

ΕΦΑΡΜΟΣΜΕΝΗ ΙΑΤΡΙΚΗ ΕΡΕΥΝΑ APPLIED MEDICAL RESEARCH

Αναγκαίος αριθμός πασχόντων στις τυχαιοποιημένες κλινικές δοκιμές

1. Εισαγωγή
2. Τυχαιοποιημένες κλινικές δοκιμές
3. Μηδενική ή άκυρη υπόθεση
4. Παράγοντες καθορισμού του αναγκαίου αριθμού πασχόντων
 - 4.1. Προκαθορισμένο επίπεδο στατιστικής σημαντικότητας
 - 4.2. Στατιστική ισχύς
 - 4.3. «Μέγεθος του αποτελέσματος»
 - 4.4. Εκτιμώμενη μεταβλητότητα των μετρήσεων
 - 4.5. Μονόπλευρος ή αμφίπλευρος έλεγχος στατιστικής σημαντικότητας
5. Αναγκαίος αριθμός πασχόντων
 - 5.1. Σύγκριση μέσων τιμών
 - 5.2. Σύγκριση ποσοστών
6. Τρόποι μείωσης του αναγκαίου αριθμού πασχόντων
7. Σύνοψη

1. ΕΙΣΑΓΩΓΗ

Ένα εξαιρετικά σημαντικό ζήτημα κατά το σχεδιασμό των τυχαιοποιημένων κλινικών δοκιμών αποτελεί ο αναγκαίος αριθμός πασχόντων για την ανάδειξη μιας σχέσης ως στατιστικά σημαντικής. Η γνώση του αριθμού των πασχόντων που χρειάζονται έχει ιδιαίτερη σημασία και αποτελεί απαραίτητη προϋπόθεση για τη διεξαγωγή μιας τυχαιοποιημένης κλινικής δοκιμής (ΤΚΔ).¹ Αυτό φαίνεται και από τον ολοένα αυξανόμενο αριθμό περιοδικών που αφορούν στις επιστήμες υγείας και που απαιτούν η διεξαγωγή των ΤΚΔ να είναι σύμφωνη με τα διεθνή κατοχυρωμένα πρότυπα (consolidated standards of reporting trials, CONSORT),² τα οποία περιλαμβάνουν την αναλυτική αναφορά των μεθόδων που χρησιμοποιούνται για τον καθορισμό του αναγκαίου αριθμού πασχόντων. Επιπλέον, οι ΤΚΔ που αφορούν στην αποτελεσματικότητα φαρμακευτικών προϊόντων είναι εξαιρετικά δαπανηρές και, συνήθως, χρηματοδοτούνται από φαρμακευτικές εταιρείες που απαιτούν τον καθορισμό του μεγέθους του «δείγματος», έτσι ώστε να εξασφαλίζεται όσο το δυνατόν περισσότερο η εγκυρότητα μιας μελέτης. Το σημαντικότερο όμως είναι ότι ένα μικρό μέγεθος «δείγματος» μπορεί να οδηγήσει στο εσφαλμένο συμπέρασμα ότι δεν υπάρχει διαφορά στην αποτελεσματικότητα μεταξύ δύο, συνήθως, θεραπευτικών –ή σπανιότερα προληπτικών– παρεμβάσεων μολονότι στην

ΑΡΧΕΙΑ ΕΛΛΗΝΙΚΗΣ ΙΑΤΡΙΚΗΣ 2009, 26(4):544–557
ARCHIVES OF HELLENIC MEDICINE 2009, 26(4):544–557

Π. Γαλάνης

Εργαστήριο Οργάνωσης και
Αξιολόγησης Υπηρεσιών Υγείας, Τμήμα
Νοσηλευτικής, Πανεπιστήμιο Αθηνών,
Αθήνα

Number of patients needed
in randomized clinical trials

Abstract at the end of the article

Λέξεις ευρητηρίου

Μέγεθος αποτελέσματος
Στατιστική ισχύς
Σφάλμα τύπου I
Σφάλμα τύπου II
Τυχαιοποιημένες κλινικές δοκιμές

Υποβλήθηκε 26.6.2008
Εγκρίθηκε 14.7.2008

πραγματικότητα υπάρχει διαφορά. Μεγαλύτερος αριθμός πασχόντων από τον αναγκαίο, εξάλλου, αποτελεί περιττό κόστος, ενώ εγείρει και ηθικούς ενδοιασμούς, επειδή περισσότεροι πάσχοντες από τους απαιτούμενους εκτίθενται σε μια πιθανόν λιγότερο αποτελεσματική θεραπευτική παρέμβαση. Πρέπει να σημειωθεί ότι ο όρος «τυχαιοποιημένες» είναι περιττός, καθώς οι κλινικές δοκιμές εξ ορισμού περιλαμβάνουν την τυχαιοποίηση.

2. ΤΥΧΑΙΟΠΟΙΗΜΕΝΕΣ ΚΛΙΝΙΚΕΣ ΔΟΚΙΜΕΣ

Μια τυχαιοποιημένη κλινική δοκιμή (randomized clinical trial)^{3–9} αποτελεί πείραμα στο οποίο τα μελετώμενα άτομα πάσχουν ήδη από μια συγκεκριμένη πάθηση,* αλλά η πάθηση αυτή δεν αποτελεί τη μελετώμενη έκβαση. Η μελετώμενη έκβαση είναι η ίαση, ο θάνατος, η εμφάνιση επιπλοκών (π.χ. μετάσταση ενός καρκίνου) κ.ά. Σκοπός της ΤΚΔ είναι η συγκριτική εκτίμηση της αποτελεσματικότητας μιας θεραπευτικής παρέμβασης. Οι μελετώμενες παρεμβάσεις σε μια ΤΚΔ δεν είναι πρωτογενή προληπτικά μέτρα, καθώς δεν χρησιμοποιούνται για την πρόληψη της εμφάνισης μιας πάθησης, αλλά για την πρόληψη των επιπλοκών

* Με τον όρο «πάθηση» αποδίδεται ο ξένος όρος "illness". Οι παθήσεις (illnesses) ταξινομούνται σε νοσήματα (diseases), μειονεκτήματα ή ελαττώματα (defects) και κακώσεις (injuries).

μιας πάθησης. Μια τροποποιημένη δίαιτα, π.χ., μετά από έμφραγμα του μυοκαρδίου μπορεί να λειτουργήσει προληπτικά για την εμφάνιση ενός δεύτερου εμφράγματος, ενώ τα χημειοθεραπευτικά φάρμακα λειτουργούν προληπτικά στην επανεμφάνιση ενός καρκίνου.

Η είσοδος των πασχόντων στη μελέτη πρέπει να γίνεται σε σημείο που να επιτρέπει την εφαρμογή της θεραπευτικής παρέμβασης για εύλογο χρονικό διάστημα. Πρέπει να αποκλείονται από τη μελέτη οι πάσχοντες που βρίσκονται σε αρχικό ή τελικό στάδιο της πάθησης, οπότε δεν επιτρέπεται η εφαρμογή της μελετώμενης θεραπευτικής παρέμβασης. Η κατανομή της θεραπευτικής παρέμβασης πρέπει να σχεδιάζεται με σκοπό την ελαχιστοποίηση της μεταβλητότητας των εξωγενών χαρακτηριστικών (συγχυτών*) που ενδεχομένως επηρεάζουν τη σύγκριση. Εάν η νέα θεραπεία, για παράδειγμα, χορηγείται σε πάσχοντες που βρίσκονται σε πιο προχωρημένο στάδιο της πάθησης, τότε αυξάνεται σημαντικά ο κίνδυνος για μη έγκυρη εκτίμηση της νέας θεραπείας. Ο καλύτερος τρόπος για να επιτευχθεί η συγκρισιμότητα είναι η ενεργητική τυχαία κατανομή των μελετώμενων ατόμων.^{4,5,19,20}

Χωρίς τη διεξαγωγή μιας ΤΚΔ, αποτελεί πρόκληση ο σχεδιασμός μιας μελέτης που να επιτρέπει την έγκυρη σύγκριση αυτών που λαμβάνουν ένα φάρμακο σε σχέση με εκείνους που λαμβάνουν κάποιο άλλο. Το πρόβλημα συνίσταται κυρίως στο φαινόμενο που αναφέρεται ως «σύγχυση λόγω ένδειξης» (confounding by indication),^{4,9} καθώς αυτοί που λαμβάνουν ένα φάρμακο γενικά διαφέρουν από εκείνους που λαμβάνουν κάποιο άλλο σύμφωνα με την ιατρική ένδειξη για την οποία το φάρμακο χορηγείται. Ακόμη και αν η συγκρινόμενη ομάδα περιλαμβάνει πάσχοντες με την ίδια πάθηση που λαμβάνουν μια διαφορετική θεραπεία, θα υπάρχουν διαφορές στη σοβαρότητα της πάθησης ή σε άλλους εξωγενείς προσδιοριστές μεταξύ των ομάδων που λαμβάνουν διαφορετικές θεραπείες. Οι διαφορές αυτές εισάγουν συστηματικό σφάλμα στη μελέτη, που ονομάζεται «σύγχυση λόγω ένδειξης».

Η «σύγχυση λόγω ένδειξης» είναι ένα συστηματικό σφάλμα που οφείλεται στις διαφορές στην πρόγνωση μεταξύ πασχόντων που λαμβάνουν διαφορετικές θεραπευτικές αγωγές και μπορεί να εξουδετερωθεί μόνο με τις ΤΚΔ.

Για παράδειγμα, ένα νέο αντιβιοτικό θεωρείται ότι είναι αποτελεσματικότερο στην αντιμετώπιση των ανθεκτικών βακτηριακών στελεχών που προκαλούν μηνιγγίτιδα, αλλά εμφανίζει ανεπιθύμητες ενέργειες και κοστίζει περισσότερο. Λογικά, το νέο αυτό αντιβιοτικό χορηγείται σε πάσχοντες που έχουν μεγαλύτερο κίνδυνο εμφάνισης μιας θανατηφόρας έκβασης. Ακόμη και αν το νέο αντιβιοτικό είναι εξαιρετικά αποτελεσματικό, η θνητότητα αυτών που το λαμβάνουν θα είναι μεγαλύτερη από τη θνητότητα εκείνων που λαμβάνουν τη συνήθη θεραπευτική αγωγή, επειδή αυτοί που λαμβάνουν το νέο αντιβιοτικό βρίσκονται σε υψηλότερο κίνδυνο εμφάνισης του θανάτου. Η έγκυρη εκτίμηση της θεραπευτικής αποτελεσματικότητας του νέου αντιβιοτικού μπορεί να επιτευχθεί μόνο εάν οι διαφορές όσον αφορά στην πρόγνωση μπορούν να ελεγχθούν ή, με άλλη διατύπωση, να εξουδετερωθούν στη σύγκριση.

Στις μη πειραματικές μελέτες, ακόμη και όταν η πληροφορία για τους δυνητικούς συγχυτές είναι επαρκής, δεν είναι δυνατόν να εξουδετερωθεί πλήρως η «σύγχυση λόγω ένδειξης». Για το λόγο αυτόν απαιτείται η διεξαγωγή ΤΚΔ στις οποίες η τυχαία κατανομή των μελετώμενων ατόμων έχει ως αποτέλεσμα να εξισώνονται οι δυνητικοί συγχυτές μεταξύ των ομάδων που λαμβάνουν διαφορετικές θεραπείες.

Το πλέον σημαντικό σε μια ΤΚΔ είναι οι πάσχοντες να κατανέμονται στις δύο –ή σπανιότερα περισσότερες– μελετώμενες ομάδες με τέτοιο τρόπο, ώστε να μεγιστοποιείται η πιθανότητα των δύο ομάδων να είναι παρόμοιες όσον αφορά στους πιθανούς συγχυτές. Ο στόχος αυτός επιτυγχάνεται με την *τυχαιοποίηση* (randomization),^{8,9,21–23} στην οποία η κατανομή κάθε πάσχοντα σε μία από τις δύο ομάδες είναι ανεξάρτητη από τους υπόλοιπους πάσχοντες και επιπλέον δεν επηρεάζεται από τις προτιμήσεις του ιατρού ή του πάσχοντα. Στην περίπτωση που υπάρχουν δύο μελετώμενες θεραπευτικές παρεμβάσεις με ισοδύναμο αριθμό ατόμων σε κάθε ομάδα, τότε η πιθανότητα κάθε πάσχοντα να ανήκει σε μία από τις δύο ομάδες είναι 50%. Η τυχαία αυτή κατανομή των μελετώμενων πασχόντων μπορεί να επιτευχθεί απλά με τη ρήψη ενός νομίσματος, χωρίς βεβαίως να αποτελεί και τη μοναδική μέθοδο τυχαιοποίησης.⁸

Στην πλειονότητα των περιπτώσεων και ιδιαίτερα όταν ο αριθμός των μελετώμενων ατόμων είναι μεγάλος, χρησιμοποιείται μια διαφορετική, μη φυσική μέθοδος τυχαιοποίησης, που είναι ο *πίνακας τυχαίων αριθμών* (random numbers table).⁸ Είναι προτιμότερο, πάντως, να καλούνται ψευδο-τυχαίοι αριθμοί, καθώς δημιουργούνται μέσω μιας μαθηματικής διαδικασίας. Οι σειρές των τυχαίων αριθμών είναι διαθέσιμες σε πίνακες ή μπορούν να δημιουργηθούν με τη βοήθεια ηλεκτρονικών υπολογιστών.

* Οι εξωγενείς προσδιοριστές (extraneous determinants) ονομάζονται δυνητικοί συγχυτές (potential confounders), όταν σχετίζονται με τη συχνότητα εμφάνισης της μελετώμενης έκβασης. Όταν οι εξωγενείς αυτοί προσδιοριστές ανισοκατανέμονται στις κατηγορίες του μελετώμενου προσδιοριστή, τότε ονομάζονται πραγματικοί συγχυτές (actual confounders).¹⁰ Οι συγχυτές^{9,11–18} μπορούν, εφόσον διαπιστωθούν, να εξουδετερωθούν (ή ελεγχθούν) κατά την ανάλυση των δεδομένων, κάτι που δεν μπορεί να γίνει στα συστηματικά σφάλματα επιλογής ή πληροφορίας.

Στις ΤΚΔ, η ομοιότητα των συγκρινόμενων ομάδων –που επιτυγχάνεται με την τυχαιοποίηση– μπορεί να ανατραπεί, αν οι πάσχοντες που ανήκουν στις ομάδες αυτές ή οι ιατροί που αξιολογούν τα αποτελέσματα των συγκρινόμενων θεραπευτικών παρεμβάσεων διαφοροποιούν συνειδητά ή υποσυνείδητα τη συμπεριφορά τους (οι πάσχοντες) ή την κρίση τους (οι ιατροί), ανάλογα με την ομάδα στην οποία ανήκουν οι πάσχοντες.³ Για να αποφευχθούν οι συνέπειες από τις υποκειμενικές προκαταλήψεις ή αντιδράσεις των πασχόντων ή των ιατρών, οι ΤΚΔ διεξάγονται κατά κανόνα με τρόπο «διπλά-τυφλό» (double-blind study).

«Τυφλή» είναι η μέθοδος κατά την οποία η κατανομή της θεραπευτικής παρέμβασης είναι άγνωστη. Όταν η κατανομή της θεραπευτικής παρέμβασης είναι άγνωστη στους πάσχοντες αλλά γνωστή στους ιατρούς τότε η μελέτη ονομάζεται «μονά-τυφλή» (single-blind study), ενώ όταν η κατανομή της θεραπευτικής παρέμβασης είναι άγνωστη τόσο στους πάσχοντες όσο και στους ιατρούς τότε η μελέτη καλείται «διπλά-τυφλή». Σε μια «διπλά-τυφλή» μελέτη, η κατανομή της θεραπευτικής παρέμβασης αποκαλύπτεται μόνον όταν υπάρχουν σοβαρές ανεπιθύμητες ενέργειες ή όταν ολοκληρωθεί η μελέτη.

Αν και οι «διπλά-τυφλές» ΤΚΔ είναι οι πλέον επιθυμητές, υπάρχουν περιπτώσεις στις οποίες δεν είναι δυνατόν να επιτευχθούν, καθώς οι διαφορές στις θεραπευτικές παρεμβάσεις είναι προφανείς, όπως η χειρουργική παρέμβαση, για παράδειγμα, έναντι της φαρμακευτικής για την αντιμετώπιση της στεφανιαίας πάθησης ή η χημειοθεραπεία έναντι της χειρουργικής θεραπείας για την αντιμετώπιση του καρκίνου του πνεύμονα.^{3,23} Υπάρχουν, βέβαια, και περιπτώσεις στις οποίες η μελετώμενη έκβαση είναι τελείως αντικειμενική (π.χ. ο θάνατος), οπότε δεν υφίσταται λόγος τυφλού ελέγχου από την πλευρά του ερευνητή.

Πρέπει, πάντως, να σημειωθεί ότι, επειδή στην κλινική πράξη δεν είναι ηθικά επιτρεπτό να αγνοεί ο θεράπων ιατρός το είδος της θεραπευτικής παρέμβασης που εφαρμόζεται, κρίνεται σκόπιμο να υπάρχουν δύο ομάδες ιατρών. Η μία ομάδα θα αφορά στην περίθαλψη των πασχόντων και θα είναι άσχετη προς τη μελέτη, ενώ η άλλη θα αξιολογεί με «τυφλό» τρόπο τα αποτελέσματα των συγκρινόμενων θεραπευτικών παρεμβάσεων. Η κατάσταση διευκολύνεται σημαντικά, όταν η αξιολόγηση του θεραπευτικού αποτελέσματος βασίζεται σε κάποιο αντικειμενικό ή ημι-αντικειμενικό εύρημα (π.χ. καρδιογράφημα, ακτινογραφία κ.ά.).³

Οι ΤΚΔ εξ ορισμού δημιουργούν ηθικούς προβληματισμούς, καθώς διαδικασίες που εμπλέκονται σ' αυτές, όπως η τυχαιοποίηση, ο «τυφλός» έλεγχος και η εφαρμογή ανενεργής ουσίας, εγείρουν ηθικά διλήμματα.⁹ Οι ΤΚΔ, όμως, είναι ο καλύτερος τρόπος τεκμηρίωσης της απο-

τελεσματικότητας μιας νέας παρέμβασης. Η καθιέρωση νέων θεραπευτικών μέτρων και η απομυθοποίηση άλλων οφείλονται σε δεδομένα από ΤΚΔ. Πέρασε μισός αιώνας, π.χ., προτού διαπιστωθεί με ΤΚΔ ότι το συντηρημένο σε φαινόλη εμβόλιο του τυφοειδούς είναι πολύ καλύτερο από το συντηρημένο σε αλκοόλη εμβόλιο, που χρησιμοποιείτο συστηματικά και ευρύτατα στις περισσότερες χώρες.

Οι ηθικές αρχές που οφείλουν να διέπουν τη διεξαγωγή μιας ΤΚΔ θα πρέπει να απασχολούν τους ερευνητές από τη φάση του σχεδιασμού έως και τη δημοσιοποίηση των αποτελεσμάτων. Η διακήρυξη του Helsinki²⁴ αναγνωρίζεται διεθνώς ως η βάση για τα ηθικά ζητήματα, τα οποία προκύπτουν κατά τη διεξαγωγή μελετών σε ανθρώπινους πληθυσμούς. Η διακήρυξη αυτή απαιτεί την ύπαρξη ενός ερευνητικού πρωτοκόλλου, το οποίο πρέπει να περιγράφει τους ηθικούς προβληματισμούς που εμπλέκονται στη μελέτη, καθώς επίσης και αν είναι σύμφωνο με τις αρχές που περιγράφονται στη διακήρυξη.⁹ Επιπλέον, το πρωτόκολλο θα πρέπει, εφόσον έχει ολοκληρωθεί, να δοθεί προς εκτίμηση, σχολιασμό, καθοδήγηση και, όπου είναι απαραίτητο, έγκριση σε μια ειδικά οριζόμενη επιτροπή διασφάλισης των ηθικών προδιαγραφών (Ethical Committee).

Τα κριτήρια και η διαδικασία επιλογής των πασχόντων που συμμετέχουν σε μια ΤΚΔ πρέπει να καθορίζονται προσεκτικά, να συμπεριλαμβάνονται στο ερευνητικό πρωτόκολλο και να τηρούνται με συνέπεια. Τα κριτήρια δεν πρέπει να είναι πολύ αυστηρά, γιατί τότε η επιλογή των πασχόντων αφορά σε ένα περιορισμένο τμήμα του νοσολογικού φάσματος, ούτε πολύ ελαστικά, επειδή τότε συμπεριλαμβάνονται πάσχοντες με έκδηλη διαγνωστική ανομοιογένεια.³ Κρίνεται σκόπιμο, το «δείγμα» των μελετώμενων πασχόντων να είναι αντιπροσωπευτικό της αντίστοιχης πάθησης, αλλά αυτό δεν είναι πάντοτε εφικτό.

Ο τρόπος με τον οποίο επιλέγονται οι πάσχοντες σε μια ΤΚΔ μπορεί να επηρεάσει τη μελετώμενη έκβαση. Στην κλινική πράξη, συνήθως, επιλέγονται πάσχοντες που είναι ευκολότερο να συμμετάσχουν σε μια ΤΚΔ. Για παράδειγμα, σε μια μελέτη που διερευνά την αποτελεσματικότητα διαφορετικών θεραπειών εμφράγματος του μυοκαρδίου, τα μελετώμενα άτομα είναι πάσχοντες που βρίσκονται στο νοσοκομείο.

Η επιλογή των πασχόντων είναι εξαιρετικής σημασίας για την ερμηνεία των αποτελεσμάτων. Είναι αρκετά παράτολμο, π.χ., να θεωρηθεί ότι το ποσοστό επιβίωσης μιας ομάδας πασχόντων με έμφραγμα του μυοκαρδίου που λαμβάνουν μια θεραπευτική αγωγή και νοσηλεύονται στην Αθήνα είναι ακριβώς το ίδιο με το ποσοστό επιβίωσης μιας ομάδας πασχόντων που λαμβάνουν την ίδια θεραπευτική αγωγή αλλά νοσηλεύονται στην Κρήτη. Και αυτό γιατί οι

δύο αυτές ομάδες πασχόντων μπορεί να διαφέρουν σε αρκετά χαρακτηριστικά, όπως π.χ. οι διατροφικές συνήθειες, που επηρεάζουν την κατάσταση των αρτηριών τους και κατ'επέκταση την πρόγνυσή τους. Αναμένεται, πάντως, ότι εάν η θεραπευτική αγωγή μειώνει τη θνητότητα των πασχόντων στην Αθήνα, τότε θα μειώνει και τη θνητότητα των πασχόντων στην Κρήτη. Σε ορισμένες ακραίες περιπτώσεις, ωστόσο, τα αποτελέσματα μπορεί να διαφέρουν σημαντικά. Πιο συγκεκριμένα, καλά σχεδιασμένες ΤΚΔ που διεξήχθησαν στο Ηνωμένο Βασίλειο έδειξαν ότι το εμβόλιο BCG μειώνει τη συχνότητα εμφάνισης της φυματίωσης σε παιδιά, ενώ παρόμοιες μελέτες που διεξήχθησαν στην Ινδία έδειξαν ότι το BCG ήταν πολύ λιγότερο αποτελεσματικό.⁸ Το εύρημα αυτό μάλλον οφειλόταν στο σημαντικά διαφορετικό βαθμό έκθεσης στο μυκοβακτηρίδιο της φυματίωσης μεταξύ των δύο πληθυσμών.

Τα κριτήρια εισόδου και αποκλεισμού των πασχόντων από μια ΤΚΔ πρέπει να στοχεύουν στη μείωση της μεταβλητότητας μεταξύ των πασχόντων. Και αυτό γιατί όσο μικρότερη είναι η μεταβλητότητα τόσο μεγαλύτερη είναι η πιθανότητα ανάδειξης μιας διαφοράς μεταξύ θεραπευτικών παρεμβάσεων ως στατιστικά σημαντικής, εάν βέβαια υπάρχει διαφορά. Το γεγονός αυτό υποδηλώνει ότι είναι επιθυμητή η όσο το δυνατόν μεγαλύτερη ομοιομορφία μεταξύ των πασχόντων. Ιδιαίτερα σημαντική είναι η ομοιομορφία των πασχόντων όσον αφορά στη βαρύτητα της πάθησης. Σε μια μελέτη, π.χ., για την εκτίμηση της αποτελεσματικότητας μιας νέας θεραπευτικής παρέμβασης για τον καρκίνο του μαστού θα πρέπει όλες οι γυναίκες να βρίσκονται στο ίδιο στάδιο της πάθησης. Σε ορισμένες περιπτώσεις, εξάλλου, διεξάγονται ΤΚΔ που αφορούν σε συγκεκριμένες ηλικιακές ομάδες, καθώς επίσης και δοκιμές στις οποίες συμμετέχουν μόνο άνδρες ή μόνο γυναίκες. Εάν όμως μια θεραπευτική παρέμβαση βρεθεί ασφαλής και αποτελεσματική στους νέους, δεν σημαίνει απαραίτητα ότι μπορεί να χρησιμοποιηθεί και στους υπερήλικες. Απαιτείται, προφανώς, η διεξαγωγή ΤΚΔ στις οποίες θα συμμετέχουν υπερήλικες.

3. ΜΗΔΕΝΙΚΗ Ή ΑΚΥΡΗ ΥΠΟΘΕΣΗ

Κατά το σχεδιασμό μιας ΤΚΔ, ο καθορισμός μίας ή περισσότερων μηδενικών υποθέσεων που πρόκειται να υποβληθούν στον κατάλληλο εμπειρικό έλεγχο είναι πρωταρχικής σημασίας. Μια στατιστική υπόθεση αποτελεί πρόβλεψη της σχέσης μεταξύ δύο ή περισσότερων χαρακτηριστικών.^{25,26} Στις ΤΚΔ, η στατιστική υπόθεση αφορά στη σχέση μεταξύ ενός προσδιοριστή και μιας έκβασης. Ο προσδιοριστής* είναι οι διάφορες θεραπευτικές παρεμβάσεις που εφαρμόζονται, ενώ η έκβαση μπορεί να είναι η ίαση, ο θάνατος, η εμφάνιση καταλοίπων κ.ά. Η στατιστική υπόθεση διακρί-

νεται στη μηδενική ή άκυρη υπόθεση (null hypothesis) και στην εναλλακτική (alternative).¹⁵ Σύμφωνα με τη μηδενική υπόθεση, δεν υπάρχει σχέση μεταξύ προσδιοριστή και συχνότητας εμφάνισης της έκβασης. Η διεξαγωγή του κατάλληλου στατιστικού ελέγχου προϋποθέτει ότι η μηδενική υπόθεση είναι αληθής, εκτός και αν βρεθούν ενδείξεις για το αντίθετο. Πρέπει να σημειωθεί ότι η μη απόρριψη της μηδενικής υπόθεσης δεν σημαίνει αποδοχή της, αλλά απλά ότι τα δεδομένα της συγκεκριμένης μελέτης δεν επαρκούν για την απόρριψή της.

Μια ΤΚΔ διερευνά την αποτελεσματικότητα της στρεπτοκινάσης και του ενεργοποιητή του ιστικού προδρόμου της πλασμίνης** στη θεραπεία του εμφράγματος του μυοκαρδίου. Ο μελετώμενος προσδιοριστής είναι η θεραπευτική παρέμβαση, ενώ η έκβαση είναι ο θάνατος στις πρώτες 30 ημέρες μετά από την προσβολή από το έμφραγμα του μυοκαρδίου. Το υπολογιζόμενο μέτρο συχνότητας είναι η θνητότητα στις πρώτες 30 ημέρες, οπότε, σύμφωνα με τη μηδενική υπόθεση, η θνητότητα στους εμφραγματίες που χορηγείται στρεπτοκινάση ισούται με τη θνητότητα σ' εκείνους που χορηγείται ενεργοποιητής του ιστικού προδρόμου της πλασμίνης. Η εναλλακτική υπόθεση είναι ότι οι θνητότητες αυτές διαφέρουν μεταξύ τους.

Έπειτα διεξάγεται ο κατάλληλος στατιστικός έλεγχος για τον έλεγχο της μηδενικής υπόθεσης, έτσι ώστε να διαπιστωθεί αν υπάρχει στατιστικά σημαντική διαφορά ανάμεσα στις δύο υπολογιζόμενες θνητότητες. Η εύρεση

* Παράγοντας κινδύνου (risk factor) ή έκθεση (exposure) ή προσδιοριστής (determinant), όπως τελικά επικράτησε να λέγεται σήμερα, είναι το χαρακτηριστικό (συγγενές, περιβαλλοντικό ή συμπεριφοράς) των ατόμων από το οποίο εξαρτάται (σχετίζεται ή συναρτάται) η συχνότητα εμφάνισης της μελετώμενης έκβασης (η οποία στην αιτιογνωστική έρευνα είναι η εμφάνιση της πάθησης).^{10,11,15,18} Ο προσδιοριστής της συχνότητας εμφάνισης μιας έκβασης περιλαμβάνει δύο κατηγορίες, την ενδεικτική κατηγορία (index category) και την κατηγορία αναφοράς (reference category). Η επιλογή της κατηγορίας αναφοράς του προσδιοριστή είναι μείζονος σημασίας στο σχεδιασμό της μελέτης. Είναι κρίσιμης σημασίας να γίνει αντιληπτό ότι η κατηγορία αναφοράς δεν είναι συμφυές χαρακτηριστικό του προσδιοριστή, αλλά αποτέλεσμα επιλογής του ερευνητή, άρα συνέπεια του σχεδιασμού της μελέτης. Από την επιλογή της κατηγορίας αναφοράς εξαρτάται η ερμηνεία των αποτελεσμάτων μιας μελέτης. Για παράδειγμα, προσδιοριστής της συχνότητας εμφάνισης της νεφρικής πάθησης δεν είναι η αρτηριακή υπέρταση, αλλά η αρτηριακή πίεση. Η αρτηριακή υπέρταση είναι μια κατηγορία και συνήθως η ενδεικτική κατηγορία του προσδιοριστή, στην οποία μελετάται η συχνότητα εμφάνισης της νεφρικής πάθησης σε σχέση πάντοτε με τη συχνότητα εμφάνισης της στην κατηγορία αναφοράς, στην προκειμένη περίπτωση στην κατηγορία των ατόμων που δεν έχουν αρτηριακή υπέρταση.

** Ο ενεργοποιητής του ιστικού προδρόμου της πλασμίνης κοστίζει αρκετά περισσότερο από τη στρεπτοκινάση, αλλά παρουσιάζει δύο σημαντικά πλεονεκτήματα έναντι αυτής. Πιο συγκεκριμένα, η στρεπτοκινάση είναι μια πρωτεΐνη που δεν προέρχεται από τον ανθρώπινο οργανισμό και για το λόγο αυτόν προκαλεί συχνότερα αλλεργικές αντιδράσεις και επιπλέον οδηγεί συχνότερα σε αιμορραγία σε σχέση με τον ενεργοποιητή του ιστικού προδρόμου της πλασμίνης.

στατιστικά σημαντικής διαφοράς υποδηλώνει ότι υπάρχει σχέση μεταξύ της θεραπευτικής παρέμβασης και της θνητότητας σύμφωνα με τα δεδομένα της συγκεκριμένης μελέτης. Στην περίπτωση αυτή, όμως, τίθεται το ερώτημα ποιος είναι ο αριθμός πασχόντων που χρειάζεται η συγκριτική αυτή μελέτη για να αναδειχθεί μια στατιστικά σημαντική σχέση; Η γνώση του αναγκαίου αριθμού πασχόντων πριν από τη διεξαγωγή μιας ΤΚΔ είναι πολύ σημαντική, καθώς επιτρέπει στους ερευνητές να γνωρίζουν αν μπορούν να συγκεντρώσουν τους πάσχοντες που απαιτούνται ή αν πρέπει να συνεργαστούν και με άλλες ερευνητικές ομάδες.

4. ΠΑΡΑΓΟΝΤΕΣ ΚΑΘΟΡΙΣΜΟΥ ΤΟΥ ΑΝΑΓΚΑΙΟΥ ΑΡΙΘΜΟΥ ΠΑΣΧΟΝΤΩΝ

Οι παράγοντες που επηρεάζουν τον αναγκαίο αριθμό πασχόντων σε μια ΤΚΔ είναι οι εξής:^{1,3,9,27-33}

- Το προκαθορισμένο επίπεδο στατιστικής σημαντικότητας
- Η πιθανότητα ότι ο κατάλληλος στατιστικός έλεγχος θα αναδείξει μια στατιστικά σημαντική σχέση, εφόσον βέβαια υπάρχει αυτή. Πρόκειται ουσιαστικά για τη στατιστική ισχύ του ελέγχου
- Η μικρότερη αναμενόμενη διαφορά στις αποτελεσματικότητες μεταξύ των συγκρινόμενων θεραπευτικών παρεμβάσεων είτε οι αποτελεσματικότητες εκφράζονται με τη μορφή μέσων τιμών είτε με τη μορφή ποσοστών
- Η εκτιμώμενη μεταβλητότητα των μετρήσεων
- Η διεξαγωγή μονόπλευρου ή αμφίπλευρου ελέγχου στατιστικής σημαντικότητας
- Το ποσοστό των πασχόντων που χάνονται στην παρακολούθηση
- Ο ερευνητικός σχεδιασμός. Στην πλέον συνηθισμένη περίπτωση της απλής τυχαίας κατανομής των πασχόντων σε δύο ομάδες, ο απαιτούμενος συνολικός αριθμός πασχόντων είναι μικρότερος όταν η κατανομή είναι ισόμετρη (1:1) παρά όταν είναι ανισόμετρη (3:1, 1:2 κ.λπ.).

Πρέπει, πάντως, να σημειωθεί ότι ορισμένοι από τους παραπάνω παράγοντες που επηρεάζουν τον αναγκαίο αριθμό πασχόντων στηρίζονται σε διάφορες στατιστικές υποθέσεις και γι' αυτό οι υπολογισμοί του αναγκαίου αριθμού πασχόντων αποτελούν εκτιμήσεις και όχι αριθμούς πασχόντων με απόλυτη ακρίβεια.³⁴ Οι στατιστικές μέθοδοι που χρησιμοποιούνται, στη συνέχεια, για τον υπολογισμό του αναγκαίου αριθμού πασχόντων βασίζονται σε δύο υποθέσεις: (α) η κατανομή των πασχόντων στις μελετώμενες ομάδες γίνεται τυχαία, κάτι που αποτελεί βασική αρχή διεξαγωγής των ΤΚΔ και (β) στις μελέτες όπου υπολογίζονται μέσες τιμές, η κατανομή των μετρήσεων θεωρείται κανονική.²⁷

Υπάρχουν στατιστικές μέθοδοι που επιτρέπουν τον υπολογισμό του αναγκαίου αριθμού πασχόντων σε μια ΤΚΔ, αν καθοριστούν οι παραπάνω παράγοντες. Με βάση τις στατιστικές αυτές μεθόδους έχουν καταρτιστεί πίνακες που παρέχουν απευθείας τον απαιτούμενο κατά περίπτωση αριθμό πασχόντων.³

4.1. Προκαθορισμένο επίπεδο στατιστικής σημαντικότητας

Καταρχήν, απαιτείται να γίνει διάκριση ανάμεσα στο προκαθορισμένο επίπεδο στατιστικής σημαντικότητας, που στο εξής θα ονομάζεται τιμή α , και στο παρατηρούμενο επίπεδο στατιστικής σημαντικότητας, που στο εξής θα καλείται τιμή P . Η τιμή α καθορίζεται αυθαίρετα από τους ερευνητές πριν από τη διεξαγωγή μιας μελέτης και λαμβάνει τιμές από 0,01–0,10, με συνηθέστερη την τιμή 0,05. Η τιμή P προκύπτει έπειτα από την εφαρμογή του κατάλληλου στατιστικού ελέγχου ανάλογα με τα δεδομένα μιας μελέτης και με τη χρήση των κατάλληλων πινάκων πιθανοτήτων.¹

Ακόμη και αν στην πραγματικότητα δεν υπάρχει διαφορά στις αποτελεσματικότητες μεταξύ των θεραπευτικών παρεμβάσεων, ενδεχομένως οι στατιστικοί έλεγχοι να αναδείξουν μικρές διαφορές που οφείλονται καθαρά στην τύχη. Έτσι, οι στατιστικοί έλεγχοι διεξάγονται με σκοπό να καθοριστεί η πιθανότητα η παρατηρηθείσα διαφορά μεταξύ των μελετώμενων παρεμβάσεων να οφείλεται στην τύχη και να μην αποτελεί πραγματική διαφορά που οφείλεται στις θεραπευτικές παρεμβάσεις. Η τιμή P είναι η πιθανότητα η διαφορά που βρέθηκε να οφείλεται στην τύχη και όχι στη μελετώμενη παρέμβαση. Εάν η τιμή P είναι πολύ μικρή, τότε η πιθανότητα ότι η παρατηρηθείσα διαφορά οφείλεται στην τύχη είναι πολύ μικρή.

Ο έλεγχος της μηδενικής υπόθεσης γίνεται με τη σύγκριση των τιμών P και α . Εάν η τιμή P που προκύπτει από τα δεδομένα μιας μελέτης είναι μικρότερη από την τιμή α που επιλέγεται αυθαίρετα από τους ερευνητές, τότε απορρίπτεται η μηδενική υπόθεση και υπάρχει στατιστικά σημαντική σχέση μεταξύ προσδιοριστή και συχνότητας εμφάνισης της έκβασης. Αντίθετα, εάν η τιμή P είναι μεγαλύτερη από την τιμή α , τότε τα δεδομένα της μελέτης δεν επαρκούν για την απόρριψη της μηδενικής υπόθεσης.^{1,28} Εάν, π.χ., η τιμή α καθοριστεί στο επίπεδο του 0,05, τότε δέχεται κανείς ότι υπάρχει πιθανότητα 5% να απορρίψει λανθασμένα τη μηδενική υπόθεση, ενώ αυτή είναι αληθής. Χρησιμοποιώντας το επίπεδο του 0,05 για την τιμή α και λαμβάνοντας υπόψη το μεγάλο αριθμό μελετών που διεξάγονται στο χώρο της υγείας, είναι πιθανό να προκύπτουν αρκετά «ψευδώς θετικά» συμπεράσματα για τη σχέση μεταξύ προσδιοριστή

και έκβασης. Προκύπτει δηλαδή στατιστικά σημαντική σχέση μεταξύ προσδιοριστή και συχνότητας εμφάνισης της έκβασης, αλλά η σχέση αυτή οφείλεται στην τύχη και όχι στην πραγματικά αποτελεσματικότερη δράση μιας θεραπευτικής παρέμβασης έναντι μιας άλλης. Για το λόγο αυτό συνιστάται η τιμή α να καθορίζεται στο επίπεδο του 0,01 και όχι στο επίπεδο του 0,05, έτσι ώστε να περιορίζεται το ποσοστό των «ψευδώς θετικών» συμπερασμάτων. Στην περίπτωση, όμως, που μειώνεται η τιμή α —οπότε μειώνεται και η πιθανότητα η παρατηρηθείσα σχέση να οφείλεται στην τύχη— αυξάνεται ο αναγκαίος αριθμός πασχόντων για την ανάδειξη μιας στατιστικά σημαντικής σχέσης.

Η πιθανότητα να απορριφθεί η μηδενική υπόθεση ενώ είναι αληθής (δηλαδή η τιμή α) ονομάζεται *σφάλμα τύπου I* (type I error) (πίν. 1).^{15,23,35} Ουσιαστικά, το σφάλμα τύπου I είναι το ποσοστό των «ψευδώς θετικών» συμπερασμάτων, όπου λανθασμένα συμπεραίνεται ότι υπάρχει διαφορά στις θεραπευτικές παρεμβάσεις, καθώς στην πραγματικότητα η διαφορά αυτή δεν υπάρχει. Όταν η τιμή α καθορίζεται στο επίπεδο του 0,05, γίνεται δεκτό ότι υπάρχει πιθανότητα 5% η παρατηρηθείσα διαφορά μεταξύ των μελετώμενων ομάδων να συμβαίνει περισσότερο από τύχη παρά να οφείλεται στο μελετώμενο προσδιοριστή (δηλαδή στις διαφορετικές θεραπευτικές παρεμβάσεις).

4.2. Στατιστική ισχύς

Η πιθανότητα ένας στατιστικός έλεγχος να αναδείξει μια στατιστικά σημαντική σχέση, εφόσον βέβαια αυτή υπάρχει πραγματικά, ονομάζεται *στατιστική ισχύς* (statistical power).^{1,27,36} Έτσι, η ισχύς είναι η ικανότητα ενός ελέγχου να απορρίψει τη μηδενική υπόθεση, ενώ αυτή είναι ψευδής.³⁷ Ορισμένοι θεωρούν ότι η στατιστική ισχύς είναι ανάλογη του ποσοστού των αληθώς θετικών αποτελεσμάτων μιας διαγνωστικής δοκιμασίας, που είναι το ποσοστό των πασχόντων οι οποίοι έχουν θετικό το αποτέλεσμα της δοκιμασίας.³⁸

Πίνακας 1. Σύγκριση των αποτελεσμάτων μιας μελέτης και της «πραγματικότητας».

		«Πραγματικότητα»	
		Διαφορά στις παρεμβάσεις	Μη διαφορά στις παρεμβάσεις
Αποτελέσματα μελέτης	Διαφορά στις παρεμβάσεις	Αληθώς θετικά αποτελέσματα (ισχύς)	Ψευδώς θετικά αποτελέσματα (σφάλμα τύπου I)
	Μη διαφορά στις παρεμβάσεις	Ψευδώς αρνητικά αποτελέσματα (σφάλμα τύπου II)	Αληθώς αρνητικά αποτελέσματα

Σε μια ΤΚΔ όπου συγκρίνονται δύο ομάδες, η ισχύς ενός στατιστικού ελέγχου πρέπει να έχει τη δυνατότητα να αναγνωρίζει μια στατιστικά σημαντική σχέση μεταξύ προσδιοριστή και συχνότητας εμφάνισης της έκβασης, εφόσον βέβαια η σχέση αυτή υπάρχει πραγματικά. Το γεγονός αυτό είναι εξαιρετικά σημαντικό όταν ο έλεγχος οδηγεί σε μη απόρριψη της μηδενικής υπόθεσης, οπότε δεν υφίσταται στατιστικά σημαντική σχέση. Στην περίπτωση αυτή όπου το αποτέλεσμα είναι αρνητικό, υπάρχουν δύο πιθανές ερμηνείες. Η πρώτη ερμηνεία είναι ότι το αποτέλεσμα του ελέγχου είναι σωστό, καθώς στην πραγματικότητα δεν υπάρχει στατιστικά σημαντική σχέση. Πρόκειται, ουσιαστικά, για το ποσοστό των αληθώς αρνητικών αποτελεσμάτων ενός ελέγχου. Η δεύτερη ερμηνεία είναι ότι το αποτέλεσμα του ελέγχου είναι εσφαλμένο, καθώς στην πραγματικότητα υπάρχει στατιστικά σημαντική σχέση αλλά ο έλεγχος δεν ήταν αρκετά ισχυρός ώστε να αναδείξει τη σχέση αυτή. Η περίπτωση αυτή αφορά στο ποσοστό των ψευδώς αρνητικών αποτελεσμάτων ενός ελέγχου. Πρόκειται ουσιαστικά για το *σφάλμα τύπου II* (type II error)^{15,23,35} (πίν. 1), που είναι η πιθανότητα μη απόρριψης της μηδενικής υπόθεσης ενώ αυτή είναι ψευδής. Ένα κατάλληλο μέγεθος «δείγματος» δίνει τη δυνατότητα σ' ένα στατιστικό έλεγχο να έχει αρκετή ισχύ, έτσι ώστε η πρώτη ερμηνεία—ότι τα αποτελέσματα του ελέγχου είναι αληθώς αρνητικά— να είναι περισσότερο λογική από τη δεύτερη—ότι τα αποτελέσματα είναι ψευδώς αρνητικά, οπότε υπάρχει σφάλμα τύπου II— στην περίπτωση που βρεθεί μια μη στατιστικά σημαντική σχέση. Αρκετές δημοσιευμένες ΤΚΔ έχουν εμφανίσει περιορισμένη ισχύ εξαιτίας μη κατάλληλου μεγέθους «δείγματος» ή άλλων λανθασμένων ερευνητικών σχεδιασμών.^{32,33}

Οι παράγοντες που επηρεάζουν την ισχύ ενός ελέγχου' είναι (α) το μέγεθος του «δείγματος», (β) η διαφορά στις αποτελεσματικότητες μεταξύ των μελετώμενων παρεμβάσεων, (γ) το προκαθορισμένο επίπεδο στατιστικής σημαντικότητας (τιμή α) και (δ) η κανονική κατανομή των μετρήσεων (οι παραμετρικοί έλεγχοι έχουν μεγαλύτερη ισχύ από τους αντίστοιχους μη παραμετρικούς).

Όσο μεγαλύτερη είναι η ισχύς ενός ελέγχου τόσο μεγαλύτερος είναι και ο αναγκαίος αριθμός πασχόντων σε μια ΤΚΔ. Στις περισσότερες ΤΚΔ, η ισχύς ορίζεται αυθαίρετα από τους ερευνητές στο επίπεδο του 0,80 ή 80%, ενώ αρκετά συνηθισμένο είναι και το επίπεδο του 0,90 ή 90%.^{1,27,28,39} Αν και οι υψηλές τιμές ισχύος είναι πάντοτε επιθυμητές, δεν είναι εύκολο να επιτευχθούν, καθώς απαιτούν μεγάλο αριθμό συμμετεχόντων, γεγονός που αυξάνει σημαντικά το κόστος και το χρόνο διεξαγωγής μιας μελέτης. Η ισχύς και το σφάλμα τύπου II είναι συμπληρωματικές έννοιες, οπότε εάν η ισχύς ενός ελέγχου τεθεί στο επίπεδο του 0,80, τότε το σφάλμα τύπου II θα είναι ίσο με 0,20. Στην

περίπτωση αυτή, δέχεται κάποιος ότι υπάρχει πιθανότητα 20% να μην απορριφθεί η μηδενική υπόθεση, ενώ είναι ψευδής, δέχεται δηλαδή ότι υπάρχει πιθανότητα 20% να μη βρεθεί στατιστικά σημαντική σχέση, ενώ στην πραγματικότητα υπάρχει.

4.3. «Μέγεθος του αποτελέσματος»

Για τον υπολογισμό του αναγκαίου αριθμού πασχόντων σε μια ΤΚΔ πρέπει να καθοριστεί κατά το σχεδιασμό της μελέτης η μικρότερη αναμενόμενη διαφορά στις αποτελεσματικότητες μεταξύ των συγκρινόμενων θεραπευτικών παρεμβάσεων που θεωρείται από τους ερευνητές ως κλινικά σημαντική και την οποία επιθυμούν να διαπιστώσουν.^{1,27,34} Η μικρότερη αυτή αναμενόμενη διαφορά είναι γνωστή και ως «μέγεθος του αποτελέσματος» (effect size)¹ και αποτελεί μέτρο του πόσο «λανθασμένη» είναι η μηδενική υπόθεση. Εάν είναι ήδη γνωστό το πραγματικό «μέγεθος του αποτελέσματος» μεταξύ δύο θεραπευτικών παρεμβάσεων, τότε προφανώς δεν προσφέρει κέρδος η διεξαγωγή της αντίστοιχης μελέτης.

Για να υπολογιστεί το «μέγεθος του αποτελέσματος» πρέπει να είναι γνωστή η αποτελεσματικότητα της συγκριτικής θεραπευτικής παρέμβασης εκφρασμένης είτε με τη μορφή ποσοστού (π.χ. θνητότητα ή ποσοστό η-ετούς επιβίωσης) είτε με τη μορφή μέσης τιμής (π.χ. μέση αρτηριακή πίεση, μέση θερμοκρασία κ.ά.). Επιπλέον, χρειάζεται να προβλεφθεί η αποτελεσματικότητα της νέας θεραπευτικής παρέμβασης που δοκιμάζεται. Η πρόβλεψη αυτή είναι υποκειμενική και στηρίζεται σε δεδομένα που προέρχονται από τη βιβλιογραφία, την κλινική εμπειρία των ερευνητών που διεξάγουν τη μελέτη ή τη διεξαγωγή πιλοτικών μελετών με μικρό αριθμό συμμετεχόντων.^{1,27} Όσο μικρότερο είναι το «μέγεθος του αποτελέσματος», όσο μικρότερη δηλαδή είναι η διαφορά στις αποτελεσματικότητες μεταξύ των συγκρινόμενων θεραπευτικών παρεμβάσεων, τόσο μεγαλύτερος είναι ο αριθμός των πασχόντων που χρειάζονται σε μια ΤΚΔ για να αναδειχθεί μια στατιστικά σημαντική σχέση, εφόσον βέβαια υπάρχει (πίν. 2).¹ Αυτό ισχύει με την προϋπόθεση ότι οι τιμές της ισχύος και του προκαθορισμένου επιπέδου σημαντικότητας (τιμή α) διατηρούνται σταθερές.

Ο αναγκαίος αριθμός πασχόντων σε μια ΤΚΔ εκτιμάται με βάση ένα συγκεκριμένο «μέγεθος αποτελέσματος». Συνήθίζεται όμως, εάν ο αριθμός των πασχόντων που χρειάζονται είναι αρκετά μεγάλος και δεν μπορούν να συγκεντρωθούν, να αυξάνεται το «μέγεθος του αποτελέσματος» ώστε να μειωθεί ο αριθμός των συμμετεχόντων. Σε μια μελέτη, π.χ., όπου η τιμή α ορίζεται στο επίπεδο του 5% και η ισχύς στο επίπεδο του 80%, η ανάδειξη μιας διαφοράς της τάξης του 3% ως στατιστικά σημαντικής απαιτεί συνολικά 2.500

Πίνακας 2. Αναγκαίος αριθμός πασχόντων ανάλογα με το «μέγεθος του αποτελέσματος», τη στατιστική ισχύ και το προκαθορισμένο επίπεδο στατιστικής σημαντικότητας (τιμή α).

«Μέγεθος αποτελέσματος» (%)	Ισχύς (%)	Τιμή α (%)	Αναγκαίος αριθμός πασχόντων (σε κάθε ομάδα)
10	80	5	384
5	80	5	1.556
10	90	5	513
10	95	5	635
10	80	1	572
10	90	1	727
10	95	1	870
5	95	1	3.531

συμμετέχοντες, ενώ η ανάδειξη μιας διαφοράς της τάξης του 4% ως στατιστικά σημαντικής απαιτεί συνολικά 1.700 συμμετέχοντες. Οι αλλαγές αυτές, πάντως, στον αναγκαίο αριθμό των πασχόντων προκαλούν αρκετές διαφωνίες, με κυριότερο επιχείρημα ότι η μελέτη δεν διαθέτει πλέον αρκετή ισχύ για να αναδείξει ως στατιστικά σημαντική την αρχική διαφορά και όχι τη μετέπειτα.^{40,41}

Το κατάλληλο «μέγεθος του αποτελέσματος» ποικίλλει σημαντικά μεταξύ των διαφόρων μελετών, καθώς εκφράζει το μικρότερο αποτέλεσμα που θα μπορούσε να θεωρηθεί ως κλινικά σημαντικό και έτσι εξαρτάται από παράγοντες όπως η φύση της μελετώμενης έκβασης, η σοβαρότητα της κατάστασης, η ευκολία πραγματοποίησης μιας παρέμβασης κ.ά.¹ Μια παρέμβαση, π.χ., που μειώνει τη μητρική θνησιμότητα μόλις κατά 1%, ενδεχομένως θα μπορούσε να θεωρηθεί ως κλινικά σημαντική, ενώ μια παρέμβαση που μειώνει την επίπτωση του φυσιολογικού νεογνικού ικτέρου κατά 10% θα μπορούσε να θεωρηθεί ως κλινικά λιγότερο σημαντική. Επιπλέον, διεξάγεται μια μελέτη για να συγκριθεί μια καθιερωμένη διαγνωστική δοκιμασία που έχει ακρίβεια 80% με μια νέα διαγνωστική δοκιμασία που έχει άγνωστη αλλά πιθανόν μεγαλύτερη ακρίβεια. Εάν η νέα δοκιμασία θεωρείται ότι έχει ακρίβεια 81%, τότε η διαφορά στην ακρίβεια μεταξύ των δύο δοκιμασιών θεωρείται κλινικά μη σημαντική, ενώ εάν η νέα δοκιμασία πιστεύεται ότι έχει ακρίβεια 90%, τότε η διαφορά στην ακρίβεια θεωρείται κλινικά σημαντική. Έτσι, στη μελέτη αυτή, η μικρότερη αναμενόμενη διαφορά καθορίζεται στο επίπεδο του 10%.

4.4. Εκτιμώμενη μεταβλητότητα των μετρήσεων

Η παράμετρος αυτή επηρεάζει τον αναγκαίο αριθμό

των πασχόντων, μόνον όταν οι αποτελεσματικότητες των συγκρινομένων θεραπευτικών παρεμβάσεων εκφράζονται με τη μορφή μέσων τιμών και όχι με τη μορφή ποσοστών.¹ Η εκτιμώμενη μεταβλητότητα των μετρήσεων εκφράζεται μέσω της αναμενόμενης σταθερής (ή τυπικής) απόκλισης των μετρήσεων που πραγματοποιούνται σε κάθε μελετώμενη ομάδα. Όσο μεγαλύτερη είναι η στατιστική μεταβλητότητα, όσο μεγαλύτερη δηλαδή είναι η σταθερή απόκλιση των μετρήσεων, τόσο μεγαλύτερος είναι ο αριθμός των πασχόντων που χρειάζονται για να αναδειχθεί η μικρότερη αναμενόμενη διαφορά μεταξύ των συγκρινομένων θεραπευτικών παρεμβάσεων. Ιδανικά, η εκτιμώμενη μετρική μεταβλητότητα θα έπρεπε να καθορίζεται με πιλοτικές μελέτες που περιλαμβάνουν μικρότερο αριθμό συμμετεχόντων. Σε ορισμένες περιπτώσεις, η μετρική μεταβλητότητα εκτιμάται με βάση την υπάρχουσα βιβλιογραφία, ενώ σπανιότερα εκτιμάται υποκειμενικά με βάση την εμπειρία και τις γνώσεις των ερευνητών.

4.5. Μονόπλευρος ή αμφίπλευρος έλεγχος στατιστικής σημαντικότητας

Σε ορισμένες περιπτώσεις, είναι γνωστό πριν από τη διεξαγωγή μιας μελέτης ότι η διαφορά μεταξύ των συγκρινομένων ομάδων είναι δυνατή προς μία μόνο κατεύθυνση, είτε θετική είτε αρνητική. Είναι γνωστό, π.χ., και ευρέως αποδεκτό ότι ο κυτταρολογικός έλεγχος κατά Παπανικολάου παρουσιάζει σήμερα τη μεγαλύτερη ακρίβεια έναντι οποιασδήποτε άλλης διαγνωστικής δοκιμασίας για την προσυμπτωματική διάγνωση του καρκίνου του τραχήλου της μήτρας. Στις περιπτώσεις αυτές, μπορεί να διεξαχθεί ένας μονόπλευρος έλεγχος στατιστικής σημαντικότητας που απαιτεί μικρότερο αριθμό ατόμων για την ανάδειξη μιας σχέσης ως στατιστικά σημαντικής σε σύγκριση με τον αντίστοιχο αμφίπλευρο έλεγχο. Ο αναγκαίος αριθμός πασχόντων σ' ένα μονόπλευρο έλεγχο με προκαθορισμένο επίπεδο στατιστικής σημαντικότητας ίσο με α (π.χ. 0,05) ισούται με τον αναγκαίο αριθμό πασχόντων σ' έναν αμφίπλευρο έλεγχο, με προκαθορισμένο επίπεδο στατιστικής σημαντικότητας ίσο με 2α (0,10), εάν οι υπόλοιπες παράμετροι είναι ίδιες και στους δύο ελέγχους. Σε ελάχιστες περιπτώσεις, πάντως, διεξάγονται μονόπλευροι έλεγχοι και γι' αυτό οι ισότητες για τον υπολογισμό του αναγκαίου αριθμού πασχόντων που παρατίθενται στη συνέχεια αφορούν σε αμφίπλευρους ελέγχους.

5. ΑΝΑΓΚΑΙΟΣ ΑΡΙΘΜΟΣ ΠΑΣΧΟΝΤΩΝ

5.1. Σύγκριση μέσων τιμών

Γνωρίζοντας τους παράγοντες που επηρεάζουν τον

αναγκαίο αριθμό πασχόντων σε μια ΤΚΔ είναι σχετικά εύκολο να υπολογιστεί ο αριθμός αυτός με την εφαρμογή της κατάλληλης ισότητας. Όταν σε μια μελέτη συγκρίνονται οι μέσες τιμές σε δύο ομάδες, τότε χρησιμοποιείται η παρακάτω ισότητα:⁴²

$$N = \frac{4 \sigma^2 (Z_{σημ} + Z_{ισχύς})^2}{D^2} \quad (1)$$

Ο συνολικός αριθμός των πασχόντων που χρειάζονται συμβολίζεται με N (πρόκειται για το άθροισμα των πασχόντων και στις δύο μελετώμενες ομάδες). Η εκτιμώμενη σταθερή απόκλιση κάθε ομάδας συμβολίζεται με σ (στην ισότητα 1, υποτίθεται ότι οι σταθερές αποκλίσεις στις δύο ομάδες είναι ίσες). Η τιμή Z που αντιστοιχεί στο προκαθορισμένο επίπεδο στατιστικής σημαντικότητας (τιμή α) συμβολίζεται με $Z_{σημ}$ (πίν. 3). Η τιμή Z που αντιστοιχεί στη στατιστική ισχύ του ελέγχου συμβολίζεται με $Z_{ισχύς}$ (πίν. 4) και η μικρότερη αναμενόμενη διαφορά μεταξύ των δύο μέσων τιμών ή, με άλλη διατύπωση, το «μέγεθος του αποτελέσματος» συμβολίζεται με D . Τόσο το $Z_{σημ}$ όσο και το $Z_{ισχύς}$ αποτελούν διαχωριστικά σημεία στον άξονα των X της τυπικής κανονικής κατανομής που υποδηλώνουν τις πιθανότητες, οι οποίες αφορούν σε συγκεκριμένο επίπεδο στατιστικής σημαντικότητας και συγκεκριμένη στατιστική ισχύ, αντίστοιχα. Στον πίνακα 3, όσο μικρότερο είναι το επίπεδο στατιστικής σημαντικότητας τόσο μεγαλύτερη είναι η τιμή $Z_{σημ}$, ενώ στον πίνακα 4 όσο μεγαλύτερη είναι η στατιστική ισχύς τόσο μεγαλύτερη είναι η τιμή $Z_{ισχύς}$. Η ισότητα 1 ισχύει εφόσον ο αριθμός των ατόμων στις δύο μελετώμενες ομάδες είναι ίσος και εφόσον χρησιμοποιείται

Πίνακας 3. Οι τιμές Z ($Z_{σημ}$) της τυπικής κανονικής κατανομής που αντιστοιχούν σε διάφορα προκαθορισμένα επίπεδα στατιστικής σημαντικότητας (τιμές α).

Τιμή α	$Z_{σημ}$
0,01	2,576
0,02	2,326
0,05	1,960
0,10	1,645

Πίνακας 4. Οι τιμές Z ($Z_{ισχύς}$) της τυπικής κανονικής κατανομής που αντιστοιχούν σε διάφορα επίπεδα στατιστικής ισχύος.

Στατιστική ισχύς	$Z_{ισχύς}$
0,80	0,842
0,85	1,036
0,90	1,282
0,95	1,645

αμφίπλευρος έλεγχος στατιστικής σημαντικότητας. Παρατηρώντας την ισότητα 1, διαπιστώνεται ότι ο αναγκαίος αριθμός πασχόντων εξαρτάται μόνο από το μέγεθος της διαφοράς μεταξύ των μέσων τιμών και όχι από τις απόλυτες τιμές των μέσων τιμών.

Μια μελέτη, για παράδειγμα, διεξάγεται για να συγκριθεί η αποτελεσματικότητα της χειρουργικής διάνοιξης της νεφρικής αρτηρίας έναντι της φαρμακευτικής αγωγής για τη μείωση της συστολικής αρτηριακής πίεσης σε πάσχοντες με δευτεροπαθή υπέρταση εξαιτίας στένωσης της νεφρικής αρτηρίας. Οι ερευνητές, στηριζόμενοι στην κλινική τους εμπειρία, θεωρούν ότι η χειρουργική επέμβαση είναι δύο φορές πιο αποτελεσματική από τη φαρμακευτική αγωγή. Η χειρουργική επέμβαση, ωστόσο, κοστίζει περισσότερο και είναι περισσότερο οδυνηρή για τους πάσχοντες έναντι της φαρμακευτικής αγωγής. Η αποτελεσματικότητα των δύο θεραπευτικών παρεμβάσεων εκφράζεται με τη μορφή μέσης τιμής και, πιο συγκεκριμένα, με τη μορφή της μέσης ελάττωσης της συστολικής αρτηριακής πίεσης. Με βάση τα αποτελέσματα πιλοτικών μελετών, οι ερευνητές εκτιμούν ότι η μέση ελάττωση της αρτηριακής πίεσης στους πάσχοντες που υφίστανται τη χειρουργική επέμβαση είναι 20 mmHg, η μέση ελάττωση της αρτηριακής πίεσης στους πάσχοντες που λαμβάνουν φαρμακευτική αγωγή είναι μόλις 10 mmHg, ενώ η σταθερή απόκλιση της ελάττωσης της πίεσης και στις δύο ομάδες πασχόντων είναι 15 mmHg. Το επίπεδο στατιστικής σημαντικότητας ορίζεται στην τιμή 0,05 και η στατιστική ισχύς στην τιμή 0,80. Έτσι, $D=20-10=10$ mmHg, $\sigma=15$ mmHg, $Z_{σημ}=1,960$ (πίν. 3) και $Z_{ισχύς}=0,842$ (πίν. 4). Εφαρμόζοντας την ισότητα 1 προκύπτει ότι $N=70,6$. Επομένως, στη συγκεκριμένη ΤΚΔ χρειάζονται συνολικά 70 πάσχοντες – η προσέγγιση του N γίνεται στον πλησιέστερο άρτιο αριθμό– από τους οποίους οι 35 θα υποβληθούν στη χειρουργική επέμβαση και οι 35 θα λάβουν φαρμακευτική αγωγή.

5.2. Σύγκριση ποσοστών

Σε μια ΤΚΔ όπου υπολογίζονται ποσοστά, η σύγκριση μεταξύ τους μπορεί να γίνει είτε με τον έλεγχο χ^2 είτε με τον έλεγχο Z . Ο έλεγχος Z βασίζεται στην κανονική προσέγγιση της διωνυμικής κατανομής και στην περίπτωση αυτή η ισότητα για τον υπολογισμό του αριθμού των πασχόντων που χρειάζονται είναι η εξής:⁴³

$$N = \frac{2[Z_{σημ}\sqrt{2\bar{p}(1-\bar{p})} + Z_{ισχύς}\sqrt{p_1(1-p_1) + p_2(1-p_2)}]^2}{D^2} \quad (2)$$

Τα ποσοστά p_1 και p_2 αποτελούν εκτιμήσεις των ερευνητών με βάση τα αποτελέσματα πιλοτικών μελετών ή την υπάρχουσα βιβλιογραφία, το D είναι η απόλυτη διαφορά

μεταξύ των δύο ποσοστών p_1 και p_2 – ή, με άλλη διατύπωση, η μικρότερη αναμενόμενη διαφορά– και το \bar{p} ισούται με $\frac{(p_1+p_2)}{2}$. Τα N , $Z_{σημ}$ και $Z_{ισχύς}$ είναι αντίστοιχα με εκείνα της ισότητας 1 και έχουν προαναφερθεί εκτενώς. Η ισότητα 2 ισχύει, εφόσον ο αριθμός των συμμετεχόντων στις δύο μελετώμενες ομάδες είναι ίσος και εφόσον ο έλεγχος της στατιστικής σημαντικότητας είναι αμφίπλευρος. Στην περίπτωση αυτή, ο αναγκαίος αριθμός πασχόντων εξαρτάται όχι μόνο από τη διαφορά μεταξύ των ποσοστών p_1 και p_2 , αλλά και από τις ίδιες τις τιμές των ποσοστών. Η εφαρμογή της ισότητας 2 απαιτεί την εκτίμηση των p_1 και p_2 , όπως επίσης και της διαφοράς τους πριν από τη διεξαγωγή της μελέτης. Το πλεονέκτημα που παρουσιάζει η σύγκριση ποσοστών είναι ότι δεν απαιτείται η εκτίμηση σταθερών αποκλίσεων.

Διεξάγεται μια μελέτη για τη σύγκριση ενός νέου φαρμάκου με το καθιερωμένο φάρμακο που χρησιμοποιείται ευρύτατα. Η αποτελεσματικότητα του καθιερωμένου φαρμάκου είναι 10% (δηλαδή 0,10), ενώ η αποτελεσματικότητα του νέου θεωρείται ότι είναι 20% (δηλαδή 0,20). Έτσι, η αναμενόμενη σχετική αποτελεσματικότητα μεταξύ των δύο φαρμάκων είναι $\frac{0,20}{0,10}=2$. Οι ερευνητές επιθυμούν να ελέγξουν σε επίπεδο σημαντικότητας ίσο με 0,05 ($\alpha=0,05$) αν πραγματικά το νέο φάρμακο είναι αποτελεσματικότερο από το καθιερωμένο και επιθυμούν να είναι βέβαιοι κατά 80% (ισχύς=80%) ότι η μελέτη τους θα οδηγήσει στο εύρημα αυτό, εφόσον η φημολογούμενη διαφορά υπάρχει πραγματικά. Έτσι, $p_1=0,10$, $p_2=0,20$, $D=0,20-0,10=0,10$, $\bar{p}=\frac{(0,10+0,20)}{2}=0,15$, $Z_{σημ}=1,960$ και $Z_{ισχύς}=0,842$. Εφαρμόζοντας την ισότητα 2 προκύπτει ότι $N=398$. Επομένως, στη συγκεκριμένη ΤΚΔ χρειάζονται συνολικά 398 άτομα, από τα οποία τα 199 θα λάβουν το καθιερωμένο φάρμακο και τα 199 θα λάβουν το νέο.

Επιπλέον, μια καθιερωμένη διαγνωστική δοκιμασία εμφανίζει 80% ακρίβεια στη διάγνωση μιας συγκεκριμένης πάθησης. Διεξάγεται μια μελέτη για την εκτίμηση μιας νέας διαγνωστικής δοκιμασίας που πιθανολογείται ότι εμφανίζει μεγαλύτερη ακρίβεια. Οι ερευνητές, στηριζόμενοι στην κλινική τους εμπειρία, θεωρούν ότι η νέα διαγνωστική δοκιμασία εμφανίζει ακρίβεια τουλάχιστον 90%. Το επίπεδο στατιστικής σημαντικότητας θεωρείται ίσο με 0,05 ($\alpha=0,05$), ενώ η στατιστική ισχύς ίση με 0,95 (ισχύς=0,95). Έτσι, $p_1=0,80$, $p_2=0,90$, $D=0,90-0,80=0,10$, $\bar{p}=\frac{(0,90+0,80)}{2}=0,85$, $Z_{σημ}=1,960$ και $Z_{ισχύς}=1,645$. Εφαρμόζοντας την ισότητα 2 προκύπτει ότι $N=656,8$. Επομένως, χρειάζονται 656 άτομα,

από τα οποία τα 328 θα υποβληθούν στην καθιερωμένη διαγνωστική δοκιμασία και τα 328 στη νέα.

6. ΤΡΟΠΟΙ ΜΕΙΩΣΗΣ ΤΟΥ ΑΝΑΓΚΑΙΟΥ ΑΡΙΘΜΟΥ ΠΑΣΧΟΝΤΩΝ

Σε αρκετές περιπτώσεις, ο αναγκαίος αριθμός πασχόντων είναι αρκετά μεγάλος και πολύ δύσκολα μπορεί να επιτευχθεί. Ένας συνηθισμένος τρόπος μείωσης του αναγκαίου αριθμού πασχόντων είναι η αύξηση της ελάχιστης αναμενόμενης διαφοράς ή, αλλιώς, η αύξηση του «μεγέθους του αποτελέσματος». ^{1,44} Ο τρόπος αυτός χρησιμοποιείται ιδιαίτερα όταν η ελάχιστη αναμενόμενη διαφορά είναι εξαιρετικά μικρή, οπότε δικαιολογείται η αύξησή της, κάτι που συνηθίζεται στις πιλοτικές μελέτες.

Επιπλέον, η χρησιμοποίηση διαφορετικού αριθμού ατόμων στις μελετώμενες ομάδες μπορεί να μειώσει τον αναγκαίο αριθμό πασχόντων. ^{1,44} Αυτό συμβαίνει επειδή οι πάσχοντες, συνήθως, είναι πιο εύκολο να υποβληθούν στην καθιερωμένη θεραπευτική παρέμβαση σε σχέση με τη νέα. Οι σχετικά απλές ισότητες 1 και 2 ισχύουν όταν ο αριθμός των συμμετεχόντων είναι ίσος στις μελετώμενες ομάδες. Όταν ο αριθμός αυτός είναι άνισος, χρησιμοποιούνται περισσότερο πολύπλοκες ισότητες. ^{42,45} Πρέπει, πάντως, να σημειωθεί ότι ο ίσος αριθμός ατόμων στις μελετώμενες ομάδες εξασφαλίζει μεγαλύτερη στατιστική αποτελεσματικότητα.

Οι στατιστικοί έλεγχοι ανά ζεύγη εμφανίζουν μεγαλύτερη στατιστική ισχύ από τους αντίστοιχους ελέγχους που περιλαμβάνουν ανεξάρτητα «δείγματα», καθώς στους ελέγχους ανά ζεύγη κάθε μέτρηση σ' ένα άτομο εξομοιώνεται με μία μέτρηση στο ίδιο άτομο. ^{1,44} Μια μελέτη, π.χ., που καταγράφει τις τιμές της αρτηριακής πίεσης στα ίδια άτομα πριν και μετά από την εφαρμογή μιας δίαιτας για την ελάττωση της αρτηριακής πίεσης έχει μεγαλύτερη στατιστική ισχύ από μια μελέτη όπου συμμετέχουν δύο ομάδες, από τις οποίες η μία λαμβάνει τη δοκιμαζόμενη δίαιτα και η άλλη όχι. Η ισότητα 1 εξακολουθεί να χρησιμοποιείται και όταν οι παρατηρήσεις είναι ανά ζεύγη, οπότε το D είναι η αναμενόμενη διαφορά στις μετρήσεις και το σ είναι η σταθερή απόκλιση της διαφοράς αυτής. Η αύξηση της ισχύος και η μείωση του αναγκαίου αριθμού πασχόντων οφείλονται στο ότι η σταθερή απόκλιση είναι μικρότερη για μεταβολές που συμβαίνουν στα ίδια άτομα παρά για διαφορές μεταξύ διαφορετικών ομάδων.

Οι στατιστικοί έλεγχοι που χρησιμοποιούνται για τη σύγκριση μέσων τιμών έχουν μεγαλύτερη στατιστική ισχύ από τους αντίστοιχους ελέγχους που χρησιμοποιούνται για τη σύγκριση ποσοστών για ένα συγκεκριμένο μέγεθος

«δείγματος». Για το λόγο αυτό συνιστάται οι ποσοτικές μεταβλητές (π.χ. αρτηριακή πίεση) να μη μετατρέπονται σε ποιοτικές (π.χ. υπερτασικοί και μη πάσχοντες). ^{1,44}

Η αύξηση, εξάλλου, του προκαθορισμένου επιπέδου σημαντικότητας (τιμή α) προκαλεί μείωση του αναγκαίου αριθμού πασχόντων, αλλά με τον τρόπο αυτόν αυξάνεται η πιθανότητα η ενδεχόμενη σχέση μεταξύ προσδιοριστή και συχνότητας εμφάνισης της έκβασης να οφείλεται στην τύχη. Τέλος, η ελάττωση της ισχύος προκαλεί μείωση του αριθμού των πασχόντων που χρειάζονται, αλλά έτσι μειώνεται και η πιθανότητα να βρεθεί μια στατιστικά σημαντική σχέση, εφόσον πραγματικά υπάρχει.

7. ΣΥΝΟΨΗ

Οι ισότητες 1 και 2 μπορούν να χρησιμοποιηθούν μόνον όταν η κατανομή των ατόμων στις μελετώμενες ομάδες είναι τυχαία, κάτι που αποτελεί βασική αρχή διεξαγωγής των ΤΚΔ. Επιπλέον, η ισότητα 1 χρησιμοποιείται μόνον όταν η κατανομή των μετρήσεων υποτίθεται ότι ακολουθεί την κανονική κατανομή, οπότε είναι δυνατή η εφαρμογή παραμετρικών μεθόδων. Όταν οι μετρήσεις δεν ακολουθούν την κανονική κατανομή, τότε ο υπολογισμός του αναγκαίου αριθμού πασχόντων απαιτεί την εφαρμογή μη παραμετρικών μεθόδων (π.χ. έλεγχος αθροίσματος διατάξεων του Wilcoxon), που είναι περισσότερο πολύπλοκες και χρησιμοποιούνται σπάνια.

Επιπλέον, οι ισότητες 1 και 2 μπορούν να χρησιμοποιηθούν και για τον καθορισμό της στατιστικής ισχύος για ένα συγκεκριμένο μέγεθος «δείγματος». Για τον υπολογισμό της ισχύος, η ισότητα 1 ή 2 επιλύεται ως προς $Z_{ισχύος}$ και έπειτα η τιμή αυτή Z αντιστοιχείται σε μια τιμή ισχύος σύμφωνα με τον πίνακα 3. Με τον τρόπο αυτό, μια «παρατηρηθείσα ισχύς» μπορεί να υπολογιστεί μετά από τη διεξαγωγή μιας μελέτης, οπότε στις ισότητες 1 ή 2 η ελάχιστη αναμενόμενη διαφορά –ή αλλιώς το «μέγεθος του αποτελέσματος»– αντικαθίσταται από την παρατηρηθείσα διαφορά που προκύπτει από τα δεδομένα της μελέτης. Η περίπτωση αυτή είναι γνωστή ως *αναδρομική ανάλυση της ισχύος* και χρησιμοποιείται ορισμένες φορές για την ερμηνεία των στατιστικών αποτελεσμάτων μιας μελέτης. Η αναδρομική αυτή ανάλυση, ωστόσο, προκαλεί αρκετές αντιρρήσεις, καθώς η «παρατηρηθείσα ισχύς» καθορίζεται πλήρως από την τιμή που λαμβάνει το προκαθορισμένο επίπεδο στατιστικής σημαντικότητας (τιμή α) και έτσι δεν προσφέρει επιπλέον πληροφορίες στην ερμηνεία της ισχύος. ^{1,46} Οι υπολογισμοί της ισχύος είναι περισσότερο κατάλληλοι όταν αφορούν στην ελάχιστη αναμενόμενη διαφορά που καθορίζεται πριν από τη διεξαγωγή μιας μελέτης και όχι

την παρατηρηθείσα διαφορά που προκύπτει μετά από την ολοκλήρωση της μελέτης. Έτσι, η ανάλυση της ισχύος πρέπει να πραγματοποιείται κατά το σχεδιασμό μιας μελέτης με σκοπό η μελέτη να διαθέτει επαρκή ισχύ για την ανάδειξη ως στατιστικά σημαντικής της ελάχιστης διαφοράς που μπορεί να θεωρηθεί ως κλινικά σημαντική.

Η ακρίβεια στον υπολογισμό του αναγκαίου αριθμού πασχόντων εξαρτάται από την ακρίβεια στις εκτιμήσεις των παραμέτρων που υπάρχουν στις ισότητες 1 και 2 και επηρεάζουν τον αναγκαίο αυτόν αριθμό. Έτσι, οι αριθμοί που προκύπτουν από τις ισότητες 1 και 2 πρέπει πάντα να θεωρούνται εκτιμήσεις ενός ελάχιστου αριθμού ατόμων που πρέπει να συμμετάσχουν σε μια μελέτη. Συνήθως σχεδιάζονται μελέτες με μεγαλύτερο αριθμό ατόμων από

εκείνον που υπολογίζεται με τις ισότητες 1 και 2, καθώς τις περισσότερες φορές είναι αδύνατη η παρακολούθηση όλων των συμμετεχόντων έως την ολοκλήρωση της μελέτης. Η εκτίμηση του αναγκαίου αριθμού πασχόντων πρέπει να γίνεται νωρίς κατά το σχεδιασμό μιας μελέτης, οπότε είναι δυνατόν να πραγματοποιηθούν οι απαιτούμενες τροποποιήσεις.

Τέλος, όπως προαναφέρθηκε, χρησιμοποιούνται διάφορες στατιστικές μέθοδοι για τον υπολογισμό του αναγκαίου αριθμού πασχόντων σε μια ΤΚΔ, γεγονός που επιτρέπει την κατάρτιση πινάκων (όπως, για παράδειγμα, οι πίνακες 5 και 6) που παρέχουν απευθείας τον απαιτούμενο κατά περίπτωση αριθμό πασχόντων.

Πίνακας 5. Αριθμός περιπτώσεων που χρειάζονται σε κάθε ομάδα για να τεκμηριωθεί με πιθανότητα 80% (ισχύς=0,80) η στατιστική σημαντικότητα της διαφοράς δύο ποσοστών στο επίπεδο σημαντικότητας 0,05 ($\alpha=0,05$), όταν διεξάγεται αμφίπλευρος έλεγχος.*

		Ποσοστό στην ομάδα Α																			
		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	
Ποσοστό στην ομάδα Β	0,05																				
	0,1	435																			
	0,15	140	686																		
	0,2	75	199	906																	
	0,25	49	100	250	1.094																
	0,3	35	62	120	293	1.251															
	0,35	27	43	72	138	329	1.377														
	0,4	21	32	49	81	152	356	1.471													
	0,45	17	24	35	54	88	162	376	1.534												
	0,5	14	19	27	38	58	93	169	387	1.565											
	0,55	12	16	21	29	41	60	96	173	391	1.565										
	0,6	10	13	17	22	30	42	61	97	173	387	1.534									
	0,65	9	11	14	18	23	31	42	61	96	169	376	1.471								
	0,7	7	9	11	14	18	23	31	42	60	93	162	356	1.377							
	0,75	6	8	10	12	14	18	23	30	41	58	88	152	329	1.251						
	0,8	5	7	8	10	12	14	18	22	29	38	54	81	138	293	1.094					
	0,85	5	6	7	8	10	11	14	17	21	27	35	49	72	120	250	906				
	0,9	4	5	6	7	8	9	11	13	16	19	24	32	43	62	100	199	686			
	0,95	3	4	5	5	6	7	9	10	12	14	17	21	27	35	49	75	140	435		

* Στην περίπτωση σύγκρισης δύο θεραπευτικών παρεμβάσεων, ο αριθμός περιπτώσεων κατά ομάδα είναι ο αριθμός των περιπτώσεων σε κάθε ομάδα και τα ποσοστά εκφράζουν την αποτελεσματικότητα των συγκρινόμενων θεραπευτικών παρεμβάσεων (π.χ. ποσοστό n-ετούς επιβίωσης, θνητότητα κ.ά.).

Πίνακας 6. Αριθμός περιπτώσεων που χρειάζονται σε κάθε ομάδα για να τεκμηριωθεί με πιθανότητα 90% (ισχύς=0,90) η στατιστική σημαντικότητα της διαφοράς δύο ποσοστών στο επίπεδο σημαντικότητας 0,05 ($\alpha=0,05$), όταν διεξάγεται αμφίπλευρος έλεγχος.*

		Ποσοστό στην ομάδα Α																			
		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	
Ποσοστό στην ομάδα Β	0,05																				
	0,1	581																			
	0,15	187	918																		
	0,2	100	266	1.212																	
	0,25	65	133	334	1.464																
	0,3	46	82	161	392	1.674															
	0,35	35	57	96	184	439	1.843														
	0,4	28	42	65	108	203	476	1.969													
	0,45	22	32	47	72	117	217	502	2.053												
	0,5	19	25	36	51	77	124	226	518	2.095											
	0,55	15	21	28	38	54	80	128	231	523	2.095										
	0,6	13	17	22	29	40	56	82	129	231	518	2.053									
	0,65	11	14	18	23	30	41	56	82	128	226	502	1.969								
	0,7	9	12	15	19	24	31	41	56	80	124	217	476	1.843							
	0,75	8	10	12	15	19	24	30	40	54	77	117	203	439	1.674						
	0,8	7	8	10	12	15	19	23	29	38	51	72	108	184	392	1.464					
	0,85	6	7	8	10	12	15	18	22	28	36	47	65	96	161	334	1.212				
	0,9	5	6	7	8	10	12	14	17	21	25	32	42	57	82	133	266	918			
	0,95	4	5	6	7	8	9	11	13	15	19	22	28	35	46	65	100	187	581		

* Στην περίπτωση σύγκρισης δύο θεραπευτικών παρεμβάσεων, ο αριθμός περιπτώσεων κατά ομάδα είναι ο αριθμός των περιπτώσεων σε κάθε ομάδα και τα ποσοστά εκφράζουν την αποτελεσματικότητα των συγκρινόμενων θεραπευτικών παρεμβάσεων (π.χ. ποσοστό n-ετούς επιβίωσης, θνητότητα κ.ά.).

ABSTRACT

Number of patients needed in randomized clinical trials

P. GALANIS

Center for Health Services Management and Evaluation, Department of Nursing, University of Athens, Athens, Greece

Archives of Hellenic Medicine 2009, 26(4):544–557

Randomized clinical trials are the most widely accepted approach for comparing the benefits of alternative treatments. In designing such studies, formal estimations of sample size are required in order to ensure that the probability of missing an important difference is small, to reduce unnecessary cost and to reduce wastage. An appropriate sample size generally depends on five study design parameters: the level of statistical significance considered appropriate (alpha level), the statistical power of the test, the minimum expected difference (effect size) between the compared groups that would be regarded as clinically significant, the estimated measurement variability and whether a one- or two-tailed statistical analysis is planned. The ability of a test to identify correctly that there is a difference between the compared groups in a trial is called the power of the test, that is, the ability of a test to reject the null hypothesis when it should be rejected. In randomized clinical trials, the statistical power is customarily set to a number greater than or equal to 0.8. If the study finds a difference between treatments, when in actuality there is no difference, then a type I error is present. If the study fails to find a difference between treatments when in actuality there is a difference, then a type II error is present. Sample size is best considered early in the planning of a study, when modifications in study design can still be made.

Key words: Effect size, Randomized clinical trials, Statistical power, Type I error, Type II error

Βιβλιογραφία

1. DEVANE D, BEGLEY C, CLARKE M. How many do I need? Basic principles of sample size estimation. *J Adv Nurs* 2004, 47:297–302
2. MOHER D, SCHULZ KF, ALTMAN DG. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet* 2001, 357:1191–1194
3. ΤΡΙΧΟΠΟΥΛΟΣ Δ. *Γενική και κλινική επιδημιολογία. Εγχειρίδιο επιδημιολογίας και αρχών κλινικής έρευνας*. Εκδόσεις Παρισιάνου, Αθήνα, 2002
4. ROTHMAN KJ. *Epidemiology. An introduction*. Oxford University Press, New York, 2002
5. ROTHMAN KJ. *Modern epidemiology*. 1st ed. Little, Brown & Co, Boston, 1986
6. ROTHMAN KJ, GREENLAND S. *Modern epidemiology*. 2nd ed. Lippincott & Raven, Philadelphia, 1998
7. ΤΡΙΧΟΠΟΥΛΟΣ Δ, ΚΑΛΑΠΟΘΑΚΗ Β, ΠΕΤΡΙΔΟΥ Ε. *Προληπτική Ιατρική και δημόσια υγεία*. Εκδόσεις Ζήτα, Αθήνα, 2000
8. BLAND M. *An introduction to medical statistics*. 2nd ed. Oxford University Press, Oxford, 1996
9. ΣΠΑΡΟΣ ΛΔ, ΛΑΜΠΡΟΥ Α, ΜΕΛΛΟΥ Κ. *Επιδημιολογία II*. Εκδόσεις ΒΗΤΑ, Αθήνα, 2005
10. MIETTINEN OS. *Theoretical epidemiology. Principles of occurrence research in medicine*. John Wiley & Sons, New York, 1985
11. ΣΠΑΡΟΣ Λ, ΓΑΛΑΝΗΣ Π, ΖΑΧΟΣ Ι, ΤΣΙΛΙΔΗΣ Κ. *Επιδημιολογία I*. Εκδόσεις ΒΗΤΑ, Αθήνα, 2004
12. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Τροποποίηση του μέτρου αποτελέσματος και σύγκριση στην εφαρμοσμένη ιατρική έρευνα. *Αρχ Ελλ Ιατρ* 2005, 22:170–177
13. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Στατιστική αλληλεπίδραση και τροποποίηση του μέτρου αποτελέσματος. *Αρχ Ελλ Ιατρ* 2005, 21:137–147
14. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Διαστρωματική ανάλυση δεδομένων. *Αρχ Ελλ Ιατρ* 2005, 21:378–384
15. ΣΠΑΡΟΣ Λ, ΓΑΛΑΝΗΣ Π. *Δοκίμια επιδημιολογίας*. Εκδόσεις Παρισιάνου, Αθήνα, 2006
16. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Διαστρωμάτωση επιδημιολογικών δεδομένων. *Αρχ Ελλ Ιατρ* 2006, 23:626–637
17. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Συστηματικά σφάλματα στις επιδημιολογικές μελέτες. *Αρχ Ελλ Ιατρ* 2007, 24:373–388
18. ΣΠΑΡΟΣ ΛΔ. *Μετα-επιδημιολογία*. Εκδόσεις ΒΗΤΑ, Αθήνα, 2001
19. BYAR DP, SIMON RM, FRIEDEWALD WT, SCHLESSELMAN JJ, DEMETS DL, ELLENBERG JH ET AL. Randomized clinical trials. Perspectives on some recent ideas. *N Engl J Med* 1976, 295:74–80
20. PETO R, PIKE MC, ARMITAGE P, BRESLOW NE, COX DR, HOWARD SV ET AL. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976, 34:585–612
21. SCHULZ KF. Randomized trials, human nature, and reporting guidelines. *Lancet* 1996, 348:596–598
22. SCHULZ KF. Subverting randomization in controlled trials. *JAMA* 1995, 274:1456–1458
23. GREENBERG RS, DANIELS SR, FLANDERS DW, ELEY WJ, BORING JR. *Medical epidemiology*. Prentice-Hall International, London, 1993
24. WORLD MEDICAL ASSOCIATION DECLARATION OF HELSINKI. Ethical principles for medical research involving human subjects. The World Medical Association, Inc (October 2000 amendment)
25. BURNS N, GROVE SK. *The practice of nursing research: Conduct, critique and utilization*. 3rd ed. WB Saunders, London, 1997
26. POLIT DF, HUNGLER BP. *Nursing research: Principles and methods*. 6th ed. Lippincott, New York, 1999
27. ENG J. Sample size estimation: How many individuals should be studied? *Radiology* 2003, 227:309–313
28. MACHIN D, CAMPBELL M, FAYERS P, PINL P. *Sample size tables for clinical studies*. 2nd ed. Blackwell Science, Oxford, 1997
29. PAGANO M, GAUVREAU K. *Principles of biostatistics*. 2nd ed. Duxbury Press, Pacific Grove, 2000
30. DANIEL WW. *Biostatistics: A foundation for analysis in the health sciences*. 7th ed. Wiley, New York, 1999
31. ALTMAN DG. *Practical statistics for medical research*. Chapman & Hall, London, 1991
32. MOHER D, DULBERG CS, WELLS GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994, 272:122–124
33. FREIMAN JA, CHALMERS TC, SMITH H, KUEBLER RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: Survey of the randomized control trial: Survey of 71 “negative” trials. *N Engl J Med* 1978, 299:690–694
34. FRIEDMAN LM, FURBERG CD, DeMETS DL. *Fundamentals of clinical trials*. 3rd ed. Springer, New York, 1998
35. ΓΑΛΑΝΗΣ Π, ΣΠΑΡΟΣ Λ. Στατιστικά μοντέλα για την ανάλυση των επιδημιολογικών δεδομένων. *Αρχ Ελλ Ιατρ* 2006, 23:404–417
36. PALLANT J. *SPSS survival manual: A step guide to data analysis using SPSS*. Open University Press, Buckingham, 2001
37. ANTHONY D. *Understanding advanced statistics: A guide for nurses and health care researchers*. Churchill Livingstone, London, 1999
38. BROWNER WS, NEWMAN TB. Are all significant P values created equal? The analogy between diagnostic tests and clinical research. *JAMA* 1987, 257:2459–2463
39. COHEN J. *Statistical power analysis for the behavioural sciences*. 2nd ed. Lawrence Erlbaum Associate Publishers, London, 1988
40. MURRAY GD. Commentary: Research governance must focus on research training. *Br Med J* 2001, 322:1461–1462
41. NESHEIM B. Commentary: Approach to power calculations has to be realistic. *Br Med J* 2001, 322:1462
42. ROSNER B. *Fundamentals of biostatistics*. 5th ed. Duxbury Press, Pacific Grove, 2000
43. FEINSTEIN AR. *Principles of medical statistics*. CRC, Boca Raton, 2002
44. BROWNER WS, NEWMAN TB, CUMMINGS SR, HULLEY SB. Estimating sample size and power. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB (eds) *Designing clinical research*. 2nd ed. Lippincott Williams & Wilkins, Philadelphia, 1993

Philadelphia, 2001:65–84

45. FLEISS JL. *Statistical methods for rates and proportions*. 2nd ed. Wiley, New York, 1981

46. HOENIG JM, HEISEY DM. The abuse of power: The pervasive fallacy of power calculations for data analysis. *Am Stat*

2001, 55:19–24

Corresponding author:

P. Galanis, 14 Dikis street, GR-157 73 Athens, Greece
e-mail: pegalan@nurs.uoa.gr

