ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ
ΤΟΜΕΑΣ ΠΙΘΑΝΟΤΗΤΩΝ, ΣΤΑΤΙΣΤΙΚΗΣ
& ΕΠΙΧΕΙΡΗΣΙΑΚΗΣ ΕΡΕΥΝΑΣ

ΔΙΕΡΕΥΝΗΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΓΙΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΙΚΟΥ ΣΦΑΛΜΑΤΟΣ ΔΗΜΟΣΙΕΥΣΗΣ ΚΑΙ ΤΗΣ ΕΠΙΔΡΑΣΗΣ ΜΙΚΡΩΝ ΜΕΛΕΤΩΝ ΣΤΑ ΜΟΝΤΕΛΑ ΜΕΤΑ-ΑΝΑΛΥΣΗΣ

ΜΠΑΗ ΙΩΑΝΝΑ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννινα, 2018

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη

ΣΤΑΤΙΣΤΙΚΗ ΚΑΙ ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΕΡΕΥΝΑ

που απονέμει το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 22/06/2018 από την Τριμελή Επιτροπή Εξέτασης:

ΟΝΟΜΑΤΕΠΩΝΥΜΟ	**ΒΑΘΜΙΔΑ**
Μαυρίδης Δημήτριος	Επίκουρος Καθηγητής του Παιδαγωγικού Τμήματος Δημοτικής Εκπαίδευσης του Πανεπιστημίου Ιωαννίνων (Επιβλέπων Καθηγητής)
Λουκάς Σωτήριος	Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων
Μπατσίδης Απόστολος	Επίκουρος Καθηγητής του Τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

″Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασία της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα.″

Μπάη Ιωάννα

Αφιερώνεται στην οικογένειά μου.

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Δημήτριο Μαυρίδη για την πολύτιμη βοήθεια και υποστήριξη καθ' όλη την διάρκεια της μεταπτυχιακής διατριβής. Τον ευχαριστώ θερμά για την άριστη συνεργασία καθώς και για την εμπιστοσύνη που έδειξε στο πρόσωπό μου. Ανάλογες θερμότατες ευχαριστίες θα ήθελα να απευθύνω στα μέλη της κριτικής επιτροπής κ. Σωτήριο Λουκά και κ. Απόστολο Μπατσίδη καθώς και στους καθηγητές του Τομέα Στατιστικής και Επιχειρησιακής Έρευνας κ. Κωνσταντίνο Ζωγράφο και κ. Κωνσταντίνα Σκούρη για τις χρήσιμες συμβουλές τους κατά την διάρκεια του μεταπτυχιακού προγράμματος. Τέλος, οφείλω ένα μεγάλο ευχαριστώ στην οικογένειά μου για την ηθική συμπαράσταση σε όλα τα χρόνια των σπουδών μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1
ΕΙΣΑΓΩΓΗ

Σκοπός του κεφαλαίου είναι η περιγραφή των διαφόρων τύπων έρευνας που συνιστούν την αποδεικτική επιστήμη (evidence-based science). Συγκεκριμένα, αναφέρονται τα βασικά χαρακτηριστικά των ερευνών και δίνεται έμφαση στις μεταξύ τους διαφορές. Περιγράφεται ακόμα η βασική δομή του περιεχομένου των κεφαλαίων της μεταπτυχιακής διατριβής καθώς και ο σκοπός εκπόνησής της.

1.1 Πυραμίδα της αποδεικτικής επιστήμης

Ο απώτερος στόχος της επιστήμης αποτελεί η βελτίωση της ποιότητας της ζωής του σύγχρονου ανθρώπου. Το κύριο εργαλείο που συμβάλει στη κατεύθυνση αυτή είναι η έρευνα. Με τον όρο έρευνα αναφερόμαστε σε μια συστηματική και καλώς σχεδιασμένη διαδικασία για την επίλυση προβλημάτων (Παρασκευόπουλος, 1993). Για να θεωρηθούν τα ευρήματα μιας έρευνας έγκυρα, είναι απαραίτητη η διερεύνηση των συνθηκών κάτω από τις οποίες έχει διεξαχθεί. Κάτω από το πλαίσιο αυτό στην βιοστατιστική έχει διαμορφωθεί η πυραμίδα της αποδεικτικής επιστήμης (Isenburg, 2015) (*Εικόνα 1.1*).

Στην βάση της πυραμίδας βρίσκονται οι έρευνες σε ζώα (animal research) και οι μεμονωμένες περιπτώσεις ασθενών (case reports). Έχουν μικρή στατιστική ισχύ και δεν μπορούν να θεωρηθούν ιδιαίτερα αξιόπιστες, επειδή δεν χρησιμοποιούν ομάδες ελέγχου για την σύγκριση των αποτελεσμάτων.
Στην επόμενη θέση βρίσκονται οι μελέτες ασθενών-μαρτύρων (case control studies). Σε αυτού του τύπου τις μελέτες παρατηρούμε μία ομάδα που έχει κάποια έκβαση σε σχέση με μία άλλη ομάδα που δεν την έχει. Ο ερευνητής κατά την διάρκειά τους προσπαθεί να εντοπίσει παράγοντες, που πιθανόν να σχετίζονται με την έκβαση του ασθενούς, προκειμένου να συνθέσει μία

στατιστική σχέση μεταξύ των παραγόντων και της έκβασης. Είναι λιγότερο αξιόπιστες από τις μελέτες κοόρτης και τις τυχαιοποιημένες κλινικές δοκιμές, που θα αναφερθούν στην συνέχεια, αφού η στατιστική σχέση μεταξύ των παραγόντων δεν μπορεί να θεωρηθεί καθολική (Isenburg, 2015).

Μετά τις μελέτες ασθενών-μαρτύρων βρίσκονται οι μελέτες κοόρτης (cohort studies). Σε αυτού του τύπου τις μελέτες τα άτομα ταξινομούνται ανάλογα με την έκθεσή τους ή μη σε κάποιον παράγοντα και παρακολουθούνται για μία χρονική περίοδο, για να δούμε αν θα παρατηρήσουμε την έκβαση που μας ενδιαφέρει. Πρόκειται για ένα είδος μελετών παρατήρησης, αφού η βασική ιδέα είναι να παρατηρούμε ένα πληθυσμό, που έχει ένα κοινό χαρακτηριστικό για μεγάλο χρονικό διάστημα. Είναι λιγότερο αξιόπιστες από τις τυχαιοποιημένες κλινικές δοκιμές, που θα αναφερθούν στην συνέχεια, καθώς οι δύο ομάδες που ταξινομούνται τα άτομα μπορεί να διαφέρουν ως προς διαφορετικά χαρακτηριστικά εκτός από την υπό μελέτη έκβαση (Isenburg, 2015).

Εξέχουσα θέση στην πυραμίδα της αποδεικτικής επιστήμης κατέχουν οι τυχαιοποιημένες κλινικές δοκιμές (randomized controlled trials). Θεωρούνται από την επιστημονική κοινότητα ως ο πιο έγκυρος τύπος έρευνας (gold standard). Κατά την διεξαγωγή τους τα άτομα χωρίζονται τυχαία σε δύο ομάδες (ή ενδεχομένως και σε περισσότερες), την πειραματική ομάδα (treatment group) και την ομάδα ελέγχου (control group). Στην πειραματική ομάδα γίνεται μία παρέμβαση Α και στην ομάδα ελέγχου γίνεται μία παρέμβαση Β ή μια εικονική παρέμβαση (placebo). Αφού τα άτομα έχουν κατανεμηθεί τυχαία στις δύο ομάδες, η όποια διαφορά σε κάποιο αποτέλεσμα οφείλεται αποκλειστικά στην διαφορετική παρέμβαση που χρησιμοποιήθηκε. Οι τυχαιοποιημένες κλινικές δοκιμές, αν και για την επιστημονική κοινότητα αποτελούν τον πιο έγκυρο τύπο έρευνας, δεν κατέχουν την κορυφή της επιστημονικής πυραμίδας, αφού τίθεται το ερώτημα αν είναι ασφαλής η εξαγωγή συμπερασμάτων από την κάθε μελέτη

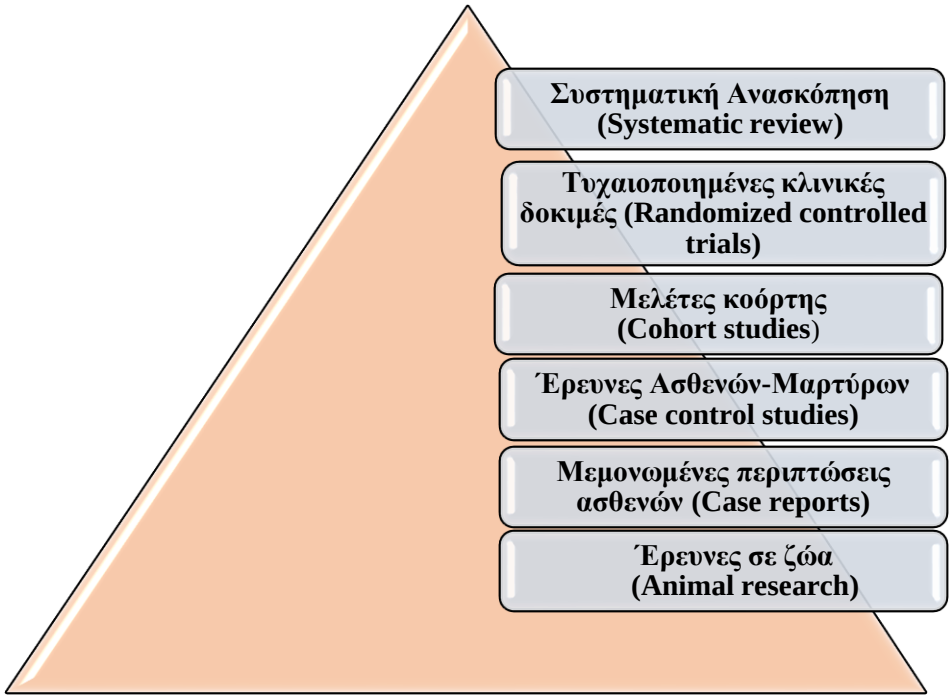
χωριστά. Για ένα επιστημονικό ζήτημα, ιδίως στο χώρο της ιατρικής, είναι σύνηθες να έχουν πραγματοποιηθεί διάφορες μελέτες. Πολλές από αυτές μπορεί να δίνουν μεροληπτικά ή αντικρουόμενα αποτελέσματα, τα οποία να οδηγούν σε παραπλανητικά συμπεράσματα (Higgins & Green, 2008).

Με βάση τα παραπάνω, προέκυψε η ανάγκη της σύνθεσης των ευρημάτων των μελετών με στόχο την έγκυρη εξαγωγή συμπερασμάτων. Στο πλαίσιο αυτό αναπτύχθηκαν οι συστηματικές ανασκοπήσεις (systematic reviews), οι οποίες αποτελούν ερευνητικές μεθόδους, που επιτρέπουν την σύνθεση των αποτελεσμάτων των μεμονωμένων ερευνών. Κατέχουν δίκαια την κορυφή της επιστημονικής πυραμίδας, αφού παρέχουν αξιόπιστα αποτελέσματα με άμεση εφαρμογή στον επιστημονικό χώρο και κυρίως στην ιατρική. Στο Κεφάλαιο 2 παρατίθενται λεπτομέρειες, που αφορούν τον σχεδιασμό και την διεξαγωγή μιας έγκυρης συστηματικής ανασκόπησης. Στο σημείο αυτό είναι ανάγκη να διευκρινιστεί, ότι η μονάδα παρατήρησης στην συστηματική ανασκόπηση δεν είναι οι συμμετέχοντες αλλά οι μελέτες. Τα δεδομένα του εκάστοτε συμμετέχοντα, πάνω στα οποία στηρίζονται οι πρωτογενείς μελέτες, σπανίως είναι διαθέσιμα. Η πληροφορία που συνήθως δίνεται σε κάποια δημοσίευση είναι τα μέτρα σχέσης/μεγέθη επίδρασης (effect sizes), για τα οποία πραγματοποιείται εκτενής αναφορά στο Κεφάλαιο 3.

Η συστηματική ανασκόπηση είναι άρρηκτα συνδεδεμένη με την έννοια της μετα-ανάλυσης (meta-analysis), η οποία αποτελεί την στατιστική διαδικασία σύνθεσης των μελετών και έχει ως στόχο την παροχή ενός συγκεντρωτικού αποτελέσματος. Για πρώτη φορά εφαρμόστηκε από το 1904 από τον Karl Pearson σε μία προσπάθεια αντιμετώπισης της μειωμένης στατιστικής ισχύος, που παρουσιάζουν μελέτες με μικρό αριθμό συμμετεχόντων (Γαλάνης, 2009). Ο όρος μετα-ανάλυση εισήχθη το 1976 από το Gene Glass (Γαλάνης, 2009). Πληροφορίες για τον τρόπο σύνθεσης των μελετών και τον υπολογισμό του συγκεντρωτικού αποτελέσματος αναφέρονται στο Κεφάλαιο 4. Επιπλέον, στο

ίδιο κεφάλαιο γίνεται αναφορά στην έννοια της ετερογένειας (heterogeneity), η οποία σχετίζεται με τον βαθμό ανομοιογένειας των μελετών, που λαμβάνουν μέρος σε μια μετα-ανάλυση καθώς και στην μέθοδο της μετα-παλινδρόμησης (meta-regression), η οποία δίνει τη δυνατότητα διερεύνησης της ετερογένειας.

Παρόλα τα πλεονεκτήματα, η στατιστική μέθοδος της μετα-ανάλυσης παρουσιάζει ορισμένες αδυναμίες. Ξεκινώντας από την συλλογή των μελετών, υπάρχει ο κίνδυνος να μην έχουν συμπεριληφθεί μελέτες, που κατέληξαν σε στατιστικώς μη σημαντικά αποτελέσματα. Πρόκειται για το συστηματικό σφάλμα δημοσίευσης (publication bias). Ακόμα, ενδέχεται οι μικρές σε μέγεθος δείγματος μελέτες να δείχνουν συστηματικά μεγαλύτερα μεγέθη επίδρασης σε σχέση με τις μεγαλύτερες μελέτες (small-study effect). Λεπτομέρειες για τις παραπάνω αδυναμίες καθώς και τρόποι εξακρίβωσής τους παρατίθενται στο Κεφάλαιο 5. Ενδεικτικά αναφέρουμε, ότι μια πρώτη ένδειξη για τον εντοπισμό τους προκύπτει από την ασυμμετρία του διαγράμματος κώνου (funnel plot), το οποίο είναι ένα απλό διάγραμμα σημείων του μεγέθους επίδρασης έναντι ενός μέτρου ακρίβειας, συνήθως του τυπικού σφάλματος της μελέτης. Έχει προταθεί ένας μεγάλος αριθμός από στατιστικά test για τον έλεγχο της ασυμμετρίας του διαγράμματος κώνου μεταξύ των οποίων τα Egger's, Begg's rank correlation και Thompson & Sharp's tests. Στο Κεφάλαιο 6 πραγματοποιείται μελέτη προσομοίωσης, που έχει ως στόχο την σύγκριση και την ανάδειξη των πιο αποτελεσματικών από τα παραπάνω test.



Εικόνα 1.1: Ενδεικτική πυραμίδα της αποδεικτικής επιστήμης

1.2 Σκοπός της μεταπτυχιακής διατριβής

Η μέθοδος της μετα-ανάλυσης κερδίζει ολοένα και περισσότερους υποστηριχτές στο χώρο της ιατρικής, αφού καθοριστική είναι η συμβολή της στην βελτίωση της κλινικής πρακτικής. Ωστόσο, η μέθοδος έχει δεχτεί έντονη κριτική. Στο πλαίσιο αυτό γίνεται αναφορά στην επίδραση του συστηματικού σφάλματος δημοσίευσης (publication bias) καθώς και στην επίδραση των μικρών σε μέγεθος μελετών στα αποτελέσματά της (small-study effect). Έχοντας την πεποίθηση ότι οι παραπάνω κριτικές δεν έχουν θέση στην περίπτωση άρτια εκτελέσιμων μετα-αναλύσεων, πραγματοποιείται ανασκόπηση και διερεύνηση των διαφόρων στατιστικών μεθόδων ανίχνευσής τους.

ΚΕΦΑΛΑΙΟ 2
ΣΥΣΤΗΜΑΤΙΚΗ ΑΝΑΣΚΟΠΗΣΗ
(SYSTEMATIC REVIEW)

Αντικείμενο του κεφαλαίου είναι η περιγραφή της μεθοδολογίας της συστηματικής ανασκόπησης (systematic review). Κάτω από το πρίσμα αυτό ακολουθεί αναλυτική περιγραφή των βασικών αρχών που διέπουν τον σχεδιασμό και την υλοποίηση μιας συστηματικής ανασκόπησης.

2.1 Είδη ανασκόπησης

Στο χώρο της ιατρικής η βελτίωση της κλινικής πρακτικής καθιστά επιτακτική την έγκυρη διεξαγωγή συμπερασμάτων από τα αποτελέσματα του μεγάλου όγκου μελετών, που έχουν διεξαχθεί πάνω σε επιστημονικά ζητήματα. Τίθεται λοιπόν η ανάγκη της επαναξιολόγησης των υπαρχόντων ερευνητικών δεδομένων, που έχουν προκύψει από τις διάφορες μελέτες. Τον σκοπό αυτό υπηρετεί η ανασκόπηση. Με τον όρο ανασκόπηση αναφερόμαστε σε μια δευτερογενή ερευνητική εργασία, που αποτελεί περίληψη της βιβλιογραφίας και έχει ως στόχο να συγκεντρώσει τα διαθέσιμα δεδομένα, που απαντούν σε ένα συγκεκριμένο ερευνητικό ερώτημα (Antman, 1992). Οι ανασκοπήσεις διακρίνονται σε αφηγηματικές (narrative review) και συστηματικές (systematic review). Όσον αφορά τις αφηγηματικές ανασκοπήσεις, δεν διαθέτουν συγκεκριμένη δομή, προσεγγίζουν ένα επιστημονικό ερώτημα με υποκειμενικό τρόπο και δεν λαμβάνουν υπόψιν παράγοντες που σχετίζονται με τον σχεδιασμό, την ποιότητα και το μέγεθος του δείγματος των πρωτογενών μελετών. Η χρήση τους έχει απορριφθεί από την πλειονότητα των επιστημόνων (Γαλάνης, 2009). Αντίθετα, οι συστηματικές ανασκοπήσεις πραγματοποιούνται με αυστηρά κριτήρια και στοχεύουν στον καθορισμό των καλύτερα μεθοδολογικά σχεδιασμένων μελετών (Γαλάνης, 2009). Το θεμελιώδες στοιχείο, που καθιστά μια ανασκόπηση συστηματική, είναι η ενδελεχής

αναζήτηση της βιβλιογραφίας (Καράσσα, 2006). Τα αποτελέσματα μιας συστηματικής ανασκόπησης είναι τα πλέον ενδεδειγμένα για εφαρμογή στην ιατρική πρακτική (Μπελλάλη, 2011). Τέλος, καθοριστική είναι η συμβολή της στην αποσαφήνιση αβέβαιων ή ελλιπών επιστημονικών θεμάτων καθώς και στην αναζήτηση νέων ερευνητικών κατευθύνσεων.

2.2 Συστηματική ανασκόπηση (Systematic Review)

Οι πρώτοι που άρχισαν να ασχολούνται συστηματικά με την κριτική αξιολόγηση των δημοσιευμένων ανασκοπήσεων και προσπάθησαν να καθορίσουν τα στάδια, που πρέπει να ακολουθούνται, προκειμένου να αυξηθεί η εγκυρότητα τους, ήταν οι κοινωνικοί επιστήμονες στις αρχές τις δεκαετίας του 1970 (Μπελλάλη, 2011). Ακολουθώντας το παράδειγμά τους, τα τελευταία 20 χρόνια οι επιστήμονες στο χώρο της υγείας άρχισαν να δίνουν ιδιαίτερη προσοχή στην ποιότητα και να εκπονούν βιβλιογραφικές ανασκοπήσεις με αυστηρά μεθοδολογικό τρόπο (Μπελλάλη, 2011). Μια ανασκόπηση, για να θεωρηθεί αυστηρά συστηματική, θα πρέπει να διεξάγεται με βάση ένα πρωτόκολλο, το οποίο περιλαμβάνει στάδια, που διέπονται από επιστημονικές αρχές και καθορίζουν τον σχεδιασμό και την υλοποίηση της. Τα στάδια της συστηματικής ανασκόπησης, όπως αυτά περιγράφονται σε εγχειρίδια για επιστήμονες του διεθνή μη κερδοσκοπικού οργανισμού Cochrane Collaboration, είναι τα παρακάτω:

2.2.1 Διατύπωση ερευνητικού ερωτήματος (Defining the review question)

Το πιο σημαντικό βήμα της συστηματικής ανασκόπησης αποτελεί η διατύπωση του ερευνητικού ερωτήματος, το οποίο καλείται να απαντηθεί από την έρευνα. Το ερευνητικό ερώτημα πρέπει να είναι σαφές και επιστημονικά τεκμηριωμένο (Πατελάρου & Μπροκαλάκη, 2010). Ακόμα, ο καθορισμός του πρέπει να γίνεται με τέτοιον τρόπο, ώστε να προσδιορίζεται ο τύπος των συμμετεχόντων,

τα είδη των παρεμβάσεων και τα είδη των αποτελεσμάτων (Higgins & Green, 2008). Συγκεκριμένα, ο τύπος των συμμετεχόντων πρέπει να περιορίζεται σε μία συγκεκριμένη ομάδα ατόμων και το εύρος του να καθορίζεται από την ποικιλομορφία των μελετών (Higgins & Green, 2008). Σχετικά με τα είδη των παρεμβάσεων οι μελετητές οφείλουν να ορίζουν τις υπό μελέτη θεραπείες είτε αυτές είναι δραστικές (χορήγηση φαρμάκων) είτε μη δραστικές (placebo) (Higgins & Green, 2008). Αναφορικά με τα είδη των αποτελεσμάτων (ωφέλιμα ή επιβλαβή) που θα αποτελέσουν το αντικείμενο της μελέτης, απαιτείται ο σαφής και ακριβής προσδιορισμός τους (Higgins & Green, 2008). Τέλος, παράλληλα με την διατύπωση του ερωτήματος είναι ανάγκη να αποσαφηνίζεται και ο σκοπός για τον οποίο θα πραγματοποιηθεί η ερευνητική ανασκόπηση (Higgins & Green, 2008).

2.2.2 Καθορισμός κριτηρίων εισαγωγής και απόκλισης μίας μελέτης (Defining eligibility criteria)

Ένα σημαντικό βήμα για την άρτια διεξαγωγή της συστηματικής ανασκόπησης αποτελεί ο καθορισμός των κριτηρίων εισαγωγής και απόκλισης μιας μελέτης. Ο αυστηρός καθορισμός των κριτηρίων ένταξης και αποκλεισμού αποσκοπεί στην επιλογή εκείνων των μελετών, που θεωρούνται πιο κατάλληλες για την πραγματοποίηση της συστηματικής ανασκόπησης. Τα κριτήρια αυτά αναφέρονται στα χαρακτηριστικά των συμμετεχόντων, στο είδος της παρέμβασης, στις μεθόδους σύγκρισης, στην έκβαση αλλά και στο είδος των μελετών (μελέτες κοόρτης, ασθενών-μαρτύρων) (Πατελάρου & Μπροκαλάκη, 2010). Πέρα από τα παραπάνω, ως κριτήρια αποκλεισμού συχνά ορίζονται και οι μελέτες, που η θεραπευτική τους παρέμβαση μεταβλήθηκε κατά τον χρόνο καθώς και οι αδημοσίευτες μελέτες, δηλαδή διδακτορικές διατριβές, μελέτες συνεδριών ή μελέτες με αρνητική έκβαση (Γαλάνης, 2009). Τα κριτήρια εισαγωγής και αποκλεισμού των μελετών παρατίθενται στο πρωτόκολλο της

συστηματικής ανασκόπησης, οφείλουν να είναι αντικειμενικά και να μην υπόκεινται στην υποκειμενικότητα των ερευνητών (Γαλάνης, 2009).

2.2.3 Αναζήτηση Βιβλιογραφίας (Searching for studies)

Βασική προϋπόθεση, για να χαρακτηριστεί μια συστηματική ανασκόπηση αξιόπιστη, είναι η δυνατότητα επανάληψη της. Στο πλαίσιο αυτό θεωρείται αναγκαία η αναλυτική περιγραφή της βιβλιογραφίας που ακολουθήθηκε (Γαλάνης, 2009). Οι ερευνητές συχνά στρέφονται στο διαδίκτυο και με λέξεις-κλειδιά αναζητούν πληροφορίες σχετικές με το επιστημονικό τους αντικείμενο πάνω σε μελέτες ακαδημαϊκών, κρατικών ή ιδιωτικών οργανισμών (Γαλάνης, 2009). Στην διευκόλυνση της αναζήτησης πληροφοριών στον χώρο της βιοϊατρικής έχει συμβάλει ο διεθνής μη κερδοσκοπικός οργανισμός Cochrane Collaboration με την ανάπτυξη της βάσης δεδομένων CENTRAL (The Cochrane Central Register of Controlled Trials), η οποία παρέχει στους ερευνητές πληθώρα ελεγχόμενων δοκιμών (Higgins & Green, 2008). Άλλες ηλεκτρονικές βάσεις δεδομένων, που μπορεί να πραγματοποιηθεί αναζήτηση βιβλιογραφίας για βιοϊατρικές μελέτες, είναι οι MEDLINE (Medical Literature Analysis and Retrieval System Online), EMBASE (The Excerpta Medical Database), CINAHL (Dissertation Abstracts International), CANCERLIT (Cancer Literature), TOXLINE (Toxicology Literature Online), DAI (Dissertation Abstracts International) και το Εθνικό Κέντρο Τεκμηρίωσης (National Documentation Centre) (Γαλάνης, 2009). Τέλος, αναζήτηση μπορεί να πραγματοποιηθεί σε ειδικές βιβλιογραφίες ανάλογα με το επιστημονικό αντικείμενο της έρευνας (Subject-specific electronic bibliographic datasets) καθώς και σε διεθνείς βάσεις δεδομένων (National and regional datasets), στις οποίες είναι διαθέσιμη η βιβλιογραφία, που παράχθηκε από την τοπική κοινωνία (Higgins & Green, 2008).

2.2.4 Αξιολόγηση και επιλογή των μελετών (Selecting studies)

Μετά την αναζήτηση της βιβλιογραφίας ακολουθεί η αξιολόγηση και η επιλογή εκείνων των ερευνών, που έχουν διεξαχθεί με τον πιο μεθοδολογικό τρόπο. Η ανάγκη αυτή απορρέει από το γεγονός, ότι η εισαγωγή κακής ποιότητας μελετών στην μετα-ανάλυση οδηγεί σε λανθασμένα συμπεράσματα (garbage in-garbage out). Για τον εντοπισμό των πιο κατάλληλων μελετών ακολουθείται από τους ερευνητές η παρακάτω διαδικασία. Αρχικά, γίνεται σύνθεση των άρθρων που προέρχονται από την ίδια την μελέτη. Στην συνέχεια, απορρίπτονται τα μη σχετικά με το ερευνητικό ερώτημα άρθρα με βάση τον τίτλο ή την περίληψή τους. Ακολουθεί η ανάκτηση του πλήρους κειμένου και πραγματοποιείται έλεγχος για την ικανοποίηση των κριτηρίων καταλληλότητας της έρευνας. Τέλος, αιτιολογείται η απόφαση επιλογής ή μη της μελέτης στην ανάλυση (Higgins & Green, 2008). Η Cochrane περιλαμβάνει μία λίστα απόρριψης μελετών με τον κύριο λόγο αποκλεισμού τους (Higgins & Green, 2008). Για την αποφυγή της μεροληψίας του ερευνητή η όλη διαδικασία πραγματοποιείται από δύο ανεξάρτητους ερευνητές και σε περίπτωση ασυμφωνίας το πρόβλημα επιλύεται είτε μετά από σύγκριση είτε από έναν τρίτο ερευνητή (Πατελάρου & Μπροκαλάκη, 2010).

2.2.5 Καταγραφή δεδομένων (Collecting data)

Στο στάδιο αυτό οι ερευνητές καταγράφουν τα βασικά χαρακτηριστικά καθώς και πληροφορίες για την αξιολόγηση του κινδύνου μεροληψίας των συμπεριλαμβανομένων ερευνών σε προκαθορισμένες πλατφόρμες, προκειμένου να εκτιμηθεί ο βαθμός ομοιότητας μεταξύ τους (Πατελάρου & Μπροκαλάκη, 2010). Ο βαθμός ομοιότητας ταυτίζεται με την έννοια της ετερογένειας μεταξύ των μελετών, η ερμηνεία της οποίας παρατίθεται στο Κεφάλαιο 4. Στα βασικά χαρακτηριστικά των μελετών συμπεριλαμβάνονται διάφορες πληροφορίες για τους συμμετέχοντες, το αντικείμενο της μελέτης καθώς και το είδος των

θεραπειών (φαρμακολογικές ή μη) (Higgins & Green, 2008). Στις φαρμακολογικές θεραπείες επικεντρωνόμαστε κυρίως στην δοσολογία των φαρμάκων καθώς και στην διάρκεια της θεραπείας, ενώ αντίστοιχα στις μη φαρμακολογικές θεραπείες μας ενδιαφέρει το περιεχόμενο και ο τρόπος διεξαγωγής τους (Higgins & Green, 2008). Πέρα από τα παραπάνω, στα βασικά χαρακτηριστικά συγκαταλέγονται οι μετρήσεις των αποτελεσμάτων, η ακεραιότητα των παρεμβάσεων καθώς και η παρουσία ανεπιθύμητων αποτελεσμάτων (adverse outcomes) (Higgins & Green, 2008). Οι πληροφορίες για την αξιολόγηση του κινδύνου μεροληψίας εστιάζονται στον έλεγχο της μελέτης ως προς την ύπαρξη ενός μη ολοκληρωμένου αποτελέσματος (incomplete outcome), στην χρήση ή μη ερωτηματολογίου στους συμμετέχοντες (bleeding) και την επιλεκτική αναφορά του αποτελέσματος (selective outcome reporting) (Higgins & Green, 2008). Ο οργανισμός Cochrane Collaboration στην προσπάθεια διευκόλυνσης των ερευνητών παρέχει τις πληροφορίες, που αφορούν το κίνδυνο μεροληψίας με την βοήθεια της χρήσης του εργαλείου Cochrane Collaboration tool (Higgins & Green, 2008). Για την αποφυγή σφαλμάτων η όλη διαδικασία πραγματοποιείται από δύο ανεξάρτητους ερευνητές (Πατελάρου & Μπροκαλάκη, 2010).

2.2.6 Στατιστική ανάλυση (Statistical analysis)

Μετά την καταγραφή των δεδομένων, το επόμενο βήμα της συστηματικής ανασκόπησης αποτελεί η στατιστική ανάλυση. Ο ρόλος της είναι ιδιαίτερα σημαντικός, αφού επιδιώκει να συνθέσει όλα τα ευρήματα των υπαρχουσών μελετών σε ένα συγκεντρωτικό αποτέλεσμα (Πατελάρου & Μπροκαλάκη, 2010). Η στατιστική μέθοδος με την οποία πραγματοποιείται η σύνθεση των μελετών, που επιλέγονται από την συστηματική ανασκόπηση, με σκοπό την δημιουργία ενός συγκεντρωτικού αποτελέσματος, καλείται μετα-ανάλυση (Γαλάνης, 2009). Η μετα-ανάλυση επιτυγχάνει την αύξηση της στατιστικής ισχύος και της ακρίβειας σε σχέση με τις επιμέρους μελέτες (Higgins & Green,

2008). Με την βοήθειά της ανιχνεύονται και ποσοτικοποιούνται οι παράμετροι, ως προς τις οποίες οι μελέτες διαφέρουν μεταξύ τους (Higgins & Green, 2008). Ακόμα, επιτρέπει να μελετώνται θέματα για τα οποία δεν θα μπορούσαμε να έχουμε πληροφορίες από την κάθε μελέτη χωριστά, ενώ ταυτόχρονα συμβάλει στην διαμόρφωση νέων ερευνητικών ερωτημάτων για μελλοντική διερεύνηση (Higgins & Green, 2008). Παρόλα τα θετικά στοιχεία που παρουσιάζει, η μετα-ανάλυση δεν είναι υποχρεωτική στο πλαίσιο της συστηματικής ανασκόπησης. Πολλές φορές η άσκοπη χρήση της οδηγεί σε παραπλανητικά συγκεντρωτικά αποτελέσματα. Συγκεκριμένα, η εφαρμογή της δεν συνίσταται σε περιπτώσεις μεγάλης ποικιλομορφίας των μελετών (ως προς τους συμμετέχοντες, τον τύπο των παρεμβάσεων κλπ.) καθώς και όταν οι υπάρχουσες μελέτες παρουσιάζουν υψηλό κίνδυνο μεροληψίας (Higgins & Green, 2008). Αν δεν πραγματοποιηθεί μετα-ανάλυση κατά την διάρκεια της συστηματικής ανασκόπησης, οι ερευνητές παρουσιάζουν τα ευρήματα των μελετών χωριστά, χωρίς να γίνεται αναφορά σε κάποιο συγκεντρωτικό αποτέλεσμα (Πατελάρου & Μπροκαλάκη, 2010).

2.2.7 Παρουσίαση αποτελεσμάτων (Presenting results)

Στο στάδιο αυτό παρουσιάζονται σε έναν πίνακα τα βασικά χαρακτηριστικά των μελετών που συμπεριελήφθησαν στην ανασκόπηση, όπως ο αριθμός συμμετεχόντων, η μέθοδος διεξαγωγής, οι διάφορες θεραπείες που χρησιμοποιήθηκαν με τα αποτελέσματά τους καθώς και το είδος μεροληψίας της εκάστοτε μελέτης (Higgins & Green, 2008). Αναφέρεται επίσης ο ακριβής αριθμός των ανακτηθέντων μελετών στο στάδιο της αναζήτησης, ο αριθμός των μελετών που δεν συμπεριελήφθησαν στην ανασκόπηση καθώς και οι αιτίες αποκλεισμού τους (Πατελάρου & Μπροκαλάκη, 2010). Τέλος, στην περίπτωση που πραγματοποιηθεί στατιστική ανάλυση κατά την διάρκεια της ανασκόπησης, απαιτείται η αναλυτική παρουσίαση των συγκεντρωτικών αποτελεσμάτων της. Για την καλύτερη ερμηνεία προτείνεται τα αποτελέσματα να συνοδεύονται από διάφορα διαγράμματα, όπως το διάγραμμα δάσους (forest plot) ή το διάγραμμα

κώνου (funnel plot), που επιτρέπουν μια επιπρόσθετη διερεύνηση (Higgins & Green, 2008). Λεπτομέρειες για τη χρήση των παραπάνω διαγραμμάτων αναφέρονται αντίστοιχα στα Κεφάλαια 4 και 5.

2.2.8 Ερμηνεία αποτελεσμάτων και διεξαγωγή συμπερασμάτων (Interpreting results and drawing conclusions)

Το τελευταίο βήμα της συστηματικής ανασκόπησης αποτελεί η ερμηνεία των αποτελεσμάτων και η εξαγωγή συμπερασμάτων. Είναι ιδιαίτερα καθοριστικό, αφού τα ευρήματα της συστηματικής ανασκόπησης έχουν άμεση εφαρμογή στο χώρο της ιατρικής και βοηθούν τους κλινικούς στην λήψη ορθών αποφάσεων. Προκύπτει λοιπόν η ανάγκη της ακριβούς και σαφούς διατύπωσης των ευρημάτων, η οποία ικανοποιείται από την πλήρη παρουσίαση όλων των πληροφοριών, που σχετίζονται με σημαντικά αποτελέσματα (Higgins & Green, 2008). Τέλος, τα αποτελέσματα της στατιστικής ανάλυσης δεν θα πρέπει να υπόκεινται στην υποκειμενικότητα των ερευνητών, για αυτό η παρουσίαση τους πρέπει να γίνεται με τη χρήση των διαστημάτων εμπιστοσύνης και των αντίστοιχων επιπέδων στατιστικής σημαντικότητας (Higgins & Green, 2008).

ΚΕΦΑΛΑΙΟ 3
ΜΕΤΡΑ ΣΧΕΣΗΣ/ΜΕΓΕΘΗ ΕΠΙΔΡΑΣΗΣ
(EFFECT SIZES)

Στο Κεφάλαιο 3 περιγράφονται τα μέτρα σχέσης/μεγέθη επίδρασης (effect sizes), τα οποία επιτρέπουν την ποσοτικοποίηση και σύγκριση της αποτελεσματικότητας των παρεμβάσεων των μελετών, που πληρούν τα κριτήρια μιας συστηματικής ανασκόπησης. Η ανάλυση εστιάζεται σε μεγέθη επίδρασης που αφορούν συνεχή και διχότομα δεδομένα.

3.1 Μέτρα σχέσης/Μεγέθη επίδρασης (effect sizes)

Σε μία τυχαιοποιημένη μελέτη το δείγμα χωρίζεται σε δύο ομάδες, την πειραματική (treatment group) και την ομάδα ελέγχου (control group). Στη συνέχεια, πραγματοποιείται μια παρέμβαση στην πειραματική ομάδα, συνήθως μια θεραπεία και μια άλλη παρέμβαση, χορήγηση εικονικού φαρμάκου στην ομάδα ελέγχου. Στόχος της μελέτης είναι ο έλεγχος της αποτελεσματικότητας της θεραπείας, η οποία πραγματοποιείται με την σύγκριση των εκβάσεων στις δύο ομάδες. Οι διάφοροι τρόποι με τους οποίους μπορεί να πραγματοποιηθεί η σύγκριση των εκβάσεων στις δύο ομάδες καλούνται μέτρα σχέσης ή μεγέθη επίδρασης (effect sizes). Η φύση των αποτελεσμάτων που εξάγονται δεν είναι ίδια για όλες τις μελέτες και συνεπώς δεν μπορούν να χρησιμοποιηθούν τα ίδια μεγέθη επίδρασης για όλους τους τύπους δεδομένων. Στο πλαίσιο της διατριβής θα ασχοληθούμε με διχότομα και συνεχή δεδομένα. Διχότομα (dichotomous) χαρακτηρίζονται τα δεδομένα, που μπορούν να λάβουν ως τιμή μόνο μία από τις δύο κατηγορίες ανταπόκρισης (γεγονός - μη γεγονός), ενώ συνεχή (continuous) χαρακτηρίζονται τα δεδομένα, που λαμβάνουν οποιαδήποτε αριθμητική τιμή.

3.2 Μεγέθη επίδρασης για διχότομα δεδομένα

Τα μεγέθη επίδρασης για τα διχότομα δεδομένα είναι οι διαφορετικοί τρόποι, με τους οποίους συμπεραίνουμε, αν ένα γεγονός είναι πιο πιθανό να συμβεί στην πειραματική ομάδα σε σχέση με την ομάδα ελέγχου με τη βοήθεια των εννοιών της αναλογίας και του κινδύνου. Με τον όρο αναλογία στη βιοστατιστική καλούμε το πηλίκο του αριθμού των γεγονότων δια του αριθμού των μη γεγονότων και αντίστοιχα κίνδυνο καλούμε το πηλίκο του αριθμού των ευνοϊκών γεγονότων προς τον συνολικό αριθμό γεγονότων. Τα πιο γνωστά μεγέθη επίδρασης είναι η διαφορά κινδύνου (Risk Difference), ο λόγος αναλογιών (Odds Ratio) και ο λόγος κινδύνου (Risk Difference).

Για τον ορισμό τους θεωρούμε μια μελέτη η οποία έχει δύο ανεξάρτητες ομάδες, την πειραματική και την ομάδα ελέγχου. Συμβολίζουμε με a, b το πλήθος των εμφανίσεων ή μη αντίστοιχα του γεγονότος, που μας ενδιαφέρει στην πειραματική ομάδα και με c, d το πλήθος των εμφανίσεων ή μη αντίστοιχα του γεγονότος στην ομάδα ελέγχου. Οι πληροφορίες αυτές απεικονίζονται στον παρακάτω 2×2 πίνακα συνάφειας.

	Γεγονός (event)	**Μη γεγονός (no event)**	**Σύνολο**
Πειραματική ομάδα	**a**	**b**	$\boldsymbol{n_1 = a + b}$
Ομάδα ελέγχου	**c**	**d**	$\boldsymbol{n_2 = c + d}$
	a + c	**b + d**	**n = a + b + c + d**

Πίνακας 3.1: 2×2 πίνακας συνάφειας μιας υποθετικής μελέτης

Για τα διχότομα δεδομένα το μέγεθος επίδρασης δίνεται από τη σχέση effect size $= \mathrm{f}(\hat{p}_\mathrm{T}) - \mathrm{f}(\hat{p}_\mathrm{C})$, όπου $\hat{p}_\mathrm{T} = \frac{\mathrm{a}}{n_1}$ η πιθανότητα ανταπόκρισης στην πειραματική ομάδα, $\hat{p}_\mathrm{C} = \frac{\mathrm{c}}{\mathrm{n}_2}$ η πιθανότητα ανταπόκρισης στην ομάδα ελέγχου και $\mathrm{f}(\cdot)$ η συνάρτηση ένωσης (link function) (Borenstein et al., 2009).

3.2.1 Διαφορά Κινδύνου (Risk Difference)

Αν η συνάρτηση ένωσης είναι η ταυτοτική τότε το μέγεθος επίδρασης καλείται διαφορά κινδύνου (risk difference) και ορίζεται ως η διαφορά του κινδύνου μεταξύ της πειραματικής ομάδας και της ομάδας ελέγχου.

$\mathrm{RD} = \mathrm{f}(\hat{p}_\mathrm{T}) - \mathrm{f}(\hat{p}_\mathrm{c}) = \hat{p}_\mathrm{T} - \hat{p}_\mathrm{c} = \frac{\mathrm{a}}{\mathrm{a+b}} - \frac{\mathrm{c}}{\mathrm{c+d}}$ (Borenstein et al., 2009)

Στη συνέχεια, ακολουθεί ο υπολογισμός της ασυμπτωτικής διακύμανσης της διαφοράς κινδύνου (RD) :

Από το γεγονός ότι $\hat{p}_\mathrm{T} \sim N\left(\mathrm{p_T}, \frac{\mathrm{P_T(1-P_T)}}{n_1}\right)$ και $\hat{p}_C \sim N\left(\mathrm{p_C}, \frac{\mathrm{P_C(1-P_C)}}{n_2}\right)$ προκύπτει ότι η ποσότητα $\hat{p}_\mathrm{T} - \hat{p}_C$ έχει ασυμπτωτική διακύμανση που εκτιμάται από τη σχέση:

$$\mathrm{Var}(\hat{p}_\mathrm{T} - \hat{p}_\mathrm{c}) \approx \frac{\hat{p}_\mathrm{T}(1-\hat{p}_\mathrm{T})}{\mathrm{n_1}} + \frac{\hat{p}_C(1-\hat{p}_\mathrm{c})}{\mathrm{n_2}} = \frac{\mathrm{ab}}{(\mathrm{a+b})^3} + \frac{\mathrm{cd}}{(\mathrm{c+d})^3}.$$

$$\text{Άρα, } \widehat{Var(\mathrm{RD})} = \mathrm{Var}\left(\hat{p}_\mathrm{T} - \hat{p}_\mathrm{c}\right) \approx \frac{\mathrm{ab}}{(\mathrm{a+b})^3} + \frac{\mathrm{cd}}{(\mathrm{c+d})^3}.$$

Συνδυάζοντας τα παραπάνω, προκύπτει το $100(1-\alpha)\%$ προσεγγιστικό διάστημα εμπιστοσύνης για την διαφορά κινδύνου RD:

$$[L_{RD}, U_{RD}] = \left[\mathrm{RD} - z_{1-\alpha/2}\sqrt{\widehat{Var(\mathrm{RD})}}\,, \mathrm{RD} + z_{1-\alpha/2}\sqrt{\widehat{Var(\mathrm{RD})}}\right]$$

Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Ερμηνεία

Αν $\mathrm{RD}=0$ τότε συμπεραίνουμε ότι δεν υπάρχει διαφορά ανάμεσα στην πειραματική και στην ομάδα ελέγχου. Αν $\mathrm{RD}>0$ είναι πιο πιθανόν το γεγονός να συμβεί στην πειραματική ομάδα από ότι στην ομάδα ελέγχου. Αντίθετα, αν $\mathrm{RD}<0$ είναι λιγότερο πιθανόν να συμβεί το γεγονός στην πειραματική από ότι στην ομάδα ελέγχου.

Παρατηρήσεις

Η διαφορά κινδύνου (RD) σαν μέτρο επίδρασης είναι εύκολη στην ερμηνεία και κατανοητή από τους κλινικούς. Παρόλα αυτά δεν χρησιμοποιείται συχνά στην πράξη, γιατί η εφαρμογή της σε διαφορετικούς πληθυσμούς οδηγεί συχνά σε μη εφικτές τιμές. Ακόμα, δεν λαμβάνει υπόψιν την αρχική και τελική πιθανότητα παρά μόνο την διαφορά τους. Για παράδειγμα, αν η πιθανότητα θανάτου ακολουθώντας μία θεραπεία Α μειώνεται από 70% σε 68% και ακολουθώντας μία θεραπεία Β μειώνεται από 3% σε 1%, τότε και στις δύο περιπτώσεις το $\mathrm{RD}=2\%$, χωρίς να λαμβάνεται υπόψιν, ότι στην δεύτερη περίπτωση η πιθανότητα μειώθηκε στο ένα τρίτο της αρχικής τιμής της.

3.2.2 Λόγος αναλογιών (Odds Ratio)

Ο λόγος κινδύνου ορίζεται ως το πηλίκο της αναλογίας στην πειραματική ομάδα προς την αναλογία στην ομάδα ελέγχου.

$$\mathrm{OR}=\frac{\text{αναλογία στην πειραματική ομάδα}}{\text{αναλογία στην ομάδα ελέγχου}}=\frac{\frac{a}{b}}{\frac{c}{d}}=\frac{ad}{bc}$$ (Borenstein et al., 2009)

Συχνά μετράμε τον λόγο αναλογίας στην λογαριθμική κλίμακα. Αποδεικνύεται ότι όταν η συνάρτηση σύνδεση είναι $f(x) = \text{logit}(x)$ με $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$, το μέγεθος επίδρασης είναι ο λόγος αναλογίας εκφρασμένος στην λογαριθμική κλίμακα. Ειδικότερα ισχύει ότι:

$$\text{effect size} = f(\hat{p}_T) - f(\hat{p}_C) = \text{logit}\left(\frac{a}{a+b}\right) - \text{logit}\left(\frac{c}{c+d}\right) =$$

$$= \log\left(\frac{\frac{a}{a+b}}{1-\frac{a}{a+b}}\right) - \log\left(\frac{\frac{c}{c+d}}{1-\frac{c}{c+d}}\right) = \log\left(\frac{a}{b}\right) - \log\left(\frac{c}{d}\right) = \log\left(\frac{\frac{a}{b}}{\frac{c}{d}}\right) = \log(OR).$$

Άρα, $\log(OR) = \text{logit}(\hat{p}_T) - \text{log it}(\hat{p}_C) = \log\left(\frac{a}{b}\right) - \log\left(\frac{c}{d}\right)$.

Στη συνέχεια, θα προσδιορίσουμε την ασυμπτωτική διακύμανση του λογαρίθμου του λόγου αναλογιών $\log(OR)$:

Είναι γνωστό ότι $\hat{p}_T = \frac{a}{a+b}$ και από το νόμο των μεγάλων αριθμών έχουμε ότι $\frac{\hat{p}_T - p_T}{\sqrt{\frac{\hat{p}_T(1-\hat{p}_T)}{n_1}}} \sim N(0,1)$ ασυμπτωτικά. Εφαρμόζοντας τη δέλτα μέθοδο για $f(\hat{p}_T) = \log\left(\frac{\hat{p}_T}{1-\hat{p}_T}\right)$ και $f'(\hat{p}_T) = \frac{1}{\hat{p}_T(1-\hat{p}_T)}$, προκύπτει ότι $\widehat{se}_{\text{log}it(\hat{p}_T)} = \frac{1}{\hat{p}_T(1-\hat{p}_T)}\sqrt{\frac{\hat{p}_T(1-\hat{p}_T)}{n_1}} = \sqrt{\frac{1}{n_1\hat{p}_T(1-\hat{p}_T)}}$ και $\text{Var}(\widehat{\text{log}it}(\hat{p}_T)) = \frac{1}{n_1\hat{p}_T(1-\hat{p}_T)}$.

Αντίστοιχα, έχουμε ότι και $\text{Var}(\widehat{\text{log}it}(\hat{p}_c)) = \frac{1}{n_2 * \hat{p}_c * (1-\hat{p}_c)}$.

Η ασυμπτωτική διακύμανση του λογαρίθμου του λόγου αναλογιών $\log(OR)$ υπολογίζεται:

$$\text{Var}(\widehat{\log(OR)}) \approx \text{Var}(\widehat{\log it}(\hat{p}_T)) + \text{Var}(\widehat{\text{log}it}(\hat{p}_c)) \approx$$

$$\approx \frac{1}{(a+b) * \left(\frac{a}{a+b}\right) * \left(1 - \frac{a}{a+b}\right)} + \frac{1}{(c+d) * \left(\frac{c}{c+d}\right) * \left(1 - \frac{c}{c+d}\right)}$$

$$= \frac{a+b}{a*b} + \frac{c+d}{cd}$$

$$= \frac{a}{a*b} + \frac{b}{a*b} + \frac{c}{c*d} + \frac{d}{c*d}$$

$$\widehat{\mathrm{Var}(\log(\mathrm{OR}))} \approx \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

Συνδυάζοντας τα παραπάνω, προκύπτει το $100(1-\alpha)\%$ προσεγγιστικό διάστημα εμπιστοσύνης για τον λογάριθμο του λόγου αναλογιών

$$\left[L_{logOR}, U_{logOR}\right] = \left[\log\mathrm{OR} - z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}(\log(\mathrm{OR}))}}\,, \log\mathrm{OR} + z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}(\log(\mathrm{OR}))}}\right]$$

και το αντίστοιχο $100(1-\alpha)\%$ προσεγγιστικό διάστημα εμπιστοσύνης για το λόγο αναλογιών $[e^{L_{logOR}}, e^{U_{logOR}}]$. Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Ερμηνεία

Αν $\mathrm{OR} = 1$ τότε συμπεραίνουμε ότι δεν υπάρχει διαφορά μεταξύ των αναλογιών στην πειραματική ομάδα και στην ομάδα ελέγχου. Αν $\mathrm{OR} > 1$ είναι $100(1-\alpha)\%$ πιο πιθανόν το γεγονός να συμβεί στην πειραματική ομάδα από ότι στην ομάδα ελέγχου. Αντίθετα, αν $\mathrm{OR} < 1$ είναι λιγότερο κατά $100(1-\alpha)\%$ πιθανόν το γεγονός να συμβεί στην πειραματική ομάδα από ότι στην ομάδα ελέγχου.

Παρατηρήσεις

Ο λόγος κινδύνου (OR) δεν είναι ιδιαίτερα κατανοητός από τους κλινικούς. Ένα άλλο μειονέκτημα είναι η ύπαρξη μηδενικών κελιών στον πίνακα συνάφειας. Διαφορετικές εναλλακτικές έχουν προταθεί σε αυτή την περίπτωση, όπως η

πρόσθεση ενός πολύ μικρού αριθμού σε όλα τα κελιά (π.χ. 0.5), αλλά δεν υπάρχει ομοφωνία προς την κατάλληλη τεχνική για αυτές τις περιπτώσεις. Ωστόσο, σε αντίθεση με τα υπόλοιπα μέτρα θέσης η εφαρμογή του σε διαφορετικούς πληθυσμούς οδηγεί πάντα σε εφικτές προβλέψεις.

3.2.3 Λόγος κινδύνου (Risk Ratio)

Ο λόγος κινδύνου ορίζεται ως το πηλίκο του κινδύνου στην πειραματική ομάδα προς τον κίνδυνο στην ομάδα ελέγχου.

$$\text{RR} = \frac{\text{κίνδυνος στην πειραματική ομάδα}}{\text{κίνδυνος στην ομάδα ελέγχου}} = \frac{\frac{\text{a}}{\text{a+b}}}{\frac{\text{c}}{\text{c+d}}} = \frac{\text{a(c+d)}}{\text{c(a+b)}}$$ (Borenstein et al., 2009)

Συχνά μετράμε τον λόγο κινδύνου στην λογαριθμική κλίμακα. Αποδεικνύεται, ότι όταν η συνάρτηση σύνδεσης είναι $\text{f(x)} = \log(\text{x})$ το μέγεθος επίδρασης είναι ο λόγος κινδύνου εκφρασμένος στην λογαριθμική κλίμακα.

$$\text{effect size} = \text{f}(\hat{p}_{\text{T}}) - \text{f}(\hat{p}_{C}) = \log\left(\frac{\text{a}}{\text{a}+\text{b}}\right) - \log\left(\frac{\text{c}}{\text{c}+\text{d}}\right) = \log\left(\frac{\frac{\text{a}}{\text{a+b}}}{\frac{\text{c}}{\text{c+d}}}\right) = \log(\text{RR}).$$

Άρα, $\log(\text{RR}) = \log(\hat{p}_{\text{T}}) - \log(\hat{p}_{C}) = \log\left(\frac{\text{a}}{\text{a+b}}\right) - \log\left(\frac{\text{c}}{\text{c+d}}\right)$.

Στη συνέχεια, θα προσδιορίσουμε την ασυμπτωτική διακύμανση του λόγου κινδύνου $\log(\text{RR})$:

Είναι γνωστό ότι $\hat{p}_{\text{T}} = \frac{\text{a}}{\text{a+b}}$ και από το νόμο των μεγάλων αριθμών έχουμε ότι $\frac{\hat{p}_{\text{T}} - \text{p}_{\text{T}}}{\sqrt{\frac{\hat{p}_{\text{T}}(1-\hat{p}_{\text{T}})}{n_1}}} \sim N(0,1)$ ασυμπτωτικά. Εφαρμόζοντας τη δέλτα μέθοδο για $f(\hat{p}_{\text{T}}) = \log(\hat{p}_{\text{T}})$ και $f'(\hat{p}_{\text{T}}) = \frac{1}{\hat{p}_{\text{T}}}$, προκύπτει ότι $\widehat{se}_{\log(\hat{p}_{\text{T}})} = \sqrt{\frac{1-\hat{p}_{\text{T}}}{n_1 \hat{p}_{\text{T}}}}$ και $\widehat{\text{Var}(\log(\hat{p}_{\text{T}}))} = \frac{1-\hat{p}_{\text{T}}}{n_1 \hat{p}_{\text{T}}}$. Κατά αντιστοιχία έχουμε ότι και $\widehat{\text{Var}(\log(\hat{p}_{\text{c}}))} = \frac{1-\hat{p}_{\text{c}}}{n_2 \hat{p}_{c}}$.

Η ασυμπτωτική διακύμανση του λογαρίθμου του λόγου κινδύνου log(RR) υπολογίζεται:

$$\mathrm{Var}(\widehat{\log(\mathrm{RR})}) \approx \mathrm{Var}(\widehat{\log(\hat{p}_T)}) + \mathrm{Var}(\widehat{\log(\hat{p}_c)}) \approx$$

$$\approx \frac{1-\frac{a}{a+b}}{(a+b)*\left(\frac{a}{a+b}\right)} + \frac{1-\frac{c}{c+d}}{(c+d)*\left(\frac{c}{c+d}\right)}$$

$$= \frac{a+b}{a*(a+b)} + \frac{-a}{a*(a+b)} + \frac{c+d}{c*(c+d)} + \frac{-c}{c*(c+d)}$$

$$= \frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}$$

$$\boxed{\mathrm{Var}(\widehat{\log(\mathrm{RR})}) \approx \frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

Συνδυάζοντας τα παραπάνω, προκύπτει το $100(1-\alpha)\%$ προσεγγιστικό διάστημα εμπιστοσύνης για τον λογάριθμο του λόγου κινδύνου

$$\left[L_{logRR}, U_{logRR}\right] = \left[\log\mathrm{RR} - z_{1-\alpha/2}\sqrt{\mathrm{Var}(\widehat{\log(\mathrm{RR})})}\,, \log\mathrm{RR} + z_{1-\alpha/2}\sqrt{\mathrm{Var}(\widehat{\log(\mathrm{RR})})}\right]$$

και αντίστοιχα το $100(1-\alpha)\%$ προσεγγιστικό διάστημα εμπιστοσύνης για τον λόγο κινδύνου $\left[e^{L_{logRR}}, e^{U_{logRR}}\right]$. Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Ερμηνεία

Αν $\mathrm{RR}=1$ τότε συμπεραίνουμε ότι δεν υπάρχει διαφορά στον κίνδυνο ανάμεσα στην πειραματική ομάδα και στην ομάδα ελέγχου. Αν $\mathrm{RR}>1$ είναι $100(1-\alpha)\%$ πιο πιθανόν το γεγονός να συμβεί στην πειραματική ομάδα από ότι στην ομάδα ελέγχου. Αντίθετα, αν $\mathrm{RR}<1$ είναι λιγότερο κατά

$100(1-\alpha)\%$ πιθανόν το γεγονός να συμβεί στην πειραματική ομάδα από ότι στην ομάδα ελέγχου.

Παρατηρήσεις

Ο λόγος κινδύνου (RR) σαν μέτρο επίδρασης χρησιμοποιείται συχνά στην πράξη, γιατί είναι εύκολος στην ερμηνεία και κατανοητός από τους κλινικούς. Ωστόσο, διακρίνεται από αδυναμίες, καθώς η εφαρμογή του σε διαφορετικούς πληθυσμούς οδηγεί πολλές φορές σε μη εφικτές τιμές και μικρές εναλλαγές στα δεδομένα δίνουν μεγάλες αποκλίσεις στις προβλέψεις. Επίσης, αν αντιστρέψουμε το γεγονός, ο λόγος κινδύνου ενδέχεται να αλλάξει ριζικά.

3.3 Μεγέθη επίδρασης για συνεχή δεδομένα

Οι διαφορετικοί τρόποι με τους οποίους συγκρίνουμε τους μέσους στην πειραματική ομάδα και στην ομάδα ελέγχου, ώστε να συμπεραίνουμε σε ποια από τις δύο ομάδες είναι πιο πιθανό να συμβεί το γεγονός αποτελούν τα μεγέθη επίδρασης (effect sizes) για τα συνεχή δεδομένα. Τα πιο γνωστά μεγέθη επίδρασης είναι η διαφορά μέσων (Mean Difference) και η τυποποιημένη διαφορά μέσων (Standardized Mean Difference). Για τον ορισμό τους θεωρούμε μία μελέτη, η οποία έχει δύο ανεξάρτητες ομάδες την πειραματική και την ομάδα ελέγχου. Συμβολίζουμε με m_t και sd_t την δειγματική μέση τιμή και την δειγματική τυπική απόκλιση στην πειραματική ομάδα και αντίστοιχα με m_c και sd_c την δειγματική μέση τιμή και την δειγματική τυπική απόκλιση στην ομάδα ελέγχου. Οι πληροφορίες αυτές απεικονίζονται στον παρακάτω πίνακα.

	Δειγματική Μέση Τιμή	**Δειγματική Τυπική Απόκλιση**	**Μέγεθος Δείγματος**
Πειραματική ομάδα	$\mathbf{m_t}$	$\mathbf{sd_t}$	$\mathbf{n_t}$
Ομάδα ελέγχου	$\mathbf{m_c}$	$\mathbf{sd_c}$	$\mathbf{n_c}$

Πίνακας 3.2: Οι ποσότητες για τα συνεχή δεδομένα

3.3.1 Διαφορά μέσων (Mean Difference)

Η διαφορά μέσων ορίζεται ως η διαφορά της δειγματικής μέσης τιμής της πειραματικής ομάδας και της ομάδας ελέγχου.

$\mathrm{MD} = \mathrm{m_t} - \mathrm{m_c}$ (Borenstein et al., 2009)

Η εκτίμηση της διακύμανσης της διαφοράς μέσων υπολογίζεται από την σχέση $\widehat{Var}(\mathrm{MD}) = \frac{\mathrm{sd_t}^2}{\mathrm{n_t}} + \frac{\mathrm{sd_c}^2}{\mathrm{n_c}}$ (Cohen, 1988). Συνδυάζοντας τα παραπάνω, προκύπτει το $100(1-\alpha)$ % διάστημα εμπιστοσύνης για την διαφορά μέσων:

$$\lceil \mathrm{L_{MD}}, \mathrm{U_{MD}} \rceil = \left[\mathrm{MD} - z_{1-\alpha/2}\sqrt{\widehat{Var}(\mathrm{MD})}\,, \mathrm{MD} + z_{1-\alpha/2}\sqrt{\widehat{Var}(\mathrm{MD})} \right]$$

Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Ερμηνεία

Αν $\mathrm{MD} = 0$ τότε συμπεραίνουμε ότι δεν υπάρχει διαφορά ανάμεσα στην πειραματική ομάδα και στην ομάδα ελέγχου. Αν $\mathrm{MD} > 0$ το γεγονός είναι πιο πιθανόν να συμβεί στην πειραματική ομάδα σε σχέση με την ομάδα ελέγχου.

Αντίθετα, αν $\mathrm{MD} < 0$ το γεγονός είναι πιο πιθανόν να συμβεί στην ομάδα ελέγχου σε σχέση με την πειραματική.

Παρατηρήσεις

Η χρήση της διαφοράς μέσων (MD) ως μέτρο επίδρασης απαιτεί όλες οι μελέτες να ακολουθούν την ίδια κλίμακα για την μέτρηση της έκβασης ειδάλλως καταλήγουμε σε παραπλανητικά συμπεράσματα.

3.3.2 Τυποποιημένη διαφορά μέσων (Standardized Mean Difference)

Η τυποποιημένη διαφορά μέσων με βάση τις πραγματικές τιμές των μελετών ορίζεται ως η διαφορά της τυποποιημένης δειγματικής μέσης τιμής της πειραματικής ομάδας και της ομάδας ελέγχου (Cohen, 1988).

$$\mathrm{SMD} = \frac{\mathrm{m_t}}{\mathrm{S_{pooled}}} - \frac{\mathrm{m_c}}{\mathrm{S_{pooled}}}, \text{ όπου } \mathrm{S_{pooled}} = \sqrt{\frac{(\mathrm{n_t}-1)\mathrm{sd_t}^2+(\mathrm{n_c}-1)\mathrm{sd_c}^2}{\mathrm{n_t}+\mathrm{n_c}-2}}$$

Η εκτίμηση της διακύμανσης της τυποποιημένης διαφοράς υπολογίζεται από την σχέση $\widehat{Var}(\mathrm{SMD}) = \frac{\mathrm{n_t}+\mathrm{n_c}}{\mathrm{n_t n_c}} + \frac{\mathrm{SMD}^2}{2(\mathrm{n_t}+\mathrm{n_c}-2)}$ (Cohen, 1988). Συνδυάζοντας τα παραπάνω προκύπτει το $100(1-\alpha)$ % διάστημα εμπιστοσύνης για την τυποποιημένη διαφορά μέσων:

$$[\mathrm{L_{SMD}}, \mathrm{U_{SMD}}] = \left[SMD - z_{1-\alpha/2}\sqrt{\widehat{Var}(\mathrm{SMD})}\,, \mathrm{SMD} + z_{1-\alpha/2}\sqrt{\widehat{Var}(\mathrm{SMD})}\right]$$

Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

Ερμηνεία

Αν $\mathrm{SMD} = 0$ τότε συμπεραίνουμε ότι δεν υπάρχει διαφορά ανάμεσα στην πειραματική ομάδα και στην ομάδα ελέγχου. Αν $\mathrm{SMD} > 0$ το γεγονός είναι πιο

πιθανόν να συμβεί στην πειραματική ομάδα σε σχέση με την ομάδα ελέγχου. Αντίθετα, αν $\mathrm{SMD} < 0$ το γεγονός είναι πιο πιθανόν να συμβεί στην ομάδα ελέγχου σε σχέση με την πειραματική.

Παρατηρήσεις

Η χρήση της τυποποιημένης διαφοράς μέσων (SMD) ως μέτρο επίδρασης έχει το πλεονέκτημα, ότι μπορεί να συνδυάσει μελέτες που έχουν το ίδιο κλινικό αποτέλεσμα αλλά διαφορετικό χρόνο ή τρόπο μέτρησης της έκβασης. Ωστόσο, υπάρχει μία βασική αδυναμία. Υποθέτει, ότι η διακύμανση μεταξύ των μελετών οφείλεται στις διαφορετικές κλίμακες μέτρησης που ενδεχομένως έχουν χρησιμοποιηθεί και όχι σε πραγματικές διαφορές της διακύμανσης των πληθυσμών, με αποτέλεσμα σε ορισμένες περιπτώσεις να καταλήγει σε παραπλανητικά συμπεράσματα.

ΚΕΦΑΛΑΙΟ 4
ΜΕΤΑ-ΑΝΑΛΥΣΗ (META-ANALYSIS)

Σκοπός του κεφαλαίου είναι η παρουσίαση της μεθόδου της μετα-ανάλυσης, η οποία αποτελεί μια αντικειμενική και ποσοτική μεθοδολογία, που χρησιμοποιείται για την σύνθεση ερευνητικών μελετών που έχουν γίνει στο παρελθόν, ώστε να οδηγήσουν σε ένα συγκεντρωτικό αποτέλεσμα. Στη συνέχεια, περιγράφονται αναλυτικά τα δύο μοντέλα, που καθορίζουν το συγκεντρωτικό αποτέλεσμα: μοντέλο σταθερών επιδράσεων (fixed effect model) και μοντέλο τυχαίων επιδράσεων (random effect model). Αναλύονται τα διαγράμματα δάσους (forest plots), που αποτελούν τον πιο χαρακτηριστικό τρόπο σχηματικής παρουσίασης των αποτελεσμάτων της μετα-ανάλυσης. Παρατίθεται ακόμα ο ορισμός της ετερογένειας των μελετών και οι τρόποι εντοπισμού της. Τέλος, για την καλύτερη ερμηνεία των αποτελεσμάτων της μετα-ανάλυσης παρουσιάζεται η μέθοδος της μετα-παλινδρόμησης (meta-regression), με την οποία εντοπίζεται η πηγή ετερογένειας των μελετών.

4.1 Μετα-ανάλυση

Στις μέρες μας ο ρυθμός με τον οποίο παράγεται και εμπλουτίζεται η επιστημονική γνώση είναι πιο ταχύς από ποτέ στην ιστορία της ανθρωπότητας (Mabe, 2009). Η πληθώρα των δημοσιευμάτων σε συνδυασμό με την εύκολη πρόσβαση στην επιστημονική βιβλιογραφία, που έχει επιφέρει η ηλεκτρονική επανάσταση, κάνει ολοένα και πιο δύσκολο το εγχείρημα των επιστημόνων να οδηγηθούν σε συγκεντρωτικά συμπεράσματα από τα ευρήματα της επιστημονικής έρευνας. Η κατάσταση γίνεται ακόμα πιο περίπλοκη στις περιπτώσεις, που τα ευρήματα των μελετών δεν είναι σύμφωνα μεταξύ τους. Συνεπώς, προκύπτει η ανάγκη της σύνθεσης πληροφοριών από τις διάσπαρτες δημοσιεύσεις. Η μέθοδος, που υπηρετεί την ανάγκη αυτή, καλείται μετα-ανάλυση. Πρόκειται για μια στατιστική τεχνική, που συνθέτει τα ευρήματα

διαφόρων μελετών με στόχο τον υπολογισμό ενός συγκεντρωτικού αποτελέσματος.

Η διαδικασία της μετα-ανάλυσης αποτελείται από δύο στάδια. Στο πρώτο στάδιο υπολογίζονται τα μεγέθη επίδρασης των μελετών και τα αντίστοιχα σφάλματά τους. Έπειτα, κατά το δεύτερο στάδιο πραγματοποιείται η σύνθεση των μεγεθών επίδρασης με στόχο την δημιουργία ενός συγκεντρωτικού αποτελέσματος. Για τον υπολογισμό της συνεισφοράς της κάθε μελέτης στο συγκεντρωτικό αποτέλεσμα έχουν αναπτυχθεί δύο διαφορετικές προσεγγίσεις, το μοντέλο σταθερών επιδράσεων (fixed effect model) και το μοντέλο τυχαίων επιδράσεων (random effect model). Η διαφορά ανάμεσα σε αυτές τις δύο προσεγγίσεις έγκειται στο τρόπο με τον οποίο ερμηνεύονται οι διαφορές στα αποτελέσματα των μελετών.

4.2 Μοντέλο σταθερών επιδράσεων (Fixed effect model)

Στο μοντέλο σταθερών επιδράσεων υποθέτουμε ότι το μέγεθος επίδρασης είναι κοινό για όλες τις μελέτες, που συμπεριλαμβάνονται στην μετα-ανάλυση. Συνεπώς, όλες οι μελέτες εκτιμούν ένα κοινό «πραγματικό» μέγεθος επίδρασης το οποίο συμβολίζεται με μ. Ακόμα, η όποια διαφορά ανάμεσα στα παρατηρούμενα μεγέθη επίδρασης (observed effect sizes) οφείλεται μόνο σε τυχαίους παράγοντες που προκύπτουν από την δειγματοληψία.

Έστω ότι έχουμε μία μελέτη και ας συμβολίσουμε με y το παρατηρούμενο μέγεθος της μελέτης και με μ το πραγματικό, τότε το y θα απέχει από το μ όσο και το τυχαίο σφάλμα ε που δημιουργείται μέσα στην μελέτη (Borenstein et al., 2009). Ακόμα, αν υποθέσουμε ότι το παρατηρούμενο μέγεθος επίδρασης ακολουθεί κανονική κατανομή με μέση τιμή μ και πληθυσμιακή διακύμανση σ^2, η σχέση που περιγράφει το παραπάνω μοντέλο σταθερών επιδράσεων είναι:

$y = \mu + \varepsilon,\ \varepsilon \sim N(0, \sigma^2)$ (Borenstein et al., 2009)

Έστω τώρα ότι έχουμε k μελέτες από τον ίδιο πληθυσμό με πληθυσμιακή διακύμανση ${\sigma_i}^2$. Αν y_i θεωρήσουμε τα παρατηρούμενα μεγέθη επίδρασης, η σχέση που περιγράφει το μοντέλο σταθερών επιδράσεων είναι:

$y_i = \mu + \varepsilon_i,\ \ \varepsilon_i \sim N(0, {\sigma_i}^2),\ \ i = 1, \dots, k$ (Borenstein et al., 2009)

Στο σημείο αυτό είναι αναγκαίος ο καθορισμός του τρόπου με τον οποίο η κάθε μελέτη θα επηρεάσει το συγκεντρωτικό αποτέλεσμα. Στην μετα-ανάλυση η διαδικασία αυτή καλείται ανάθεση βαρών. Έχουν αναπτυχθεί πολλοί μέθοδοι ανάθεσης βαρών με πιο δημοφιλή, την μέθοδο της αντίστροφης διασποράς. Σύμφωνα με την μέθοδο αυτή, το βάρος της κάθε μελέτης ισούται με το αντίστροφο της διακύμανσή της. Επειδή στην περίπτωση της μετα-ανάλυσης σταθερών επιδράσεων η διακύμανση της κάθε μελέτης ισούται με την πληθυσμιακή διακύμανση, τα βάρη περιγράφονται από την σχέση $w_{i,FE} = \frac{1}{{\sigma_i}^2}$ (Borenstein et al., 2009).

Μετά την ποσοτικοποίηση της επίδρασης της κάθε μελέτης, η οποία πραγματοποιείται με την ανάθεση των βαρών, μας ενδιαφέρει η ποσοτικοποίηση της συνολικής επίδρασης των μελετών. Στην μετα-ανάλυση η συνολική επίδραση καλείται συμβατικά «διαμάντι» και ορίζεται ως ο σταθμισμένος μέσος όρος των παρατηρούμενων μεγεθών επίδρασης. Για τον υπολογισμό της θα χρησιμοποιηθεί η μέθοδος της μέγιστης πιθανοφάνειας. Τα παρατηρούμενα μεγέθη επίδρασης y_i ακολουθούν ασυμπτωτικά κανονική κατανομή με μέση τιμή μ και διακύμανση ${\sigma_i}^2$ που θεωρείται γνωστή και περίπου ίση με ${s_i}^2$. Ισχύει δηλαδή, $y_i \sim N(\mu, {\sigma_i}^2),\ \ i = 1, \dots, k.$

Η συνάρτηση της πιθανοφάνειας των παρατηρούμενων μεγεθών επίδρασης ισούται με:

$$L = \prod_{i=1}^{k} f(y_i, \mu) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi{\sigma_i}^2}}\, e^{-\frac{(y_i-\mu)^2}{2{\sigma_i}^2}} = \frac{1}{(\sqrt{2\pi})^k \prod_{i=1}^{k}\sqrt{{\sigma_i}^2}}\, e^{-\sum_{i=1}^{k}\frac{(y_i-\mu)^2}{2{\sigma_i}^2}}.$$

Ο λογάριθμος της συνάρτησης πιθανοφάνειας ισούται αντίστοιχα με:

$$ln(L) = -\frac{k}{2}\, ln(2\pi) - \frac{1}{2}\, ln\left(\prod_{i=1}^{k} \sqrt{{\sigma_i}^2}\right) - \sum_{i=1}^{k} \frac{(y_i - \mu)^2}{2{\sigma_i}^2}.$$

Παρατηρούμε ότι ο λογάριθμος της συνάρτησης πιθανοφάνειας είναι διαφορίσιμος. Συνεπώς, παραγωγίζουμε ως προς μ.

$$\frac{dln(L)}{d\mu} = \sum_{i=1}^{k} \frac{y_i - \mu}{{\sigma_i}^2} = \sum_{i=1}^{k} \frac{y_i}{{\sigma_i}^2} - \mu \sum_{i=1}^{k} \frac{1}{{\sigma_i}^2}$$

Θέτουμε την παράγωγο ίση με το μηδέν και λύνουμε ως προς *μ.*

$$\frac{dln(L)}{d\mu} = 0 \leftrightarrow$$

$$\sum_{i=1}^{k} \frac{y_i}{{\sigma_i}^2} - \mu \sum_{i=1}^{k} \frac{1}{{\sigma_i}^2} = 0 \leftrightarrow$$

$$\mu = \frac{\sum_{i=1}^{k} \frac{y_i}{{\sigma_i}^2}}{\sum_{i=1}^{k} \frac{1}{{\sigma_i}^2}}$$

Γνωρίζοντας ότι $w_{i,FE} = \frac{1}{{\sigma_i}^2}$, προκύπτει ο εκτιμητής της συνολικής επίδρασης των μελετών $\hat{\mu}_{FE} = \frac{\sum_{i=1}^{k} w_{i,FE} y_i}{\sum_{i=1}^{k} w_{i,FE}},\ i = 1, \dots, k$ (Borenstein et al., 2009).

Για μεγαλύτερη ακρίβεια της στατιστικής σημαντικότητας οι ερευνητές πέρα από τη συνολική επίδραση παραθέτουν και το αντίστοιχο διάστημα εμπιστοσύνης. Για τον υπολογισμό του διαστήματος εμπιστοσύνης απαιτείται ο καθορισμός της διακύμανσης της συνολικής επίδρασης καθώς και το αντίστοιχο τυπικό σφάλμα. Η διακύμανση της συνολικής επίδρασης ισούται με το αντίστροφο του αθροίσματος των βαρών της μετα-ανάλυσης $V_M = \frac{1}{\sum_{i=1}^{k} w_{i,FE}}$,

ενώ το τυπικό σφάλμα με την τετραγωνική ρίζα της συνολικής διακύμανσης $s_M = \sqrt{V_M} = \sqrt{\frac{1}{\sum_{i=1}^{k} w_{i,FE}}}$. Συνδυάζοντας τα παραπάνω και υποθέτοντας ότι η συνολική επίδραση κατανέμεται ασυμπτωτικά κανονικά, το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης δίνεται ως:

$$[L, U] = [\hat{\mu}_{FE} - z_{1-a/2}\sqrt{V_M}\,, \hat{\mu}_{FE} + z_{1-a/2}\sqrt{V_M}\,]$$ (Borenstein et al., 2009)

Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

4.3 Ορισμός και είδη ετερογένειας (Heterogeneity)

Όπως έχει ήδη αναφερθεί, η μετα-ανάλυση συνθέτει τα ευρήματα των μελετών, οι οποίες προέκυψαν από την συστηματική ανασκόπηση για την εξαγωγή ενός συγκεντρωτικού αποτελέσματος. Η ακρίβεια και η αξιοπιστία της εξαρτάται από τον βαθμό στον οποίο οι μελέτες διαφέρουν μεταξύ τους. Η όποια διαφορετικότητα μεταξύ των μελετών ονομάζεται ετερογένεια. Υπάρχουν τρία διαφορετικά είδη ετερογένειας, η κλινική (clinical heterogeneity), η μεθοδολογική (methodological heterogeneity) και η στατιστική (statistical heterogeneity) (Higgins & Green, 2008). Η κλινική ετερογένεια δημιουργείται από τα διαφορετικά χαρακτηριστικά διεξαγωγής των μελετών, όπως η ποικιλομορφία των συμμετεχόντων, το είδος των παρεμβάσεων και η μεταβλητότητα των εκβάσεων (Higgins & Green, 2008). Η ποικιλομορφία των συμμετεχόντων μπορεί να οφείλεται στην διαφορά ηλικίας, το φύλο και γενικά σε οποιοδήποτε χαρακτηριστικό διαφοροποιεί τους συμμετέχοντες. Ακόμα, η διαφοροποίηση των μελετών στην διάρκεια της θεραπείας, στο είδος και στην δοσολογία των φαρμάκων οδηγεί σε μεταβλητότητα των παρεμβάσεων, ενώ ο διαφορετικός τρόπος μέτρησης των αποτελεσμάτων μπορεί να επιφέρει μεταβλητότητα στις εκβάσεις. Η μεθοδολογική ετερογένεια οφείλει την ύπαρξή της στον διαφορετικό σχεδιασμό των μελετών καθώς και στην ποιότητά τους,

δηλαδή την διαφορετικότητα στον κίνδυνο μεροληψίας (risk of bias) (Higgins & Green, 2008). Τέλος, η στατιστική ετερογένεια δημιουργείται από την παρουσία είτε της κλινικής, είτε της μεθοδολογικής ετερογένειας, είτε από τον συνδυασμό τους (Higgins & Green, 2008). Στο πλαίσιο της μετα-ανάλυσης η στατιστική ετερογένεια καλείται ετερογένεια (heterogeneity), συμβολίζεται με τ^2 και αντιπροσωπεύει την μεταβλητότητα των αποτελεσμάτων των επιμέρους μελετών που δεν οφείλονται στην τύχη. Η συμβολή της είναι καθοριστική, καθώς επηρεάζει τα αποτελέσματα της μετα-ανάλυσης κατά την σύνθεση των μελετών. Στο μοντέλο σταθερών επιδράσεων που παρουσιάστηκε, η πηγή της μεταβλητότητας οφείλεται καθαρά σε σφάλματα της δειγματοληψίας, με αποτέλεσμα η ετερογένεια να είναι μηδενική. Συνεπώς, προέκυψε η ανάγκη να αναπτυχθεί ένα ακόμα μοντέλο που θα αναδεικνύει την συμβολή της ετερογένειας στην μετα-ανάλυση. Το μοντέλο που υπηρετεί την ανάγκη αυτή καλείται μοντέλο τυχαίων επιδράσεων (random effect model) και προτάθηκε από τους DerSimonian & Laird το 1986 (DerSimonian & Laird, 1986).

4.4 Μοντέλο τυχαίων επιδράσεων (Random effect model)

Το μοντέλο τυχαίων επιδράσεων υποθέτει ότι οι μελέτες που συμπεριλαμβάνονται στην μετα-ανάλυση αποτελούν τυχαίο δείγμα του πληθυσμού όλων των πιθανών μελετών. Αντίστοιχα υποθέτει ότι και τα μεγέθη επίδρασης αποτελούν τυχαίο δείγμα του πληθυσμού όλων των πιθανών μεγεθών επίδρασης. Στοχεύει συνεπώς στο να εκτιμηθεί το εύρος τιμών που μπορεί να πάρει το συγκεντρωτικό αποτέλεσμα. Από τα παραπάνω προκύπτει ότι οι μελέτες εκτιμούν διαφορετικά αποτελέσματα, τα οποία σχετίζονται μεταξύ τους και ότι οι διαφορές στα παρατηρούμενα μεγέθη επίδρασης οφείλονται όχι μόνο σε τυχαίους παράγοντες που προκύπτουν από την δειγματοληψία, αλλά και στην στατιστική ετερογένεια, δηλαδή σε διαφορές των πραγματικών μεγεθών επίδρασης (Higgins & Green, 2008).

Έστω ότι έχουμε μία μελέτη και ας συμβολίσουμε με y το παρατηρούμενο μέγεθος της μελέτης και με θ το πραγματικό τότε το y θα απέχει από το θ όσο και το τυπικό σφάλμα ε που δημιουργείται μέσα στην μελέτη (Borenstein et al., 2009). Αν υποθέσουμε ακόμα ότι το παρατηρούμενο μέγεθος y ακολουθεί κανονική κατανομή με μέση τιμή θ και διακύμανση σ^2 και ότι το πραγματικό μέγεθος επίδρασης θ κατανέμεται γύρω από κανονική κατανομή με μέση τιμή μ και διακύμανση τ^2, η σχέση που περιγράφει το παραπάνω μοντέλο τυχαίων επιδράσεων είναι:

$$y = \theta + \varepsilon_1,\ \varepsilon_1 \sim N(0, \sigma^2)$$

$$\theta = \mu + \varepsilon_2,\ \varepsilon_2 \sim N(0, \tau^2) \quad \text{(Borenstein et al., 2009)}$$

Έστω τώρα ότι έχουμε k μελέτες που προέρχονται από τον ίδιο πληθυσμό με πληθυσμιακή διακύμανση ${\sigma_i}^2$. Αν y_i θεωρήσουμε τα παρατηρούμενα μεγέθη επίδρασης, η σχέση που περιγράφει το μοντέλο τυχαίων επιδράσεων είναι:

$$y_i = \theta_i + \varepsilon_{1i},\ \ \varepsilon_{1i} \sim N(0, {\sigma_i}^2)$$

$$\theta_i = \mu + \varepsilon_{2i},\ \ \varepsilon_{2i} \sim N(0, \tau^2) \quad i = 1, \dots, k \quad \text{(Borenstein et al., 2009)}$$

Για την ανάθεση των βαρών θα ακολουθήσουμε τη μέθοδο της αντίστροφης διασποράς. Σύμφωνα με τη μέθοδο αυτή το βάρος της κάθε μελέτης είναι ίσο με το αντίστροφο της συνολικής διακύμανσης. Στην περίπτωση όμως του μοντέλου τυχαίων επιδράσεων η συνολική διακύμανση ισούται με το άθροισμα της πληθυσμιακής διακύμανσης της κάθε μελέτης και της ετερογένειας. Συνεπώς, τα βάρη περιγράφονται από την σχέση:

$$w_{i,RE} = \frac{1}{{\sigma_i}^2 + \tau^2} \text{ (Borenstein et al., 2009)}$$

Για να είναι εφικτός όμως ο υπολογισμός τους, κρίνεται απαραίτητη η εκτίμηση της ετερογένειας των μελετών τ^2. Έχουν προταθεί πολλές μέθοδοι για τον

προσδιορισμό της ετερογένειας. Μερικές από αυτές είναι: η μέθοδος της μέγιστης πιθανοφάνειας, των DerSimonian-Laid, η μέθοδος REML καθώς και οι εμπειρικές μέθοδοι Bayes. Εναλλακτικά, έχουν αναπτυχθεί τρόποι εκτίμησης της ετερογένειας που στηρίζονται στους εκτιμητές Hedges, Sidik-Jonkman και Hunter-Schmidt (Rukhin, 2013). Μετά την ανάθεση των βαρών έχουμε όλη την διαθέσιμη πληροφορία για τον καθορισμό της συνολικής επίδρασης, η οποία αποτελεί τον σταθμισμένο μέσο όρο των παρατηρούμενων μεγεθών επίδρασης

$$\hat{\mu}_{RE} = \frac{\sum_{i=1}^{k} w_{i,RE} y_i}{\sum_{i=1}^{k} w_{i,RE}}$$ (Borenstein et al., 2009)

Επιπλέον, η διακύμανση της συνολικής επίδρασης ισούται με το αντίστροφο του αθροίσματος των βαρών της μετα-ανάλυσης $V_M = \frac{1}{\sum_{i=1}^{k} w_{i,RF}}$, ενώ το αντίστοιχο τυπικό σφάλμα με την τετραγωνική ρίζα της συνολικής διακύμανσης $s_M = \sqrt{V_M} = \sqrt{\frac{1}{\sum_{i=1}^{k} w_{i,RE}}}$. Συνδυάζοντας τα παραπάνω και υποθέτοντας ότι η συνολική επίδραση κατανέμεται ασυμπτωτικά κανονικά, προκύπτει ότι το $100(1-\alpha)\%$ διάστημα εμπιστοσύνης είναι:

$$[L, U] = [\hat{\mu}_{RE} - z_{1-a/2}\sqrt{V_M}\,, \hat{\mu}_{RE} + z_{1-a/2}\sqrt{V_M}\,]$$ (Borenstein et al., 2009)

Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1-\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής.

4.5 Σύγκριση των δύο μοντέλων

Η κύρια διαφορά των δύο μοντέλων είναι η διαφορετική πηγή μεταβλητότητας. Στο μοντέλο σταθερών επιδράσεων υποθέτουμε ότι οι διαφορές στα παρατηρούμενα μεγέθη οφείλονται στην ύπαρξη τυχαίου δειγματοληπτικού σφάλματος (within-studies variance), σε αντίθεση με το μοντέλο τυχαίων επιδράσεων που αποδίδουμε τις όποιες διαφορές πέρα από την ύπαρξη

δειγματοληπτικού σφάλματος και σε πραγματικές διαφορές των μελετών (between-studies variance).

Ακόμα, η διαφορετική μεταβλητότητα επιτάσσει διαφορετικά βάρη στα δύο μοντέλα. Συγκεκριμένα, στο μοντέλο σταθερών επιδράσεων τα βάρη ορίζονται ως το αντίστροφο της πληθυσμιακής διακύμανσης, ενώ στο μοντέλο τυχαίων επιδράσεων ορίζονται ως το αντίστροφο του αθροίσματος της πληθυσμιακής διακύμανσης και της ετερογένειας. Η ύπαρξη της ετερογένειας στα βάρη του μοντέλου τυχαίων επιδράσεων μειώνει τις σχετικές διαφορές τους. Όσο αυξάνεται η ετερογένεια, τόσο διαφορετικά είναι τα βάρη για τα δύο μοντέλα (Nikolakopoulou et al., 2014).

Από τα παραπάνω προκύπτει, ότι η συνολική εκτίμηση που ορίζεται ως ο σταθμισμένος μέσος όρος των παρατηρούμενων μεγεθών δεν είναι ίδια για τα δύο μοντέλα. Σε γενικές γραμμές όταν η ετερογένεια δεν είναι μεγάλη, οι δύο μέθοδοι δίνουν παρόμοιες εκτιμήσεις αλλά διαφορετικά διαστήματα εμπιστοσύνης (Nikolakopoulou et al., 2014). Συγκεκριμένα, το μοντέλο τυχαίων επιδράσεων δίνει πιο συντηρητικά αποτελέσματα και μεγαλύτερα διαστήματα εμπιστοσύνης σε σχέση με το μοντέλο σταθερών επιδράσεων (Nikolakopoulou et al., 2014). Τα δύο μοντέλα δίνουν ακριβώς τα ίδια αποτελέσματα, όταν η ετερογένεια είναι ίση με το μηδέν. Στις περιπτώσεις αυτές, παρότι τα αποτελέσματα είναι ίδια, δεν παύει η ερμηνεία της συνολικής επίδρασης να προσεγγίζεται με διαφορετικό τρόπο.

Τέλος, οι μεγάλες σε μέγεθος μελέτες έχουν μεγαλύτερη επιρροή στο συνολικό αποτέλεσμα στο μοντέλο σταθερών επιδράσεων, σε αντίθεση με τις μικρές που ευνοούνται περισσότερο από το μοντέλο τυχαίων επιδράσεων (Nikolakopoulou et al., 2014). Είναι προφανές, ότι οι μεγάλες σε μέγεθος μελέτες έχουν μεγαλύτερη ακρίβεια και επομένως μικρότερη διακύμανση σε σχέση με τις μικρές. Στο μοντέλο σταθερών επιδράσεων τα βάρη ανατίθενται ως το

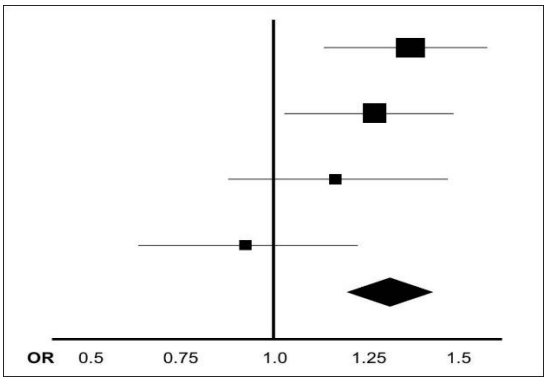
αντίστροφο της διακύμανσης των μελετών και συνεπώς είναι μεγαλύτερα σε σχέση με τα αντίστοιχα βάρη των μικρών μελετών. Στο μοντέλο τυχαίων επιδράσεων όμως τα βάρη ανατίθενται ως το αντίστροφο του αθροίσματος της διακύμανσης και της ετερογένειας των μελετών. Η ύπαρξη της ετερογένειας στον υπολογισμό των βαρών, η οποία είναι κοινή για όλες τις μελέτες, οδηγεί σε πιο ισορροπημένα αποτελέσματα, τα οποία υποβαθμίζουν την επιρροή των μεγάλων μελετών.

4.6 Διάγραμμα δάσους (forest plot)

Στην πλειονότητα των περιπτώσεων οι μετα-αναλύσεις συνοδεύονται από οπτικές παρουσιάσεις των αποτελεσμάτων τους, που είναι γνωστές ως διαγράμματα δάσους (forest plots). Πρόκειται για διαγράμματα που απεικονίζουν, τους εκτιμητές και τα διαστήματα εμπιστοσύνης της κάθε μελέτης καθώς και τα αντίστοιχα της συνολικής επίδρασης τους. Ο οριζόντιος άξονας αντιστοιχεί στο εκτιμώμενο μέτρο επίδρασης, ενώ ο κατακόρυφος άξονας αντιπροσωπεύει την γραμμή πάνω στην οποία το εκτιμώμενο μέγεθος επίδρασης δεν επιφέρει κανένα αποτέλεσμα (no effect line) (Nikolakopoulou et al., 2014a). Αν το μέτρο επίδρασης είναι ο λόγος κινδύνου (RR) ή ο λόγος αναλογιών (OR), τότε ο κατακόρυφος άξονας τέμνει τον οριζόντιο στην μονάδα. Αντίστοιχα, αν το μέτρο επίδρασης είναι η διαφορά κινδύνου (RD), η μέση διαφορά (MD) ή η τυποποιημένη μέση διαφορά (MSD), τότε ο κατακόρυφος άξονας τέμνει τον οριζόντιο στο μηδέν.

Πέρα από τον οριζόντιο άξονα που αντιπροσωπεύει το εκτιμώμενο μέτρο επίδρασης στο διάγραμμα, υπάρχουν και άλλες οριζόντιες γραμμές διαφορετικού μήκους, κάθε μία από τις οποίες διαπερνά ένα τετράγωνο. Οι γραμμές αντιστοιχούν στις μελέτες, ενώ τα τετράγωνα παριστάνουν τις εκτιμήσεις τους. Το μέγεθος των τετραγώνων διαφέρει, αφού καθορίζεται από το αντίστοιχο μέγεθος της μελέτης. Ακόμα, το μήκος των οριζόντιων γραμμών

αντιπροσωπεύει το διάστημα εμπιστοσύνης της εκτίμησης για αυτό και ποικίλει (Nikolakopoulou et al., 2014a). Τέλος, στη βάση του διαγράμματος υπάρχει μία οριζόντια γραμμή πάνω στη οποία βρίσκεται ένας ρόμβος. Ο ρόμβος απεικονίζει τη συνολική επίδραση της μετα-ανάλυσης και συμβατικά καλείται «διαμάντι», ενώ το μήκος της οριζόντιας γραμμής αναπαριστά το αντίστοιχο διάστημα εμπιστοσύνης. Η χρήση των διαγραμμάτων δάσους είναι ιδιαίτερα διαδεδομένη, αφού προσφέρουν μια σαφή και άμεση εικόνα της μετα-ανάλυσης, ενώ η ερμηνεία τους δεν απαιτεί ιδιαίτερες γνώσεις.



Εικόνα 4.1: Διάγραμμα δάσους (forest plot)

4.7 Τρόποι ελέγχου και υπολογισμός ετερογένειας

Η αξιόπιστη διεξαγωγή της μετα-ανάλυσης απαιτεί τον εντοπισμό πιθανής ετερογένειας των μελετών. Έχουν αναπτυχθεί πολλοί τρόποι για τον έλεγχο της ύπαρξης και τον υπολογισμό της ετερογένειας. Πιο αναλυτικά:

Έλεγχος ετερογένειας με το διάγραμμα δάσους (forest plot)

Μία πρώτη οπτική ένδειξη για την ύπαρξη ή μη ετερογένειας προσφέρει το διάγραμμα δάσους. Συγκεκριμένα, ετερογένεια έχουμε σε περιπτώσεις που τα διαστήματα εμπιστοσύνης δεν έχουν κοινά σημεία, δηλαδή οι οριζόντιες γραμμές των μελετών στο διάγραμμα δεν αλληλοκαλύπτονται (Higgins &

Green, 2008). Μάλιστα όσο περισσότερο δεν αλληλοκαλύπτονται οι οριζόντιες γραμμές, τόσο περισσότερη ετερογένεια έχουμε. Αντίθετα, σε περιπτώσεις απουσίας ετερογένειας τα διαστήματα εμπιστοσύνης έχουν κοινά σημεία. Η χρήση των διαγραμμάτων δάσους για τον έλεγχο της ετερογένειας των μελετών είναι απλή και δεν απαιτεί ιδιαίτερες γνώσεις. Ωστόσο, είναι μία οπτική μέθοδος και υπόκειται στην υποκειμενικότητα του ερευνητή.

Q-test

Το Q-test αποτελεί έναν στατιστικό τρόπο ελέγχου της ετερογένειας. Έστω ότι έχουμε *k* μελέτες σε μία μετα-ανάλυση και θέλουμε να ελέγξουμε την μηδενική υπόθεση ***H_o: οι μελέτες είναι ομοιογενείς*** έναντι της εναλλακτικής υπόθεσης ***H_a: οι μελέτες δεν είναι ομοιογενείς.***

Για τον έλεγχο χρησιμοποιείται το στατιστικό $Q = \sum_{i=1}^{k} w_{i,FE}(y_i - \hat{\mu}_{FE})^2$, όπου $w_{i,FE} = \frac{1}{{\sigma_i}^2}$ τα βάρη και $\hat{\mu}_{FE} = \frac{\sum_{i=1}^{k} w_{i,FE} y_i}{\sum_{i=1}^{k} w_{i,FE}}$ η συνολική επίδραση του μοντέλου σταθερών επιδράσεων. Το στατιστικό προτάθηκε από τον Cochran το 1954 γι' αυτό και είναι γνωστό ως Q στατιστικό του Cochran. Αποδεικνύεται εύκολα ότι ακολουθεί X^2 κατανομή με *k*-1 βαθμούς ελευθερίας (Higgins & Thompson, 2002).

Έχουμε $y_i \sim N(\mu, {\sigma_i}^2) \leftrightarrow$

$$\frac{y_i - \mu}{\sigma_i} \sim N(0,1) \leftrightarrow$$

$$\frac{(y_i - \mu)^2}{{\sigma_i}^2} \sim N^2(0,1) \equiv {\chi_1}^2 \leftrightarrow$$

$$\sum_{i=1}^{k} \frac{(y_i - \mu)^2}{{\sigma_i}^2} \sim {\chi_k}^2 \leftrightarrow$$

$$\sum_{i=1}^{k} w_i (y_i - \mu)^2 \sim {\chi_k}^2$$

Άρα, $Q = \sum_{i=1}^{k} w_i (y_i - \hat{\mu})^2 \sim {\chi_{k-1}}^2$

Περιπτώσεις όπου η p-τιμή που προκύπτει από το στατιστικό είναι μεγαλύτερη του επιπέδου σημαντικότητας, δηλώνουν ύπαρξη ετερογένειας, το μέγεθος της οποίας υποδεικνύεται από την τιμή του στατιστικού (Higgins & Thompson, 2002).

Το Q-test αποτελεί τον πιο διαδεδομένο τρόπο ελέγχου ύπαρξης ετερογένειας. Ανταποκρίνεται καλά για μεγάλο αριθμό μελετών. Ωστόσο η ισχύς του ελέγχου είναι χαμηλή, όταν τα δεδομένα είναι αραιά (sparse data) ή όταν κάποια μελέτη έχει πολύ μεγαλύτερο βάρος από τις υπόλοιπες (Hardy & Tompson, 1988).

Το μέτρο I^2

Ένας τρόπος υπολογισμού της ετερογένειας πραγματοποιείται με το μέτρο $I^2 = \frac{Q-k+1}{Q} * 100\%$. Το μέτρο I^2 προτάθηκε από τους Higgins & Thompson το 2002 και εκφράζει το ποσοστό της μεταβλητότητας, που οφείλεται στην ετερογένεια και όχι στην δειγματική διακύμανση. Είναι φανερό ότι όταν το I^2 είναι ίσο με μηδέν δεν υπάρχει ετερογένεια, δηλαδή το τ^2 είναι ίσο με μηδέν και συνεπώς το μοντέλο σταθερών επιδράσεων και το μοντέλο τυχαίων επιδράσεων ταυτίζονται. Τιμές του I^2 μικρότερες του 20% δηλώνουν μικρή ετερογένεια, μεταξύ 20% - 60% μέτρια, ενώ τιμές μεγαλύτερες του 60% αντιστοιχούν σε μεγάλο βαθμό ετερογένειας των μελετών. Στην πράξη τις περισσότερες φορές το I^2 είναι μεγαλύτερο του 50%, καθώς το ποσοστό της μεταβλητότητας οφείλεται περισσότερο στην παρουσία ετερογένειας από ότι στην δειγματική διακύμανση. Η χρήση του μέτρου I^2 για τον έλεγχο της ετερογένειας δεν συνίσταται σε περιπτώσεις που υπάρχουν λίγες μελέτες στην μετα-ανάλυση.

4.8 Μετα-παλινδρόμηση (Meta-regression)

Συχνά μια συστηματική ανασκόπηση σταματά μετά την διεξαγωγή της μετα-ανάλυσης και την λήψη του συνολικού μεγέθους επίδρασης των μελετών. Τίθεται λοιπόν το ζήτημα, αν είναι απαραίτητη μια περαιτέρω στατιστική ανάλυση για τον εντοπισμό των παραγόντων που επηρεάζουν τα μεγέθη επίδρασης και το συγκεντρωτικό αποτέλεσμα. Στις περιπτώσεις όπου η ετερογένεια των μελετών είναι μηδενική, η μετα-ανάλυση είναι αρκετή για να συνοψίσει τη δημοσιευμένη πληροφορία. Ωστόσο, όταν υπάρχει σημαντική ετερογένεια στα αποτελέσματα που μας ενδιαφέρουν, είναι σκόπιμο να συνεχιστεί η διερεύνηση και να διασαφηνιστεί αν η πηγή της ετερογένειας οφείλεται στα χαρακτηριστικά (methodological diversity) ή στον πληθυσμό της μελέτης (clinical diversity). Η στατιστική μέθοδος με την οποία επιτυγχάνεται η διερεύνηση της ετερογένειας των μελετών καλείται μετα-παλινδρόμηση (meta-regression). Εφαρμόζεται μετά την μετα-ανάλυση και θεωρείται επέκτασή της, αφού βοηθάει στην καλύτερη ερμηνεία των αποτελεσμάτων.

Η μέθοδος της μετα-παλινδρόμησης στοχεύει στο να συσχετίσει τα μεγέθη επίδρασης με ένα ή περισσότερα χαρακτηριστικά των μελετών που περιλαμβάνονται στην μετα-ανάλυση (Thomson & Higgins, 2002). Σε αυτή την κατεύθυνση έχουν αναπτυχθεί δύο διαφορετικές προσεγγίσεις: τα μοντέλα σταθερών επιδράσεων μετα- παλινδρόμησης (fixed effect meta-regression) και τα μοντέλα τυχαίων επιδράσεων μετα-παλινδρόμησης (random effect meta-regression). Η χρήση ενός μοντέλου μετα-παλινδρόμησης σταθερών επιδράσεων στην πράξη φαίνεται ακατάλληλη, αφού τα μοντέλα σταθερών επιδράσεων υποθέτουν ότι η ετερογένεια είναι μηδενική. Παρόλα αυτά υπάρχουν αναφορές τους στην βιβλιογραφία. Στο πλαίσιο της διατριβής θα επικεντρωθούμε σε μοντέλα μετα-παλινδρόμησης τυχαίων επιδράσεων. Ακολουθεί η περιγραφή του πιο διαδεδομένου μοντέλου μετα-παλινδρόμησης τυχαίων επιδράσεων. Έστω ότι y_i είναι η εκτίμηση του μεγέθους επίδρασης της

i-οστής μελέτης. Υποθέτουμε ότι το y_i ακολουθεί κανονική κατανομή με μέση τιμή το πραγματικό μέγεθος επίδρασης θ_i και διακύμανση το σφάλμα δειγματοληψίας ${\sigma_i}^2$. Υποθέτουμε ακόμα ότι και το θ_i ακολουθεί κανονική κατανομή με μέση τιμή μια συνάρτηση των επεξηγηματικών μεταβλητών της i-οστής μελέτης και διακύμανση την ετερογένεια τ^2. Επομένως υποθέτουμε ότι:

$y_i \sim N(\theta_i, {\sigma_i}^2)$ και

$\theta_i \sim N(\chi_i \beta, \tau^2),\ i = 1, \dots, k$

Συνδυάζοντας τις παραπάνω κατανομές έχουμε ότι:

$$y_i \sim N\left(\chi_i \beta, \tau^2 + {\sigma_i}^2\right),\ i = 1, \dots, k$$

Το μοντέλο γραμμικής παλινδρόμησης που προκύπτει κάτω από την υπόθεση της κανονικής κατανομής είναι:

$y_i = \chi_i \beta + u_i + \varepsilon_i$ με $u_i \sim N(0, \tau^2),\ \varepsilon_i \sim N(0, {\sigma_i}^2)$ τα αντίστοιχα σφάλματα.

Η μετα-παλινδρόμηση αποτελεί αδιαμφισβήτητα ένα ισχυρό εργαλείο που συμβάλει στη διερευνητική ανάλυση της ετερογένειας με στόχο την καλύτερη κατανόηση των αποτελεσμάτων της μετα-ανάλυσης. Στην πράξη υπάρχουν πολλοί περιορισμοί που μπορεί να διαστρεβλώσουν την εγκυρότητα των αποτελεσμάτων των μοντέλων της μετα-παλινδρόμησης. Αρχικά, το μέγεθος του δείγματος μπορεί να είναι ανεπαρκές για την πραγματοποίηση μιας μετα-παλινδρόμησης. Ακόμα, οι δημοσιευμένες μελέτες ενδέχεται να μην λαμβάνουν υπόψιν πληροφορίες που είναι απαραίτητες για το μοντέλο και αφορούν τη συσχέτιση των παραγόντων, δημιουργώντας έτσι αδυναμία προσαρμογής του σε συγχυτικούς παράγοντες. Τέλος, ακόμα και όταν οι δημοσιευμένες μελέτες περιέχουν πληροφορίες που αφορούν τη συσχέτιση των παραγόντων, μπορεί η φύση των χαρακτηριστικών των μελετών να είναι τέτοια, που να ευνοεί τη συσχέτιση μεταξύ τους δημιουργώντας πρόβλημα πολυσυγγραμμικότητας

(Thomson & Higgins, 2002). Από τα παραπάνω προκύπτει ότι η αλόγιστη χρήση της μετα-παλινδρόμησης, σε περιπτώσεις που δεν είναι αναγκαία, οδηγεί σε παρερμηνείες. Συνεπώς, η διεξαγωγή της δεν είναι πάντοτε απαραίτητη μετά το πέρας της συστηματικής ανασκόπησης.

ΚΕΦΑΛΑΙΟ 5
ΣΥΣΤΗΜΑΤΙΚΟ ΣΦΑΛΜΑ ΔΗΜΟΣΙΕΥΣΗΣ
(PUBLICATION BIAS)

Στο Κεφάλαιο 5 γίνεται εκτενής αναφορά στην έννοια του συστηματικού σφάλματος δημοσίευσης (publication bias) καθώς και στην πιθανή επίδραση των μικρών σε μέγεθος μελετών στα αποτελέσματα της μετα-ανάλυσης (small-study effect). Δίνεται ακόμα λεπτομερής περιγραφή του διαγράμματος κώνου (funnel plot) και της πιο εξελιγμένης έκδοσή του που είναι γνωστή ως contour enhanced funnel plot. Στη συνέχεια, εστιάζουμε στην ερμηνεία της πιθανής ασυμμετρίας των παραπάνω διαγραμμάτων. Στο πλαίσιο αυτό γίνεται αναφορά στα διάφορα στατιστικά test και κυρίως στα test γραμμικής παλινδρόμησης που έχουν αναπτυχθεί για τον έλεγχο της ασυμμετρίας. Τέλος, έχοντας ως στόχο την εξακρίβωση του συστηματικού σφάλματος δημοσίευσης, αναλύονται οι μέθοδοι Fail-Safe N, Trim and Fill καθώς και η στατιστική μέθοδος, που στηρίζεται στο μοντέλο επιλογής μελετών προς δημοσίευση, που προτάθηκε από τον Copas (Copas, 1999, Copas & Shi, 2000) (Copas selection model).

5.1 Είδη μεροληψίας αναφορών (Reporting bias)

Η μετα-ανάλυση ως τεχνική έχει δεχτεί κριτική, καθώς εμπεριέχει ορισμένες μεθοδολογικές αδυναμίες (Wilson, 1999). Στο πλαίσιο της παρούσης διατριβής θα μας απασχολήσει η αδυναμία προσαρμογής της μεθόδου απέναντι στην μεροληπτική συλλογή μελετών κατά το στάδιο της συστηματικής ανασκόπησης, η οποία αναφέρεται στην βιβλιογραφία με τον όρο μεροληψία αναφορών (reporting bias). Συγκεκριμένα, όταν ο ερευνητής κατά την διεξαγωγή μιας ανασκόπησης συμπεριλαμβάνει επιλεκτικά στην ανάλυση μελέτες, κατευθύνει τα αποτελέσματα δημιουργώντας εσφαλμένα

συμπεράσματα. Οι παράγοντες που διαμορφώνουν τον μεροληπτικό τρόπο επιλογής των μελετών από την πλευρά του ερευνητή είναι πολλοί. Για να διευκολυνθεί η επικοινωνία μεταξύ των επιστημόνων διακρίνονται σε κατηγορίες.

Η πιο γνωστή κατηγορία είναι το συστηματικό σφάλμα δημοσίευσης (publication bias). Πρόκειται για την τάση να δημοσιεύονται με μεγαλύτερη πιθανότητα οι μελέτες που καταλήγουν σε στατιστικά σημαντικά αποτελέσματα από εκείνες που δεν καταλήγουν. Μειώνει σημαντικά την εγκυρότητα μιας μετα-ανάλυσης (Smith R, 1999). Ένα άλλο είδος μεροληψίας είναι η μεροληψία γλώσσας (language bias). Μελέτες που δημοσιεύονται στην αγγλική γλώσσα έχουν μεγαλύτερη πιθανότητα να εντοπιστούν από ερευνητές, που διεξάγουν μία συστηματική ανασκόπηση, από αυτές που δημοσιεύονται σε οποιαδήποτε άλλη γλώσσα. Επισημαίνεται ακόμα η μεροληψία ετεροαναφορών (citation bias). Πρόκειται για την τάση να εντοπίζονται και να συμπεριλαμβάνονται κατά την διάρκεια μιας συστηματικής έρευνας μελέτες που αναφέρονται συχνότερα σε εργασίες τρίτων.

Είδος μεροληψίας αποτελεί και η επιλεκτική παρουσίαση αποτελεσμάτων (selective reporting). Συχνά οι ερευνητές επιδιώκουν να αποστέλλουν προς κρίση στα περιοδικά μόνο ένα μικρό μέρος των ερευνητικών τους αποτελεσμάτων. Με την πρακτική αυτή αποκρύπτεται το πραγματικό σύνολο των πληροφοριών. Παράδειγμα αυτού του σφάλματος είναι όταν ο ερευνητής στέλνει προς δημοσίευση τα αποτελέσματα από μια υπό-ομάδα μελετών ή όταν στέλνει τα αποτελέσματα από μια έκβαση (π.χ. αποτελεσματικότητα) και αποκρύπτει σκοπίμως τα αποτελέσματα από μια άλλη (π.χ. παρενέργειες). Επιπλέον, θα πρέπει να αναφερθεί και η μεροληψία του μέσου δημοσίευσης (location bias). Πολλές φορές τα μικρά περιοδικά, στοχεύοντας στον εντυπωσιασμό της κοινής γνώμης, επιλέγουν να δημοσιεύουν άρθρα με στατιστικώς σημαντικά αποτελέσματα. Τέλος, γίνεται αναφορά στην

μεροληψία χορηγού (sponsorship bias). Μελέτες που χρηματοδοτούνται από χορηγούς (π.χ. φαρμακευτικές εταιρίες) με μη στατιστικά σημαντικά αποτελέσματα δεν καταλήγουν συχνά προς δημοσίευση με αποτέλεσμα να μην μπορούν να εντοπιστούν κατά την διαδικασία μιας συστηματικής ανασκόπησης. Το ίδιο συμβαίνει και με πολλές μεταπτυχιακές ή διδακτορικές διατριβές καθώς και με έρευνες που παρουσιάζονται σε συνέδρια, δημιουργώντας έτσι αυτό που στη στατιστική αποκαλείται «γκρίζα βιβλιογραφία».

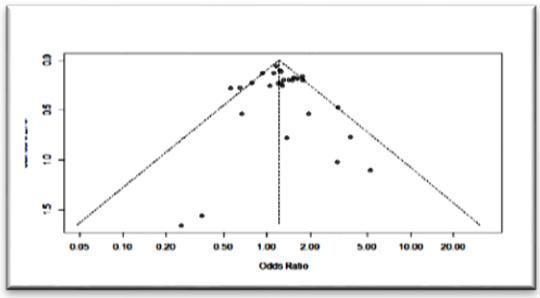
Από τα παραπάνω είδη της μεροληψίας θα εστιάσουμε στο συστηματικό σφάλμα δημοσίευσης και θα αναζητήσουμε τρόπους εξακρίβωσης του με στόχο τη μεγαλύτερη εγκυρότητα των μετα-αναλύσεων. Στη συνέχεια, ακολουθούν γραφικοί τρόποι που επιτρέπουν την πιθανή ανίχνευση του συστηματικού σφάλματος δημοσίευσης. Στο σημείο αυτό θα πρέπει να τονιστεί ότι οι διάφοροι μέθοδοι για την ανίχνευση του συστηματικού σφάλματος δημοσίευσης δεν μπορούν να ελεγχθούν μιας και τα δεδομένα που χρειάζονται για τον έλεγχό τους δεν έχουν δημοσιευτεί. Μόνο υποθέσεις μπορούμε να κάνουμε και η πλέον συνηθισμένη υπόθεση είναι ότι οι μικρές σε μέγεθος μελέτες με μη στατιστικά σημαντικά αποτελέσματα δεν δημοσιεύονται. Για τον λόγο αυτό, η ύπαρξη σχέσεως μεγέθους δείγματος και μεγέθους επίδρασης αποτελεί μια ένδειξη (και όχι απόδειξη) ότι υπάρχει συστηματικό σφάλμα δημοσίευσης.

5.2 Διάγραμμα κώνου (funnel plot)

Το διάγραμμα κώνου (funnel plot) προτάθηκε το 1984 από τους Light & Pillemer (Light & Pillemer, 1984). Πρόκειται για ένα διάγραμμα διασποράς των μελετών, του μεγέθους επίδρασης και ενός μέτρου ακρίβειας, συνήθως του τυπικού σφάλματος (Mavridis & Salanti, 2014). Πιο αναλυτικά, στο κέντρο του διαγράμματος βρίσκεται η εκτίμηση του μεγέθους επίδρασης. Ο κάθετος άξονας αντιστοιχεί στο τυπικό σφάλμα και ο οριζόντιος στο μέγεθος επίδρασης

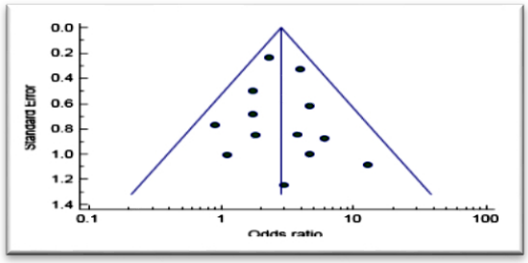
(Sterne & Egger, 2001). Σε αντίθεση με τα κοινά διαγράμματα το τυπικό σφάλμα αυξάνεται από το πάνω μέρος του κάθετου άξονα προς το κάτω (Γαλάνης, 2009).

Με τη βοήθεια του διαγράμματος κώνου μπορούμε να έχουμε μια πρώτη ένδειξη για την ύπαρξη ή μη συστηματικού σφάλματος δημοσίευσης. Συγκεκριμένα, όταν δεν παρατηρείται συστηματικό σφάλμα δημοσίευσης, τα σημεία που αντιστοιχούν στις διάφορες μελέτες δημιουργούν την οπτική απεικόνιση ενός συμμετρικού αντεστραμμένου χωνιού. Στην περίπτωση αυτή οι μεγάλες σε μέγεθος μελέτες έχουν μικρή τυπική απόκλιση, δηλαδή μεγάλη ακρίβεια και βρίσκονται στην κορυφή του χωνιού (Sterne & Harbord, 2004). Αντίστοιχα, οι μικρές σε μέγεθος μελέτες έχουν μεγάλη τυπική απόκλιση, δηλαδή μικρή ακρίβεια και βρίσκονται στο κάτω μέρος του διαγράμματος. Το διάγραμμα κώνου όμως δεν είναι πάντοτε συμμετρικό. Η ασυμμετρία του διαγράμματος δεν συνδέεται αποκλειστικά με την έννοια του συστηματικού σφάλματος δημοσίευσης. Συχνά οι μικρές σε μέγεθος μελέτες διεξάγονται με πιο προσεγμένες πειραματικές συνθήκες σε σχέση με τις μεγάλες, δημιουργώντας έτσι έντονη παρεμβολή στα δεδομένα και καταλήγοντας σε παραποιημένες εκτιμήσεις των μεγεθών επίδρασης (Sterne & Egger, 2001). Η επιρροή αυτή των μικρών μελετών δηλώνεται με τον όρο small-study effect. Υπάρχει λοιπόν σχέση μεταξύ ασυμμετρίας διαγράμματος κώνου, μεγέθους δείγματος μελετών και εκτιμήσεων των μεγεθών επίδρασης. Συνεπώς, κάθε παράγοντας, που σχετίζεται με το μέγεθος των μελετών ή το μέγεθος επίδρασης καθώς και τον συνδυασμό τους, μπορεί να ευθύνεται για την ασυμμετρία του διαγράμματος.



Διάγραμμα 5.1: Ασύμμετρο funnel plot

Παρουσιάζεται έντονη ασυμμετρία. Οι δημοσιευμένες μελέτες παρατηρούνται στο δεξιό μέρος του διαγράμματος, ενώ απουσιάζουν στο κάτω αριστερό.



Διάγραμμα 5.2: Συμμετρικό funnel plot

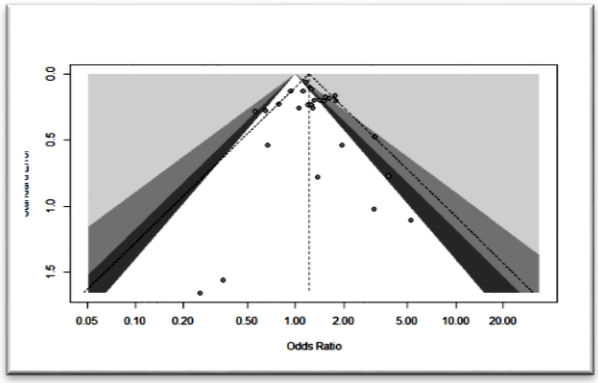
Το διάγραμμα χαρακτηρίζεται συμμετρικό. Οι δημοσιευμένες μελέτες δημιουργούν την οπτική απεικόνιση ενός αντεστραμμένου χωνιού.

5.3 Contour enhanced funnel plot

Με βάση τα παραπάνω γίνεται άμεσα αντιληπτό ότι η χρήση των διαγραμμάτων κώνου από μόνη της δεν αρκεί για να αποφανθούμε, αν η ασυμμετρία που ενδεχομένως να εμφανίζεται στο διάγραμμα οφείλεται σε συστηματικό σφάλμα δημοσίευσης ή σε άλλους παράγοντες. Τα τελευταία χρόνια έχουν προταθεί πιο εξελιγμένα οπτικά διαγράμματα που διευκολύνουν τον ερευνητή στην ερμηνεία

μιας πιθανής ασυμμετρίας. Στην βιβλιογραφία είναι γνωστά με τον όρο contour enhanced funnel plots. Όπως και στα απλά διαγράμματα κώνου ο κάθετος άξοναs αντιστοιχεί στο τυπικό σφάλμα και ο οριζόντιος στο μέγεθος επίδρασης. Το τυπικό σφάλμα αυξάνεται από το πάνω μέρος του κάθετου άξονα προς το κάτω. Κέντρο του διαγράμματος θεωρείται η τιμή μη επίδρασης κάτω από την μηδενική υπόθεση. Συγκεκριμένα, αν το μέγεθος επίδρασης είναι η διαφορά κινδύνου (RD), η μέση διαφορά (MD) ή η τυποποιημένη μέση διαφορά (SMD), τότε το κέντρο ταυτίζεται με το μηδέν. Αντίστοιχα, αν το μέτρο επίδρασης είναι ο λόγος κινδύνου (RR) ή ο λόγος αναλογιών (OR), τότε το κέντρο ταυτίζεται με τη μονάδα.

Το contour enhanced funnel plot αποτελείται από σκιασμένες περιοχές, που υποδεικνύουν τα διάφορα επίπεδα στατιστικής σημαντικότητας (Peters et al., 2008). Συγκεκριμένα, η μη σκιασμένη λευκή περιοχή στη μέση αντιστοιχεί σε τιμές $p > 0.10$, η γκρίζα σκιασμένη περιοχή σε $0.05 < p < 0.10$, η πιο σκούρα γκρίζα περιοχή αντιστοιχεί σε $0.01 < p < 0.05$ και η περιοχή έξω από το διάγραμμα σε τιμές $p < 0.01$. Η ερμηνεία της ασυμμετρίας στηρίζεται στον παρακάτω κανόνα. Αν οι περισσότερες μικρές μελέτες είναι στατιστικά σημαντικές, αποτελεί ένδειξη ότι η ασυμμετρία του διαγράμματος οφείλεται σε συστηματικό σφάλμα δημοσίευσης. Για την οπτική κατανόηση του contur enhanced funnel plot παρατίθεται το διάγραμμα 5.3.



Διάγραμμα 5.3: Contour enhanced funnel plot

Στο διάγραμμα είναι αισθητή η έλλειψη συμμετρίας. Οι μελέτες που ευθύνονται για την ασυμμετρία είναι στη λευκή περιοχή και δεν δείχνουν στατιστικά σημαντικά αποτελέσματα. Η ασυμμετρία μπορεί να οφείλεται στην ετερογένεια των μελετών.

Παρατηρήσεις

Η χρήση των contour enhanced funnel plots είναι ευρέως διαδεδομένη. Η εφαρμογή τους είναι απλή και υποστηρίζεται από διάφορα στατιστικά πακέτα. Ακόμα, βελτιώνουν σημαντικά την ερμηνεία της ασυμμετρίας του διαγράμματος κώνου, με αποτέλεσμα να προτείνεται η συστηματική τους χρήση από την έρευνα. Ωστόσο, σε ορισμένες περιπτώσεις οδηγούν σε παραπλανητικά συμπεράσματα, αφού η ασυμμετρία που ενδεχομένως παρατηρείται μπορεί να είναι αποτέλεσμα του τρόπου δειγματοληψίας και τίποτα περισσότερο.

5.4 Στατιστικά test για το έλεγχο της ασυμμετρίας του funnel plot

Η οπτική απεικόνιση ενός διαγράμματος κώνου δίνει στον ερευνητή μια πρώτη εντύπωση για την ύπαρξη ή μη ασυμμετρίας στο διάγραμμα. Πολλές φορές όμως, δεν είναι σαφής ο διαχωρισμός και η ερμηνεία τίθεται στην υποκειμενική ευχέρεια του μελετητή. Για να αποφευχθεί η σύγχυση σε ότι αφορά την ύπαρξη ή μη συμμετρίας σε ένα διάγραμμα κώνου, έχουν αναπτυχθεί διάφορα στατιστικά test. Στο σημείο αυτό είναι ανάγκη να διευκρινιστεί ότι τα test εστιάζουν στον χαρακτηρισμό ενός διαγράμματος ως συμμετρικό ή μη και όχι αν η πιθανή ασυμμετρία οφείλεται σε συστηματικό σφάλμα δημοσίευσης ή άλλους παράγοντες.

5.4.1 Begg's rank correlation test

Το μη παραμετρικό test που προτάθηκε από τους Begg & Mazumdar (Begg & Mazumdar, 1994) εξετάζει την συσχέτιση μεταξύ των εκτιμητών των μεγεθών επίδρασης και των αντίστοιχων διακυμάνσεών τους. Ο έλεγχος ταυτίζεται με το γνωστό Kendall' tau rank correlation test. Έστω ότι έχουμε k μελέτες. Συμβολίζουμε με $\widehat{Y_i}$ την εκτίμηση του μεγέθους επίδρασης της i-οστής μελέτης και με $\widehat{v_i}$ την αντίστοιχη δειγματική διακύμανση. Στη συνέχεια τυποποιούμε τα μεγέθη επίδρασης, τα οποία παίρνουν την μορφή $Y_i^* = \frac{Y_i - \overline{Y}.}{\sqrt{\tilde{v}_i}}$, όπου $\overline{Y}. = \left(\sum_{j=1}^{k} \frac{\widehat{Y_j}}{\widehat{v_j}}\right) \Big/ \left(\sum_{j=1}^{k} \frac{1}{\widehat{v_j}}\right)$ η εκτίμηση του συνολικού μεγέθους επίδρασης που προέρχεται από το μοντέλο τυχαίων επιδράσεων και $\tilde{v}_i = \widehat{v_i} - \left(\sum_{j=1}^{k} \frac{1}{\widehat{v_j}}\right)^{-1}$ η διακύμανση της διαφορά $\widehat{Y_i} - \overline{Y}.$ Συμβολίζουμε με $Y^* = (Y_1^*, Y_2^*, \dots, Y_k^*)$ το σύνολο των τυποποιημένων εκτιμητών των μεγεθών επίδρασης και με $\hat{v} = (\hat{v}_1, \hat{v}_2, \dots, \hat{v}_k)$ το σύνολο των διακυμάνσεων των εκτιμητών των μεγεθών επίδρασης. Θα μας απασχολήσει ο έλεγχος της μηδενικής υπόθεσης H_0: *$Y^*, \hat{v}$ είναι ασυσχέτιστα* έναντι της εναλλακτικής H_a: *$Y^*, \hat{v}$ δεν είναι ασυσχέτιστα*. Από

τα παραπάνω σύνολα προκύπτουν k ζεύγη της μορφής $(Y_i^*, \hat{v}_i)$, $i = 1, \dots, k$. Δύο ζεύγη $(Y_i^*, \hat{v}_i)$ και $(Y_j^*, \hat{v}_j)$ θα ονομάζονται εναρμονισμένα (concordant) αν και τα δύο μέλη του ενός ζεύγους είναι μεγαλύτερα (αντίστοιχα μικρότερα) από τα αντίστοιχα μέλη του άλλου. Δηλαδή, αν $Y_i^* > Y_j^*$ (αντίστοιχα $Y_i^* < Y_j^*$) τότε $\hat{v}_i > \hat{v}_j$ (αντίστοιχα $\hat{v}_i < \hat{v}_j$). Με P θα συμβολίζουμε το συνολικό αριθμό των εναρμονισμένων ζευγών. Ακόμα, δύο ζεύγη $(Y_i^*, \hat{v}_i)$ και $(Y_j^*, \hat{v}_j)$ θα ονομάζονται μη εναρμονισμένα (discordant) αν η διάταξη των πρώτων μελών είναι αντίθετη από την διάταξη των δεύτερων. Δηλαδή αν $Y_i^* > Y_j^*$ (αντίστοιχα $Y_i^* < Y_j^*$) τότε $\hat{v}_i < \hat{v}_j$ (αντίστοιχα $\hat{v}_i > \hat{v}_j$). Αντίστοιχα με Q θα συμβολίζουμε το συνολικό αριθμό των μη εναρμονισμένων ζευγών. Τέλος, ισοβαθμούνται (tied) θα ονομάζονται τα ζεύγη που δεν είναι ούτε εναρμονισμένα ούτε μη εναρμονισμένα. Στον έλεγχο που προτάθηκε από τους Begg & Mazumdar γίνεται η υπόθεση ότι δεν υπάρχουν ισοβαθμούντα ζεύγη. Η στατιστική συνάρτηση του ελέγχου $Z = \frac{P-Q}{\sqrt{\frac{k(k-1)(2k+5)}{18}}}$ ακολουθεί ασυμπτωτικά τυπική κανονική κατανομή. Η μηδενική υπόθεση απορρίπτεται για επίπεδο σημαντικότητας α, αν $|Z| > z_{1-\alpha/2}$. Το $z_{1-\alpha/2}$ αντιπροσωπεύει το $1 - \alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής. Η απόρριψη της μηδενικής υπόθεσης συνδέεται με ασυμμετρία του διαγράμματος κώνου (Begg & Mazumdar, 1994).

Αξιολόγηση

Το Begg's rank correlation test είναι ευρέως διαδεδομένο για τον έλεγχο της ασυμμετρίας του διαγράμματος κώνου. Έχει το πλεονέκτημα, ότι μπορεί να εφαρμόζεται τόσο σε συνεχή όσο και σε διχότομα δεδομένα. Ωστόσο, δε δίνει έγκυρα αποτελέσματα για μετα-αναλύσεις με μικρό αριθμό μελετών (Peters et al., 2006). Ένα άλλο μειονέκτημα είναι ότι δεν είναι τόσο ισχυρό όσο τα

διάφορα test γραμμικής παλινδρόμησης (linear regression tests) που θα αναφερθούν στη συνέχεια (Peters et al., 2006).

5.4.2 Στατιστικά test γραμμικής παλινδρόμησης

5.4.2.1 Egger´s et al. test

Ένα μη παραμετρικό test, που βασίζεται στην γραμμική παλινδρόμηση, έχει προταθεί για τον έλεγχο της ασυμμετρίας του διαγράμματος κώνου από τους Matthias Egger et al. (Egger et al., 1997). Αν Y_i είναι η εκτίμηση του μεγέθους επίδρασης της i-οστής μελέτης και v_i η αντίστοιχη δειγματική διακύμανση, το σταθμισμένο μοντέλο γραμμικής παλινδρόμησης που στηρίχτηκε το test είναι:

$$\mathrm{Y}_i = \mathrm{b}_1 + \mathrm{b}_0 \sqrt{v_i} + \varepsilon_i, \text{ με βάρη } \frac{1}{v_i}, \varepsilon_i \sim N(0, v_i), \quad \mathrm{i} = 1, \dots, k.$$

Με βάση το παραπάνω μοντέλο γραμμικής παλινδρόμησης ελέγχεται η μηδενική υπόθεση H_0: $b_0 = 0$ έναντι της εναλλακτικής H_a: $b_0 \neq 0$. Η στατιστική συνάρτηση του ελέγχου $\mathrm{T} = \frac{\widehat{b_0}}{se(\widehat{b_0})}$ ακολουθεί t κατανομή με k-2 βαθμούς ελευθερίας. Η μηδενική υπόθεση απορρίπτεται για επίπεδο σημαντικότητας α, αν $|\mathrm{T}| > t_{k-2,1-a/2}$, όπου $t_{k-2,1-a/2}$ το $1 - a/2$ ποσοστιαίο σημείο της t κατανομής με k-2 βαθμούς ελευθερίας. Η απόρριψη της μηδενικής υπόθεσης υποδεικνύει ασυμμετρία στο διάγραμμα κώνου.

Αξιολόγηση

Η χρήση του Egger’s test είναι ευρέως διαδεδομένη για τον έλεγχο της ασυμμετρίας του διαγράμματος κώνου. Εφαρμόζεται τόσο σε συνεχή όσο και σε διχότομα δεδομένα. Όμως η προσαρμογή του πάνω σε διχότομα δεδομένα δίνει υψηλό σφάλμα τύπου Ι λόγω της συσχέτισης των εκτιμητών των μεγεθών επίδρασης (λογάριθμος του λόγου αναλογιών/κινδύνων) και των αντίστοιχων

τυπικών σφαλμάτων τους (Peters et al., 2006). Έχουν προταθεί πολλά εναλλακτικά test προκειμένου να ξεπεραστεί η παραπάνω αδυναμία.

5.4.2.2 Tang & Lui's test

Το 2000 στην προσπάθεια αντιμετώπισης της αδυναμίας του Egger΄s test πάνω στα διχότομα δεδομένα οι Tang & Lui πρότειναν ένα εναλλακτικό έλεγχο που θα στηρίζεται στην προσαρμογή του παρακάτω σταθμισμένου μοντέλου γραμμικής παλινδρόμησης (Tang & Lui, 2000):

$\mathrm{Y}_i = \mathrm{b}_1 + \mathrm{b}_0 \sqrt{N_i} + \varepsilon_i,$ με βάρη $\frac{1}{N_i}, \varepsilon_i \sim N(0, N_i), \quad \mathrm{i} = 1, \dots, k.$

Στην παραπάνω σχέση με Y_i αναπαρίσταται εκτίμηση του μεγέθους επίδρασης της i-οστής μελέτης και με N_i το συνολικό μέγεθος του δείγματος. Με βάση το προηγούμενο μοντέλο γραμμικής παλινδρόμησης ελέγχεται η μηδενική υπόθεση H_0: $b_0 = 0$ έναντι της εναλλακτικής H_a: $b_0 \neq 0$. Η στατιστική συνάρτηση του ελέγχου $\mathrm{T} = \frac{\widehat{b_0}}{se(\widehat{b_0})}$ ακολουθεί t κατανομή με k-2 βαθμούς ελευθερίας. Η μηδενική υπόθεση απορρίπτεται για επίπεδο σημαντικότητας α, αν $|\mathrm{T}| > t_{k-2,1-a/2}$, όπου $t_{k-2,1-a/2}$ το $1 - a/2$ ποσοστιαίο σημείο της t κατανομής με k-2 βαθμούς ελευθερίας. Η απόρριψη της μηδενικής υπόθεσης υποδεικνύει ασυμμετρία στο διάγραμμα κώνου.

5.4.2.3 Thompson & Sharp's test

Μια τροποποιημένη έκδοση του Egger΄s test προτάθηκε το 1999 από τους Thompson & Sharp (Thompson & Sharp, 1999). Η βασική ιδέα πάνω στην οποία στηρίχτηκαν είναι η τροποποίηση του σταθμισμένου μοντέλου γραμμικής παλινδρόμησης που προτάθηκε από τους Egger et al. Συγκεκριμένα πρότειναν την προσαρμογή του ακόλουθου σταθμισμένου μοντέλου γραμμικής παλινδρόμησης:

$Y_i = b_1 + b_0 \sqrt{v_i} + \varepsilon_i$, με βάρη $\frac{1}{v_i + \widehat{\tau^2}}$, $\varepsilon_i \sim N(0, v_i + \widehat{\tau^2})$, $i = 1, \ldots, k$.

Στην παραπάνω σχέση με Y_i αναπαρίσταται η εκτίμηση του μεγέθους επίδρασης της i-οστής μελέτης, v_i η αντίστοιχη δειγματική διακύμανση και με $\widehat{\tau^2}$ η εκτίμηση της διακύμανσης μεταξύ των μελετών με την μέθοδο των ροπών. Κατά αντιστοιχία με το Egger΄s test ελέγχεται η μηδενική υπόθεση H_0: $b_0 = 0$ έναντι της εναλλακτικής H_a: $b_0 \neq 0$. Η στατιστική συνάρτηση του ελέγχου $T = \frac{\widehat{b_0}}{se(\widehat{b_0})}$ ακολουθεί t κατανομή με k-2 βαθμούς ελευθερίας. Η μηδενική υπόθεση απορρίπτεται για επίπεδο σημαντικότητας α, αν $|T| > t_{k-2,1-a/2}$, όπου $t_{k-2,1-a/2}$ το $1 - a/2$ ποσοστιαίο σημείο της t κατανομής με k-2 βαθμούς ελευθερίας. Η απόρριψη της μηδενικής υπόθεσης υποδεικνύει ασυμμετρία στο διάγραμμα κώνου.

5.5 Τρόποι εξακρίβωσης συστηματικού σφάλματος δημοσίευσης

Όπως έχει ήδη αναφερθεί, ο εντοπισμός της ασυμμετρίας του διαγράμματος κώνου δεν συνδέεται απαραίτητα με το συστηματικό σφάλμα δημοσίευσης. Προκύπτει λοιπόν η ανάγκη της αναζήτησης νέων στατιστικών μεθόδων, με στόχο την έγκυρη εξακρίβωση της ύπαρξης ή μη συστηματικού σφάλματος δημοσίευσης. Κάτω από αυτό το πλαίσιο ακολουθεί λεπτομερής περιγραφή των μεθόδων Fail-safe N, Trim and Fill καθώς και της στατιστικής μεθόδου, που στηρίζεται στο μοντέλο επιλογής μελετών προς δημοσίευση που προτάθηκε από τον Copas (Copas, 1999, Copas & Shi, 2000) (Copas selection models).

5.5.1 Fail-safe N method

Η μέθοδος Fail-safe N αποτελεί μία από τις πρώτες τεχνικές με στόχο την εξακρίβωσης της ύπαρξης συστηματικού σφάλματος δημοσίευσης. Η βασική ιδέα είναι να εντοπίσει πόσες μελέτες με μη μηδενικό μέγεθος επίδρασης πρέπει

να προστεθούν, ώστε να προκύψει συνολικό μέγεθος επίδρασης μη στατιστικά σημαντικό.

Η μέθοδος βασίζεται στην στατιστική συνάρτηση $Z = \frac{\sum_{i=1}^{k} Z_i}{\sqrt{k}}$, που ακολουθεί ασυμπτωτικά τυπική κανονική κατανομή και έχει ως κρίσιμη περιοχή $Z \geq z_{\alpha/2}$. Με $z_{\alpha/2}$ συμβολίζουμε το $\alpha/2$ ποσοστιαίο σημείο της τυπικής κανονικής κατανομής. Το Z_i αντιπροσωπεύει το τυποποιημένο μέγεθος επίδρασης της i-οστής μελέτης και το k το συνολικό αριθμό των μελετών. Συμβολίζεται ακόμα με k_0 ο αριθμός των μελετών που πρέπει να προστεθούν, ώστε να προκύψει συνολικό μέγεθος επίδρασης μη στατιστικά σημαντικό. Στο σημείο αυτό γίνεται η υπόθεση ότι υπάρχουν αδημοσίευτες οι k_0 μελέτες και ότι δεν έχουν σημαντικά αποτελέσματα, δηλαδή το μέγεθος επίδρασής τους είναι μηδέν. Συνδυάζοντας τα παραπάνω υπολογίζουμε το k_0 από την σχέση:

$$\frac{\sum_{i=1}^{k} Z_i}{\sqrt{k + k_0}} < z_{\alpha/2} \leftrightarrow \left(\sum_{i=1}^{k} Z_i \right)^2 \Big/ \left(z_{\alpha/2} \right)^2 - k < k_0$$

Καταλήγουμε στην ύπαρξη συστηματικού σφάλματος δημοσίευσης στις περιπτώσεις όπου είναι μικρός ο αριθμός των μελετών που πρέπει να προστεθούν, ώστε να προκύψει συνολικό μέγεθος επίδρασης στατιστικά μη σημαντικό. Πριν χαρακτηριστεί «μικρή» η τιμή του k_0, απαιτείται σύγκριση αναλογικά με το συνολικό αριθμό των μελετών, ώστε να αποφευχθούν σφάλματα (Orwin, 1983).

Συγκεντρωτικά

Η μέθοδος υποθέτει ότι υπάρχουν k_0 αδημοσίευτες μελέτες με μη σημαντικά αποτελέσματα, δηλαδή με μεγέθη επίδρασης μηδέν που πρέπει να προστεθούν, ώστε να προκύψει συνολικό μέγεθος επίδρασης στατιστικά μη σημαντικό. Για τον υπολογισμό του αριθμού αυτών των μελετών γίνεται χρήση της σχέσης:

$$\mathbf{k_0} > \left(\sum_{\mathbf{i=1}}^{\mathbf{k}} \mathbf{Z_i}\right)^{\mathbf{2}} \Big/ \left(\mathbf{z_{\alpha/2}}\right)^{\mathbf{2}} - \mathbf{k}$$

Καταλήγουμε σε συστηματικό σφάλμα δημοσίευσης για «μικρές», αναλογικά με το συνολικό αριθμό των μελετών, τιμές του k_0.

Αξιολόγηση της μεθόδου

Η μέθοδος Fail-safe N θεωρείται μια απλή τεχνική για την εξακρίβωση του συστηματικού σφάλματος δημοσίευσης. Παρότι είναι εύκολη η κατανόησή της από τους ερευνητές, η χρήση της δεν είναι διαδεδομένη. Εμπεριέχει αρκετές μεθοδολογικές αδυναμίες. Αρχικά, το στατιστικό Z δεν ευθύνεται άμεσα για τον προσδιορισμό του αριθμού των μελετών που πρέπει να προστεθούν (Orwin, 1983). Επιπλέον, η μέθοδος δεν προσαρμόζεται πάνω στο μέγεθος επίδρασης παρά μόνο στη p-τιμή (Orwin, 1983). Επομένως, δεν υπάρχουν καθόλου ενδείξεις του «πραγματικού» αποτελέσματος του μεγέθους επίδρασης. Από την περιγραφή της μεθόδου γίνεται άμεσα αντιληπτό, ότι δεν λαμβάνεται καθόλου υπόψιν το σχήμα του διαγράμματος κώνου και αγνοείται η πιθανή ύπαρξη ετερογένειας των μελετών (Orwin, 1983). Τέλος, επιλέγει αυθαίρετα το μέγεθος επίδρασης για τις αδημοσίευτες μελέτες να είναι το μηδέν και σε ορισμένες περιπτώσεις δημιουργεί μελέτες που δεν υπάρχουν (make up data) με υποθετικά μεγέθη δείγματος, ειδικά αν $k_0 > 5k + 10$ (Iyengar & Greenhouse, 1988).

5.5.2 Trim and Fill

Στη συνέχεια, γίνεται αναφορά στη μη παραμετρική μέθοδο Trim and Fill, που έχει ως στόχο την εξακρίβωση του συστηματικού σφάλματος δημοσίευσης (Duval & Tweedie, 2000b). Η τεχνική βασίζεται στην επανόρθωση της συμμετρίας του διαγράμματος κώνου. Υποθέτουμε ότι λείπουν μελέτες στο ένα άκρο του διαγράμματος κώνου και η βασική ιδέα είναι να προστεθούν νέες,

ώστε να προκύψει διάγραμμα με συμμετρικό κέντρο (Duval & Tweedie, 2000a).

Τα βασικά βήματα της μεθόδου:

1. Εκτιμάται ο αριθμός των μελετών που θα πρέπει να αφαιρεθούν από την μία πλευρά του διαγράμματος κώνου, ώστε να προκύψει διάγραμμα με συμμετρικό κέντρο. Ακολουθεί αναλυτική περιγραφή της μεθόδου εκτίμησης.
2. Αφαιρούνται οι μελέτες και στη συνέχεια γίνεται μετα-ανάλυση πάνω σε αυτές.
3. Θεωρείται ως «πραγματικό» κέντρο του διαγράμματος κώνου η εκτίμηση που προέκυψε από την μετα-ανάλυση των μελετών που αφαιρέθηκαν.
4. Για κάθε μελέτη που αφαιρείται, δημιουργείται ένα ακριβές αντίγραφο στο κέντρο του διαγράμματος.
5. Γίνεται μετα-ανάλυση με όλες τις μελέτες, δηλαδή με αυτές που υπήρχαν αλλά και με αυτές που προστέθηκαν.

Οι μετα-αναλύσεις που απαιτούνται κατά την εφαρμογή της μεθόδου μπορούν να προσαρμοστούν τόσο με το μοντέλο σταθερών όσο και με το μοντέλο τυχαίων επιδράσεων. Δεδομένα προσαρμογής έδειξαν, ότι αν χρησιμοποιηθεί το μοντέλο σταθερών επιδράσεων στα βήματα 1-4 και το μοντέλο τυχαίων επιδράσεων στο βήμα 5, θα προκύψουν καλύτερα αποτελέσματα από το να χρησιμοποιηθεί το μοντέλο σταθερών επιδράσεων σε όλα τα βήματα (Schwarzer et al., 2009). Αντίστοιχα, με τη χρήση του μοντέλου σταθερών επιδράσεων στα βήματα 1-4 και τυχαίων επιδράσεων στο βήμα 5 προκύπτουν οριακά καλύτερα αποτελέσματα από το να χρησιμοποιηθεί το μοντέλο τυχαίων επιδράσεων σε όλα τα βήματα (Schwarzer et al., 2009).

Ακολουθεί η ανάλυση της μεθόδου που εκτιμά τον αριθμό των μελετών που θα πρέπει να αφαιρεθούν από την μια πλευρά του διαγράμματος κώνου. Έστω ότι έχουμε k μελέτες. Αρχικά, δημιουργούμε για κάθε μια από τις μελέτες την διαφορά της εκτίμησης του μεγέθους επίδρασης από τη συνολική εκτίμηση. Οι διαφορές έχουν πρόσημο «+» όταν η εκτίμηση του μεγέθους επίδρασης της κάθε μελέτης είναι μεγαλύτερη από τη συνολική και πρόσημο «-» στην αντίθετη περίπτωση. Στη συνέχεια διατάσσουμε τις διαφορές κατά αύξουσα σειρά με βάση τις απόλυτες τιμές τους και υπολογίζουμε τις τάξεις τους. Στο σημείο αυτό επισημαίνεται ότι αν δύο ή περισσότερες διαφορές είναι ίσες κατά απόλυτο τιμή, τότε η τάξη που παίρνει η καθεμία από αυτές είναι ίση με το μέσο όρο των τάξεων που θα είχαν αν ήταν διαφορετικές μεταξύ τους. Η εκτίμηση του αριθμού των μελετών που θα πρέπει να αφαιρεθούν από την μία πλευρά του διαγράμματος κώνου δίνεται από τον τύπο $L_0 = \frac{4T_k - k(k+1)}{2k-1}$, όπου k ο συνολικός αριθμός των μελετών και T_k το άθροισμα των τάξεων των παραπάνω διαφορών που έχουν πρόσημο «+». Απομακρύνονται λοιπόν οι L_0 μελέτες με το πιο μεγάλο μέγεθος επίδρασης και η διαδικασία επαναλαμβάνεται έως ότου επιτευχθεί συμμετρία στο διάγραμμα κώνου (Gleser & Olkin, 1996).

Αξιολόγηση της μεθόδου

Η μέθοδος κατασκευάστηκε πάνω στην ισχυρή υπόθεση της συμμετρίας του διαγράμματος κώνου, χωρίς όμως να απαιτεί υποθέσεις που σχετίζονται με το μηχανισμό του συστηματικό σφάλματος δημοσίευσης (Schwarzer et al., 2009). Ακόμα, οι εκτιμητές και τα συμπεράσματα πολλές φορές αποκλίνουν, γιατί βασίζονται σε δεδομένα τα οποία έχουν υποστεί παρέμβαση (Schwarzer et al., 2009). Άλλωστε η μέθοδος κατασκευάζει δεδομένα που δεν υπάρχουν (make up data). Τέλος, κρίνεται ανεπαρκής αν υπάρχει σημαντική ετερογένεια μεταξύ των μελετών (Schwarzer et al., 2009).

5.5.3 Μέθοδοι επιλογής μελετών προς δημοσίευση (Selection modeling)

Μία εξέχουσα κατηγορία από τεχνικές εντοπισμού συστηματικού σφάλματος δημοσίευσης αποτελούν οι μέθοδοι που στηρίζονται σε μοντέλα επιλογής μελετών προς δημοσίευση (selection model). Η βασική ιδέα των τεχνικών αυτών είναι ο συνδυασμός ενός μοντέλου μεγεθών επίδρασης (effect size model), το οποίο εμπεριέχει πληροφορίες για τα αποτελέσματα της μετα-ανάλυσης, με ένα μοντέλο επιλογής (selection model), που περιγράφει τον μηχανισμό καθορισμού των μελετών που τίθενται προς δημοσίευση. Πιο αναλυτικά, το μοντέλο μεγεθών επίδρασης περιγράφει την κατανομή των μεγεθών επίδρασης χωρίς την ύπαρξη συστηματικού σφάλματος δημοσίευσης, ενώ το μοντέλο επιλογής καθορίζει την πιθανότητα δημοσίευσης της κάθε μελέτης. Στην πράξη τα μοντέλα επιλογής περιέχουν ένα μεγάλο αριθμό άγνωστων παραμέτρων, που καθιστά περίπλοκη την εφαρμογή τους. Ωστόσο, για να αντιμετωπιστεί η αδυναμία αυτή εκτιμούμε τις παραμέτρους από άλλα σύνολα δεδομένων μετα-ανάλυσης ή εναλλακτικά εισάγουμε μεταβλητές και διεξάγουμε ανάλυση ευαισθησίας (Vevea & Woods, 2005).

5.5.3.1 Copas selection model

Η πιο διαδεδομένη μέθοδος εντοπισμού συστηματικού σφάλματος δημοσίευσης είναι αυτή που στηρίζεται στο μοντέλο επιλογής μελετών που προτάθηκε από τους Copa και Shi (Copas, 1999, Copas & Shi, 2000). Η μέθοδος συνδυάζει το μοντέλο τυχαίων επιδράσεων με το μοντέλο επιλογής του Copas (Copas selection model), που δίνει την πιθανότητα δημοσίευσης της *i*-οστής μελέτης. Η ύπαρξη συστηματικού σφάλματος δημοσίευσης συνδέεται με την παράμετρο συσχέτισης ρ των δύο μοντέλων. Όσο πιο ισχυρή είναι η συσχέτιση τόσο περισσότερο σχετίζεται το μέγεθος επίδρασης με την πιθανότητα δημοσίευσης με αποτέλεσμα να είναι πιο πιθανός ο εντοπισμός του συστηματικό σφάλμα δημοσίευσης.

Πιο αναλυτικά για κάθε i-οστή μελέτη της μετα-ανάλυσης υποθέτουμε ότι το πραγματικό μέγεθος επίδρασης θ_i ακολουθεί κανονική κατανομή με μέση τιμή θ και διακύμανση τ^2 που οφείλεται στην ετερογένεια (between-studies variance). Συμβολίζουμε ακόμα με $\hat{y}_i$ την εκτίμηση του μεγέθους επίδρασης θ_i και με ${\sigma_i}^2$ την διακύμανση του $\hat{y}_i$ που οφείλεται στην δειγματοληψία της κάθε μελέτης (within-study variance). Το ${\sigma_i}^2$ θεωρείται γνωστό για κάθε μελέτη και περίπου ίσο με ${s_i}^2$. Δηλαδή ισχύει: ${\sigma_i}^2 \approx {s_i}^2$.

Συνδυάζοντας τα παραπάνω προκύπτει ότι:

$$\hat{y}_i = \theta_i + (\tau^2 + {\sigma_i}^2)^{1/2}\varepsilon_i,\ \ \varepsilon_i \sim N(0,1),\ \ i = 1, \dots, k.$$

Η επιλογή των μελετών προς δημοσίευση πραγματοποιείται από το μοντέλο

$$Z_i = \gamma_0 + \frac{\gamma_1}{s_i} + \delta_i,\ \ \delta_i \sim N(0,1),\ \ i = 1, \dots, k.$$

Ισοδύναμα ακολουθώντας τον συμβολισμό των Copas & Shi έχουμε:

$$Z_i = u_i + \delta_i,\quad \delta_i \sim N(0,1), \text{ όπου } u_i = \gamma_0 + \frac{\gamma_1}{s_i}$$

Οι παράμετροι γ_0, γ_1 δεν εκτιμώνται. Είναι σταθερές που περιγράφουν την έκταση της «επιλογής» του μοντέλου. Συγκεκριμένα, αν υποθέσουμε ότι η μελέτη με το μικρότερο τυπικό σφάλμα s_{small} δημοσιεύεται με πιθανότητα p_{small} και αντίστοιχα η μελέτη με το μεγαλύτερο τυπικό σφάλμα s_{large} δημοσιεύεται με πιθανότητα p_{large} οι παράμετροι γ_0, γ_1 υπολογίζονται από τις σχέσεις:

$$p_{small} = \Phi\left(\gamma_0 + \frac{\gamma_1}{s_{small}}\right) \quad \text{και} \quad p_{large} = \Phi\left(\gamma_0 + \frac{\gamma_1}{s_{large}}\right), \quad \text{όπου} \quad \Phi(\cdot) \text{ η}$$

αθροιστική συνάρτηση της τυπικής κανονικής κατανομής.

Όσον αφορά την επιλογή των μελετών προς δημοσίευση υποθέτουμε ότι η μελέτη i θα δημοσιεύεται μόνο όταν η μεταβλητή $Z_i > 0$. Με βάση τα παραπάνω, τα $(\varepsilon_i, \delta_i)$ είναι τυχαία σφάλματα που προκύπτουν από τα δύο μοντέλα και ακολουθούν διδιάστατη κανονική κατανομή με μέση τιμή 0 και πίνακα διακυμάνσεων συνδιακυμάνσεων $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Συνεπώς, η παράμετρος συσχέτισης των δύο μοντέλων ρ ισούται με τη συνδιακύμανση των $(\varepsilon_i, \delta_i)$, δηλαδή $\rho = cov(\varepsilon_i\ , \delta_i)$.

Συνδυάζοντας τις πληροφορίες που παρέχουν τα παραπάνω μοντέλα, η πιθανότητα δημοσίευσης της κάθε μελέτης υπολογίζεται:

$$\Pr(Z_i > 0) = \Pr\left(\delta_i > -\gamma_0 - \frac{\gamma_1}{s_i}\right) = \Phi\left(\gamma_0 + \frac{\gamma_1}{s_i}\right)$$

όπου $\Phi(\cdot)$ η αθροιστική συνάρτηση της τυπικής κανονικής κατανομής.

Αν $p = 0$, τα $Z_i, \hat{y}_i$ είναι ανεξάρτητα και η μετα-ανάλυση των παρατηρούμενων μεγεθών δίνει μία εκτίμηση του πραγματικού μεγέθους επίδρασης χωρίς σφάλμα δημοσίευσης. Αντίθετα, αν $p > 0$ οι επιλεγόμενες από το μοντέλο μελέτες θα έχουν $Z_i > 0$ με αποτέλεσμα να είναι πιο πιθανόν τα δ_i να παρουσιάζουν μεγάλές τιμές. Στην περίπτωση αυτή και τα ε_i αναμένεται να λαμβάνουν μεγάλες τιμές, αφού συσχετίζονται θετικά με τα δ_i. Από τα παραπάνω προκύπτει, ότι η εκτίμηση του μεγέθους επίδρασης $\hat{y}_i$ υπερεκτιμά την πραγματική τιμή θ. Συνεπώς, όσο πιο ισχυρή είναι η συσχέτιση των δύο μοντέλων τόσο πιο πιθανή είναι η ύπαρξη σφάλματος δημοσίευσης (Copas & Shi, 2000).

ΚΕΦΑΛΑΙΟ 6
ΠΡΟΣΟΜΟΙΩΣΗ

6.1 Σκοπός της έρευνας προσομοίωσης

Η αξιοπιστία των αποτελεσμάτων μιας μετα-ανάλυσης συνδέεται άμεσα με την δυνατότητα εντοπισμού του συστηματικού σφάλματος δημοσίευσης και της πιθανής επίδρασης των μικρών σε μέγεθος μελετών. Μια πρώτη ένδειξη για τον εντοπισμό τους προκύπτει από την ασυμμετρία του διαγράμματος κώνου. Όπως αναφέρθηκε και στο Κεφάλαιο 5, για να αποφευχθεί η υποκειμενική ερμηνεία από την πλευρά των ερευνητών για την ύπαρξη ή μη ασυμμετρίας στο διάγραμμα κώνου, έχει αναπτυχθεί διάφορα στατιστικά test. Κάτω από το πλαίσιο αυτό κρίνεται αναγκαία η μεταξύ τους σύγκριση με στόχο την ανάδειξη των πιο αποτελεσματικών. Στην παρούσα μελέτη προσομοίωσης πραγματοποιείται σύγκριση της αποτελεσματικότητας μεταξύ των Egger's, Begg's rank correlation και Thompson & Sharp's tests.

6.2 Σενάρια προσομοίωσης

Η προσομοίωση πραγματοποιήθηκε στο στατιστικό πακέτο R (R Development Core Team, 2008) http://www.r-project.org/ (r-project.org), το οποίο αποτελεί μια γλώσσα προγραμματισμού με κύριο στόχο την επίλυση στατιστικών προβλημάτων. Η διεξαγωγή της περιορίστηκε σε συνεχή δεδομένα. Έχοντας ως στόχο την λεπτομερή σύγκριση της αποτελεσματικότητάς τους στην προσομοίωση, συμπεριελήφθησαν διάφορες τιμές για τον αριθμό των μελετών $k = 10, 20, 50$. Οι τιμές αυτές θεωρούνται οι πιο συνήθεις και αντιστοιχούν σε μικρό, μεσαίο και μεγάλο αριθμό μελετών. Ακόμα, για τον συντελεστή που αντιπροσωπεύει τη σχέση μεταξύ του τυπικού σφάλματος και του μεγέθους επίδρασης θεωρήσαμε τις τιμές $b = 0, 0.3, 0.5$, ενώ για την διακύμανση της ετερογένειας $\tau^2 = 0, 0.1, 0.3$. Όσον αφορά το συνολικό μέγεθος επίδρασης

θεωρήσαμε τις τιμές $m = 0, 0.3, 0.5$, οπότε συνολικά έχουμε 81 διαφορετικά σενάρια. Το κάθε σενάριο πραγματοποιήθηκε για B= 1000 τον αριθμό επαναλήψεις.

6.3 Βήματα προσομοίωσης

6.3.1 Δημιουργία δεδομένων

Το πρώτο βήμα για την διεξαγωγή μιας προσομοίωσης αποτελεί η δημιουργία δεδομένων. Τα δεδομένα που δημιουργούνται θα πρέπει να αναπαριστούν με όσο το δυνατό πιο ρεαλιστικό τρόπο τα αποτελέσματα μιας πραγματικής έρευνας. Στο πλαίσιο της παρούσης προσομοίωσης ακολουθήθηκε για την κατασκευή των δεδομένων το μοντέλο:

$\theta_i \sim N(m, \tau^2)$

$y_i \sim N(\theta_i + b * s_i\ , {s_i}^2),\ i = 1, \dots, k$ (Moreno et al., 2009)

Με θ_i αναπαρίσταται το πραγματικό μέγεθος επίδρασης της κάθε μελέτης, με m το συγκεντρωτικό μέγεθος επίδρασης και με τ^2 η διακύμανση της ετερογένειας. Συμβολίζεται ακόμα με y_i το μέγεθος επίδρασης της κάθε μελέτης, με s_i το τυπικό σφάλμα της i μελέτης και με b ο συντελεστής που αντιπροσωπεύει την μεταξύ τους σχέση. Ο αριθμός των μελετών που συμπεριλαμβάνονται στην μετα-ανάλυση είναι ίσος με k.

Αρχικά, προσομοιώσαμε το τυπικό σφάλμα ως τη τετραγωνική ρίζα μιας X^2 κατανομής με k βαθμούς ελευθερίας (Moreno et al., 2009). Στην συνέχεια, κατασκευάσαμε το πραγματικό μέγεθος επίδρασης θ_i για τις k μελέτες από την κανονική κατανομή με μέση τιμή το m και διακύμανση τ^2. Τέλος, δημιουργήσαμε τα μεγέθη επίδρασης y_i για κάθε μια από τις k μελέτες από την κανονική κατανομή με μέση τιμή $\theta_i + b * s_i$.

6.3.2 Μετρήσεις

Όπως έχει αναφερθεί και στο Κεφάλαιο 5, τα Egger's, Begg's rank correlation και Thompson & Sharp's test ελέγχουν την μηδενική υπόθεση H_0: *το διάγραμμα κώνου είναι συμμετρικό* έναντι της εναλλακτικής H_a: *το διάγραμμα κώνου είναι ασύμμετρο*. Πραγματοποιούνται B= 1000 επαναλήψεις. Σε κάθε μια από αυτές παράγονται και αποθηκεύονται οι αντίστοιχες p-τιμές τους. Μετά το πέρας των $B = 1000$ επαναλήψεων υπολογίζεται η πιθανότητα απόρριψης της μηδενικής υπόθεσης για κάθε ένα από τα παραπάνω test. Συνεπώς, από κάθε σενάριο λαμβάνουμε ως αποτέλεσμα την πιθανότητα απόρριψης της μηδενικής υπόθεσης για συγκεκριμένη τιμή του αριθμού των μελετών k, του συντελεστή b που αντιπροσωπεύει την σχέση μεταξύ του τυπικού σφάλματος και του μεγέθους επίδρασης, της διακύμανσης της ετερογένειας τ^2 και του συνολικού μεγέθους επίδρασης m. Στο σημείο αυτό είναι ανάγκη να διευκρινιστεί ότι στα σενάρια όπου $b = 0$ η πιθανότητα απόρριψης της μηδενικής υπόθεσης ταυτίζεται με το σφάλμα τύπου I, δηλαδή με την πιθανότητα απόρριψης της μηδενικής υπόθεσης όταν παράγουμε δεδομένα από αυτή. Αντίστοιχα, στα σενάρια όπου $b \neq 0$ η πιθανότητα απόρριψης της μηδενικής υπόθεσης ταυτίζεται με την ισχύ, δηλαδή την πιθανότητα να απορρίπτουμε την μηδενική υπόθεση όταν παράγουμε δεδομένα από την εναλλακτική.

6.3.3 Αποτελέσματα προσομοίωσης

Στην συνέχεια παρατίθενται κατάλληλοι πίνακες με στόχο την εξαγωγή συμπερασμάτων για την αποτελεσματικότητα των Egger's test, Begg's rank correlation test και Thompson & Sharp's tests για τα διάφορα σενάρια**.**

Αριθμός μελετών $k = 10$					
Συντελεστής b	**Συνολικό μέγεθος επίδρασης m**	**Ετερογένεια τ^2**	**Egger's method linreg**	**Begg's Rank method rank**	**Thompson & Sharp's method mm**
$b = 0$	**0**	**0**	0.037	0.024	0.014
		0.1	0.049	0.027	0.018
		0.3	0.066	0.008	0.055
	0.3	**0**	0.054	0.015	0.018
		0.1	0.052	0.023	0.02
		0.3	0.058	0.027	0.029
	0.5	**0**	0.038	0.031	0.012
		0.1	0.054	0.035	0.011
		0.3	0.078	0.033	0.023
	0	**0**	0.1	0.033	0.046
		0.1	0.07	0.042	0.034
		0.3	0.113	0.016	0.063

$\boldsymbol{b = 0.3}$	**0.3**	**0**	0.098	0.037	0.053
		0.1	0.081	0.015	0.058
		0.3	0.049	0.006	0.063
	0.5	**0**	0.094	0.029	0.048
		0.1	0.116	0.011	0.082
		0.3	0.052	0.025	0.043
$\boldsymbol{b = 0.5}$	**0**	**0**	0.177	0.069	0.103
		0.1	0.137	0.057	0.085
		0.3	0.202	0.027	0.118
	0.3	**0**	0.117	0.054	0.048
		0.1	0.194	0.036	0.12
		0.3	0.331	0.019	0.154
	0.5	**0**	0.276	0.046	0.185
		0.1	0.115	0.053	0.045
		0.3	0.194	0.028	0.127

Πίνακας 6.1: Δεδομένα προσομοίωσης για k=10 μελέτες

Αριθμός μελετών $k = 20$					
Συντελεστής b	**Συνολικό μέγεθος επίδρασης m**	**Ετερογένεια τ^2**	**Egger's method linreg**	**Begg's Rank method rank**	**Thompson & Sharp's method mm**
$b = 0$	**0**	**0**	0.047	0.023	0.023
		0.1	0.108	0.02	0.053
		0.3	0.255	0.043	0.085
	0.3	**0**	0.044	0.016	0.03
		0.1	0.105	0.025	0.047
		0.3	0.06	0.129	0.043
	0.5	**0**	0.051	0.023	0.028
		0.1	0.057	0.017	0.036
		0.3	0.142	0.033	0.057
	0	**0**	0.158	0.044	0.132
		0.1	0.08	0.06	0.056
		0.3	0.164	0.022	0.117

$\boldsymbol{b = 0.3}$	**0.3**	**0**	0.091	0.042	0.068
		0.1	0.292	0.035	0.195
		0.3	0.248	0.02	0.187
	0.5	**0**	0.113	0.055	0.077
		0.1	0.145	0.047	0.119
		0.3	0.172	0.026	0.126
$\boldsymbol{b = 0.5}$	**0**	**0**	0.318	0.076	0.266
		0.1	0.472	0.049	0.389
		0.3	0.227	0.031	0.223
	0.3	**0**	0.357	0.088	0.311
		0.1	0.398	0.057	0.321
		0.3	0.223	0.154	0.269
	0.5	**0**	0.347	0.087	0.289
		0.1	0.322	0.051	0.294
		0.3	0.234	0.059	0.284

Πίνακας 6.2: Δεδομένα προσομοίωσης για k=20 μελέτες

Αριθμός μελετών $k = 50$					
Συντελεστής b	**Συνολικό μέγεθος επίδρασης m**	**Ετερογένεια τ^2**	**r**	**Begg's Rank method rank**	**Thompson & Sharp's method mm**
$b = 0$	**0**	**0**	0.058	0.022	0.044
		0.1	0.136	0.06	0.062
		0.3	0.085	0.041	0.05
	0.3	**0**	0.054	0.01	0.035
		0.1	0.143	0.086	0.057
		0.3	0.595	0.391	0.047
	0.5	**0**	0.042	0.026	0.025
		0.1	0.065	0.108	0.065
		0.3	0.301	0.174	0.04
	0	**0**	0.466	0.048	0.423
		0.1	0.301	0.057	0.346
		0.3	0.316	0.219	0.241

$\boldsymbol{b = 0.3}$	**0.3**	**0**	0.481	0.044	0.438
		0.1	0.178	0.164	0.398
		0.3	0.19	0.141	0.252
	0.5	**0**	0.402	0.045	0.326
		0.1	0.479	0.107	0.388
		0.3	0.34	0.269	0.25
$\boldsymbol{b = 0.5}$	**0**	**0**	0.831	0.129	0.805
		0.1	0.75	0.084	0.689
		0.3	0.601	0.316	0.581
	0.3	**0**	0.901	0.069	0.884
		0.1	0.715	0.112	0.655
		0.3	0.187	0.39	0.587
	0.5	**0**	0.927	0.058	0.904
		0.1	0.734	0.063	0.794
		0.3	0.298	0.303	0.563

Πίνακας 6.3: Δεδομένα προσομοίωσης για k=50 μελέτες

6.3.4 Συμπεράσματα

Η εξαγωγή συμπερασμάτων πραγματοποιήθηκε έπειτα από τη σύγκριση της πιθανότητας απόρριψης της μηδενικής υπόθεσης για τα διάφορα σενάρια της προσομοίωσης. Συγκεκριμένα, όταν ο αριθμός των μελετών που λαμβάνουν μέρος στην μετα-ανάλυση είναι μικρός ($k = 10$) και ο συντελεστής που δηλώνει τη σχέση μεταξύ τυπικού σφάλματος και μεγέθους επίδρασης είναι $b = 0$, παρατηρούμε ότι το Begg's rank correlation test απορρίπτει την μηδενική υπόθεση λιγότερο συχνά από ότι θα έπρεπε για επίπεδο σημαντικότητας $\alpha = 5\%$. Παρουσιάζει δηλαδή χαμηλό σφάλμα τύπου Ι. Σε αντίστοιχο συμπέρασμα καταλήγουμε για το Thompson & Sharp's test, όταν ο αριθμός των μελετών είναι $k = 10$ και $b = 0$. Όσο πιο μικρές είναι οι τιμές της ετερογένειας, τόσο πιο σπάνια απορρίπτεται η μηδενική υπόθεση. Το test του Egger απορρίπτει σε ικανοποιητικά επίπεδα τη μηδενική υπόθεση για $k = 10$ και $b = 0$. Στο σημείο αυτό είναι ανάγκη να διευκρινιστεί ότι σε κανένα από τα παραπάνω test η ισχύς δεν είναι υψηλή για μικρό αριθμό μελετών.

Παρόμοια εικόνα έχουμε και για $k = 20$ μελέτες. Τα Begg's rank correlation και Thompson & Sharp's tests δεν απορρίπτουν τόσο συχνά όσο θα έπρεπε. Το test του Egger απορρίπτει ικανοποιητικά όταν δεν υπάρχει ετερογένεια. Στις περιπτώσεις όμως, όπου υπάρχει ετερογένεια απορρίπτει πολύ συχνότερα από το αντίστοιχο επίπεδο σημαντικότητας του 5%. Όσον αφορά την ισχύ, είναι σαφώς υψηλότερη από ότι στα αντίστοιχα σενάρια με μικρό αριθμό μελετών. Ωστόσο, δεν μπορεί να θεωρηθεί επαρκής.

Όταν ο αριθμός των μελετών είναι $k = 50,$ οι τιμές της ετερογένειας είναι μεγάλες και το πραγματικό μέγεθος επίδρασης $m \neq 0$ το test του Egger απορρίπτει συχνά την μηδενική υπόθεση για επίπεδο σημαντικότητας $\alpha = 5\%$. Παρόμοια συμπεριφορά έχει το Begg's rank correlation test. Συγκριτικά με τα προηγούμενα, στα αντίστοιχα σενάρια το Thomson & Sharp's test δεν

παρουσιάζει μεγάλο σφάλμα τύπου Ι. Ακόμα, όταν $b = 0.3$ η ισχύς δεν είναι ιδιαίτερα υψηλή σε κανένα από τα test. Ειδικά στα σενάρια όπου η ετερογένεια είναι ίση με το 0.3. Τέλος, για μεγάλο αριθμό μελετών $k = 50$ και $b = 0.5$ η ισχύς των Egger και Thomson & Sharp's tests είναι υψηλή. Σε αντίθεση το Begg's rank correlation test δεν παρουσιάζει ιδιαίτερα υψηλή ισχύ στα αντίστοιχα σενάρια.

6.4 Γενικό Συμπέρασμα

Σκοπός της παρούσης προσομοίωσης αποτέλεσε η σύγκριση της πιθανότητας απόρριψης της μηδενικής υπόθεσης στα Egger's, Begg's rank correlation και Thompson & Sharp's test για τα διάφορα σενάρια. Έπειτα από την πραγματοποίηση της μελέτης εξαγάγαμε τα παρακάτω συμπεράσματα. Κανένα από τα τρία test δεν κρίνεται ικανοποιητικό ως προς το σφάλματα τύπου Ι. Συγκριτικά με τα υπόλοιπα το test των Thomson & Sharp φαίνεται να είναι καλύτερο όσον αφορά το σφάλμα τύπου Ι για μεγάλο αριθμό μελετών. Η ισχύς δεν θεωρείται επαρκής σε κανένα από τα test, όταν ο αριθμός των μελετών $k \leq 20$. Στις περιπτώσεις όπου ο αριθμός των μελετών είναι μεγάλος, η ισχύς των Egger και Thomson & Sharp's tests χαρακτηρίζεται υψηλή σε αντίθεση με το Begg's rank correlation test. Το αποτέλεσμα ήταν σε κάποιο βαθμό αναμενόμενο μιας και είναι δύσκολο να προσαρμοστεί ένα μοντέλο παλινδρόμησης σε τόσο μικρό αριθμό μελετών. Στην πράξη, οι έλεγχοι αυτοί παρερμηνεύονται μιας και οι περισσότερες συστηματικές ανασκοπήσεις έχουν μικρό αριθμό μελετών και οι αναλυτές στηρίζονται πολύ στη p-τιμή του ελέγχου.

ΠΑΡΑΡΤΗΜΑ

Κώδικας Μελέτης Προσομοίωσης

Στη συνέχεια, περιγράφεται αναλυτικά ο κώδικας που χρησιμοποιήθηκε στο στατιστικό πακέτο της R για τη διεξαγωγή της μελέτη προσομοίωσης στο Κεφάλαιο 6. Ενδεικτικά, παραθέτουμε τον κώδικα για την πραγματοποίηση του σεναρίου $k = 10,\ b = 0.5,\ \tau^2 = 0.1,\ m = 0.3$. Η προετοιμασία απαιτεί την εγκατάσταση των παρακάτω βιβλιοθηκών στο πακέτο λογισμικού της R:

install.packages("meta")

install.packages("metafor")

Κώδικας στην R

```
# Βιβλιοθήκες

library("meta")

library("metafor")

B=1000 # Αριθμός επαναλήψεων

k=10   # Αριθμός μελετών

b=0.5      # Συντελεστής που αντιπροσωπεύει τη σχέση μεταξύ μεγέθους επίδρασης και τυπικού σφάλματος

t=0.1      # Ετερογένεια μεταξύ των μελετών

m=0.3    # Συνολικό μέγεθος επίδρασης

s=sqrt(rchisq(k,1)) # Προσομοίωση τυπικού σφάλματος
```

```
pv1=c()

pv2=c()

pv3=c()

for(i in 1:B)

{

 theta= rnorm(k,m,t)  # Δημιουργία του πραγματικού μεγέθους επίδρασης της κάθε μελέτης

y= rnorm(k, theta+b*s,s) # Δημιουργία του εκτιμώμενου μεγέθους επίδρασης της κάθε μελέτης

meta1=metagen(y,s) # Μετα-ανάλυση

test.SSE= metabias(meta1,method.bias="linreg")  # Egger's test

pv1[i]=test.SSE$p.value # Διάνυσμα για τις p-τιμές του Egger's test

test.SSE2=metabias(meta1,method.bias="rank") # Begg’s rank correlation test

 pv2[i]=test.SSE2$p.value # Διάνυσμα για τις p-τιμές του Begg’s rank correlation test

test.SSE3=metabias(meta1,method.bias="mm") # Thompson & Sharp’s test

pv3[i]=test.SSE3$p.value # Διάνυσμα για τις p-τιμές του Thompson & Sharp’s test

}

length(which(pv1<0.05))/B # Πιθανότητα απόρριψης της μηδενικής υπόθεσης για το Egger's test
```

length(which(pv2<0.05))/B # Πιθανότητα απόρριψης της μηδενικής υπόθεσης για το Begg’s rank correlation test

length(which(pv3<0.05))/B # Πιθανότητα απόρριψης της μηδενικής υπόθεσης για το Thompson & Sharp’s test

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια η μέθοδος της μετα-ανάλυσης καθιερώνεται ολοένα και περισσότερο στον χώρο της ιατρικής. Βρίσκεται στην κορυφή της αποδεικτικής πυραμίδας και θεωρείται από πολλούς οργανισμούς, όπως ο Παγκόσμιος Οργανισμός Υγείας (Π.Ο.Υ), η πλέον ενδεδειγμένη μέθοδος για την βελτίωση της κλινικής πρακτικής. Ωστόσο, η μέθοδος εμπεριέχει ορισμένες μεθοδολογικές αδυναμίες. Στο πλαίσιο της παρούσης μεταπτυχιακής διατριβής θα εστιάσουμε στην επίδραση του συστηματικού σφάλματος δημοσίευσης (publication bias) καθώς και στην πιθανή επιρροή των μικρών σε μέγεθος μελετών στα αποτελέσματά της (small-study effect).

Αρχικά στο Κεφάλαιο 1 περιγράφονται οι μέθοδοι της αποδεικτικής επιστήμης (evidence-based science). Συγκεκριμένα, αναφέρονται οι διάφοροι τύποι έρευνας με τα χαρακτηριστικά και τις διαφορές τους. Προσδιορίζεται ακόμα η κεντρική ιδέα κάθε κεφαλαίου της μεταπτυχιακής διατριβής καθώς και ο σκοπός εκπόνησης της. Στο Κεφάλαιο 2 γίνεται αναφορά στην μεθοδολογία της συστηματικής ανασκόπησης (systematic review). Στο πλαίσιο αυτό περιγράφονται αναλυτικά οι βασικές αρχές και τα στάδια που διέπουν την υλοποίηση της. Στο Κεφάλαιο 3 τίθεται η ανάγκη ποσοτικοποίησης και σύγκρισης της αποτελεσματικότητας των παρεμβάσεων των μελετών, που πληρούν τα κριτήρια για τη διεξαγωγή μιας συστηματικής ανασκόπησης. Παρατίθενται λοιπόν οι ορισμοί των μέτρων σχέσης/μεγεθών επίδρασης (effect sizes) για συνεχή και διχότομα δεδομένα.

Αντικείμενο του Κεφαλαίου 4 είναι η περιγραφή της στατιστικής μεθόδου της μετα-ανάλυσης (meta-analysis), η οποία συνθέτει ποσοτικά τα αποτελέσματα των μελετών της συστηματικής ανασκόπησης, για να καταλήξει σε ένα συγκεντρωτικό αποτέλεσμα. Γίνεται αναφορά στα δύο μοντέλα μετα-ανάλυσης, το μοντέλο σταθερών επιδράσεων (fixed effect model) και το μοντέλο τυχαίων

επιδράσεων (random effect model). Παρατίθεται ακόμα ο ορισμός της ετερογένειας (heterogeneity), δηλαδή της διακύμανσης των πραγματικών μεγεθών επίδρασης κάθε μελέτης καθώς και οι τρόποι εντοπισμού της. Για την καλύτερη ερμηνεία των αποτελεσμάτων και τη διερεύνηση της ετερογένειας παρουσιάζεται η μέθοδος της μετα-παλινδρόμησης (meta-regression).

Στο Κεφάλαιο 5 αναλύεται η έννοια του συστηματικού σφάλματος δημοσίευσης (publication bias) και η πιθανή επίδραση των μικρών σε μέγεθος μελετών στα αποτελέσματα της μετα-ανάλυσης (small-study effect). Περιγράφεται αναλυτικά το διάγραμμα κώνου (funnel plot) και η πιο εξελιγμένη έκδοσή του, γνωστή ως contour enhanced funnel plot. Στη συνέχεια, επικεντρωνόμαστε στην ερμηνεία της πιθανής ασυμμετρίας του διαγράμματος κώνου. Για το λόγο αυτό γίνεται αναφορά στα διάφορα στατιστικά test και κυρίως στα test γραμμικής παλινδρόμησης, που έχουν αναπτυχθεί για τον έλεγχο της ασυμμετρίας. Τέλος, για την εξακρίβωση του συστηματικού σφάλματος δημοσίευσης γίνεται περιγραφή των μεθόδων Fail-safe N, Trim and Fill καθώς και της στατιστικής μεθόδου, που στηρίζεται στο μοντέλο επιλογής μελετών προς δημοσίευση, που προτάθηκε από τον Copas (Copas, 1999, Copas & Shi, 2000) (Copas selection model).

Όπως έχει αναφερθεί, έχει αναπτυχθεί ένας μεγάλος αριθμός από στατιστικά test για τον έλεγχο της ασυμμετρίας του διαγράμματος κώνου μεταξύ των οποίων τα πιο γνωστά είναι τα Egger’s, Begg’s rank correlation και Thompson & Sharp’s tests. Τίθεται η ανάγκη σύγκρισης της αποτελεσματικότητάς τους. Κάτω από το πλαίσιο αυτό πραγματοποιείται μια μελέτη προσομοίωσης στο Κεφάλαιο 6, που έχει ως στόχο την ανάδειξη των πιο αποτελεσματικών μεθόδων για την ανίχνευση της επίδρασης μικρών μελετών στο συγκεντρωτικό αποτέλεσμα. Η διατριβή ολοκληρώνεται με το παράρτημα, στο οποίο παρατίθενται ο κώδικας που δημιουργήθηκε για την πραγματοποίηση της μελέτης προσομοίωσης στο Κεφάλαιο 6.

ABSTRACT

Meta-analysis has been an established evidence synthesis method in the field of medicine and is increasingly employed for assessing the effectiveness of interventions. It ranks at the top of the evidence-based hierarchy and is considered by many, including the World Health Organization (WHO) to be the most appropriate method for improving clinical practice. However, there are some methodological weaknesses, which may compromise the results form meta-analysis. In the context of this postgraduate dissertation, we will focus on the effect of publication bias and on the possible influence of small size studies on estimating a meta-analytic treatment effect (small-study effect).

Initially, Chapter 1 describes the study designs of evidence-based science. More specifically, the different types of research with their characteristics and differences are described. Additionally, it states the purpose of postgraduate thesis and the main contents of the chapters. Chapter 2 refers to the systematic review methodology, describing its basic principles and stages its implementation in detail. In Chapter 3 arises the need to quantify and compare the effectiveness of the interventions satisfying the eligibility criteria in systematic review. We describe effect sizes for dichotomous and continuous outcomes. Chapter 4 describes the statistical method of meta-analysis, which synthesizes quantitatively the effect sizes from the included studies to get a pooled result. We present the two most popular meta-analysis models, the fixed effect model and the random effect model. We also define the heterogeneity, the variance of the underlying true effects, and present methods to identify it. We also present meta-regression models that are commonly applied to explore heterogeneity.

In Chapter 5, we present the problem of publication bias and the possible effect of small size studies on the results of meta-analysis (small-study effect). We

describe the funnel plot and its more advanced version, contour enhanced funnel plot. We then concentrate on interpreting the possible asymmetry of the funnel plot. For this reason, reference is made to the various statistical tests and especially to the linear regression tests developed for asymmetry control. Finally, for the identification of the publication bias the Fail-safe N, Trim and Fill methods are described as well as a statistical method based on the Copas selection model. As reported, a large number of statistical tests have been developed to control the asymmetry of the funnel plot. The most prominent ones are Egger's, Begg's rank correlation and Thompson & Sharp's tests. We carry out a simulation study to evaluate these tests for detecting small-study effects and publication bias in Chapter 6. The dissertation ends with an appendix describing the R code used to carry out the simulation study.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Η βιβλιογραφία δίνεται με την σειρά εμφάνισης μέσα στο κείμενο.

1. Παρασκευόπουλος, Ι. (1993). Μεθοδολογία επιστημονικής έρευνας. Αθήνα: Εκδόσεις Γρηγόρη.
2. Isenburg, M. (2015). LibGuides: Introduction to Evidence-Based Practice: Types of studies.
3. Higgins, J., Green, S. (2008). Cochrane Handbook for Systematic Reviews of Interventions. Wiley Blackwell.
4. Γαλάνης, Π. (2009). Συστηματική Ανασκόπηση Και Μετα-Ανάλυση. Αρχεία Ελληνικής Ιατρικής, 26, 826-841.
5. Antman, E.M., Lau, J., Kupelnick, B., Mosteller, F., Chalmers, T.C. (1992). A comparison of results of meta- analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA, 268, 240-248.
6. Καράσσα, Φ. (2006). Αρχές και μεθοδολογία της συστηματικής ανασκόπησης της βιβλιογραφίας. Ελληνική Ρευματολογία, 17, 289-297.
7. Μπελλάλη, Θ. (2011). Βασικές Αρχές και Μεθοδολογία της Συστηματικής Ανασκόπησης Ποσοτικών Μελετών. Νοσηλευτική, 50,10-22.
8. Πατελάρου, Ε., Μπροκαλάκη, Η. (2010). Μεθοδολογία της Συστηματικής Ανασκόπησης και Μετα-ανάλυσης. Νοσηλευτική, 49, 122-130.
9. Borenstein, M., Hedges, L., Higgins, J., Rothstein H. (2009). Introduction to Meta-Analysis (first). Wiley Blackwell.
10. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd). Hillsdale, NJ: Lawrence Erlbaum.

11. Mabe, M.A. (2009). Scholarly publishing. European Review, 17, 3-22.
12. Dersimonian, R., Laird, N. (1986). Meta-Analysis in Clinical Trials. Statistics in Medicine, 188, 177-188.
13. Rukhin, A.L. (2013). Estimating heterogeneity variance in meta-analysis. Journal of the Royal Statistical Society, Series B: Statistical Methodology, 75, 451-469.
14. Nikolakopoulou, A., Mavridis, D., Salanti, G. (2014) Demystifying fixed and random effects meta-analysis. Evidence-Based Mental Health, 17, 53-57.
15. Higgins, J.P., Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. Statistics in Medicine, 21, 1539-1558.
16. Hardy, R.J., Thompson, S.G. (1998). Detecting and describing heterogeneity in meta-analysis. Statistics in Medicine, 17, 841-856.
17. Thompson, S.G., Higgins, J.P. (2002). How should meta-regression analyses be undertaken and interpreted? Statistics in Medicine, 21, 1559-1573.
18. Wilson, D.B. (1999). Practical meta-analysis. American Evaluation Association.
19. Light, R.J., Pillemer, D.B. (1984). Summing up: the science of reviewing research. Cambridge, MA: Harvard University Press.
20. Mavridis, D., Salanti, G. (2014). Exploring and accounting for publication bias in mental health: A brief overview of methods. Evidence Based Mental Health, 17, 11-15.
21. Sterne, J.A, Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. Journal of Clinical Epidemiology, 54, 1046-1055.

22. Sterne, J.A., Harbord, R.M. (2004). Funnel plots in meta-analysis. The Stata Journal, 4, 127-141.
23. Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. Journal of Clinical Epidemiology, 61, 991-996.
24. Begg, C.B., Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. Biometrics, 50, 1088-1101.
25. Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. JAMA, 295, 676-80.
26. Egger, M., Smith, G.D., Schneider, M., Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. BMJ, 315, 629-634.
27. Tang, J.L., Liu, J.L. (2000). Misleading funnel plot for detection of bias in meta-analysis. Journal of Clinical Epidemiology, 53, 477-484.
28. Thompson, S.G., Sharp, S.J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. Statistics in Medicine, 18, 2693-708.
29. Rücker, G., Schwarzer, G., Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. Statistics in Medicine, 27, 746-63.
30. Orwin, R.G. (1983). A fail-safe N for effect size in meta-analysis. Journal of Educational Statistics, 8, 157-159.
31. Iyengar, S., Greenhouse, J.B. (1988). Selection models and the file drawer problem. Statistical Science, 3, 109-117.

32. Duval, S.J, Tweedie, R.L. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. Journal of the American Statistical Association, 95, 89-98.
33. Duval, S.J, Tweedie, R.L. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics, 56, 455-463.
34. Gleser, L.J, Olkin, I. (1996). Models for estimating the number of unpublished studies. Statistics in Medicine, 15, 2493-2507.
35. Schwarzer, G., Carpenter, J., Rücker, G. (2010). Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. Journal of Clinical Epidemiology, 63, 282-288.
36. Vevea, J.L., Woods, C.M. (2005). Publication bias in research synthesis: sensitivity analysis using a priori weight functions. Psychological Methods, 10, 428-43.
37. Copas, J. (1999). What works? selectivity models and meta-analysis. Journal of the Royal Statistical Society Series A, 169, 95-109.
38. Copas, J., Shi, J. (1999). Reanalysis of epidemiological evidence on lung cancer and passive smoking. BMJ, 320, 417-418.
39. Moreno, S.G., Sutton, A.J., Ades, A.E., Stanley, T.D., Abrams, K.R., Peters, J.L., Cooper, N.J. (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study. BMC Medical Research Methodology, 9, 2.