.1743534	.0112198	12	.046758
12	12	Schwartz et al	-.9416085	.0363344	10.9	.0436226
13	13	Pancharunity et al	-.4307829	.0540563	12.3	.0736306
14	14	Joubran et al	-.2744369	.0193102	12.9	.1282648
15	15	Verhoef et al25	-.4620355	.0264284	9.5	.0451867
16	16	Genest et al	-.6733446	.0143757	10.8	.0453186
17	17	Hopkins et al	-.7765288	.028552	10.3	.0283912
18	18	von Eckardstein et al	-.7339692	.0340177	8.1	.0090313
19	19	Siberberg et al	-.2613648	.0156648	13.1	.0594286
20	20	Graham et al	-.2357223	.0059463	10.9	.0130111

Αυτή τη φορά θα παρατεθούν και θα αναλυθούν τα αποτελέσματα απο το μοντέλο Β και τη μέθοδο ML.

```
mrmeta logor vor md vmd, model(b) method(ml) study(author)
```

Τα αποτελέσματα του ml maximize:

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Πίνακας 13. Μεγιστοποιημένα αποτελέσματα του προγράμματος mrmeta με χρήση του μοντέλου B ML.

```
                                                  Number of obs   =         25
                                                  Wald chi2(0)    =          .
Log likelihood =   -4.81125                       Prob > chi2     =          .

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ratio        |
       _cons |  -.0272861   .0044884    -6.08   0.000    -.0360832    -.018489
-------------+----------------------------------------------------------------
mu           |
       _cons |   11.60127   .3270676    35.47   0.000     10.96023    12.24231
-------------+----------------------------------------------------------------
taut         |
       _cons |  -7.876696   .4089155   -19.26   0.000    -8.678156   -7.075236
-------------+----------------------------------------------------------------
tauy         |
       _cons |   .9664274    .287075     3.37   0.001     .4037707    1.529084
------------------------------------------------------------------------------
```

Η εκτίμηση του ratio από την εντολή ml maximize είναι -0.0272861. Η μέση διαφορά, από πλευράς της, εκτιμάται 11.60127, με τυπικό σφάλμα 0.3270676, και στατιστικά σημαντικό διάστημα εμπιστοσύνης [10.96023 12.24231] (δεν περιλαμβάνει την τιμή 0).

Το nlcom στη συνέχεια υπολογίζει τον τελικό λογάριθμου του λόγου σχετικών πιθανοτήτων:

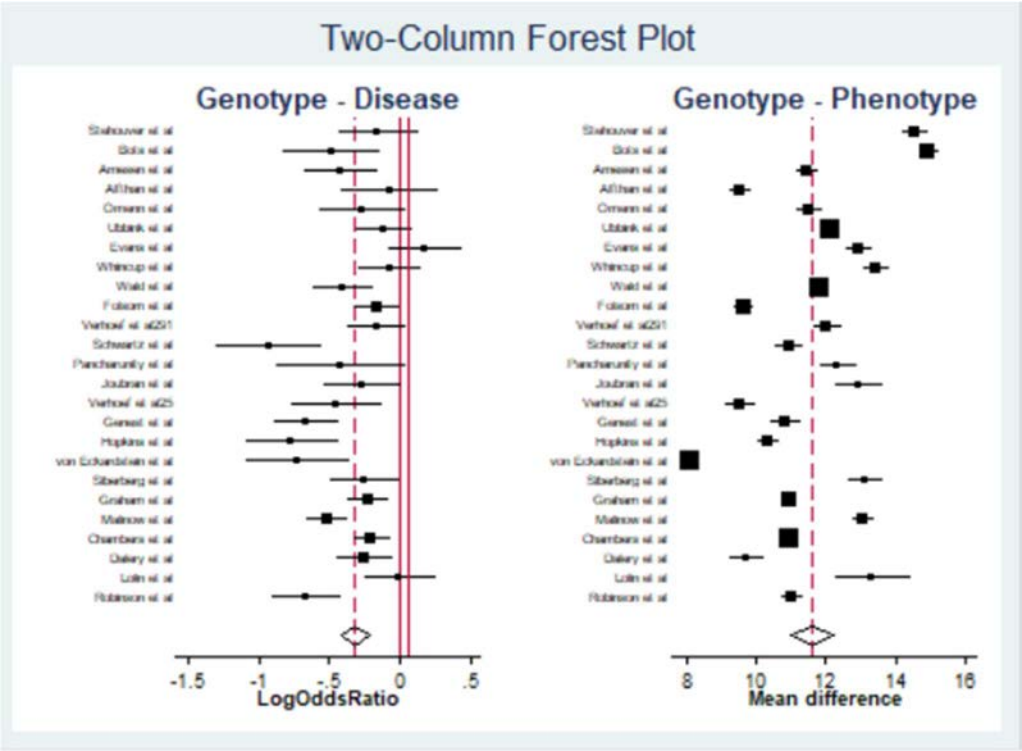
Πίνακας 14. Τελικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων με το πρόγραμμα mrmeta με χρήση του μοντέλου B και της μεθόδου ML.

```
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
LogOddsRatio |  -.3165534   .0527995    -6.00   0.000    -.4200385   -.2130683
------------------------------------------------------------------------------

Pooled log odds ratio= -0.3166 and its standard error se= 0.0528
95% Conf. Interval= [-0.4200 -0.2131]
```

Το τελικό αποτέλεσμα του λογαριθμικού λόγου σχετικών πιθανοτήτων εκτιμάται -0.3166 με τυπικό σφάλμα 0.0528. Είναι στατιστικά σημαντικό αποτέλεσμα με τιμη p-value μηδενική, κατά πολυ μικρότερη από το 0.05, και ένα σχετικά μικρό διάστημα εμπιστοσύνης.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

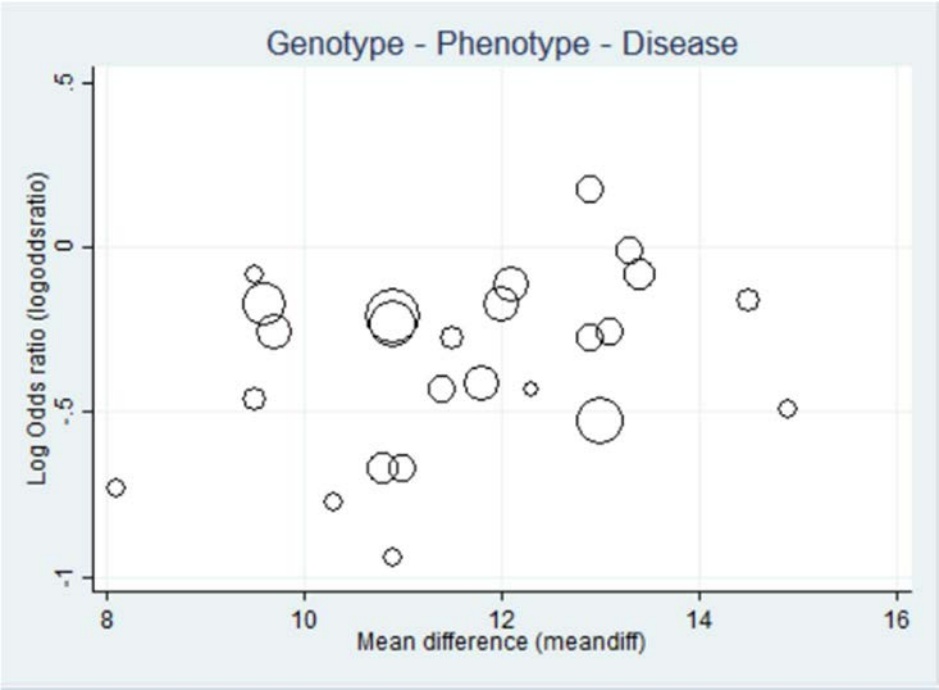


Γράφημα 13. Διπλό δενδροδιάγραμμα του προγράμματος mrmeta και χρήση του μοντέλου B ML

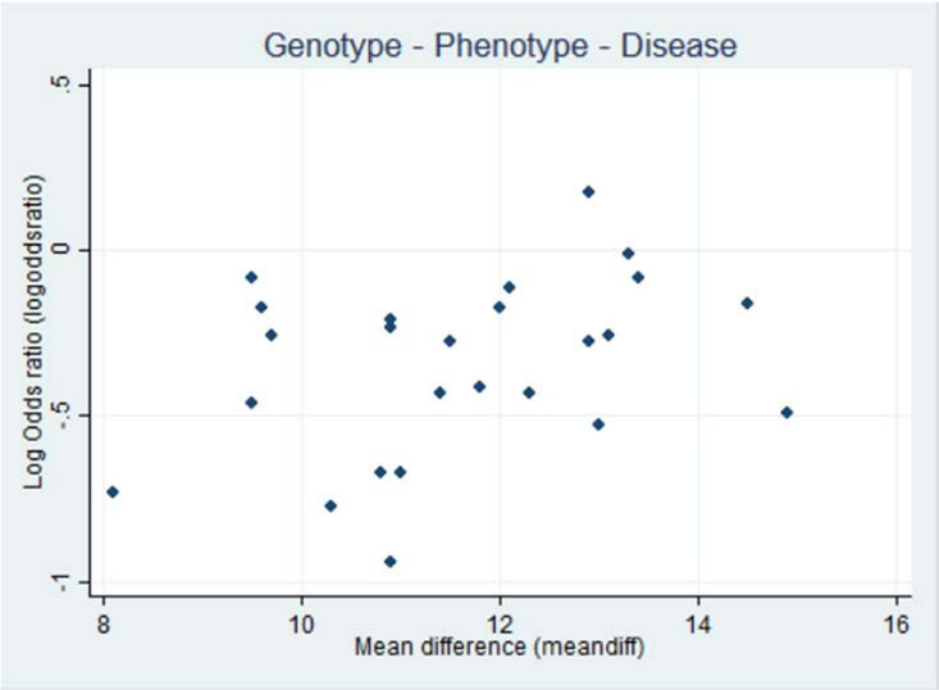
Σε αυτό το παράδειγμα επιλέχθηκε το `study(author)`, με αποτέλεσμα στα αριστερά του δενδροδιαγράμματος να εμφανίζονται τα ονόματα των συγγραφέων της κάθε μελέτης. Στο αριστερό δενδροδιάγραμμα όλες οι μελέτες εκτός από μία (οι 24 δηλαδή), βρίσκονται αριστερά από το 0, υποδηλώνοντας μία θετική επίδραση του γονότυπου στη νόσο. Από αυτές τις μελέτες οι 8 έχουν διαστήματα εμπιστοσύνης που περιλαμβάνουν το 0, με αποτέλεσμα να μην είναι στατιστικά σημαντικές. Στις υπόλοιπες 16 παρατηρείται ένας μεγάλος βαθμός επίδρασης, με σχετικά μικρά διαστήματα εμπιστοσύνης. Η μία και μοναδική που βρίσκεται δεξιά από το 0 έχει μικρή επίδραση με διάστημα εμπιστοσύνης που περιλαμβάνει το 0, καθιστώντας τη τελικά μη στατιστικά σημαντική. Το διαμάντι, λοιπόν, σε αυτό το δενδροδιάγραμμα υποδηλώνει μία τελική θετική επίδραση των επιπέδων ομοκυστεΐνης στην IHD, αφού βρίσκεται αριστερά του 0, με ένα αρκετά μικρό διάστημα εμπιστοσύνης. Στο δεξί δενδροδιάγραμμα όλες οι μελέτες ανεξαιρέτως παρουσιάζουν μία, είτε μικρή είτε μεγάλη, αρνητική επίδραση του γονότυπου στο φαινότυπο με διαστήματα εμπιστοσύνης κατά βάση στατιστικά σημαντικά και πολύ στενά. Το συγκεντρωτικό αποτέλεσμα, λοιπόν, όπως αναμένεται παρουσιάζει μία αρνητική επίδραση του γονότυπου στο φαινότυπο, καθώς βρίσκεται σε τιμή πιο κοντά στο 12 παρά στο 0 και ένα μικρό διάστημα εμπιστοσύνης.

Στην περίπτωση που ο χρήστης επιλέξει την εμφάνιση ενός γραφήματος φυσαλίδων ή διασποράς:

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 14. Γράφημα φυσαλίδων του λογαρίθμου του λόγου σχετικών πιθανοτήτων και της μέσης διαφοράς, του προγράμματος mrmeta και χρήση του μοντέλου B ML



Γράφημα 15. Γράφημα διασποράς του λογαρίθμου του λόγου σχετικών πιθανοτήτων και της μέσης διαφοράς, του προγράμματος mrmeta με χρήση του μοντέλου B ML

Από τα γραφήματα αυτά διαφαίνεται μία ασθενής γραμμική συσχέτιση του logOR και του mean difference, υποδηλώνοντας μία αντίστοιχη ασθενή συσχέτιση των επιπέδων ομοκυστεΐνης στην IHD. Από το Γράφημα 14, οι 5 μελέτες από τις 25 φαίνεται να έχουν το μεγαλύτερο βάρος στη μετα-ανάλυση αυτή.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Τελικά παραθέτονται τα συγκεντρωτικά αποτελέσματα από όλα τα μοντέλα και τις μεθόδους, για μία πιο εύκολη σύγκρισή τους.

Πίνακας 15. Συγκεντρωτικά αποτελέσματα του λογαρίθμου του λόγου σχετικών πιθανοτήτων με το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Log (Odds Ratio)	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	-0.3308694	0.0498349	-0.428544	-0.2331947	0.000
A REML	-0.3315401	0.0509777	-0.4314544	-0.2316257	0.000
B ML	-0.3165534	0.0527995	-0.4200385	-0.2130683	0.000
B REML	-0.3166032	0.0540954	-0.4226283	-0.2105781	0.000
mvmeta ML	-0.3308694	0.0498349	-0.428544	-0.2331947	0.000
mvmeta REML	-0.3315401	0.0509777	-0.4314545	-0.2316257	0.000

Πίνακας 16. Συγκεντρωτικά αποτελέσματα της μέσης διαφοράς με το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Mean Difference	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	11.60465	0.3281102	10.96157	12.24773	0.000
A REML	11.60493	0.3348369	10.94866	12.2612	0.000
B ML	11.60127	0.3270676	10.96023	12.24231	0.000
B REML	11.60167	0.3334386	10.94814	12.2552	0.000
mvmeta ML	11.60465	0.3281102	10.96157	12.24773	0.000
mvmeta REML	11.60493	0.3348369	10.94866	12.2612	0.000

Πίνακας 17. Συγκεντρωτικά αποτελέσματα λόγου logOR προς mean difference από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Ratio	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	-0.0285118	0.0046246	-0.0375758	-0.0194478	0.000
A REML	-0.0285689	0.004731	-0.0378415	-0.0192963	0.000
B ML	-0.0272861	0.0044884	-0.360832	-0.018489	0.000
B REML	-0.0272895	0.004599	-0.0363033	-0.0182756	0.000
mvmeta ML	-0.0285118	0.0046246	-0.0375758	-0.0194478	0.000
mvmeta REML	-0.0285689	0.004731	-0.0378415	-0.0192963	0.000

Όλες οι τιμές είναι στατιστικά σημαντικές, με τιμή p-value μηδενική και διαστήματα εμπιστοσύνης σχετικά μικρά. Όλες οι εκτιμήσεις του τελικού λογαριθμικού λόγου σχετικών πιθανοτήτων είναι πολύ κοντά στο 0.32 με τυπικά σφάλματα 0.05. Τα τελικά αποτελέσματα της μέσης διαφοράς παρουσιάζουν εξίσου κοντινές τιμές με εκτιμήσεις 11.60 και τυπικά σφάλματα στο 0.33. Το ratio εκτιμάται -0.02 με se 0.004. Συπερασματικά, σε αυτό το σετ δεδομένων παρατηρείται ότι τα συγκεντρωτικά αποτελέσματα του μοντέλου A ML και της mvmeta ML είναι ολόιδια, αλλά και του μοντέλου A REML με της mvmeta REML.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

3.2.3 Παράδειγμα 3: Μοντέλο Β- Μέθοδος REML

Σαν τελικό παράδειγμα θα χρησιμοποιηθούν τα δεδομένα από την ίδια μετα-ανάλυση του παραδείγματος 2 αλλά αυτή τη φορά με τις 13 μελέτες που εξετάζουν την πιθανή συσχέτιση μεταξύ αυξημένων επιπέδων ομοκυστεΐνης στο αίμα και του κινδύνου εγκεφαλικού επεισοδίου σε υγιείς πληθυσμούς.

Τα δεδομένα έχουν τη μορφή:

Πίνακας 18. Δεδομένα παραδείγματος 3 που περιλαμβάνουν επίπεδα ομοκυστεΐνης - εγκεφαλικό επεισόδιο.

	id	study	logor	varor	mnd	varmnd
1	1	Evans et al	-2.407946	1.667916	12.9	4.417999
2	2	Ubbink et al	.0099503	.1694987	12.1	1.400833
3	3	Omenn et al	-.5276328	.3061742	11.5	.9284208
4	4	Alfthan et al	-.5447272	.0676216	9.5	.1764706
5	5	Stehouwer et al	.1570037	.0329232	14.5	.3297015
6	6	Bots et al	-.4942963	.0272104	14.9	.1552688
7	7	Perry et al	-.4462871	.0332808	12.3	.152
8	8	Stampfer / Verhoef et al	-.1053605	.0271488	10.7	.1071074
9	9	Schwartz et al	-.3424903	.0389045	10.9	.3211111
10	10	Graham et al	-.1743534	.0104448	10.9	.105625
11	11	Lindgren et al	.0676587	.0312665	14.3	.1290244
12	12	Coull et al	.1570037	.1476158	10.3	.0385714
13	13	Evers et al	-1.203973	.0869303	11.3	.0508475

Η εκτέλεση του μοντέλου Β με τη μέθοδο REML συντάσσεται ως εξής:

```
mrmeta logor varor mnd varmnd, model(b) method(reml) study(study)
```

Τα μεγιστοποιημένα αποτελέσματα που προκύπτουν:

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Πίνακας 19. Μεγιστοποιημένα αποτελέσματα του προγράμματος mrmeta με χρήση του μοντέλου B REML.

```
                                                Number of obs   =         13
                                                Wald chi2(0)    =          .
Log likelihood = -24.212309                     Prob > chi2     =          .
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ratio						
_cons	-.0220592	.0087636	-2.52	0.012	-.0392355	-.0048829
mu						
_cons	11.95789	.505114	23.67	0.000	10.96789	12.9479
taut						
_cons	-7.494579	.7970042	-9.40	0.000	-9.056679	-5.93248
tauy						
_cons	1.059962	.4415627	2.40	0.016	.1945152	1.925409

Η εκτίμηση του ratio είναι -0.0220592. Η μέση διαφορά, από πλευράς της, εκτιμάται 11.95789, με τυπικό σφάλμα 0.505114 και διάστημα εμπιστοσύνης [10.96789 12.9479].

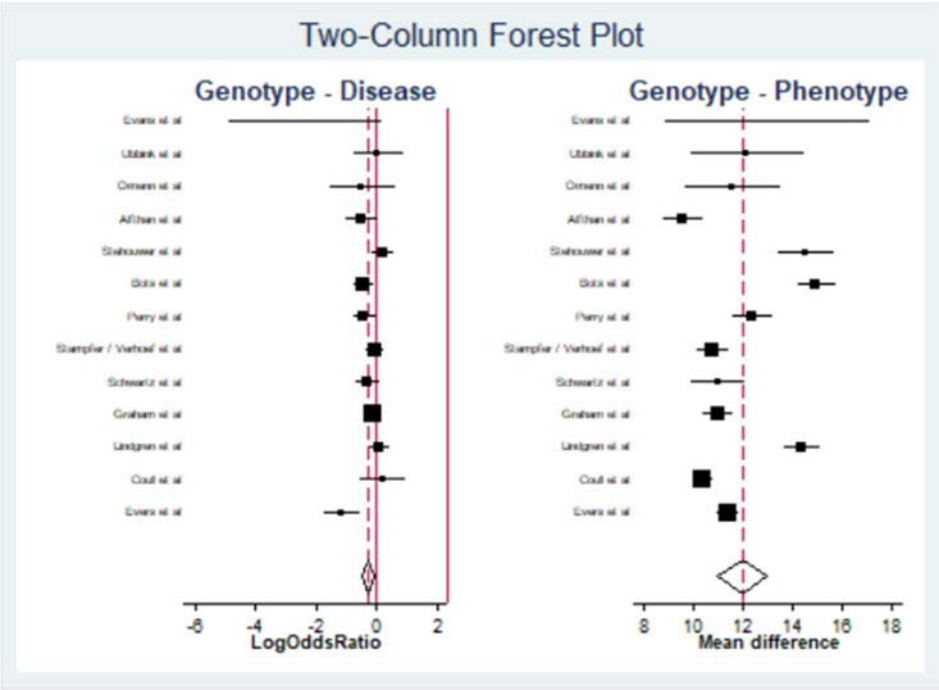
Πίνακας 20. Συγκεντρωτικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων με το πρόγραμμα mrmeta και χρήση του μοντέλου B και της μεθόδου REML.

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LogOddsRatio	-.2637815	.1053149	-2.50	0.012	-.470195	-.0573681

```
Pooled log odds ratio= -0.2638 and its standard error se= 0.1053
95% Conf. Interval= [-0.4702 -0.0574]
```

Ο τελικός λογαριθμικός λόγος σχετικών πιθανοτήτων υπολογίζεται -0.2638, με τυπικό σφάλμα 0.1053. Το αποτέλεσμα είναι στατιστικά σημαντικό, καθώς το διάστημα εμπιστοσύνης [-0.4702 -0.0574] είναι στενό και δεν περιλαμβλανει την τιμή 0. Επίσης, η τιμή p-value υπολογίστηκε 0.012, τιμή μικρότερη του 0.05.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

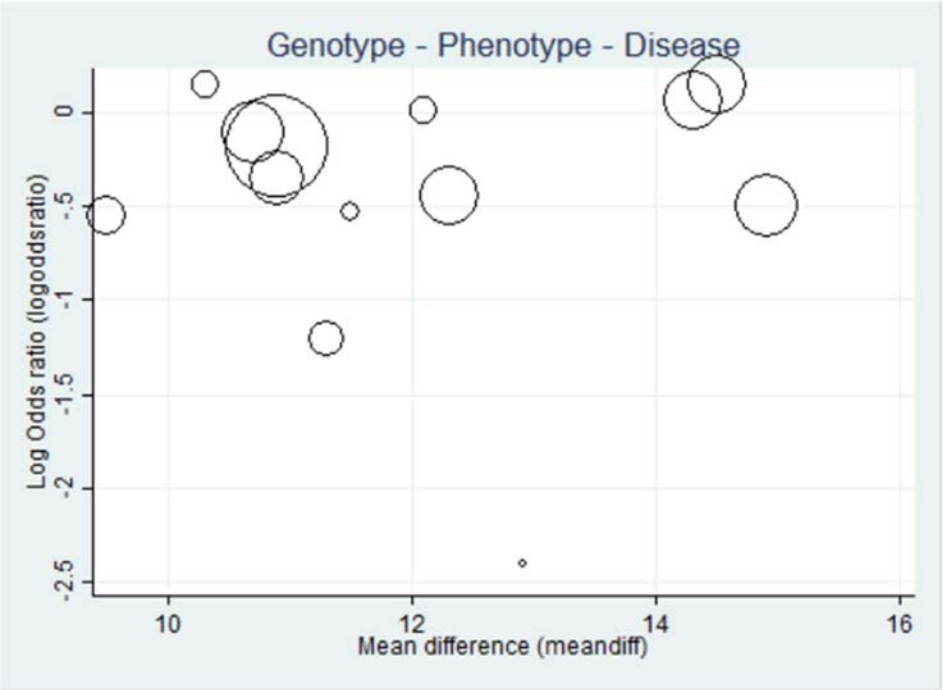


Γράφημα 16. Διπλό δενδροδιάγραμμα μελετών με το πρόγραμμα mrmeta και χρήση του μοντέλου B REML

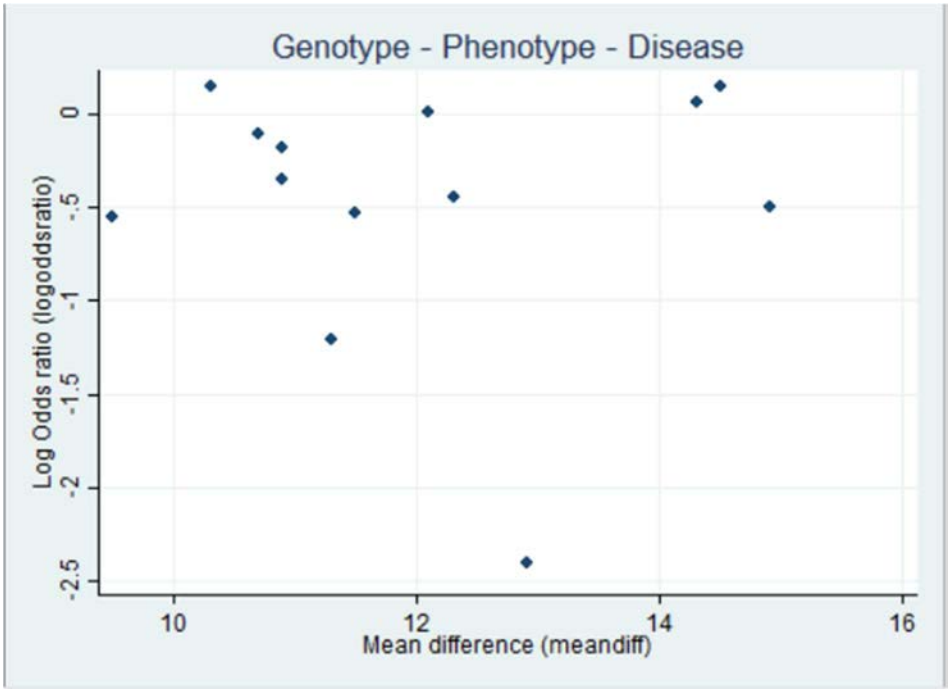
Το αριστερό δενδροδιάγραμμα αναδεικνύει μία ισχυρή συσχέτιση μεταξύ γονότυπου και νόσου, καθώς σχεδόν οι 5 μόνο μελέτες που το διάστημα εμπιστοσύνης τους δεν περιλαμβάνει το 0, βρίσκονται αριστερά από αυτό. Οι υπόλοιπες 8 μελέτες περιλαμβάνουν το 0, με αποτέλεσμα να μην είναι στατιστικά σημαντικές. Αναμενόμενα, λοιπόν, το συγκεντρωτικό αποτέλεσμα βρίσκεται αριστερά από την τιμή 0 και έχει πολύ μικρό διάστημα εμπιστοσύνης, που δεν το περιλαμβάνει. Στο δεξί δενδροδιάγραμμα, όποια και να είναι επίδραση και τα διαστήματα εμπιστοσύνης της εκάστοτε μελέτης, απέχουν κατά πολύ από το 0. Συνεπώς, το αναμενόμενο συγκεντρωτικό αποτέλεσμα εκφράζει μία αδιαμφισβήτητη αρνητική επίδραση του γονότυπου στο φαινότυπο και είναι στατιστικά σημαντικό.

Στην περίπτωση που ο χρήστης επιλέξει την εμφάνιση ενός γραφήματος φυσαλίδων ή διασποράς:

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156



Γράφημα 17. Γράφημα φυσαλίδων του λογάριθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta και χρήση του μοντέλου B REML



Γράφημα 18. Γράφημα διασποράς του λογάριθμου του λόγου σχετικών πιθανοτήτων με τη μέση διαφορά του προγράμματος mrmeta και χρήση του μοντέλου B REML

Από τα γραφήματα διασποράς και φυσαλίδων παρατηρείται μία ασθενής έως μηδενική γραμμική συσχέτιση του logOR και του mean difference, υποδηλώνοντας αντίστοιχα ασθενή έως μηδενική συσχέτιση των επιπέδων ομοκυστεϊνης μετο εγκεφαλικό.

Τα συγκεντρωτικά αποτελέσματα από όλα τα μοντέλα και τις μεθόδους:

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Πίνακας 21. Συγκεντρωτικά αποτελέσματα λόγου σχετικής πιθανότητας από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Log(Odds Ratio)	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	-0.281987	0.0961164	-0.4703718	-0.0936023	0.003
A REML	-0.2857658	0.1022214	-0.486116	-0.0854156	0.005
B ML	-0.2603487	0.0986623	-0.4537233	-0.0669741	0.008
B REML	-0.2637815	0.1053149	-0.470195	-0.573681	0.012
mvmeta ML	-0.281987	0.0961165	-0.4703718	-0.0936023	0.003
mvmeta REML	-0.2857658	0.1022214	-0.486116	-0.0854156	0.005

Πίνακας 22. Συγκεντρωτικά αποτελέσματα μέσης διαφοράς από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Mean Difference	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	11.95112	0.4883843	10.99391	12.90834	0.000
A REML	11.95293	0.5072387	10.95876	12.9471	0.000
B ML	11.95516	0.4864384	11.00176	12.90857	0.000
B REML	11.95789	0.505114	10.96789	12.9479	0.000
mvmeta ML	11.95112	0.4883843	10.99391	12.90834	0.000
mvmeta REML	11.95293	0.5072387	10.95876	12.9471	0.000

Πίνακας 23. Συγκεντρωτικά αποτελέσματα λόγου logOR προς mean difference από το πρόγραμμα mrmeta.

Models			95% Confidence interval		
	Ratio	*Standard error*	*Lower limit*	*Upper limit*	*p-value*
A ML	-0.023595	0.0082996	-0.039862	-0.0073281	0.004
A REML	-0.0239076	0.0088212	-0.0411969	-0.0066183	0.007
B ML	-0.0217771	0.0082112	-0.0378707	-0.0056835	0.008
B REML	-0.0220592	0.0087636	-0.0392355	-0.0048829	0.012
mvmeta ML	-0.023595	0.0082996	-0.039862	-0.0073281	0.004
mvmeta REML	-0.0239076	0.0088212	-0.0411969	-0.0066183	0.007

Όλα τα αποτελέσματα είναι στατιστικά σημαντικά. Τα συγκεντρωτικά αποτελέσματα του λογαριθμικού λόγου σχετικών πιθανοτήτων κυμαίνονται σε τιμές από -0.26 έως και -0.28 και όλα έχουν σχετικά μικρό τυπικό σφάλμα. Τα μοντέλα Β, από πλευράς τους εμφανίζουν κοντινά αποτελέσματα, πάλι στατιστικά σημαντικά, με λίγο μεγαλύτερη τιμή p-value = 0.008 ή 0.012. Τα συγκεντρωτικά αποτελέσματα της μέσης διαφοράς είναι, αν όχι ίδια, σχεδόν ολόιδια μεταξύ όλων των μοντέλων και των μεθόδων. Υπολογίζουν μία μέση διαφορά 11.95 με μικρό τυπικό σφάλμα από 0.49 μέχρι 0.50. Το ratio , αντίστοιχα εκτιμάται 0.02 με se 0.008, σε όλα τα μοντέλα. Τελικά, παρατηρούνται και σε αυτό το σετ δεδομένων ολόιδιες τιμές στο μοντέλο A ML με την mvmeta ML και στο μοντέλο A REML με την mvmeta REML.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Κεφάλαιο 4

Συζήτηση

Η εργασία αυτή αξιοποίησε μετα-αναλύσεις που εξέτασαν, με Μεντελιανή τυχαιοποίηση, τη συσχέτιση γονότυπου, φαινότυπου και ασθένειας ενσωματώνοντας τα μοντέλα Α και Β, μέγιστης πιθανοφάνειας και περιορσμένης μέγιστης πιθανοφάνειας. Όπως αναφέρθηκε και στο δεύτερο κεφάλαιο, το πρώτο μοντέλο Α βασίζεται σε διμεταβλητές κανονικές κατανομές και υποθέτει ότι η συσχέτιση εντός της μελέτης είναι μηδενική. Το δεύτερο μοντέλο Β, με τη σειρά του βελτιώνει την παραμετροποίηση των ετερογενειών στις συσχετίσεις μεταξύ γονότυπου-φαινότυπου, φαινότυπου-νόσου και γονότυπου-νόσου. Το μοντέλο Β παρέχει πιο ακριβείς εκτιμήσεις, λαμβάνοντας υπόψη την ετερογένεια μεταξύ των μελετών.

Συμπερασματικά, τα αποτελέσματα των δύο μοντέλων υποδεικνύουν ότι οι παρατηρούμενες σχέσεις μεταξύ γονότυπου, φαινότυπου και ασθένειας επηρεάζονται σημαντικά από την ετερογένεια μεταξύ των μελετών. Η χρήση της Μεντελιανής τυχαιοποίησης σε συνδυασμό με τα μοντέλα Α και Β παρέχει μια ισχυρή μεθοδολογική προσέγγιση για την κατανόηση των αιτιολογικών σχέσεων και των επιδράσεων των γονιδίων στη νόσο.

Στην εργασία αυτή, λοιπόν, κατασκευάστηκε ένα πρόγραμμα, το οποίο έχει συγκεντρώσει τα παραπάνω μοντέλα και παρέχει την δυνατότητα στο χρήστη να επιλέξει ένα από αυτά για να πραγματοποίηση της μετα-ανάλυσής του. Το πρόγραμμα αυτό κατασκευάστηκε με τη χρήση του στατιστικού πακέτου Stata (έκδοση 13.1), στο οποίο δεν υπήρχε άλλο τέτοιο πρόγραμμα που να κάνει την παραπάνω διεργασία. Για τον έλεγχο της λειτουργίας του κώδικα και των αποτελεσμάτων του εκτελέστηκαν τρία παραδείγματα. Με βάση τα αποτελέσματα των παραδειγμάτων αυτών, παράγονται αρκετές παρατηρήσεις για τα μοντέλα που χρησιμοποίηθηκαν.

Το πρώτο παράδειγμα χρησιμοποίησε τα δεδομένα από μία μετα-ανάλυση με 64 μελέτες. Πιο συγκεκριμένα εξέταζε τη συσχέτιση μεταξύ του γονιδίου MTHFR, του επιπέδου ομοκυστεΐνης και της CHD. Το τελικό αποτέλεσμα του λογαρίθμου του λόγου σχετικών πιθανοτήτων του MTHFR στη CHD είναι 0.23. Όλα τα αποτελέσματα είναι στατιστικά σημαντικά, με το μοντέλο Β και τη μέθοδο ML/REML να βγάζουν τη μικρότερη τιμή p-value. Η τελική μέση διαφορά στη CHD υπολογίζεται στο 2.64. Όλα τα αποτελέσματα είναι πάλι στατιστικά σημαντικά με μηδενικές τιμές p-value και διαστήματα εμπιστοσύνης που δεν περιλαμβάνουν το 0. Τα μοντέλα Β ML/REML, βέβαια εμφανίζουν μία διαφοροποίηση στα διαστήματα εμπιστοσύνης τα οποία είναι λίγο πιο μεγάλα. Στη συνέχεια, στο δενδροδιάγραμμα του ΓραφήματοςΓράφημα 10 αναπαρίστανται οι μελέτες των δεδομένων με τα συγκεντρωτικά αποτελέσματα, τα οποία τελικά αποδεικνύουν ότι μία αλλαγή στα επίπεδα ομοκυστεΐνης δεν επηρρεάζει απαραίτητα τον κίνδυνο εμφάνισης CHD.

Στη συνέχεια χρησιμοποιήθηκαν δεδομένα από μία μετα-ανάλυση με 25 μελέτες. Η συγκεκριμένη μετα-ανάλυση βασίστηκε στην πρόταση ότι οι συγκεντρώσεις ομοκυστεΐνης στο αίμα σχετίζονται με τον κίνδυνο ισχαιμικής καρδιοπάθειας (IHD) και εγκεφαλικού επεισοδίου.

Στο δεύτερο παράδειγμα, λοιπόν χρησιμοποιήθηκαν τα δεδομένα από αυτήν τη μετα-

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

ανάλυση με τις 25 μελέτες, οι οποίες εξέτασαν τη συσχέτιση των επιπέδων ομοκυστεΐνης στο αίμα με τον κίνδυνο IHD. Ο λογαριθμικός λόγος σχετικών πιθανοτήτων υπολογίστηκε -0.32, με όλα τα αποτελέσματα να είναι στατιστικά σημαντικά, με μηδενικές τιμές p-value. Το μικρότερο διάστημα εμπιστοσύνης υπολογίστηκε από το μοντέλο A με τη μέθοδο REML. Η τελική μέση διαφορά υπολογίστηκε 11,56, με όλα τα αποτελέσματα να είναι στατιστικά σημαντικά. Τα μοντέλα B ML/REML, πάλι εμφανίζουν μία διαφοροποίηση στα διαστήματα εμπιστοσύνης τα οποία είναι λίγο πιο στενά αυτή τη φορά. Το Γράφημα 13 συνοψίζει τις μελέτες που συμπεριλήφθηκαν στη μετα-ανάλυση και εμφανίζει τα αντίστοιχα συγκεντρωτικά αποτελέσματα.

Στο τρίτο παράδειγμα χρησιμοποιήθηκαν τα δεδομένα από την ίδια μετα-ανάλυση, τα οποία όμως περιείχαν τις 13 μελέτες που εξέτασαν τη συσχέτιση των επιπέδων ομοκυστεΐνης στο αίμα με τον κίνδυνο εγκεφαλικού επεισοδίου. Ο λογαριθμικός λόγος σχετικών πιθανοτήτων υπολογίστηκε -0.28, με όλα τα αποτελέσματα να είναι στατιστικά σημαντικά. Τα μοντέλα B ML/REML, για άλλη μια φορά, ξεχωρίζουν από τις υπόλοιπες διαδικασίες καθώς έχουν τιμές p-value 0.008 και 0.012. Η μέση διαφορά εκτιμήθηκε 11.95, με τιμή p-value από την κάθε εκτέλεση 0. Στα μοντέλα B ML/REML παρατηρούμε κάποιες μικρές, αλλά αμελητέες αποκλίσεις στις τιμές τους. Στο Γράφημα 16 εμφανίζονται οι μελέτες της μετα-ανάλυσης και τα αντίστοιχα συγκεντρωτικά αποτελέσματα.

Η μετα-ανάλυση που αναφέρεται στο δεύτερο και στο τρίτο παράδειγμα υποδηλώνει ότι η αυξημένη ομοκυστεΐνη είναι, το πολύ, ένας μέτριος ανεξάρτητος προγνωστικός δείκτης κινδύνου IHD και εγκεφαλικού σε υγιείς πληθυσμούς (Clarke κ.α., 2002). Τα ευρήματα υποδεικνύουν ότι τα υψηλά επίπεδα ομοκυστεΐνης μπορεί να παίζουν ρόλο στην ανάπτυξη καρδιαγγειακών νοσημάτων, αλλά οι ισχυρότερες συσχετίσεις στις αναδρομικές μελέτες υποδηλώνουν ότι τα επίπεδα ομοκυστεΐνης μπορεί επίσης να επηρεαστούν από την παρουσία της νόσου. Συνολικά, η μείωση των επιπέδων ομοκυστεΐνης κατά 25% φαίνεται να μειώνει τον κίνδυνο για ισχαιμική καρδιοπάθεια και εγκεφαλικό, γεγονός που υποστηρίζει την ανάγκη για περαιτέρω έρευνα και πιθανές παρεμβάσεις για τη μείωση των επιπέδων ομοκυστεΐνης (Clarke κ.α., 2002).

Τελικά, λήφθηκαν πολλές πληροφορίες από τη σύγκριση των αποτελεσμάτων μετα-ανάλυσης από την mvmeta και τα μοντέλα A και B ML/REML. Παρ' όλα αυτά χρησιμοποιήθηκαν μόνο τρία σετ δεδομένων, τα οποία περιείχαν ένα μικρό ως μέτριο αριθμό μελετών, οπότε δεν θα μπορούσαν να γενικευτούν τα αποτελέσματά των μοντέλων με απόλυτη σιγουριά. Σίγουρα θα ήταν χρήσιμα περισσότερα παραδείγματα και με περισσότερες μελέτες που να διαφέρουν, με σκοπό την εξέτασή τους σε κάθε μοντέλο. Παρ' όλα αυτά επαληθεύτηκε ότι τα μοντέλα κατασκευάστηκαν σωστά στο στατιστικό πακέτο της Stata και έδωσαν κάποια σημαντικά αποτελέσματα για τη σύγκρισή τους, όπως αναλύθηκε παραπάνω.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Βιβλιογραφικές αναφορές

Attia, J., Thakkinstian, A., & D'Este, C. (2003). Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *Journal of clinical epidemiology*, *56*(4), 297–303. https://doi.org/10.1016/s0895-4356(03)00011-8

Bagos P. G. (2015). Meta-analysis in Stata using gllamm. *Research synthesis methods*, *6*(4), 310–332. https://doi.org/10.1002/jrsm.1157

Bailey, L. B., & Gregory III, J. F. (1999). Polymorphisms of methylenetetrahydrofolate reductase and other enzymes: Metabolic significance, risks and impact on folate requirement. *Journal of Nutrition, 129*(5), 919–922. doi:10.1093/jn/129.5.919

Belbasis, L., & Bellou, V. (2018). Introduction to Epidemiological Studies. *Methods in molecular biology (Clifton, N.J.)*, *1793*, 1–6. https://doi.org/10.1007/978-1-4939-7868-7_1

Clayton, D., & McKeigue, P. M. (2001). Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet (London, England)*, *358*(9290), 1356–1360. https://doi.org/10.1016/S0140-6736(01)06418-2

Colhoun, H. M., McKeigue, P. M., & Davey Smith, G. (2003). Problems of reporting genetic associations with complex outcomes. *Lancet (London, England)*, *361*(9360), 865–872. https://doi.org/10.1016/s0140-6736(03)12715-8

Davey Smith G (2006). Capitalising on Mendelian randomization to assess the effects of treatments. JLL Bulletin: Commentaries on the history of treatment evaluation (https://www.jameslindlibrary.org/articles/capitalising-on-mendelian-randomization-to-assess-the-effects-of-treatments/)

Didelez, V., & Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical methods in medical research*, *16*(4), 309–330. https://doi.org/10.1177/0962280206077743

Elston, R. C., Satagopan, J. M., & Sun, S. (2012). Genetic terminology. *Methods in molecular biology (Clifton, N.J.)*, *850*, 1–9. https://doi.org/10.1007/978-1-61779-555-8_1

Fieller E. C., (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B(Methodogical)* (Vol. 16) 175-185. https://doi.org/10.1111/j.2517-6161.1954.tb00159.x

Fisch G. S. (2017). Whither the genotype-phenotype relationship? An historical and methodological appraisal. *American journal of medical genetics. Part C, Seminars in medical genetics*, *175*(3), 343–353. https://doi.org/10.1002/ajmg.c.31571

Galanis P. (2010). *The wrong application of P values and hypotheses test in biomedical research*.

Haidich A. B. (2010). Meta-analysis in medical research. *Hippokratia*, *14*(Suppl 1), 29–37.

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Hajat C. (2011). An introduction to epidemiology. *Methods in molecular biology (Clifton, N.J.)*, *713*, 27–39. https://doi.org/10.1007/978-1-60327-416-6_3

Hirschhorn, J. N., Lohmueller, K., Byrne, E., & Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in medicine : official journal of the American College of Medical Genetics*, *4*(2), 45–61. https://doi.org/10.1097/00125817-200203000-00002

Ho, C. H., Huang, Y. J., Lai, Y. J., Mukherjee, R., & Hsiao, C. K. (2022). The misuse of distributional assumptions in functional class scoring gene-set and pathway analysis. *G3 (Bethesda, Md.)*, *12*(1), jkab365. https://doi.org/10.1093/g3journal/jkab365

Homocysteine Studies Collaboration. (2002). Homocysteine and risk of ischemic heart disease and stroke: A meta-analysis. *Journal of the American Medical Association, 288*(15), 2015–2022. doi:10.1001/jama.288.16.2015

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (3rd ed.). Monash University, Australia

Keavney B. (2002). Genetic epidemiological studies of coronary heart disease. *International journal of epidemiology*, *31*(4), 730–736. https://doi.org/10.1093/ije/31.4.730

Kendall, M., & Stuart, A. (1973). *The advanced theory of statistics: Distribution theory* (Vol. 1). London: Griffin. https://doi.org/10.1038/159148a0

Kingdom, R., & Wright, C. F. (2022). Incomplete Penetrance and Variable Expressivity: From Clinical Studies to Population Cohorts. *Frontiers in genetics*, *13*, 920390. https://doi.org/10.3389/fgene.2022.920390

Klerk, M., Verhoef, P., Clarke, R., Blom, H. J., Kok, F. J., & Schouten, E. G. (2002). MTHFR 677C→T polymorphism and risk of coronary heart disease: A meta-analysis. *Journal of the American Medical Association, 288*(15), 2023–2031. doi:10.1001/jama.288.16.2023

Midgley, J. P., Matthew, A. G., Greenwood, C. M., & Logan, A. G. (1996). Effect of reduced dietary sodium on blood pressure: a meta-analysis of randomized controlled trials. *JAMA*, *275*(20), 1590–1597. https://doi.org/10.1001/jama.1996.03530440070039

Minelli, C. (2004). An integrated approach to the meta-analysis of genetic association studies using Mendelian randomization. *American Journal of Epidemiology, 160*(5), 445–452. https://doi.org/10.1093/aje/kwh228

Mounier, N., & Kutalik, Z. (2023). Bias correction for inverse variance weighting Mendelian randomization. *Genetic epidemiology*, *47*(4), 314–331. https://doi.org/10.1002/gepi.22522

Rader J. I. (2002). Folic acid fortification, folate status and plasma homocysteine. *The Journal of nutrition*, *132*(8 Suppl), 2466S–2470S. https://doi.org/10.1093/jn/132.8.2466S

Smith, G. D., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

epidemiology contribute to understanding environmental determinants of disease?. *International journal of epidemiology*, *32*(1), 1–22. https://doi.org/10.1093/ije/dyg070

Smith, G. D., & Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human molecular genetics*, *23*(R1), R89–R98. https://doi.org/10.1093/hmg/ddu328

Smith, G. D., Song, F., & Sheldon, T. A. (1993). Cholesterol lowering and mortality: the importance of considering initial level of risk. *BMJ (Clinical research ed.)*, *306*(6889), 1367–1373. https://doi.org/10.1136/bmj.306.6889.1367

Sterne, J. A., Egger, M., & Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ (Clinical research ed.)*, *323*(7304), 101–105. https://doi.org/10.1136/bmj.323.7304.101

Storey, E. C., & Fuller, H. R. (2022). Genotype-phenotype correlations in human diseases caused by mutations of LINC complex-associated genes: A systematic review and meta-summary. *Cells, 11*(24), 4065. https://doi.org/10.3390/cells11244065

Sutton, A. J., Abrams, K. R., & Jones, D. R. (2002). Generalized synthesis of evidence and the threat of dissemination bias. the example of electronic fetal heart rate monitoring (EFM). *Journal of clinical epidemiology*, *55*(10), 1013–1024. https://doi.org/10.1016/s0895-4356(02)00460-2

Sutton, A., J., Abrams, K., R., Jones, D., R., Sheldon, T., & Song, F. (2000). *Methods for meta-analysis in medical research*. Chichester, UK: John Wiley and Sons Ltd.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0127-1

Tennant, P. W. G., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., Harrison, W. J., Keeble, C., Ranker, L. R., Textor, J., Tomova, G. D., Gilthorpe, M. S., & Ellison, G. T. H. (2021). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International journal of epidemiology*, *50*(2), 620–632. https://doi.org/10.1093/ije/dyaa213

Thomson, G. (1947). The Advanced Theory of Statistics. *Nature 159*, 148. https://doi.org/10.1038/159148a0

Thompson, J. R., Minelli, C., Abrams, K. R., Tobin, M. D., & Riley, R. D. (2005). Meta-analysis of genetic studies using Mendelian randomization--a multivariate approach. *Statistics in medicine*, *24*(14), 2241–2254. https://doi.org/10.1002/sim.2100

Wald, D. S., Law, M., & Morris, J. K. (2002). Homocysteine and cardiovascular disease: Evidence on causality from a meta-analysis. *British Medical Journal, 325*(7374), 1202. doi: 10.1136/bmj.325.7374.1202

Welz, T., Doebler, P., & Pauly, M. (2022). Fisher transformation based confidence intervals of correlations in fixed- and random-effects meta-analysis. *The British journal of mathematical and statistical psychology*, *75*(1), 1–22. https://doi.org/10.1111/bmsp.12242

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156

Zou, L., Guo, H., & Berzuini, C. (2022). Bayesian mendelian randomization with study heterogeneity and data partitioning for large studies. *BMC medical research methodology*, *22*(1), 162. https://doi.org/10.1186/s12874-022-01619-4

Institutional Repository - Library & Information Centre - University of Thessaly
02/09/2026 09:50:04 EEST - 216.73.216.156