

ΕΙΔΙΚΟ ΑΡΘΡΟ SPECIAL ARTICLE

Το φαινόμενο της μεροληψίας δημοσίευσης στις επιστήμες υγείας Παρουσίαση στατιστικών μεθόδων για την αντιμετώπισή του

Η μεροληψία δημοσίευσης είναι ένα φαινόμενο, το οποίο προκαλείται από την επιρροή των αποτελεσμάτων μιας έρευνας αναφορικά με τις πιθανότητες να δημοσιευτεί η εν λόγω έρευνα. Συνήθως, έρευνες με θετικά αποτελέσματα είναι πιθανότερο να δημοσιευτούν σε σχέση με αυτές που καταλήγουν σε αρνητικά συμπεράσματα, γεγονός που οδηγεί σε μια κυριαρχία των ψευδώς θετικών αποτελεσμάτων στη βιβλιογραφία. Η υποκειμενική πλευρά της μεροληψίας έγκειται στο γεγονός της δημοσιοποίησης των ευρημάτων από πλευράς του ερευνητή, ενώ η αντικειμενική πλευρά της είναι περισσότερο δυσδιάκριτη και απορρέει από το ότι η απόφαση προς δημοσίευση μιας έρευνας ή όχι εξαρτάται από τα αποτελέσματά της. Οι πρωταρχικές αιτίες της αντικειμενικής μεροληψίας είναι οι προτιμήσεις των εκδοτών και των συγγραφέων, συμπληρωματικά των διαφόρων αναλυτικών χειρισμών και άλλων αντικειμενικών επιρροών. Στην παρούσα ανασκόπηση θα παρατεθούν επί πλέον στατιστικές μέθοδοι για την ταυτοποίηση, την ποσοτικοποίηση και τον υπολογισμό της επίδρασης της μεροληψίας δημοσίευσης στις μετα-αναλύσεις. Επιπρόσθετα, θα αναφερθούν οι στρατηγικές οι οποίες, αν υιοθετηθούν, θα συμβάλουν σημαντικά στην πρόληψη της μεροληψίας, όπως για παράδειγμα η αναζήτηση και η συγκέντρωση της γκριζας βιβλιογραφίας, η εκ των προτέρων καταγραφή ερευνών κ.ά.

1. ΕΙΣΑΓΩΓΗ

Ένας κοινός λόγος για την έναρξη διεξαγωγής μιας επιστημονικής έρευνας είναι η πρόθεση των ερευνητών να (απο)δείξουν την παρουσία ή την έλλειψη μιας συσχέτισης (για παράδειγμα, μεταξύ έκθεσης και ασθένειας) ή διαφορές στις εκτιμήσεις παραμέτρων (για παράδειγμα, μέσος, τυπική απόκλιση, αναλογίες) μεταξύ πληθυσμών. Τέτοιου είδους σχέσεις ή διαφορές μπορεί να είναι πραγματικές, μπορεί να οφείλονται στην τύχη ή μπορεί να σχετίζονται με άλλους παράγοντες όπως η μεροληψία.¹ Η μεροληψία είναι μια μορφή συστηματικού σφάλματος, η οποία μπορεί να επηρεάσει την επιστημονική έρευνα και να αλλοιώσει τη διαδικασία των μετρήσεων. Μια μεροληπτική έρευνα υπολείπεται εγκυρότητας ανάλογα με το βαθμό της μεροληψίας της. Ενώ κάποιες κατηγορίες ερευνών είναι περισσότερο ευάλωτες στη μεροληψία, η παρουσία της είναι παγκόσμια. Είναι δύσκολο ή ακόμη και ακατόρθωτο να εξαλείψει κάποιος τη μεροληψία. Στη διαδικασία απαλοιφής της, νέες μορφές μεροληψίας μπορεί να εισαχθούν ή ακόμη υπάρχει

η περίπτωση οι έρευνες στην ερμηνεία τους να μην είναι γενικεύσιμες. Οι αιτίες εμφάνισής της είναι αναρίθμητες. Σχετικά ανάλογος είναι και ο αριθμός των παραγόντων που είναι εμπειρικά γνωστοί για την εισαγωγή και την υπόθαψη της μεροληψίας μέσα στη βιβλιογραφία και οι οποίοι συμβάλλουν στον κίνδυνο των ψευδώς θετικών αποτελεσμάτων. Στον τομέα της Επιδημιολογίας, η μεροληψία μπορεί να οδηγήσει σε λανθασμένες εκτιμήσεις των σχέσεων ή, απλούστερα, τα παρατηρούμενα αποτελέσματα των ερευνών θα τείνουν να είναι λανθασμένα και αρκετά διαφορετικά από τα πραγματικά. Η μεροληψία θα έπρεπε να θεωρείται πρωτίστως μια λειτουργία της διαδικασίας μελέτης (για παράδειγμα, σχεδιασμός και μέθοδοι) και όχι των αποτελεσμάτων.²

Είναι σημαντικό να διαχωριστεί η μεροληψία δημοσίευσης από άλλες μορφές μεροληψίας, για να είναι εφικτή η απομόνωση και η μέτρησή της. Προτού όμως μετρηθεί, πρέπει να προσδιοριστεί ο όρος αυτός με τρόπο μονοσήμαντο. Ο προσδιορισμός του όρου οφείλει επί πλέον να

ΑΡΧΕΙΑ ΕΛΛΗΝΙΚΗΣ ΙΑΤΡΙΚΗΣ 2013, 30(3):340-354
ARCHIVES OF HELLENIC MEDICINE 2013, 30(3):340-354

Ο. Αρραγγελίδη

Department of Epidemiology and Public Health, University College of London, London, UK

The publication bias phenomenon in health sciences: Exploration of statistical methods of correction

Abstract at the end of the article

Λέξεις ευρετηρίου

Δημοσίευση
Εγκυρότητα
Επιστήμη
Έρευνα
Μεροληψία
Ποιότητα

Υποβλήθηκε 4.11.2012
Εγκρίθηκε 28.11.2012

υπόκειται σε ποσοτική ερμηνεία. Το συγκεκριμένο στοιχείο είναι εξαιρετικής σημασίας, καθώς ο όρος μεροληψία είναι ευρέως χρησιμοποιούμενος σε διαφορετικά φάσματα της επιστήμης και όχι μόνο. Η ιατρική κυρίως βιβλιογραφία περιέχει πιθανόν ένα μεγάλο τμήμα πληροφορίας αλλά και γνωμών οι οποίες πάσχουν από μεροληψία, κυρίως στο εκδοτικό πεδίο, σε ανασκοπικές δημοσιεύσεις κ.λπ., και γι' αυτόν το λόγο ο όρος μεροληψία δημοσίευσης αντιπροσωπεύει επιτυχώς το σχετικό φαινόμενο.³

2. ΠΤΥΧΕΣ ΜΕΡΟΛΗΨΙΑΣ

Ιδιαίτερη προσοχή έδωσαν οι Begg και Berlin στο θέμα της διοχέτευσης και της παρουσίασης των αποτελεσμάτων. Σε μια έρευνά τους το 1989 διατύπωσαν τη θεωρία ότι το σφάλμα δημοσίευσης αποτελείται από δύο διακριτά και συχνά συσχετισμένα μεταξύ τους συστατικά, ένα υποκειμενικό και ένα αντικειμενικό. Το αντικειμενικό χαρακτηρίζει ένα περισσότερο δυσδιάκριτο και σοβαρό πρόβλημα, καθώς αντικατοπτρίζεται στα «αντικειμενικά» δεδομένα. Είναι αυτό ακριβώς το πρόβλημα που αναλύθηκε παραπάνω και αντιπροσωπεύει τη δημοσίευση ερευνών ανάλογα με τα αποτελέσματά τους. Η μεροληψία δημοσίευσης επηρεάζει ακόμη και άμεσα τα περιληπτικά δεδομένα, όπως η μέση επιβίωση και τα ποσοστά ανταπόκρισης. Κατά συνέπεια, η απόσπαση περιληπτικών δεδομένων από δημοσιευμένες αναφορές είναι επικίνδυνη.⁴ Οι κύριες αιτίες της αντικειμενικής μεροληψίας δημοσίευσης είναι οι προτιμήσεις συγγραφέων και εκδοτών, επί πλέον των επιλεκτικών αναλυτικών χειρισμών και άλλων υποκειμενικών επιρροών. Γενικά, υπάρχει η αντίληψη ότι πολλοί εκδότες περιοδικών δείχνουν την προτίμησή τους σε αναφορές αποτελεσμάτων ερευνών τα οποία παρουσιάζουν στατιστική σημαντικότητα και αυτό έχει διατυπωθεί και αναλυθεί εκτενώς υπό τον όρο εκδοτική πολιτική (editorial policy).⁵

Το δεύτερο συστατικό και υποκειμενικό εμπερικλείει επίσης κάποια σοβαρά φαινόμενα, όπως για παράδειγμα υπερβολικούς και ακραίους ισχυρισμούς, τα οποία πυροδοτούνται από παρουσιάσεις που έχουν τη μορφή της συνηγορίας στα αποτελέσματα και είναι κοινές στα επιστημονικά άρθρα. Αυτού του τύπου η μεροληψία, με άλλα λόγια, ενέχει τη μεροληπτική γνώμη του ερευνητή και συχνά αντανάκλαται στον τόνο της παρουσίασης, την παράθεση της βιβλιογραφίας και την ερμηνεία των στοιχείων, παρά στα αντικειμενικά δεδομένα.⁴ Το φαινόμενο αυτό δεν είναι εύκολο να μετρηθεί, αλλά έχει ερευνηθεί από τον Chalmers, ο οποίος συνδύασε και διασταύρωσε τις γνώμες που παρουσιάστηκαν όπως και την ειδικότητα του συγγραφέα σε ποικίλους τύπους ερευνών, περιλαμβανομένων των ερευνών της προφυλακτικής ακτινοθεραπείας για τον καρκίνο του μαστού, και πρόβαλε σημαντική συσχέτιση.⁶

Συγχρόνως, ο επιπολάζων ενθουσιασμός της ολοκλήρωσης και της δημοσίευσης των ερευνών με στατιστικώς σημαντικά αποτελέσματα πιθανόν ενισχύεται σε κάποιες περιπτώσεις από την άγνοια της φύσης της στατιστικής μεταβλητότητας, οδηγώντας τους ερευνητές να συνάγουν θετικά συμπεράσματα όπου το μέγεθος του δείγματος είναι πολύ μικρό για να δικαιολογήσει τέτοιου είδους συμπεράσματα ή όπου άλλες πλευρές του σχεδιασμού της έρευνας οδηγούν σε ακραία συμπεράσματα.

2.1. Μεροληψία διοχέτευσης των αποτελεσμάτων (dissemination bias)

Στον έλεγχο των στατιστικών υποθέσεων, η μηδενική υπόθεση (H_0) αναφέρεται κατά το πλείστον στην απουσία διαφοράς μεταξύ δύο θεραπευτικών αποτελεσμάτων ή στην έλλειψη συσχέτισης μεταξύ μεταβλητών. Δυστυχώς, όμως, σε κάποιες επιστήμες, περιλαμβανομένης της Ιατρικής και των συμπεριφορικών επιστημών, η μηδενική υπόθεση των ερευνητών συχνά ικανοποιεί τον ορισμό της μηδενικής ως «ασήμαντης» και ενδεχομένως και «μη έγκυρης».⁷ Να σημειωθεί ότι στο παραπάνω σχήμα οι όροι σφάλμα τύπου I και σφάλμα τύπου II αναφέρονται στην απόρριψη της μηδενικής υπόθεσης όταν αυτή είναι αληθής, και στη μη απόρριψη της όταν δεν είναι αληθής, αντίστοιχα.⁸

Μια ευκόλως εννοούμενη λοιπόν πηγή μεροληψίας στις αναφερόμενες εκτιμήσεις του μεγέθους του αποτελέσματος απορρέει από την αποτυχία να αναφερθούν οι λεπτομέρειες μιας στατιστικής ανάλυσης, όταν οι διαφορές των μέσων δεν είναι στατιστικά σημαντικές. Ο υπολογισμός της εκτίμησης του μεγέθους ενός αποτελέσματος απαιτεί τις αξίες των στατιστικών δειγμάτων. Οι συγγραφείς μερικές φορές αποτυγχάνουν να αναφέρουν είτε τις στατιστικές δοκιμασίες είτε τις περιγραφικές στατιστικές και καταλήγουν να αναφέρουν απλά ότι δεν παρατήρησαν «κάποια στατιστικά σημαντική διαφορά». Αυτή η τάση να αναφέρουν τις λεπτομέρειες των στατιστικών αναλύσεων μόνον όταν ληφθούν σημαντικά αποτελέσματα αποτελεί τμήμα ενός φαινομένου που έχει ονομαστεί «προκατάληψη ενάντια στη μηδενική υπόθεση» (prejudice against the null hypothesis). Όμως, αν και αυτή η πρακτική είναι εύκολο να ανιχνευτεί, αναμφίβολα οδηγεί σε αποτυχία να υποβληθούν προς δημοσίευση ολόκληρες ερευνητικές αναφορές που εξετάζουν μια κύρια δοκιμασία υπόθεσης. Για παράδειγμα, ενώ το 60% των ψυχολόγων ερευνητών σε μια έρευνα έδειξαν ότι θα υπέβαλαν μια μελέτη προς δημοσίευση αν τα αποτελέσματά της ήταν στατιστικώς σημαντικά, εν τούτοις μόνο το 6% φάνηκε να επιθυμεί να πράξει το ίδιο αν τα αποτελέσματα δεν ήταν το ίδιο σημαντικά.⁷ Παρόμοια αποτελέσματα εξήχθησαν όταν σε

μια μελέτη ταυτοποιήθηκαν 125 έρευνες οι οποίες ήταν ολοκληρωμένες, αλλά όχι δημοσιευμένες. Η πλειοψηφία αυτών (82%) δεν είχε καν υποβληθεί προς δημοσίευση. Η έλλειψη ευεργετικών θεραπευτικών αποτελεσμάτων εντοπίστηκε στο 34% των περιπτώσεων και ήταν η πιο συχνή αιτία για την αποτυχία της δημοσίευσης.⁹

Το 1979, ο Rosenthal επέστησε την προσοχή σε ένα πρόβλημα, το επονομαζόμενο "file drawer problem", που θα μπορούσε να μεταφραστεί στα Ελληνικά ως «πρόβλημα του συρταριού με τους φακέλους», αφού όπως φαντάστηκε αυτές οι μη στατιστικά σημαντικές έρευνες φυλάσσονται στα ερμάρια των ερευνητών. Όπως διατύπωσε λοιπόν, επεξηγώντας το πρόβλημα, τα περιοδικά είναι γεμάτα με το 5% των ερευνών που εμφανίζουν τύπου I σφάλματα, ενώ τα ερμάρια των εργαστηρίων είναι γεμάτα με το 95% των ερευνών που παρουσιάζουν μη στατιστικά σημαντικά αποτελέσματα (π.χ. $p > 0,05$).¹⁰ Το 1985, ο Hedges πρόσθεσε ότι η μεροληψία δημοσίευσης δεν πυροδοτείται μόνο από την πολιτική αναζήτησης στατιστικά σημαντικών αποτελεσμάτων των εκδοτών, αλλά και από τη διστακτικότητα των ερευνητών να αναφέρουν αποτελέσματα όταν το παρατηρούμενο επίπεδο στατιστικής σημαντικότητας (τιμή p) δεν είναι σημαντικό.¹¹ Έτσι, η στατιστική σημαντικότητα είναι ένας καθοριστικός παράγοντας στη δημοσίευση κάποιας έρευνας, αφού ερευνητές είναι πιθανόν να μην την υποβάλλουν προς δημοσίευση εάν δεν πληροί την απαιτούμενη σημαντικότητα, ενώ και οι εκδότες με τη σειρά τους, ακόμη κι αν υποβληθεί, ενδέχεται να μην τη δημοσιεύσουν.¹²

Ο Sterling¹³ ήταν πιθανόν ο πρώτος που έδωσε έμφαση στο γεγονός ότι η τάση προς τη δημοσίευση θετικών αποτελεσμάτων και την απόρριψη αρνητικών ευρημάτων είναι ένα σοβαρό πρόβλημα. Ο ίδιος συγκέντρωσε όλα τα άρθρα που είχαν δημοσιευτεί σε τέσσερα περιοδικά στη διάρκεια ενός έτους (1955 ή 1956) και διαπίστωσε ότι το 97% των άρθρων που χρησιμοποιούσαν δοκιμασίες στατιστικής σημαντικότητας είχαν απορρίψει τη μηδενική υπόθεση στο επίπεδο $\alpha = 0,05$. Και άλλοι στις κοινωνικές και στις συμπεριφορικές επιστήμες κατέληξαν σε παρόμοια αποτελέσματα για τη μεροληψία δημοσίευσης. Εάν οι έρευνες που αξιολογήθηκαν από τον Sterling είναι αντιπροσωπευτικές όλων των ερευνών που διεξήχθησαν στην Ψυχολογία, τα αποτελέσματά τους υπονοούν είτε ένα μεγάλο αριθμό σφαλμάτων τύπου I, είτε μια μέτρια δύναμη στην υπερβολή του 0,90, είτε κάποιο συνδυασμό αυτών των πιθανοτήτων.¹⁴

Η διαχείριση και η χρήση όμως μιας μη αντιπροσωπευτικής μερίδας στατιστικά σημαντικών ερευνών ή ερευνών που εξάγουν αποτελέσματα στραμμένα προς μια θετική

κατεύθυνση συνεπάγεται ένα σύνολο αντιπροσωπευτικών ερευνών μέσα στη διαδικασία μιας μετα-ανάλυσης. Ένα πρότυπο μετα-αναλυτικό μοντέλο θα καταλήξει σε ένα συμπέρασμα μεροληπτικό υπέρ της σημαντικότητας ή της «θετικότητας». Αυτό είναι ιδιαίτερα επίφοβο και επικίνδυνο για μια μετα-ανάλυση, τα δεδομένα της οποίας προέρχονται αποκλειστικά από τη δημοσιευμένη επιστημονική βιβλιογραφία, αλλά οι έρευνες μπορεί να είναι «συγκυμμένες» για επί πλέον λόγους εκτός της αποτυχίας τους να δημοσιευτούν.¹⁵

2.2. Μεροληψία στη φάση της εκδοτικής απόφασης (publication bias in editorial decision making)

Η οπτική γωνία ενός εκδότη σχετικά με το τι είδους έρευνες παρέχουν τις πλέον χρήσιμες πληροφορίες, ποιες καινοτομίες είναι πιθανότερο να αποδειχθούν αποτελεσματικές και ποιο μέγεθος αποτελέσματος είναι άξιο δημοσίευσης, μπορεί εξ ίσου με αυτή των συγγραφέων να συνεισφέρει στη συστηματική διαλογή των δεδομένων.¹⁶ Για παράδειγμα, μια γενικότερη ανησυχία αιωρείται αναφορικά με το γεγονός ότι, εφ' όσον ο μισθός ενός εκδότη πληρώνεται εν μέρει μέσω των διαφημιστικών εσόδων των περιοδικών, οι εκδοτικές τους αποφάσεις είναι δυνατό να διαφοροποιούνται και να παραλλάσσονται συνειδητά ή ασυνείδητα.¹⁷

Οι ερευνητές πρέπει να καθορίζουν για τους ίδιους ποια μεγέθη αποτελεσμάτων είναι σημαντικά προκειμένου να σχεδιάσουν κατάλληλα για ένα επαρκές μέγεθος δείγματος. Οι οπτικές των εκδοτών σχετικά με τα χρήσιμα μεγέθη αποτελεσμάτων μπορούν να θεωρηθούν ως μια παρέχουσα προοπτική στην κρίση των ερευνητών. Όμως, επειδή τυπικά οι εκδότες «προσφέρουν» αυτές τις κρίσεις εκ των προτέρων, η κρίση τους μπορεί να οδηγήσει σε μεροληπτικό δείγμα μεγέθους αποτελέσματος το οποίο δημοσιεύεται στη βιβλιογραφία.

Σε κάθε περίπτωση όμως υπάρχουν ενδείξεις γι' αυτό που παρατίθεται στη συγκεκριμένη ενότητα ως εκδοτική μεροληψία. Το 1983, σε ένα σημείο μιας στήλης του γνωστού περιοδικού *British Medical Journal* φιλοξενήθηκε η συμβουλή ότι «αυτοί που αναζητούν ταχεία δημοσίευση μιας έρευνας (ειδικά με αρνητικά αποτελέσματα) θα έπρεπε να υποβάλλουν τα άρθρά τους σε περιοδικό που λειτουργεί επί πληρωμή». Και τελειώνοντας «... όσο για το πόσοι άνθρωποι θα το δουν τότε ... λοιπόν, αρνητικά αποτελέσματα ποτέ δεν προκάλεσαν καθηλωτική ανάγνωση».¹⁸ Η σχετική κατάσταση επέφερε συζήτηση για το εάν είναι λογικό να υπάρχει μια εφημερίδα με αρνητικά αποτελέσματα: «Η επιστημονική ανάγκη για δημοσίευση των αρνητικών ή των μηδενικής διαφοράς αποτελεσμάτων είναι εμφανής

και το περιοδικό των αρνητικών αποτελεσμάτων (*Journal of Negative Results*) αποτελεί αστείο. Τέτοιο περιοδικό είναι όχι μόνο προαποφασισμένα βαρετό, αλλά αποτελεί και οικονομική καταστροφή.¹⁹ Ο ίδιος ο Maxwell, επί πλέον, πρόσθεσε ότι τουλάχιστον οι εκδότες θα έπρεπε να παρέχουν μια καταγραφή των αρνητικών αποτελεσμάτων και του θέματος, καθώς και τα ονόματα των συγγραφέων και τις διευθύνσεις, τα οποία να είναι προσβάσιμα.¹⁹ Αυτό αποτελεί μια εξ ίσου μη ικανοποιητική λύση στο πρόβλημα.²⁰ Ο Edward Huth, εκδότης του περιοδικού *Annals of Internal Medicine*, δήλωσε ότι ένα ηλεκτρονικό ιατρικό περιοδικό θα μπορούσε να δημοσιευτεί σε χαμηλότερο κόστος έναντι αυτού των τυπωμένων περιοδικών, κι έτσι θα ήταν σε καλύτερη θέση να δημοσιεύει αρνητικά ή μηδενικά αποτελέσματα. Όμως, προειδοποίησε ότι «πρέπει να είναι σαφές από τον τίτλο ή από την περίληψη ότι η έρευνα κατέληξε σε αρνητικό συμπέρασμα, από φόβο μήπως οι συγγραφείς ή οι ερευνητές νομίζουν ότι λαμβάνουν θετικά αποτελέσματα».²¹ Η εφημερίδα *The Journal of American Medical Association* κάποτε περιελάμβανε στις σελίδες της ένα αρχείο που ονομαζόταν «αρνητικά αποτελέσματα», το οποίο φιλοξενούνταν στις σελίδες της τακτικά, περίπου μία φορά το μήνα από το 1962–1968. Τα άρθρα είχαν έκταση 1, 1½ ή και 2 σελίδες ακόμη. Δυστυχώς, όμως, δεν υπάρχουν περαιτέρω πληροφορίες για την τεκμηρίωση της έναρξής της ή τη συνέχιση αυτού του τμήματος της εφημερίδας.²⁰

Από την άλλη πλευρά, στο στάδιο του εκδότη, δεν θα πρέπει να παραλειφθεί ότι υπάρχει μια γενικότερη ανάγκη για περικοπή των υποβληθέντων χειρογράφων, ως αποτέλεσμα της έλλειψης διαθέσιμου χώρου στα περιοδικά. Ενώ οι κύριοι προβληματισμοί αφορούν συνήθως στην ορθότητα των μεθόδων που χρησιμοποιήθηκαν και στη σημασία ή στη σχετικότητα του διαπραγματευόμενου θέματος, η αποκλειστικότητα των ευρημάτων είναι επίσης σημαντικό να επηρεάσει την απόφαση μια έρευνα να γίνει δεκτή ή να απορριφθεί.

Ως αποτέλεσμα αυτού του είδους πολιτικής, φαίνεται πλέον βέβαιο ότι άρθρα που προσφέρουν στατιστική ση-

μαντικότητα έχουν αυξημένες πιθανότητες αποδοχής (105 από τα 106 άρθρα στο προαναφερόμενο περιοδικό ήταν σημαντικά στο επίπεδο 5% στην έρευνα που διεξήχθη από τον Sterling και τα οποία παρουσιάζονται στον πίνακα 1).¹³

2.3. Οικονομικά κίνητρα

Ο Davidson συγκέντρωσε 107 δοκιμές, δημοσιευμένες το 1984, και τις ταξινομήσε σύμφωνα με την κατεύθυνση των αποτελεσμάτων (αυτές που υποστήριζαν τη νέα θεραπεία και εκείνες που υποστήριζαν τη συμβατική), καθώς και την πηγή της χρηματοδότησής τους (αυτές που υποστηρίζονταν από τις φαρμακευτικές εταιρείες και αυτές που ήταν γενικά υποστηριζόμενες). Εβδομήντα έξι από τις 107 δοκιμές (71%) υποστήριζαν τη νέα θεραπεία και 31 συμβατική (29%). Από αυτές που υποστήριζαν τη νέα, οι 33 (43%) ήταν χρηματοδοτούμενες μέσω φαρμακευτικών εταιρειών, ενώ από εκείνες της συμβατικής θεραπείας μόνο 4 (13%). Η εν λόγω αναφορά μεταφράζεται σε ποσοστό 89% (33/37) των ερευνών ότι ήταν χρηματοδοτούμενες από φαρμακευτικές εταιρείες, ενώ 61% (43/70) των χρηματοδοτούμενων από γενικούς φορείς ήταν υπέρ της νέας θεραπείας. Το ποσοστό των άρθρων που υποστήριζαν τις νέες έναντι των συμβατικών θεραπειών διακυμαινόταν αισθητά, σε εξάρτηση με το περιοδικό που το αξιολογούσε²² (πίν. 2). Ο Davidson κατέληγε στην έρευνά του ως εξής: «Ενώ φαίνεται απίθανο ότι υπάρχουν συνομωσίες που προσπαθούν να συγκαλύψουν αποτελέσματα, ένας δεδομένος αποκλεισμός των αρνητικών αποτελεσμάτων είναι πιθανόν να υφίσταται».²² Το ενδεχόμενο μιας συνομωσιακής προσπάθειας συγκαλύψης των αποτελεσμάτων έχει ήδη εγερθεί, αλλά αποκλείστηκε.²³ Η πιθανότητα ότι η χρηματοδότηση μπορεί να επηρεάσει τον τρόπο με τον οποίο παρουσιάζονται τα αποτελέσματα εκτείνεται πέρα από τον τομέα των κλινικών δοκιμών.⁴

Επιδημιολογικές έρευνες έχουν δείξει ότι μελέτες των περιβαλλοντικών κινδύνων που δεν παρουσιάζουν συσχέτιση μεταξύ ασθένειας και έκθεσης συχνά ήταν δημο-

Πίνακας 1. Ανασκόπηση δημοσιευμένων άρθρων, χρησιμοποιώντας δοκιμασία στατιστικής σημαντικότητας.¹³

Περιοδικό	Αριθμός ερευνών σε 5% επίπεδο στατιστικής σημαντικότητας			
	Έρευνες με δοκιμές στατιστικής σημαντικότητας	Στατιστικά σημαντικές έρευνες	Στατιστικά μη σημαντικές έρευνες	Αριθμός επιβεβαιωτικών ερευνών
Journal of Experimental Psychology	106	105	1	0
Journal of Comparative and Physiological Psychology	94	91	3	0
Journal of Clinical Psychology	62	59	3	0
Journal of Social Psychology	32	31	1	0

Πίνακας 2. Κατεύθυνση αποτελεσμάτων κλινικών δοκιμών δημοσιευμένων το 1984 σε επιλεγμένα περιοδικά.²⁰

Περιοδικό	n (%)		Σύνολο
	Υπέρ της νέας θεραπείας	Υπέρ της συμβατικής	
Annals of Internal Medicine	12 (86)	2 (14)	14
Archives of Internal Medicine	11 (79)	3 (21)	14
Lancet	30 (75)	10 (25)	40
New England Journal of Medicine	18 (67)	9 (33)	27
American Journal of Medicine	5 (42)	7 (58)	12

σιευμένες από ομάδες με δεδομένο ιδιαίτερο ενδιαφέρον στην απόδειξη έλλειψης υπαιτιότητας, όπως εταιρείας που εισήγαγε τους κινδύνους.²⁴ Υπάρχουν όντως εικασίες ότι η μεροληψία δημοσίευσης είναι αρκετά πιο σοβαρή σε επιδημιολογικές μελέτες, παρά στην κλινική έρευνα.²⁵ Στις περιβαλλοντικές μελέτες υπάρχει μια πιθανή μεροληψία υπέρ των ερευνών που δεν αποδεικνύουν κάποια σημαντική συσχέτιση μεταξύ ασθένειας-έκθεσης, ενώ το αντίθετο ισχύει για τις κλινικές μελέτες. Για παράδειγμα, εάν υπήρχαν μόνο αδύναμα στοιχεία για το ρόλο του ασβέστη στη διαδικασία της καρκινογένεσης, θα αντιπροσώπευε έναν ικανοποιητικό παράγοντα, τόσο για το κοινό όσο και για τη βιομηχανία, αλλά όχι για τα κανονικά επίπεδα ασβέστη.³ Ο Kotelchuck πρότεινε ότι ο συγκεκριμένος λόγος, ιστορικά, έχει παράσχει ένα κίνητρο για τη βιομηχανία ώστε εκείνη να δημοσιεύει έρευνες αρνητικές για τον ασβέστη.²⁶ Ένα επί πλέον παράδειγμα αρνητικής μεροληψίας περιγράφηκε το 1980 σε μια ανασκόπηση αναφορών υποβεβλημένων στις κυβερνήσεις τους από σκανδιναβικές φαρμακευτικές εταιρείες. Το ποσοστό των ελεγχόμενων δοκιμών που παρείχαν στοιχεία για ανεπιθύμητες ενέργειες φαρμάκων ήταν συνεχώς χαμηλότερο μεταξύ των δημοσιευμένων ερευνών, παρά μεταξύ των αδημοσιευτών.²⁷

2.4. Τα χαρακτηριστικά των ερευνών

Όπως ήδη αναφέρθηκε και παραπάνω, είναι πιθανόν κάποια χαρακτηριστικά της έρευνας να συντελούν και να συνδέονται με την εμφάνιση του σφάλματος δημοσίευσης. Κάποια από τα διακριτά χαρακτηριστικά που έχουν μελετηθεί είναι η παρουσία ή η απουσία τυχαιοποίησης, το μέγεθος του δείγματος, οι διερευνητικές σε σχέση με τις επιβεβαιωτικές έρευνες, ο καθορισμός πρωτοκόλλου, η φύση του περιοδικού και ο ημερολογιακός χρόνος.

Ένας άλλος παράγοντας ο οποίος είναι πιθανόν να ενισχύει τη μεροληψία δημοσίευσης είναι η απουσία ενός σαφώς καθορισμένου πρωτοκόλλου. Αυτό συνήθως ισχύει σε πολλές ιστορικές, ελεγχόμενες έρευνες, αλλά σχετίζεται επίσης με τις μη ελεγχόμενες έρευνες, για παράδειγμα έρευνες στη φάση II νέων παραγόντων καρκίνου ή συνδυασμών. Χιλιάδες τέτοιου είδους έρευνες συντάσσονται ετησίως, και ενώ κάποιες είναι προσεκτικά σχεδιασμένες με αντικειμενικά ακριβή δείγματα, σε άλλες περιπτώσεις είναι πιθανόν νέοι ασθενείς να εισέλθουν στο πρωτόκολλο. Δεν είναι ασύνηθες σε αυτές τις περιπτώσεις τα αποτελέσματα να δημοσιεύονται σε μια σειρά ασθενών οι οποίοι χαρακτηρίζονται από αυθαίρετες ημερομηνίες έναρξης και παύσης της συμμετοχής τους, επιτρέποντας έτσι επιλεκτικότητα στα αποτελέσματα που θα δημοσιευτούν.³ Ένα αντιπροσωπευτικό παράδειγμα αντιθετικών αποτελεσμάτων που απορρέουν από τέτοιου είδους έρευνα παρουσιάστηκε το 1984.²⁸ Το 1987, σε μια έρευνα των Furberg και Morgan παρουσιάστηκε η χρήση του μη θνησιγόνου μυοκαρδιακού εμφράγματος ως επιθυμητού σημείου, το οποίο βρέθηκε να είναι σημαντικό σε μια μετα-ανάλυση της ενδοφλέβιας στρεπτοκινάσης, ένα αποτέλεσμα που δεν τεκμηριώθηκε σε μια μεγάλη πολυκεντρική δοκιμή. Σε γενικές γραμμές, η χρήση πολλαπλών συγκρίσεων είναι πιθανόν να ενισχύσει τη μεροληψία δημοσίευσης.³

Ο ημερολογιακός χρόνος επίσης έχει θεωρηθεί ότι επηρεάζει τη μεροληψία δημοσίευσης από την άποψη ότι οι αρχικές έρευνες τείνουν να έχουν ένα πιο διερευνητικό ύφος, ενώ οι πλέον πρόσφατες περισσότερο επιβεβαιωτικό. Επί πλέον, όσο συσσωρεύεται η γνώση συμπαραγόντων, οι πληροφορίες είναι πιθανόν να χρησιμοποιηθούν αποτελεσματικά στις επόμενες έρευνες. Αυτή η εξήγηση έχει δοθεί από τον Greenwald για να ερμηνεύσει προσωρινές αλλαγές στις έρευνες στάσεων και επιλεκτικής μάθησης,⁷ καθώς και από τους Pocock et al σχετικά με έρευνες της χαμηλής λιποπρωτεϊνικής πυκνότητας χοληστερόλης, καθώς και των καρδιακών διαταραχών.²⁹

Το μέγεθος του δείγματος θεωρείται ότι διαδραματίζει ένα σημαντικό ρόλο, άλλοτε άμεσο και άλλοτε έμμεσο, στην απόφαση δημοσίευσης. Σε μια έρευνα, το 1989, ανασκοπήθηκαν 246 δημοσιευμένες έρευνες για τη θεραπεία του καρκίνου και εντοπίστηκε ισχυρή συσχέτιση μεταξύ του μεγέθους του δείγματος και του θεραπευτικού αποτελέσματος.^{4,30} Πιο συγκεκριμένα, έρευνες με μικρότερα μεγέθη δειγμάτων παρουσίαζαν μεγαλύτερα θεραπευτικά αποτελέσματα. Η σχετική τάση είναι πιο ισχυρή σε τυχαιοποιημένες έρευνες έναντι μη τυχαιοποιημένων ερευνών. Αυτό υποδηλώνει ότι μικρές δοκιμές με μεγάλα αποτελέσματα τείνουν να είναι περισσότερο επιθυμητές προς δημοσίευση, ενώ μεγάλες δοκιμές είναι πιθανόν να

δημοσιευτούν ανεξαρτήτως αποτελέσματος.²⁰ Σε μια έρευνα που αναφέρθηκε και πιο πάνω, όπου ανασκοπήθηκαν 7 μελέτες παρατήρησης σχετικά με την προστατευτική δράση της υψηλής πυκνότητας λιποπρωτεΐνης (HDL) στα ισχαιμικά καρδιακά επεισόδια, παρατηρήθηκαν τα εξής:²⁹ Έξι από τις έρευνες παρουσίαζαν μικρή προστατευτική επίδραση, με διαφορές στη μέση ποσότητα HDL που κυμαινόταν από 0,04–0,09 mmol/L και μέγεθος δείγματος το οποίο κυμαινόταν από 55–193 περιπτώσεις. Η μια ακραία τιμή έφερε ισχυρά στατιστικό υπολογισμό της τάξης του 0,036 mmol/L βασισμένο σε ένα δείγμα 17 περιπτώσεων. Όχι μόνο η εν λόγω έρευνα ήταν μακράν η μικρότερη, αλλά ήταν επίσης και η πρώτη από τις έρευνες που δημοσιεύτηκαν και κατά συνέπεια θεωρείται διερευνητικού ύφους, τη στιγμή που οι υπόλοιπες ήταν επιβεβαιωτικές έρευνες. Εάν ανατρέξει κάποιος σε ένα διαφορετικό παράδειγμα παρόμοιας όμως φύσης, θα διαπιστωθεί και πάλι η επιρροή του μεγέθους του δείγματος, αυτή τη φορά σε επιδημιολογική μελέτη για τους εργαζόμενους στην επεξεργασία βινυλοχλωριδίου. Παρατηρήθηκε λοιπόν σε μια περιληπτική μελέτη η ύπαρξη σχέσης μεταξύ τυποποιημένου λόγου θνησιμότητας και αναμενόμενου αριθμού θανάτων.³¹

3. ΜΕΘΟΔΟΙ ΔΙΑΠΙΣΤΩΣΗΣ ΚΑΙ ΕΠΙΔΙΟΡΘΩΣΗΣ ΤΗΣ ΥΠΑΡΞΗΣ ΜΕΡΟΛΗΨΙΑΣ ΚΑΤΑ ΤΗ ΔΗΜΟΣΙΕΥΣΗ

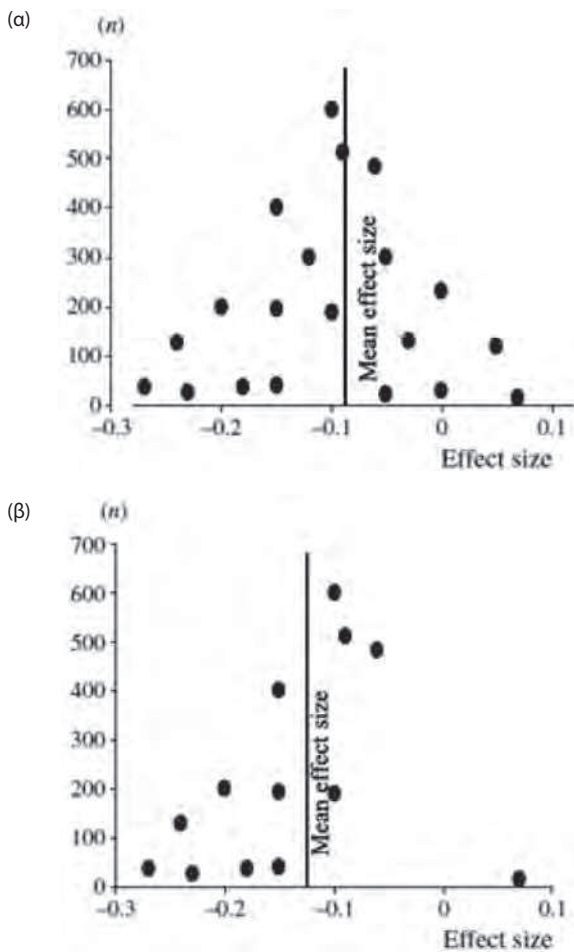
3.1. Τρόποι ανίχνευσης της μεροληψίας

Οι μέθοδοι αναγνώρισης και περιορισμού της μεροληψίας ταξινομούνται διαδοχικά ως εξής: Πρόληψη της μεροληψίας δημοσίευσης που προηγείται μιας συστηματικής ανασκόπησης (π.χ. προοπτική καταγραφή των δοκιμών), περιορισμός της μεροληψίας που συμβαίνει κατά τη διάρκεια της συστηματικής ανασκόπησης (π.χ. εντοπισμός αδημοσίευτων ερευνών ή γκριζας βιβλιογραφίας, όπως θα παρατηρηθεί πιο κάτω), ανίχνευση μεροληψίας κατά τη διάρκεια της ανασκόπησης (π.χ. διάγραμμα χωνί [funnel plot], σχετικές δοκιμασίες και μοντέλο ανάλυσης ευαισθησίας), καθώς και περιορισμός επίδρασης της μεροληψίας μετά από μια συστηματική ανασκόπηση (για παράδειγμα, επιβεβαιωτικές δοκιμές μεγάλης κλίμακας).³²

3.1.1. *Διάγραμμα χωνί.* Μια από τις μεθόδους που αναφέρθηκαν παραπάνω ως μέθοδος ανίχνευσης του σφάλματος δημοσίευσης είναι και το λεγόμενο διάγραμμα χωνί (funnel plot). Το όνομά του βασίζεται στο γεγονός ότι η ακρίβεια στους υπολογισμούς του υποκείμενου θεραπευτικού αποτελέσματος ενισχύεται όσο αυξάνει το μέγεθος του δείγματος των ερευνών στην ανασκόπηση. Έτσι, όταν ένα μέτρο του μεγέθους της έρευνας τοποθετείται στον κάθετο άξονα, τα αποτελέσματα από τις μικρές έρευνες θα

είναι διασκορπισμένα στην κάτω πλευρά του γραφήματος, με τη διασπορά να παρατηρείται στενότερη ανάμεσα στις μεγαλύτερες έρευνες.¹⁶ Στην απουσία μεροληψίας (και όταν οι έρευνες εκτιμούν το ίδιο υποκείμενο αποτέλεσμα) το γράφημα θα ομοιάζει με ένα ανεστραμμένο συμμετρικό χωνί, όπως γίνεται αντιληπτό και από την εικόνα 1α. Σε αυτό το μοντέλο, όλες οι έρευνες της ανάλυσης προϋποτίθεται ότι υπολογίζουν το ίδιο αποτέλεσμα. Γι' αυτόν το λόγο, τα εκτιμώμενα αποτελέσματα θα πρέπει να κατανέμονται γύρω από το άγνωστο πραγματικό αποτέλεσμα, με το εύρος των εκτιμήσεων να αντανakλούν τις διασπορές τους. Συνεπώς, όπως αναφέρθηκε και παραπάνω, οι μικρές έρευνες στην κάτω πλευρά του γραφήματος πρέπει να είναι αρκετά διασκορπισμένες, ενώ όσο αυξάνει το μέγεθος του δείγματός μας προς τα επάνω το γράφημα πρέπει να στενεύει, αντιπροσωπεύοντας τη μικρότερη διασπορά. Εάν όμως υπάρχει επίδραση της μεροληψίας δημοσίευσης, αυτό θα πρέπει επίσης να αντανakλάται στο γράφημα. Για παράδειγμα, εάν αρνητικές έρευνες είναι λιγότερο πιθανό να δημοσιευτούν, το γράφημα θα τείνει να είναι επικλινές ή ασύμμετρο, παράγοντας μια αρνητική συσχέτιση στο γράφημα ή, αλλιώς, μια θετική συσχέτιση μεταξύ των εκτιμήσεων των αποτελεσμάτων και των διασπορών τους.³³ Στην εικόνα 1β απεικονίζεται ένα μη συμμετρικό γράφημα. Σε αυτή την περίπτωση, το συνδυασμένο αποτέλεσμα των μετα-αναλύσεων θα υπερεκτιμήσει το μέσο αποτέλεσμα ή τη δύναμη της σχέσης. Όσο μεγαλύτερη είναι η ασυμμετρία στο γράφημα, τόσο πιο πιθανό θεωρείται το μέγεθος της μεροληψίας να είναι μεγάλο. Fail-safe (αποτυχίας-ασφάλειας) αποκαλούνται οι μέθοδοι που χρησιμοποιούνται για την εκτίμηση του αριθμού των μη δημοσιευμένων ερευνών με αρνητικά αποτελέσματα, ο οποίος πρέπει να ισχύει έτσι ώστε το αποτέλεσμα μιας μετα-ανάλυσης να μην είναι πλέον σημαντικό ή το μέγεθος του αποτελέσματος να είναι αμελητέας ποσότητας.³⁴

Τα funnel plots χρησιμοποιήθηκαν αρχικά στην εκπαιδευτική έρευνα και στην Ψυχολογία.³⁵ Παρά το γεγονός όμως ότι πρωτοεμφανίστηκαν στη βιβλιογραφία των κοινωνικών επιστημών, ακόμη δεν χρησιμοποιούνται ευρέως σε αυτό το επιστημονικό πεδίο. Το 1994 εξετάστηκαν 74 μετα-αναλύσεις που είχαν δημοσιευτεί από το 1985–1991 στο περιοδικό *Psychological Bulletin* και βρέθηκε ότι μόνο μία είχε χρησιμοποιήσει funnel plot.³⁶ Ακόμη πιο πρόσφατα, εντοπίστηκε μία μόνο ανασκόπηση που χρησιμοποίησε funnel plot ανάμεσα στις 28 μετα-αναλύσεις οι οποίες εμφανίστηκαν το 1999–2002 σε δύο επιφανή επιστημονικά περιοδικά κοινωνικού περιεχομένου (*Review of Educational Research, Psychological Bulletin*). Οι επιστήμονες που διεξήγαγαν την έρευνα ανέφεραν ότι, αντίθετα, 6 από τις 19 μετα-αναλύσεις που δημοσιεύτηκαν στο γνωστό περιοδικό



Εικόνα 1. Funnel plot απουσία μεροληψίας (α), funnel plot παρουσία μεροληψίας (ασυμμετρία) (β).⁶¹

Journal of the American Medical Association από το 2001 έως τον Απρίλιο του 2002 συζήτησαν την εφαρμογή αυτών των γραφημάτων για τα αποτελέσματά τους (αν και μόνο δύο από αυτά εμφανίστηκαν δημοσιευμένα).³⁷

Οι επιλογές των μονάδων των δύο αξόνων του funnel plot είναι διάφορες. Η διάταξη των πιθανών επιλογών οδηγεί στον προβληματισμό και στη σκέψη ότι η οπτική εντύπωση που προσδίδει το γράφημα μπορεί να έχει επηρεαστεί από την επιλογή των αξόνων, ενώ επί πλέον οι ερευνητές ενδέχεται να έχουν επιλέξει τους συγκεκριμένους άξονες ακριβώς για να σχηματίσουν την επιθυμητή εντύπωση, ή ότι οι επιλεγμένοι άξονες μπορεί να μην είναι κατάλληλοι για τον εντοπισμό της μεροληψίας.

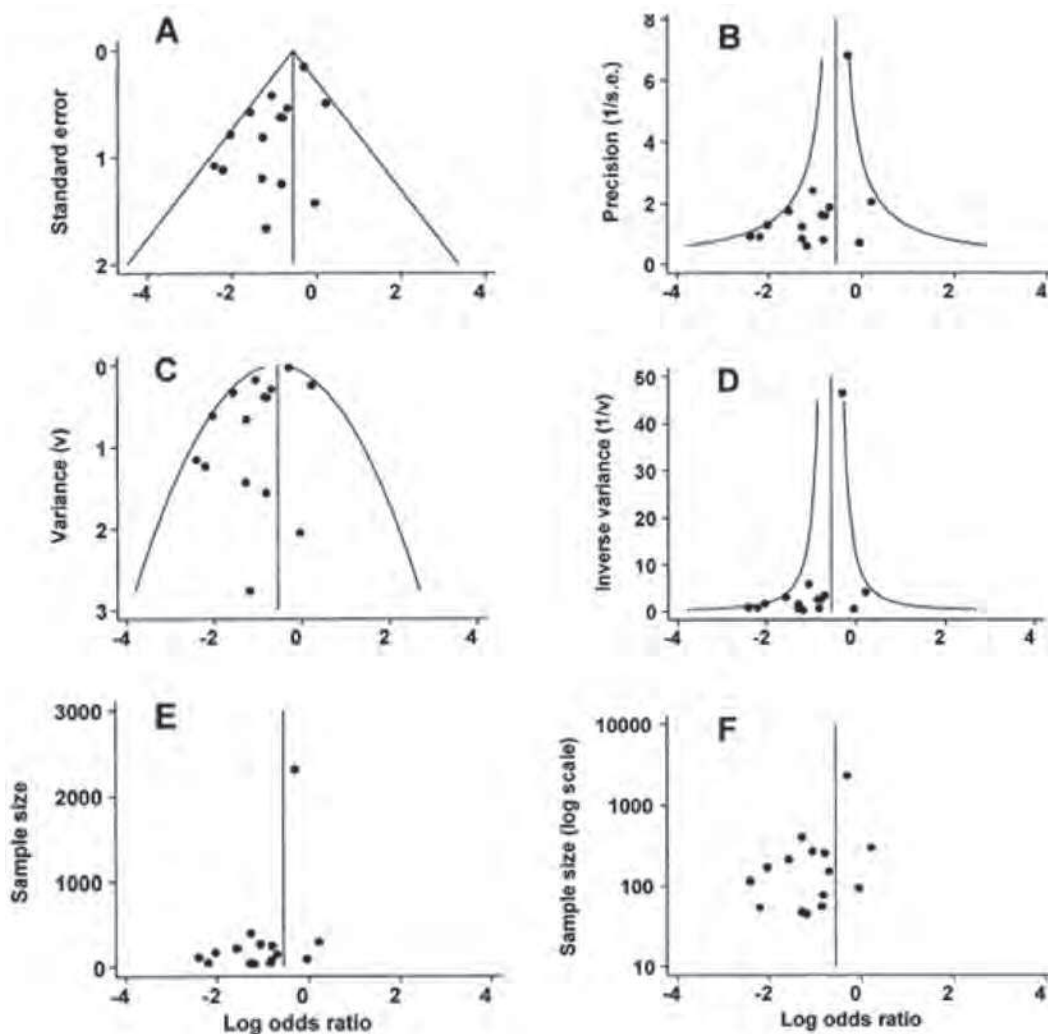
3.1.2. Επιλογή αξόνων του funnel plot

α. Έρευνες με διχότομα αποτελέσματα. Με το θέμα αυτό ασχολήθηκαν εκτεταμένα οι Sterne και Egger το 2001, προβάλλοντας την επιλογή του κάθετου άξονα του funnel

plot, στο πλαίσιο τυχαιοποιημένων ελεγχόμενων δοκιμών σχετικά με τη χρήση μαγνησίου για την πρόληψη θανάτου που προκαλείται από έμφραγμα μυοκαρδίου.³⁸ Παρουσίασαν λοιπόν έξι επιλογές. Πιο συγκεκριμένα, παρουσιάζονται 6 funnel plots, το καθένα από τα οποία απεικονίζει δοκιμές χρησιμοποιώντας σε κάθε κάθετο άξονα ένα διαφορετικό μέτρο μεγέθους δείγματος (εικ. 2). Σε όλες τις περιπτώσεις, χρησιμοποιήθηκε ως μέτρο μεγέθους του αποτελέσματος της παρέμβασης ο λογάριθμος του λόγου συμπληρωματικών πιθανοτήτων (logOR), του οποίου η μηδενική τιμή (OR=1) αντιστοιχεί σε καμιά επίδραση της παρέμβασης, ενώ αρνητικές τιμές (OR<1) ανταποκρίνονται σε ευεργετικά αποτελέσματα της παρέμβασης.

Έτσι λοιπόν, το πρώτο (Α) από τα έξι funnel plots χρησιμοποιεί στον κάθετο άξονά του ως μέτρο μεγέθους δείγματος το τυπικό σφάλμα (standard error). Όμως, προκειμένου να καταλήξουν οι μεγαλύτερες έρευνες να είναι στην κορυφή του γραφήματος εφ' όσον έχουν τα μικρότερα τυπικά σφάλματα, ο κάθετος άξονας χρειάστηκε να αναστραφεί, έχοντας τη μηδενική τιμή του τυπικού σφάλματος στην κορυφή του άξονα. Οι διαγώνιες γραμμές αντιπροσωπεύουν τα αναμενόμενα 95% διαστήματα εμπιστοσύνης γύρω από τη συνολική εκτίμηση, τα οποία υπολογίζονται με βάση τον εξής τύπο: Σύνοψη εκτίμησης αποτελέσματος $\pm(1,96 \text{ SE})$ για κάθε τυπικό σφάλμα του κάθετου άξονα. Αυτά δείχνουν το βαθμό ετερογένειας μεταξύ των ερευνών. Στην απουσία ετερογένειας, 95% των ερευνών θα βρίσκονται μέσα στο διάγραμμα που ορίζεται από αυτές τις γραμμές. Στο συγκεκριμένο γράφημα, αρκετά σημεία πέφτουν αριστερά της γραμμής, αντανakλώντας OR χαμηλότερα των αναμενόμενων υπό την υπόθεση του μοντέλου σταθερών επιδόσεων. Το funnel plot (B) χρησιμοποιεί ως μέτρο μεγέθους δείγματος την ακρίβεια (1/SE). Το (C) χρησιμοποιεί τη διακύμανση (variance), που είναι το τετράγωνο του τυπικού σφάλματος, ενώ το (D) το αντίστροφο της (1/διακύμανση). Στα funnel plots (E) και (F) χρησιμοποιήθηκε το μέγεθος του δείγματος και ο λογάριθμος του μεγέθους δείγματος, αντίστοιχα.

β. Έρευνες με συνεχή αποτελέσματα. Μέχρι στιγμής από τα στοιχεία που διατίθενται, καμιά εμπειρική έρευνα δεν έχει εξετάσει και παρουσιάσει την καλύτερη επιλογή αξόνων για funnel plots, τα αποτελέσματα των οποίων από τις συγκεντρωτικές έρευνες να εμφανίζονται μετρημένα σε ονομαστική κλίμακα, όπως για παράδειγμα οι τιμές της αρτηριακής πίεσης. Τα αποτελέσματα σε αυτές τις έρευνες είναι συνήθως τυποποιημένοι ονομαστικοί δείκτες, όπως η τυποποιημένη διαφορά μέσω ή η συσχέτιση. Όσον αφορά στην επιλογή του οριζόντιου άξονα, λιγότερο αμφιλεγόμενες είναι οι μετρήσεις διαφορών σχετικά με τα θεραπευτικά αποτελέσματα. Ενώ υπάρχουν αρκετοί



Εικόνα 2. Funnel plots με διαφορετικούς κάθετους άξονες δοκιμών μαγνησίου. Τα σημεία ανταποκρίνονται στις θεραπευτικές επιδράσεις ανεξάρτητων δοκιμών και οι διαγωνίες ή οι κοίλες γραμμές δείχνουν το 95% διάστημα εμπιστοσύνης γύρω από τη συνολική εκτίμηση.¹⁶

λόγοι για να θεωρούνται ευκολότερα ερμηνεύσιμα τα θεραπευτικά αποτελέσματα που μετρώνται ως διαφορές των μέσων παρά τυποποιημένες διαφορές μέσων ή συντελεστές συσχέτισης, αυτές οι επιλογές δεν έχουν εμφανείς επιπτώσεις στο σχήμα των funnel plots.¹⁶

Στις κοινωνικές ιδιαίτερα επιστήμες, οι οποίες χρησιμοποιούν τουλάχιστον δύο τυπικούς δείκτες αποτελέσματος, τη διαφορά μέσων “d” και τη συσχέτιση “r”, τα τυπικά σφάλματα των αποτελεσμάτων των ερευνών εξαρτώνται από το μέγεθος των ερευνών (n_{obs}) και από τα αποτελέσματά τους. Για παράδειγμα, το τυπικό σφάλμα της συσχέτισης είναι

$$SE(r) = \frac{1 - \rho^2}{\sqrt{n_{obs}}}$$

το οποίο εξαρτάται από το μέγεθος του πληθυσμού συσχέτισης, καθώς και από το n_{obs} . Όπως ακριβώς το μέγεθος του τυπικού

σφάλματος του λογαρίθμου του λόγου συμπληρωματικών πιθανοτήτων (logOR) εξαρτάται από τον αριθμό των συμβάντων, το μέγεθος του τυπικού σφάλματος της συσχέτισης r εξαρτάται από τη συσχέτιση που παρατηρείται στο δείγμα. Για παράδειγμα, το τυπικό σφάλμα της συσχέτισης $r=0,2$ από ένα δείγμα παρατηρήσεων $n_{obs}=178$ είναι 0,072. Στο ίδιο τυπικό σφάλμα καταλήγουμε με μια συσχέτιση $r=0,8$ και $n_{obs}=25$. Γ’ αυτόν το λόγο, είναι και δύσκολο να υπολογιστεί μια σταθερή περιοχή εμπιστοσύνης για συσχετίσεις σε funnel plots όταν το r βρίσκεται σε άξονα μαζί με το πλήθος των παρατηρήσεων ή τα τυπικά τους σφάλματα.

Μια εναλλακτική για τη συσχέτιση και το d είναι η δημιουργία γραφημάτων, χρησιμοποιώντας σταθεροποιητές μεταλλαγών διακύμανσης. Αυτοί μπορούν να χρησιμοποιηθούν για την απόκτηση δεικτών με τυπικά σφάλματα, που είναι σχετικά απλές λειτουργίες του πλήθους παρατηρήσεων

n_{obs} . Για τη συσχέτιση μπορεί να χρησιμοποιηθεί η οικεία μετατροπή z του Fisher:

$$z = 1/2 \ln \left[\frac{1+r}{1-r} \right]$$

Η τιμή z έχει τυπικό σφάλμα $\frac{1}{\sqrt{n_{obs}-3}}$, το οποίο μπορεί να χρησιμοποιηθεί για τη δημιουργία μιας περιοχής εμπιστοσύνης για τιμές του z εναντίον τιμών n_{obs} ή τα τυπικά τους σφάλματα.³⁹ Γενικά, πάντως, όσον αφορά στην αξία που δίνουν οι αναγνώστες στα αποτελέσματα τα οποία εμφανίζονται στα funnel plots, οι σταθεροποιημένες διακύμανσης μορφές που λαμβάνουν τα r και d μπορεί να μην είναι τόσο κατατοπιστικές όσο οι αυθεντικές τους μορφές.¹⁶

3.2. Στατιστικές μέθοδοι παλινδρόμησης

Μέχρι στιγμής, από την προηγούμενη ενότητα προσδιορίσαμε μια μέθοδο για την εξέταση του σφάλματος δημοσίευσης, η οποία είναι μάλλον αρκετά υποκειμενική ώστε να περιοριστούμε σε αυτή και να στηρίξουμε τα συμπεράσματά μας, αφού το ίδιο γράφημα και η ασυμμετρία του είναι δυνατόν να ερμηνευτούν με διαφορετικούς τρόπους από διαφορετικούς ερευνητές.⁴⁰ Έτσι λοιπόν, περισσότερο επίσημοι και υποκειμενικοί τρόποι χρειάστηκε να επιστρατευτούν ώστε με ποσοτικές και επομένως πιο αξιόπιστες οδούς να οδηγούμαστε σε περισσότερο ασφαλή συμπεράσματα. Αυτές οι στατιστικές διερευνητικές μέθοδοι παρουσιάζονται παρακάτω.

3.2.1. Μέθοδος Megg-Mazumdar. Οι Begg και Mazumdar εισήγαγαν μια μη παραμετρική σταθμισμένη δοκιμασία, η οποία βασίζεται στη βαθμιαία συσχέτιση μεταξύ των εκτιμήσεων του αποτελέσματος των παρεμβάσεων (t) και των δειγματικών διακυμάνσεών τους (v), για να διερευνήσουν το γεγονός ότι το σφάλμα δημοσίευσης τείνει να δημιουργεί ένα είδος σχέσης μεταξύ αυτών των δύο χαρακτηριστικών.³³ Η συγκεκριμένη προσέγγιση είναι ελκυστική λόγω του γεγονότος ότι ως διαδικασία υπολογισμών αλλά και λογικής είναι απλή, αν και έχουν εγερθεί κάποιες ανησυχίες σχετικά με την πιθανή έλλειψη ισχύος της. Από τα αποτελέσματα λοιπόν που θα προκύψουν από τη διεξαγωγή μιας έρευνας εξαρτάται, σύμφωνα με αυτά που έχουν καταγραφεί μέχρι στιγμής, η πιθανότητα αυτή να επιλεγεί προς δημοσίευση ή όχι. Η εν λόγω πιθανότητα αναφέρεται από τους συγγραφείς ως λειτουργία βάρους. Το συγκεκριμένο άρθρο από το οποίο ελήφθησαν τα στοιχεία, ασχολήθηκε με δύο τύπους επιλογής: Μια λειτουργία βάρους που εξαρτάται μόνο από το t, το παρατηρούμενο αποτέλεσμα, και μια που εξαρτάται από το p, δηλαδή το

παρατηρούμενο επίπεδο σημαντικότητας p-value, όπου το p θεωρείται ως μια λειτουργία του $\frac{t}{v^{1/2}}$. Παρόμοιες έρευνες

με διαφορετικά παρατηρούμενα αποτελέσματα και διακυμάνσεις δημιουργήθηκαν, έως ότου επιλεγεί ένας αριθμός k ερευνών. Οι διακυμάνσεις λοιπόν για την εφαρμογή αυτής της μεθόδου πρέπει να είναι σταθεροποιημένες μέσω της τυποποίησης του μεγέθους του αποτελέσματος $\{t_i^*\}$ που

προηγήθηκε της δοκιμασίας και προσδιορίζεται από τον

εξής τύπο: $t_i^* = \frac{t_i - \bar{t}}{\sqrt{v_i^*}}$, όπου $\bar{t} = \frac{(\sum t_j v_j^{-1})}{\sum v_j^{-1}}$ είναι η συνήθης

εκτίμηση μέσω του μοντέλου σταθερών αποτελεσμάτων του περιληπτικού αποτελέσματος, ενώ $v_i^* = v_i - (\sum v_j^{-1})^{-1}$ είναι η διακύμανση του $(t_i - \bar{t})$.

Χρησιμοποίησαν λοιπόν για τη διεξαγωγή της δοκιμασίας τη μέθοδο της ταξινομημένης συσχέτισης Kendall's tau (τ). Αυτή περιλαμβάνει την αρίθμηση του πλήθους των ζευγών των ερευνών, οι οποίες ταξινομήθηκαν στην ίδια σειρά, σε αντιστοιχία με δύο παράγοντες (για παράδειγμα t^* και v). Εάν αυτός ο αριθμός είναι ο x, και εάν y είναι ο αριθμός των ζευγαριών των ερευνών που ταξινομήθηκαν στην αντίθετη σειρά, τότε η κανονική στατιστική δοκιμασία είναι:

$$z = \frac{(x-y)}{[k(k-1)(2k+5)/18]^{1/2}}$$

Είναι μια εμπειρική δοκιμασία βασισμένη στο γεγονός ότι οι σχεδιασμοί της κλίσης που αναμένεται να προκύψουν στο funnel plot οδηγούν σε συσχέτιση μεταξύ των εκτιμήσεων των αποτελεσμάτων και των διακυμάνσεών τους. Να αναφερθεί επίσης ότι επειδή η σειρά των τυπικών σφαλμάτων είναι ίδια με τη σειρά των διακυμάνσεων, η συγκεκριμένη δοκιμασία είναι ισοδύναμη με τη συσχέτιση μεταξύ του τυποποιημένου μεγέθους αποτελεσμάτων και των διακυμάνσεών τους.

3.2.2. Μέθοδος γραμμικής παλινδρόμησης του Egger. Οι Egger et al, προκειμένου να διαπιστώσουν εάν υπάρχει συσχέτιση μεταξύ της αυθεντικότητας των εκτιμήσεων των δοκιμασιών και της ακριβείας τους, διενήργησαν μια ανάλυση παλινδρόμησης ακολουθώντας την εξής σχέση:⁴¹

$$SND = a + b \times SE(d)^{-1}$$

Εξαρτημένη μεταβλητή στην προκειμένη μονο-παραγοντική παλινδρόμηση είναι το αποτέλεσμα της επίδρασης (diagnostic d) διαιρούμενο με το τυπικό του σφάλμα (SND: standard normal deviate) και ανεξάρτητη μεταβλητή η ακρίβεια (precision), το αντίστροφο του τυπικού σφάλματος (1/SE). Ο όρος a είναι το σημείο τομής και b είναι η κλίση. Ο σταθερός όρος a παρέχει το μέτρο ασυμμετρίας όσο μεγαλύτερη είναι η απόκλιση του από το μηδέν τόσο με-

γαλύτερη ασυμμετρία υφίσταται. Πιο συγκεκριμένα, θετικές τιμές $\alpha > 0$ υποδηλώνουν μια τάση προς υψηλότερα επίπεδα δοκιμαστικής ακρίβειας σε έρευνες με μικρότερα μεγέθη δείγματος,⁴² ενώ αρνητικές τιμές υποδεικνύουν ότι οι λιγότερο ακριβείς έρευνες έχουν περισσότερο έκδηλες επιδράσεις σε σχέση με τις πλέον ακριβείς, και αυτό ενδεχομένως οδηγεί στο συμπέρασμα της συνύπαρξης σφάλματος επιλογής. Εάν το παρατηρούμενο επίπεδο σημαντικότητας p -value του α είναι 0,10 ή μικρότερο, η ασυμμετρία θεωρείται στατιστικά σημαντική και γίνεται αποδεκτή.⁴³ Όσο μικρότερη είναι η δοκιμή, τόσο μικρότερη θα είναι η ακρίβειά της. Στην απουσία ασυμμετρίας του funnel plot, τα σημεία του γραφήματος που απεικονίζει την τυπική κανονική απόκλιση εναντίον της ακρίβειας θα εστιάζονται γύρω από μια γραμμή που θα περνάει από την αρχή των αξόνων ($SND=0$ και $\alpha=0$) και θα έχει κλίση β , η οποία θα προσδίδει το μέγεθος και την κατεύθυνση της επίδρασης.¹⁶

Η γραμμή της παλινδρόμησης, συνεπώς, δεν θα περνάει από τις πραγματικές τιμές, κι έτσι ο σταθερός όρος α θα παρέχει ένα μέτρο ασυμμετρίας (όσο μεγαλύτερη η απόκλιση του από το μηδέν τόσο πιο εμφανής είναι η ασυμμετρία). Η δοκιμασία λοιπόν του Egger ακολουθεί την ιδέα ότι με ένα στατιστικό έλεγχο της μηδενικής υπόθεσης:

$$H_0: \alpha=0,$$

δηλαδή δεν υπάρχει ασυμμετρία, καθίσταται δυνατόν να ελεγχθεί αν όντως ισχύει ή όχι και ανάλογα να γίνει αποδεκτή η ύπαρξη ασυμμετρίας ή να απορριφθεί. Να σημειωθεί ότι είναι απαραίτητο να πραγματοποιηθεί αμφίπλευρος έλεγχος του στατιστικού επιπέδου σημαντικότητας p -value.

Εάν υπάρχει ασυμμετρία, οι μικρές μελέτες θα δείχνουν επιδράσεις που θα διαφέρουν συστηματικά από τις μεγάλες μελέτες, οπότε η γραμμή παλινδρόμησης δεν θα περνάει από την αρχή των αξόνων. Να συμπληρωθεί ότι το παραπάνω γράφημα ανταποκρίνεται στο ακτινικό γράφημα, το οποίο περιγράφηκε από τον *Galbraith* και είναι ένα διμεταβλητό στικτόγραμμα (scatter plot) (x, y) του τυποποιημένου αιτιολογικού δείκτη t/s στον κατακόρυφο άξονα, έναντι της ακρίβειας (precision) $1/s$, όπου s είναι το τυπικό σφάλμα του αιτιολογικού δείκτη t , για κάθε μελέτη στον οριζόντιο άξονα.⁴⁴

Να προστεθεί επίσης ότι στο αρχικό άρθρο του Egger περιλαμβάνονταν μια εναλλακτική μορφή της σχετικής δοκιμασίας, όπου η παλινδρόμηση της μεταβλητής απόκρισης σε σχέση με την ανεξάρτητη (ακρίβεια) σταθμιζόταν με το αντίστροφο της διακύμανσης της επίδρασης, δηλαδή με $1/v$. Αυτή βέβαια η μορφή υπολειπόταν θεωρητικής τεκμηρίωσης, κι έτσι έπαψε να υποστηρίζεται από τους συγγραφείς. Σήμερα, συστήνεται η διεξαγωγή μόνο της αστάθμιστης παλινδρόμησης των δύο μεταβλητών.¹⁶ Η

αστάθμιστη παλινδρόμηση μπορεί να εμφανιστεί πανομοιότυπη με μια σταθμισμένη παλινδρόμηση μιας παρεμβατικής επίδρασης t πάνω σε τυπικό σφάλμα s : $E(t_i) = b + as_i$, όπου τα βάρη ($1/v_i$) είναι αντιστρόφως ανάλογα της διακύμανσης της υπολογιζόμενης επίδρασης της παρέμβασης σε κάθε έρευνα. Όσο μεγαλύτερη είναι η τιμή του συντελεστή b , τόσο πιο έντονη είναι η ένδειξη για επιρροή των μικρών ερευνών (small-study effect).⁴⁵ Να σημειωθεί επίσης ότι ο παράγοντας α στην παλινδρόμηση αντιπροσωπεύει την κλίση στη σταθμισμένη παλινδρόμηση μεταξύ του t_i και του s_i . Με αυτόν τον τρόπο, η δοκιμασία παλινδρόμησης για την ασυμμετρία του funnel plot ανταποκρίνεται σε μια δοκιμασία γραμμικής τάσης σε ένα τέτοιο γράφημα παρεμβατικής επίδρασης προς τυπικού σφάλματος. Η στάθμιση με το αντίστροφο της διακύμανσης λαμβάνει υπ' όψη της την ετερογένεια μεταξύ των μελετών, περιλαμβάνοντας μια πολλαπλασιαστική παράμετρο διασποράς ϕ , με την οποία η διακύμανση πολλαπλασιάζεται σε κάθε έρευνα.⁴⁶

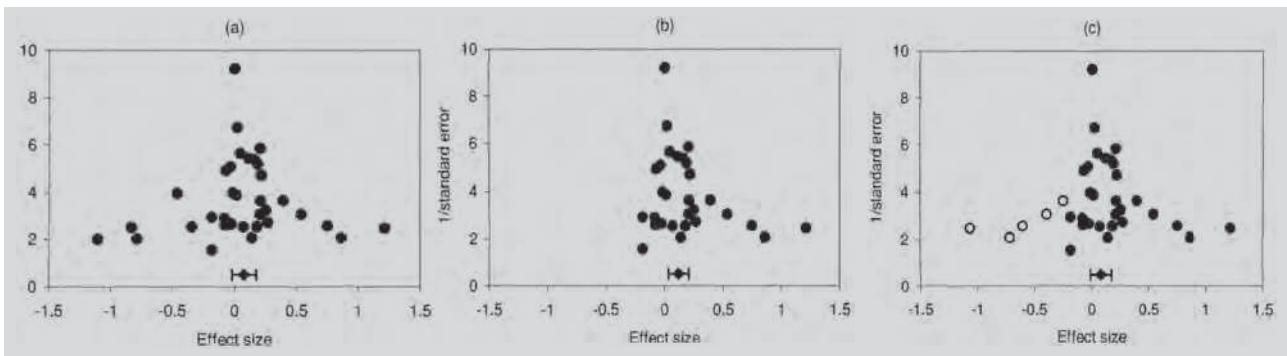
Η σχέση $t_i \sim N(t_i, \sigma_i^2)$ δηλώνει ότι η παρεμβατική επίδραση i μελετών ακολουθεί την κανονική κατανομή με μέσο t_i διακύμανση σ_i^2 . Το στατιστικό μοντέλο για αυτή την παλινδρόμηση είναι:

$$t_i \sim N(as_i + b, v_i\phi)$$

4. ΜΕΘΟΔΟΙ ΑΝΑΤΑΞΗΣ ΤΗΣ ΜΕΡΟΛΗΨΙΑΣ ΔΗΜΟΣΙΕΥΣΗΣ

4.1. Η μέθοδος Trim & Fill

Οι Duval και Tweedie περιέγραψαν ένα νέο τρόπο μέτρησης του μεγέθους του προβλήματος της μεροληψίας δημοσίευσης. Η προσέγγισή τους αποτελείται από δύο βήματα. Αρχικά, υπολογίζουν τον αριθμό των ερευνών που απουσιάζουν χρησιμοποιώντας μεθόδους βασισμένες σε παραδοχές συμμετρίας. Αυτές είναι απλές στο να τεθούν σε εφαρμογή, σε αντίθεση με άλλες υπάρχουσες μεθόδους, και στην πράξη εμφανίζονται να εντοπίζουν τις έρευνες που απουσιάζουν, οι οποίες υποδεικνύονται οπτικά μέσω του γραφήματος χοάνης (funnel plot). Σε δεύτερη φάση, εναποθέτουν τις τιμές που λείπουν χρησιμοποιώντας την προσέγγιση "trim & fill" (δηλαδή κόβοντας και επαναποθετώντας αυτές), η οποία καθιστά δυνατή τη λήψη εκτιμήσεων της συνολικής επίδρασης στα αποτελέσματα της μετα-ανάλυσης λόγω της μεροληψίας. Αντί να ελέγξουν απλά την πιθανότητα της ύπαρξης της μεροληψίας, ο αλγόριθμος της δεδομένης μεθόδου όπως διατύπωσαν στην έρευνά τους προσδίδει έναν τρόπο αξιολόγησης και επί πλέον προσαρμογής στη μεροληψία, ώστε κάποιος να μπορεί να γνωρίζει εάν το μέγεθος του προβλήματος χρήζει ή όχι προσοχής.⁴⁷



Εικόνα 3. Funnel plots από (a) 35 έρευνες προσομοίωσης, με ένα συνολικό μέγεθος επίδρασης 0,080 και διάστημα εμπιστοσύνης 95% με (-0,018, 0,178). (b) Οι 5 πιο αριστερά τοποθετημένες έρευνες απομακρύνονται με ένα συνολικό μέγεθος επίδρασης 0,124 με 95% CI (0,037, 0,210) και (c) τα «ξαναγεμισμένα» funnel plot με ένα συνολικό μέγεθος επίδρασης, υπολογισμένο με τη μέθοδο trim & fill, να ισούται με 0,082 και διάστημα εμπιστοσύνης 95% με (-0,011, 0,176).³⁶

Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο, ενδεχομένως η πλέον κοινή μέθοδος που χρησιμοποιείται για τον εντοπισμό της μεροληψίας σε μια μετα-ανάλυση είναι το γράφημα χοάνης. Προς υπενθύμιση, πρόκειται για ένα γράφημα το οποίο απεικονίζει την επίδραση που βρέθηκε σε κάθε έρευνα σε σχέση με το μέγεθός της, μετρημένα με ένα σταθερό τρόπο. Με τον τρόπο που επιλέγεται κάθε φορά να καθοριστεί η μονάδα του κάθε άξονα, επιτυγχάνεται σε κάθε περίπτωση η απεικόνιση των πλέον ακριβών εκτιμήσεων (αυτές από τις μεγαλύτερες έρευνες) στην κορυφή του γραφήματος, ενώ οι λιγότερο ακριβείς απεικονίζονται στη βάση. Θεωρητικά, το γράφημα πρέπει να σχηματίζει μια «χοάνη» γύρω από τον πραγματικό μέσο που προσπαθεί κάποιος να υπολογίσει. Το γεγονός ότι υπάρχει μεροληψία δημοσίευσης σε μια συγκεκριμένη βάση δεδομένων πιθανόν συνάγεται υποκειμενικά, καθώς το σχήμα το οποίο θα έπρεπε να μοιάζει με χοάνη είναι ασύμμετρο. Έτσι λοιπόν, όταν εμφανίζεται η έλλειψη κάποιων ερευνών στην κάτω γωνία, υπάρχει η υπόθεση ότι είτε λόγω εκδοτικής πολιτικής, είτε λόγω παρεμβάσεων του συγγραφέα, είτε λόγω άλλων παραγόντων οι οποίοι αναφέρθηκαν παραπάνω, οι συγκεκριμένες έρευνες (που, ας υποθεθεί, ότι δεν παρουσιάζουν σημαντικότητα ή ίσως το αντίθετο αποτέλεσμα από αυτό που στόχευαν οι συγγραφείς όταν άρχιζαν τις έρευνες) είναι αυτές που δεν δημοσιεύονται. Η οπτική εντύπωση της μεροληψίας σε ένα funnel plot είναι αμιγώς υποκειμενική. Η μέθοδος του trim & fill παρέχει μια περισσότερο αντικειμενική προσέγγιση βασισμένη στην τεκμηρίωση πίσω από την ερμηνεία ενός funnel plot.

Η μέθοδος trim & fill λοιπόν είναι μια μη παραμετρική μέθοδος εκτίμησης της μεροληψίας στις μετα-αναλύσεις. Η βασική ιδέα της μεθόδου είναι η προσθήκη ερευνών στο γράφημα, μέχρι εκείνο να καταστεί συμμετρικό. Πιο αναλυτικά, αποτελείται από τα εξής στάδια:⁴⁸

- Υπολογίζει τον αριθμό των ερευνών στο ασύμμετρο εξωτερικό τμήμα του funnel plot, χρησιμοποιώντας μεθόδους βασισμένες σε αναλογίες (rank-based)
- Απομακρύνει (trim) αυτές τις έρευνες και εφαρμόζει μετα-ανάλυση σε εκείνες που απέμειναν
- Η εκτίμηση που προκύπτει από την «ελλιμματική» μετα-ανάλυση θεωρείται ως πραγματικό κέντρο της χοάνης
- Για κάθε αφαιρούμενη έρευνα, δημιουργείται μια επιπρόσθετη έρευνα (filled) ως είδωλο αντικατοπτρισμού γύρω από το κέντρο του γραφήματος
- Διεξάγεται μετα-ανάλυση στις αυθεντικές και στις αφαιρούμενες έρευνες.

Για να παρουσιαστεί η επίδραση της μεροληψίας δημοσίευσης, στο γράφημα 3b απεικονίζεται το ίδιο γράφημα με το γράφημα 3a αλλά με τις κάτω αριστερά έρευνες να έχουν αποσιωπηθεί. Αυτό δείχνει το τυπικό σχέδιο το οποίο χρησιμοποιείται για την προβολή της μεροληψίας.³⁵

5. ΜΕΘΟΔΟΙ ΠΡΟΛΗΨΗΣ ΤΗΣ ΜΕΡΟΛΗΨΙΑΣ ΔΗΜΟΣΙΕΥΣΗΣ

5.1. Γκρίζα βιβλιογραφία

Η ταυτοποίηση παρόμοιων ερευνών μέσω μιας αμερόληπτης αναζήτησης αποτελεί καθοριστικό τμήμα της διαδικασίας μιας συστηματικής ανασκόπησης. Ο σκοπός μιας καλής αναζήτησης σε μια ανασκόπηση είναι να επιφέρει μια όσο το δυνατόν περισσότερο διεξοδική συλλογή ερευνών, τόσο των δημοσιευμένων όσο και των αδημοσίευτων, η οποία να είναι κατάλληλη για την απάντηση του ερευνητικού ερωτήματος που τέθηκε στην ανασκόπηση. Όμως, υπάρχει ένας αριθμός διαφορετικών τρόπων με τους οποίους μπορεί να εισαχθεί μεροληψία μέσα σε μια ανασκόπηση κατά τη

διάρκεια της διαδικασίας αναζήτησης. Οι ερευνητές εκείνοι που διεξάγουν συστηματικές ανασκοπήσεις οφείλουν να διασφαλίζουν μια κατά το δυνατόν πιο λεπτομερειακή έρευνα, ώστε να αποφύγουν την εισαγωγή μεροληψίας στην ανασκόπησή τους. Είναι σημαντικό το γεγονός ότι οι συγγραφείς της ανασκόπησης δεν περιορίζουν την ανασκόπησή τους στις έρευνες που είναι δημοσιευμένες στην αγγλική γλώσσα, καθώς τα στοιχεία δείχνουν ότι δοκιμές δημοσιευμένες σε γλώσσες εκτός της αγγλικής μπορεί να έχουν διαφορετικά αποτελέσματα από τις δημοσιευμένες στα Αγγλικά έρευνες.^{49,50} Είναι επίσης σημαντικό να μη βασίζονται αποκλειστικά στις αναφορές που παρατίθενται με βάση τις λέξεις-κλειδιά στις σχετικές έρευνες, εφ' όσον όσο περισσότερο έχει επικληθεί μια έρευνα, τόσο πιο εύκολο είναι να εντοπιστεί και να συμπεριληφθεί σε μια συστηματική ανασκόπηση.⁵¹

Η γκρίζα βιβλιογραφία γενικά περιλαμβάνει βιβλιογραφία που δεν έχει δημοσιευτεί στο παρελθόν, έχει περιορισμένη κατανομή ή δεν είναι διαθέσιμη μέσω των συμβατικών οδών.⁵² Παραδείγματα γκρίζας βιβλιογραφίας περιλαμβάνουν περιλήψεις συνεδρίων, ερευνητικές αναφορές, κεφάλαια βιβλίων, αδημοσίευτα δεδομένα, φοιτητικές εργασίες-θέσεις και προσωπική αλληλογραφία.

5.2. Ηλεκτρονικές πηγές γκρίζας βιβλιογραφίας

Μια προσπάθεια εύρεσης σχετικών ερευνών για ένα συγκεκριμένο θέμα συνήθως αρχίζει με την αναζήτηση σε ηλεκτρονικές βιβλιογραφικές βάσεις δεδομένων. Η διεθνής οργάνωση Cochrane Collaboration, η οποία συντελεί στη διαμόρφωση αποφάσεων σχετικά με παρεμβάσεις υγείας, έχει ασχοληθεί ευρέως με την ταυτοποίηση αναφορών τυχαιοποιημένων δοκιμών, όπου με αυτόν τον τρόπο ελαχιστοποιεί τον κίνδυνο εισαγωγής μεροληψίας στις συστηματικές ανασκοπήσεις. Η Κεντρική Καταγραφή Ελεγχόμενων Δοκιμών της Cochrane (CENTRAL), η οποία δημοσιεύεται ως τμήμα (προϊόν) της Cochrane Collaboration, είναι πλέον η πιο περιεκτική πηγή καταγραφών που σχετίζονται με ελεγχόμενες δοκιμές παρεμβάσεων υγείας.⁵³ Από τον Ιανουάριο του 2004, η συγκεκριμένη καταγραφή περιέχει περισσότερες από 400.000 παραθέσεις αναφορών τυχαιοποιημένων δοκιμών και άλλων ερευνών, ενδεχομένως σχετικών ώστε να περιληφθούν σε συστηματικές ανασκοπήσεις παρεμβάσεων υγείας.¹⁶

Υπάρχει ένας σεβαστός αριθμός διαθέσιμων ηλεκτρονικών πηγών, ο οποίος είναι σχετικά προσβάσιμος και μπορεί να αποδειχθεί χρήσιμος στην αναζήτηση της γκρίζας βιβλιογραφίας. Υπάρχουν τέσσερις μορφές-κλειδιά των ηλεκτρονικών αυτών πηγών: Βάσεις δεδομένων της γκρίζας βιβλιογραφίας, οι οποίες περιλαμβάνουν αποκλει-

στικά έρευνες που εντοπίζονται στην γκρίζα βιβλιογραφία, γενικές βάσεις βιβλιογραφικών δεδομένων, οι οποίες περιλαμβάνουν έρευνες δημοσιευμένες στην γκρίζα αλλά και στη συμβατική βιβλιογραφία, βάσεις συγκεκριμένων βιβλιογραφικών δεδομένων, οι οποίες περιλαμβάνουν έρευνες δημοσιευμένες στην γκρίζα αλλά και στη συμβατική βιβλιογραφία και ανήκουν σε συγκεκριμένη θεματολογία και, τέλος, καταγραφές δοκιμών σε εξέλιξη, οι οποίες περιλαμβάνουν πληροφορίες σχετικά με προγραμματισμένες και εν εξελίξει έρευνες.

Η διενέργεια ερευνών σε ηλεκτρονικές βιβλιογραφικές βάσεις δεδομένων, όπως οι Medline, Embase, Eric και PsychInfo, αποτελεί ένα σημαντικό τρόπο εντοπισμού ερευνών που πιθανόν να περιληφθούν σε μια συστηματική ανασκόπηση. Όμως, είναι σημαντικό μια έρευνα να μην περιοριστεί αποκλειστικά σε ηλεκτρονικές βάσεις δεδομένων, καθώς αυτό από μόνο του ενέχει την πιθανότητα εισαγωγής μεροληψίας σε μια ανασκόπηση. Μια πρόσφατη συστηματική ανασκόπηση ερευνών συνέκρινε τα αποτελέσματα της χειροκίνητης αναζήτησης με αυτά της αναζήτησης ηλεκτρονικών βάσεων δεδομένων, όπως το Medline και Embase που προαναφέρθηκαν, ώστε να ταυτοποιήσει αναφορές από τυχαιοποιημένες δοκιμές. Η χειροκίνητη αναζήτηση αναγνώρισε το 92–100% του συνολικού αριθμού των αναφορών που βρέθηκαν. Η μέθοδος Cochrane Highly Sensitive Search Strategy, η οποία χρησιμοποιείται για την αναγνώριση αναφορών τυχαιοποιημένων δοκιμών στη βάση Medline,⁵⁴ ταυτοποίησε το 80% του συνολικού αριθμού αναφορών των δοκιμών που βρέθηκαν. Ηλεκτρονικές αναζητήσεις της κατηγορίας «σύνθετες» εντόπισαν το 65% ενώ αυτές της κατηγορίας «απλές» εντόπισαν το 42%. Η ανάκτηση από μια ηλεκτρονική αναζήτηση ήταν περισσότερη όταν η αναζήτηση περιορίστηκε σε περιοδικά της αγγλικής γλώσσας: Ποσοστό 62% έναντι 39% στα περιοδικά που δημοσιεύουν σε άλλες γλώσσες. Ένας κύριος λόγος που αφορά στην αποτυχία των ηλεκτρονικών αναζητήσεων να εντοπίσουν όλες τις αναφορές των τυχαιοποιημένων δοκιμών ήταν ότι κάποιες αναφορές ήταν δημοσιευμένες ως περιλήψεις ή και παραρτήματα περιοδικών και δεν αναφέρονταν συστηματικά από τις ηλεκτρονικές βάσεις δεδομένων. Όταν η αναζήτηση περιοριζόταν σε πλήρεις αναφορές, η ανάκτηση ερευνών στις σύνθετες αναζητήσεις αυξανόταν στο 82%. Ένας επί πλέον λόγος της αποτυχίας των ηλεκτρονικών αναζητήσεων στην ταυτοποίηση όλων των αναφορών των τυχαιοποιημένων δοκιμών ήταν ότι κάποιες αναφορές δημοσιεύτηκαν πριν από το 1991, όταν το σύστημα της Εθνικής Βιβλιοθήκης Ιατρικής (National Library of Medicine) για την καταγραφή και την παράθεση αναφορών δοκιμών δεν ήταν τόσο πλήρως αναπτυγμένο, όπως σήμερα.⁵²

5.3. Αναζήτηση σε συνεδριακές καταχωρήσεις

Οι συνεδριακές καταχωρήσεις είναι επίσης μια σημαντική πηγή για την ταυτοποίηση ερευνών, οι οποίες θα συμπεριληφθούν σε μια συστηματική ανασκόπηση. Οι περιλήψεις αυτές συνήθως δεν περιλαμβάνονται σε ηλεκτρονικές βάσεις δεδομένων, ενώ σχετικές έρευνες ταυτοποιούνται μόνο μέσω αναζήτησης σε συνεδριακές καταχωρήσεις. Υπάρχουν ικανά στοιχεία ώστε να δείξουν ότι η αποτυχία στην πλήρη δημοσίευση μιας έρευνας, η οποία παρουσιάστηκε σε συνέδριο, είναι άμεσα συνδεδεμένη με τη σημαντικότητα των ευρημάτων της. Μια σχετικά πρόσφατη συστηματική ανασκόπηση 79 ξεχωριστών εκτιμήσεων αυτής της υπόθεσης έδειξε ότι το 44,5% των ερευνητικών μελετών, οι οποίες παρουσιάστηκαν με μορφή περιλήψεων σε επιστημονικά συνέδρια, ήταν πλήρως δημοσιευμένες. Αυτό το ποσοστό είναι υψηλότερο (63,1%) για τις περιλήψεις εκείνες που περιορίστηκαν μόνο στην παράθεση των αποτελεσμάτων των τυχαίοποιημένων δοκιμών. Οι περισσότερες έρευνες, και εφ' όσον αυτό εφαρμοζόταν, δημοσιεύονταν πλήρως μέσα σε 3 χρόνια από την παρουσίασή τους σε κάποια συνάντηση.⁵⁵ Αυτό σημαίνει ότι για ένα μεγάλο ποσοστό δοκιμών και των αποτελεσμάτων τους η μοναδική πηγή πληροφοριών μπορεί να είναι μια περιλήψη δημοσιευμένη στις καταχωρήσεις κάποιου συνεδρίου. Τα εν λόγω ευρήματα ισχύουν και για τομείς έρευνας εκτός της φροντίδας υγείας. Η περισσότερο κοινή εξήγηση που δόθηκε από τους ερευνητές ως προς το γιατί η έρευνα που υποβάλλει επιστημονικές συναθροίσεις δεν δημοσιεύεται πλήρως στη συνέχεια, είναι η έλλειψη του χρόνου. Στις αιτίες περιλαμβάνεται επίσης το γεγονός ότι οι ερευνητές πιστεύουν πως το περιοδικό δεν θα δημοσιεύσει την έρευνά τους, ή ακόμη και ότι οι ίδιοι οι συγγραφείς είναι της άποψης ότι τα ευρήματά τους δεν είναι αρκετά σημαντικά.^{56,57}

5.4. Άμεση επικοινωνία με τους συγγραφείς

Κάποιες έρευνες σε οποιαδήποτε μορφή είναι πιθανόν να μη δημοσιευτούν ποτέ, και αν κατανοηθεί ότι οι αδημοσίευτες έρευνες είναι επίσης πιθανόν να έχουν διαφορετικά αποτελέσματα από τις δημοσιευμένες, είναι σημαντικό να ταυτοποιηθούν και να συμπεριληφθούν στη διαδικασία της ανασκόπησης ώστε να ελαχιστοποιηθεί η μεροληψία.⁵⁸ Η επικοινωνία με τους ίδιους τους ερευνητές και με ειδικούς σε συγκεκριμένα αντικείμενα μπορεί να αποτελέσει σημαντική οδό ανίχνευσης ημιτελών ή και αδημοσίευτων ερευνών. Αυτό βέβαια στηρίζεται στην καλή βούληση των συγγραφέων και στο χρόνο που διαθέτουν στην ανταπόκριση στην παράκληση του ερευνητή. Δυστυχώς, υπάρχουν στοιχεία που υποστηρίζουν ότι παρακλήσεις για περαιτέρω πληροφορίες σχετικά με έρευνες δεν είναι πάντα επιτυχείς. Για την αντιμετώπιση του συγκεκριμένου

προβλήματος, κάποιες εταιρείες φαρμάκων καταγράφουν κάθε έρευνα σε εξέλιξη ή ολοκληρωμένη δοκιμή οι οποίες είναι διαθέσιμες για δημοσίευση, είτε μέσω της εφαρμογής Μετα-καταγραφή Τυχαίοποιημένων Δοκιμών είτε παρέχοντας καταγραφές ερευνών στις ιστοσελίδες τους.¹⁶

5.5. Αναζήτηση σε διαδικτυακούς τόπους

Το διαδίκτυο (Internet) αποτελεί μια σημαντική πηγή πληροφοριών σχετικά με τις εν εξέλιξη ή ολοκληρωμένες έρευνες, ειδικότερα αυτές που δεν έχουν δημοσιευτεί επίσημα. Παρ' όλα αυτά, η αναζήτηση στο διαδίκτυο μπορεί να αποτελέσει σημαντική ανάληψη. Αρκετές από τις μηχανές γενικής αναζήτησης δεν επιτρέπουν μια περισσότερο εξεζητημένη και συγκεκριμένη διερεύνηση, και κατά συνέπεια εντοπίζουν και προτείνουν χιλιάδες ιστότοπους προς έλεγχο. Μόνο λιγοστές έρευνες έχουν διεξαχθεί για την εξακρίβωση του κατά πόσο οι αναζητήσεις σε διεθνείς διαδικτυακούς τόπους είναι χρήσιμος τρόπος εντοπισμού περαιτέρω αδημοσίευτων ή σε εξέλιξη ερευνών.¹⁶

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο σκοπός της παρούσας ανασκοπικής έρευνας ήταν η καταγραφή αναλυτικά και με σαφήνεια αλλά και ο καθορισμός των κύριων αιτίων μεροληψίας κατά τη διαδικασία συλλογής των ερευνών, εξαγωγής δεδομένων από αυτές, καθώς και διασποράς των αποτελεσμάτων τα οποία εξαγονται στη διάρκεια της διαδικασίας δημοσίευσης και οι οποίες απειλούν την εγκυρότητα μιας μετα-ανάλυσης. Το φαινόμενο της μεροληψίας αποτελούσε ένα παραμελημένο και δυσνόητο θέμα, κυρίως ακαδημαϊκού ενδιαφέροντος, μέχρι πρόσφατα, όπου οι μετα-αναλύσεις και οι συστηματικές ανασκοπήσεις εισχώρησαν ενεργά στον ακαδημαϊκό κόσμο. Ο τεράστιος όγκος των συνδυασμένων δεδομένων δεν διευκόλυνε την ορθή εξαγωγή συμπερασμάτων, καθώς τα «αντικειμενικά» αναμενόμενα αποτελέσματα στηρίζονταν σε υποκειμενικούς τρόπους συλλογής δεδομένων, σύνθεσης και παρουσίασής τους.

Η μεροληψία δημοσίευσης δεν είναι απλά ένα φαινόμενο που καθίσταται εμφανές με τη συλλογή δεδομένων από τη βιβλιογραφία. Συναντάται επί πλέον σε αναφορές μεμονωμένων ερευνών σε όλα τα πεδία των επιστημών.⁵⁹ Μια ουσιώδης συνέπεια είναι η προσωρινή διάκριση που μπορεί να δοθεί σε έρευνες οι οποίες σταδιακά ανακαλύφθηκε μέσω διαδοχικών επαναληπτικών μελετών ότι παρουσιάζουν εσφαλμένως θετικά αποτελέσματα.⁴ Αυτό έχει αντίκτυπο στην ίδια την ιατρική πρακτική από ερευνητές οι οποίοι αποδίδουν στη βιβλιογραφία μεγάλη αξιοπιστία. Με τη σκέψη αυτή ως δεδομένη, εσφαλμένα αποτελέσματα και

ισχυρισμοί μπορεί να επηρεάσουν σημαντικά την ιατρική πρακτική, αν και τα ελλιπή στοιχεία που υπάρχουν στη διοχέτευση των αποτελεσμάτων μεμονωμένων ερευνών μέσα στην ιατρική πρακτική φαίνεται να δείχνουν ότι μεμονωμένες έρευνες έχουν περιορισμένο άμεσο αντίκτυπο.⁶⁰

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ την κα Δαφνή, Καθηγήτρια Βιοστατιστικής ΕΚΠΑ για την πολύτιμη συμβολή της στη διεξαγωγή της παρούσας ανασκόπησης.

ABSTRACT

The publication bias phenomenon in health sciences: Exploration of statistical methods of correction

O. ARCHANGELIDI

*Clinical Epidemiology Group, Department of Epidemiology and Public Health,
University College London, London, UK*

Archives of Hellenic Medicine 2013, 30(3):340–354

Publication bias is a term used to refer to a tendency towards the preparation, submission and publication of research findings based on the nature and direction of the research results. It is a phenomenon reflected in increased publication rates among studies that show a statistically significant effect of treatment, leading to a higher prevalence of false positive published results. The subjective component of publication bias involves the “biased” opinion of the investigator and is often reflected in the tone of the presentation, while the objective component is more subtle and depends on the fact that the decision to publish is often influenced by positive results of a study. The primary causes of objective publication bias are author and editorial preferences, in addition to selective analytical maneuvers and other subjective influences. The present review explores several statistical methods that have been developed to identify and quantify the impact of publication bias on meta-analyses. Basic strategies are discussed that if widely adopted would go a long way towards alleviating publication bias, including the search for, and inclusion of “gray” literature and prospective registration of studies.

Key words: Bias, Publication, Quality, Science, Study, Validity

Βιβλιογραφία

- HANLEY JA. The place of statistical methods in radiology (and in the bigger picture). *Invest Radiol* 1989, 24:10–16
- HENNEKENS C, BURING JE. *Analysis of epidemiological studies: Evaluating the role of bias*. Little Brown & Co, Boston, Mass, 1987:272–286
- BEGG CB, BERLIN JA. Publication bias: A problem in interpreting medical data. *J Roy Stat Soc* 1988, 151:419–463
- BEGG CB, BERLIN JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989, 81:107–115
- MELTON AW. Editorial. *J Exp Psychol* 1962, 64:553–557
- CHALMERSTC. Informed consent, clinical research and the practice of medicine. *Trans Am Clin Climatol Assoc* 1983, 94:204–212
- GREENWALD AG. Consequences of prejudice against the null hypothesis. *Psychol Bull* 1975, 82:1–20
- ΓΑΛΑΝΗΣ Π. Στατιστικές μέθοδοι ανάλυσης δεδομένων. *Αρχ Ελλ Ιατρ* 2008, 26:699–711
- DICKERSIN K, CHAN S, CHALMERS TC, SACKS HS, SMITH H JR. Publication bias and clinical trials. *Control Clin Trials* 1987, 8:343–353
- ROSENTHAL R. The “file drawer problem” and tolerance for null results. *Psychol Bull* 1979, 86:638–641
- HEDGES L, OLKIN I. *Statistical methods for meta-analysis*. Academic Press, New York, 1985
- EASTERBROOK PJ, BERLIN JA, GOPALAN R, MATTHEWS DR. Publication bias in clinical research. *Lancet* 1991, 337:867–872
- STERLING TD. Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *J Am Stat Assoc* 1959, 54:30–34
- HEDGES LV. Modelling publication selection effects in meta-analysis. *Stat Sci* 1992, 7:246–255
- COOPER H. *Synthesizing research*. Thousand Oaks, CA, Sage, 1998
- ROTHSTEIN H, SUTTON AJ, BORESTEIN M. *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley, Chichester, UK, 2005
- KOSHLAND DE Jr. Conflict of interest policy. *Science* 1992, 257:595
- MINERVA. Views. *Br Med J* 1983, 287:1886
- MAXWELL C. Clinical trials, reviews, and the Journal of Negative Results. *Br J Clin Pharmacol* 1981, 1:15–18
- DICKERSIN K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990, 263:1385–1389
- NEW YORK TIMES. 1986, C1–C7

22. DAVIDSON RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986, 1:155–158
23. NICHOLSON PA. Information for drug trial participants. *Lancet* 1987, ii:396
24. BROSS IDJ. *Scientific strategies to save your life: A statistical approach to primary prevention*. Marcel Dekker, New York, 1981
25. HIGGINSON J. Discussion of the paper by Begg and Berlin. *J Roy Stat Soc* (in press)
26. KOTELCHUCK D. Asbestos research: Winning the battle but losing the war. *Health/PAC Bull* 1974, 61:1–27
27. HEMMINKI E. Study of information submitted by drug companies to licensing authorities. *Br Med J* 1980, 280:833–836
28. MOERTEL CG. Improving the efficiency of clinical trials: A medical perspective. *Stat Med* 1984, 3:455–468
29. POCOCK SJ, SHAPER AG, PHILLIPS AN, WALKER M, WHITEHEAD TP. High density lipoprotein cholesterol is not a major risk factor for ischaemic heart disease in British men. *Br Med J (Clin Res Ed)* 1986, 292:515–519
30. BERLIN JA, BEGG CB, LOUIS TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989, 84:381–392
31. BEAUMONT JJ, BRESLOW NE. Power considerations in epidemiologic studies of vinyl chloride workers. *Am J Epidemiol* 1981, 114:725–734
32. SONG F, PAREKH S, HOOPER L, LOKE YK, RYDER J, SUTTON AJ ET AL. Dissemination and publication of research findings: An updated review of related biases. *Health Technol Assess* 2010, 14:iii, ix–xi, 1–193
33. BEGG CB, MAZUMDAR M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994, 50:1088–1101
34. ORWIN RG. A fail safe N for effect size in meta-analysis. *J Educ Stat* 1983, 8:157–159
35. LIGHT RJ, PILLEMER DB. *Summing-up: The science of reviewing research*. Harvard University Press, Cambridge, MA, 1984
36. LIGHT RJ, SINGER JD, WILLETT JB. *The visual presentation and interpretation of meta-analyses*. Russel Sage Foundation Publ, New York, 1994
37. BECKER BJ, MORTON SC. *Publication bias: Methods and practice in evidence-based medicine and social science*. Mathematical Science Research Institute, Berkley-California, 2002
38. STERNE JA, EGGER M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *J Clin Epidemiol* 2001, 54:1046–1055
39. FISHER RA. On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* 1921, 1:3–32
40. GREENLAND S. Invited commentary: A critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994, 140:290–296
41. EGGER M, DAVEY SMITH G, SCHNEIDER M, MINDER C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997, 315:629–634
42. SONG F, KHAN KS, DINNES J, SUTTON AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *Int J Epidemiol* 2002, 31:88–95
43. TANG JL, LIU JLY. Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 2000, 53:477–484
44. GALBRAITH RF. A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med* 1988, 7:889–894
45. STERNE JA, GAVAGHAN D, EGGER M. Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000, 53:1119–1129
46. THOMSON SG, SHARP SJ. Explaining heterogeneity in meta-analysis: A comparison of methods. *Stat Med* 1999, 18:2693–2708
47. DUVAL S, TWEEDIE R. A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *J Am Stat Assoc* 2000, 95:89–98
48. SCHWARZER G, CARPENTER J, RÜCKER G. Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *J Clin Epidemiol* 2010, 63:282–288
49. EGGER M, ZELLWEGER-ZÄHNERT, SCHNEIDER M, JUNKER C, LENGELER C, ANTES G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997, 350:326–329
50. JÜNI P, HOLENSTEIN F, STERNE J, BARTLETT C, EGGER M. Direction and impact of language bias in meta-analyses of controlled trials: Empirical study. *Int J Epidemiol* 2002, 31:115–123
51. GØTZSCHE PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)* 1987, 295:654–656
52. AUGER CP. *Information sources in grey literature (guides to information sources)*. 4th ed. Bowker Saur, London, 1998
53. DICKERSIN K, MANHEIMER E, WIELAND S, ROBINSON KA, LEFEBVRE C, McDONALD S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002, 25:38–64
54. DICKERSIN K, SCHERER R, LEFEBVRE C. Identifying relevant studies for systematic reviews. *Br Med J* 1994, 309:1286–1291
55. SCHERER RW, LANGENBERG P, VON ELM E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007, 2:MR000005
56. CALLAHAM ML, WEARS RL, WEBER EJ, BARTON C, YOUNG G. Positive outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1998, 280:254–257
57. WEBER EJ, CALLAHAM ML, WEARS RL, BARTON C, YOUNG G. Unpublished research from a medical specialty meeting: Why investigators fail to publish. *JAMA* 1998, 280:257–259
58. DICKERSIN K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997, 9(Suppl 1):15–21
59. LAIRD NM, MOSTELLER F. Discussion of the paper by Begg and Berlin. *J Roy Stat Soc* (in press)
60. FINEBERG HV. Clinical evaluation: How does it influence medical practice? *Bull Cancer* 1987, 74:333–346
61. LEUCHT S, KISSLING W, DAVIS JM. How to read and understand and use systematic reviews and meta-analyses. *Acta Psychiatrica Scand* 2009, 119:443–450

Corresponding author:

O. Archangelidi, Clinical Epidemiology Group, Department of Epidemiology and Public Health, University College London, Room 112, 1st Floor, 1–19 Torrington Place, London WC1E 7HN, UK
 e-mail: o.archangelidi@ucl.ac.uk