

Η Αξιοπιστία και η Εγκυρότητα των Εργαλείων Μέτρησης σε Ποσοτικές Μελέτες

Χριστίνα Ουζούνη, Κωνσταντίνος Νακάκης

Validity and Reliability of Measurement Instruments in Quantitative Studies

Abstract at the end of the article

Καθηγητές Εφαρμογών,
Τμήμα Νοσηλευτικής,
Τεχνολογικό Εκπαιδευτικό Ίδρυμα Λαμίας
Τμήμα Νοσηλευτικής, Τεχνολογικό
Εκπαιδευτικό Ίδρυμα Λαμίας

Υποβλήθηκε: 13.4.2011
Εγκρίθηκε: 24.5.2011

Υπεύθυνη αλληλογραφίας:
Χριστίνα Ουζούνη
Τμήμα Νοσηλευτικής
Τεχνολογικό Εκπαιδευτικό Ίδρυμα Λαμίας
3ο χλμ. Π.Ε.Ο. Λαμίας-Αθηνών
351 00 Λαμία
Τηλ.: 2231 060 178
e-mail: christinaouzouni@yahoo.com

Για να υπάρχουν ερευνητικές μελέτες με έγκυρα αποτελέσματα, είναι αναγκαίο να χρησιμοποιούνται για τη διεξαγωγή τους ερευνητικά εργαλεία, τα οποία προηγουμένως έχουν ελεγχθεί για την εγκυρότητα και την αξιοπιστία τους. Γι' αυτό, κάθε ερευνητής όταν μελετά μια έρευνα, εστιάζει στην αξιοπιστία και στην εγκυρότητα του οργάνου μέτρησης της μελέτης. Η αξιοπιστία αναφέρεται στη συνέπεια μέσω της επαναληπτικότητας και της αναπαραγωγισιμότητας μιας σειράς μετρήσεων που οδηγούν στο ίδιο αποτέλεσμα, στη συνοχή και στην ομοιογένεια ενός εργαλείου μέτρησης, καθώς και στο βαθμό που είναι απαλλαγμένο από το τυχαίο σφάλμα. Η εκτίμηση της αξιοπιστίας γίνεται με τον έλεγχο-επανάλεγχο, την αξιοπιστία εναλλακτικών τύπων, την αξιοπιστία των ημίκλαστων ή ημίσεων, την εσωτερική συνοχή, τις μετρήσεις του ίδιου παρατηρητή/βαθμολογητή και τις μετρήσεις μεταξύ παρατηρητών/βαθμολογητών. Η εγκυρότητα αφορά στην εκτίμηση, του κατά πόσο ένα εργαλείο μέτρησης μετράει αυτό για το οποίο υποστηρίζει ότι σκοπεύει να μετρήσει και συνδέεται με την ύπαρξη του συστηματικού σφάλματος. Ανάλογα με τον τρόπο προσέγγισης που επιλέγεται να εκτιμηθεί η εγκυρότητα ενός εργαλείου μέτρησης χρησιμοποιούνται διαφορετικοί τύποι εγκυρότητας, όπως «περιεχομένου», «όψης», «συντρέχουσα-προβλεπτική» (με βάση κάποιο κριτήριο) και «εννοιολογικής κατασκευής» (παραγοντική, μέθοδος των γνωστών ομάδων, συγκλίνουσα, αποκλίνουσα ή διακρίνουσα) εγκυρότητα. Ο σκοπός του παρόντος άρθρου συνεχιζόμενης εκπαίδευσης είναι να παρουσιάσει με παραδείγματα τους πλέον σημαντικούς τύπους αξιοπιστίας και εγκυρότητας των εργαλείων μέτρησης σε ποσοτικές μελέτες.

Λέξεις ευρετηρίου: Αξιοπιστία, εγκυρότητα, εργαλεία μέτρησης, κλίμακες, όργανα μέτρησης

1. Εισαγωγή
2. Αξιοπιστία
 - 2.1 Αξιοπιστία επαναληπτικών μετρήσεων
 - 2.2 Αξιοπιστία εναλλακτικών τύπων ή εναλλακτικής φόρμας
 - 2.3 Αξιοπιστία των ημίκλαστων ή ημίσεων
 - 2.4 Αξιοπιστία εσωτερικής συνοχής ή συνάφειας
 - 2.5 Αξιοπιστία μετρήσεων του ίδιου παρατηρητή/βαθμολογητή και αξιοπιστία μετρήσεων μεταξύ παρατηρητών/βαθμολογητών
3. Εγκυρότητα
 - 3.1 Εγκυρότητα περιεχομένου
 - 3.2 Φαινομενική εγκυρότητα ή εγκυρότητα όψης
 - 3.3 Εγκυρότητα κριτηρίου
 - 3.3.1 Συντρέχουσα ή ταυτόχρονη εγκυρότητα
 - 3.3.2 Προβλεπτική εγκυρότητα
 - 3.4 Εγκυρότητα εννοιολογικής κατασκευής ή δομική εγκυρότητα
 - 3.4.1 Παραγοντική εγκυρότητα
 - 3.4.2 Μέθοδος των γνωστών ομάδων
 - 3.4.3 Συγκλίνουσα εγκυρότητα
 - 3.4.4 Αποκλίνουσα ή διακρίνουσα εγκυρότητα
4. Συμπεράσματα

1. Εισαγωγή

Η αναζήτηση της αξιοπιστίας και της εγκυρότητας των οργάνων μέτρησης είναι δύο βασικά κριτήρια για την εξασφάλιση έγκυρων αποτελεσμάτων, στις έρευνες που πραγματοποιούνται στο χώρο της υγείας.¹

Ένα όργανο μέτρησης δεν μπορεί να χρησιμοποιηθεί σε ερευνητική μελέτη, αν δεν έχει παρουσιάσει ένα ικανοποιητικό επίπεδο αξιοπιστίας (reliability) και εγκυρότητας (validity).² Στο παρελθόν, οι ερευνητές αντιμετώπιζαν δυσκολίες στην εύρεση εργαλείων μέτρησης για κλινικά θέματα ή φαινόμενα που μελετούσαν και διαμόρφωναν οι ίδιοι όργανα μέτρησης. Σε αυτή την περίπτωση, ήταν αναγκαίο να υποστηρίξουν ότι το όργανο μέτρησης που είχαν διαμορφώσει διέθετε αξιοπιστία και εγκυρότητα.³ Ωστόσο, σήμερα υπάρχει πληθώρα διαθέσιμων οργάνων μέτρησης και κατά συνέπεια παρέχεται η δυνατότητα στους ερευνητές να χρησιμοποιούν ήδη υπάρχοντα και δοκιμασμένα ερευνητικά εργαλεία που είναι σταθμισμένα. Η χρήση ερευνητικών εργαλείων που επανειλημμένα

έχουν χρησιμοποιηθεί εμφανίζει πολλά πλεονεκτήματα, επειδή μέσα από την εφαρμογή τους σε έρευνες έχουν βελτιωθεί και έχει επαναξιολογηθεί η αξιοπιστία και η εγκυρότητά τους.³

Είναι σημαντικό να κατανοηθεί από ένα μελλοντικό ερευνητή η διαφορά ανάμεσα στις έννοιες «αξιοπιστία» και «εγκυρότητα».

Η υψηλή αξιοπιστία ενός εργαλείου μέτρησης συνδέεται με την ελαχιστοποίηση του τυχαίου σφάλματος.² Επιπρόσθετα, αναφέρεται στη «συνοχή», στη «συνέπεια» και στη «σταθερότητα» που εμφανίζει ένα ερευνητικό εργαλείο, ώστε η μεταβλητότητα των αποτελεσμάτων να είναι μικρή, αν επαναληφθεί η μέτρηση κάτω από όμοιες ή σχεδόν όμοιες συνθήκες.

Η εγκυρότητα αναφέρεται στο κατά πόσο ένα όργανο μέτρησης μετράει ό,τι υποστηρίζει ότι μετράει. Ο Oppenheim χρησιμοποιεί το παράδειγμα του ρολογιού για να αποσαφηνίσει τη σχέση μεταξύ εγκυρότητας και αξιοπιστίας.⁴ Ένα ρολόι θεωρείται ότι είναι έγκυρο όταν δείχνει τη σωστή ώρα και αξιόπιστο όταν δείχνει σταθερά σωστή ώρα καθόλη τη διάρκεια της λειτουργίας του. Ο παραλληλισμός της εγκυρότητας και της αξιοπιστίας με την ώρα που δείχνει ένα ρολόι, βοηθάει επίσης να κατανοηθεί ότι ένα εργαλείο μέτρησης μπορεί να έχει υψηλή αξιοπιστία αλλά σε μερικές περιπτώσεις μειωμένη εγκυρότητα. Για παράδειγμα, ένα ρολόι που δείχνει σταθερά την ώρα κατά 10 min πιο μπροστά είναι αξιόπιστο, γιατί με συνέπεια και σταθερότητα δείχνει 10 min πάντα πιο μπροστά από την ακριβή ώρα. Ωστόσο, δεν είναι έγκυρο γιατί δε δείχνει ποτέ τη σωστή ώρα. Συνεπώς, ο βαθμός αξιοπιστίας ενός εργαλείου μέτρησης δε διασφαλίζει και το βαθμό εγκυρότητάς του.

Η προσεκτική δημιουργία ενός εργαλείου μέτρησης μπορεί να του προσδώσει υψηλή αξιοπιστία, αλλά είναι επίσης υποχρέωση του ερευνητή που θα το δομήσει να διασφαλίσει ταυτόχρονα και την εγκυρότητά του.

2. Αξιοπιστία

Η αξιοπιστία είναι το πρώτο χαρακτηριστικό που θα πρέπει να διαθέτει ένα εργαλείο μέτρησης και αναφέρεται στη σταθερότητα που εμφανίζει σε διαδοχικές μετρήσεις.⁵⁻⁷ Ένα εργαλείο μέτρησης θεωρείται αξιόπιστο όταν σε επαναλαμβανόμενες μετρήσεις σε ίδιο δείγμα και σε διαφορετικές χρονικές στιγμές, εμφανίζει σταθερά τα ίδια αποτελέσματα, εκτός εάν έχει συμβεί μια σημαντική αλλαγή μεταξύ των μετρήσεων. Η αξιοπιστία μπορεί να αποδοθεί εννοιολογικά με τους όρους «σταθερότητα» και «εσωτερική συνοχή» που αφορά στις παραμέτρους τις οποίες θα πρέπει οι ερευνητές να εξετάζουν, προκειμένου

να χρησιμοποιήσουν ένα εργαλείο μέτρησης στην πράξη.⁸ Επίσης, ένα όργανο μέτρησης θεωρείται αξιόπιστο στο βαθμό που είναι απαλλαγμένο από το τυχαίο σφάλμα.^{2,8}

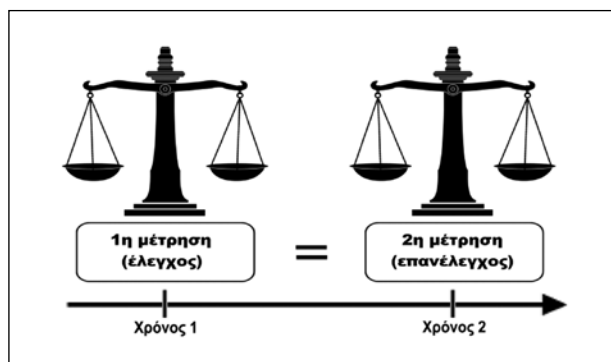
Στατιστικά, η αξιοπιστία εκτιμάται με το συντελεστή συσχέτισης r (correlation coefficient).⁹ Ο συντελεστής συσχέτισης (r) κυμαίνεται από την τιμή 0 σύμφωνα με την οποία το εργαλείο μέτρησης δεν είναι αξιόπιστο, μέχρι την τιμή 1.0 που δείχνει ότι διαθέτει τη μέγιστη αξιοπιστία. Όσο ο συντελεστής συσχέτισης προσεγγίζει την τιμή 1.0 ($r=1.0$) τόσο μεγαλύτερη αξιοπιστία θεωρείται ότι διαθέτει ένα όργανο μέτρησης. Ως αποδεκτό επίπεδο αξιοπιστίας θεωρείται ο συντελεστής συσχέτισης $r \geq 0.70$.⁹

Παράδειγμα: Ένα άτομο ανεβαίνει σε μια ζυγαριά που δείχνει ότι είναι 89 kg. Επαναλαμβάνει άμεσα δύο επόμενες διαδοχικές μετρήσεις του βάρους του στην ίδια ζυγαριά, η οποία δείχνει ότι είναι 97 kg και 72 kg. Σε αυτή την περίπτωση θεωρείται ότι η ζυγαριά δεν είναι αξιόπιστη. Αντίθετα, εάν πραγματοποιηθούν επαναλαμβανόμενες διαδοχικές μετρήσεις και η ζυγαριά δείχνει σταθερά το ίδιο βάρος σώματος, τότε θεωρείται ότι είναι αξιόπιστη. (Ωστόσο, εάν η ένδειξη που δείχνει η ζυγαριά επαναληπτικά δεν είναι το ακριβές πραγματικό βάρος, τότε δε θεωρείται έγκυρη.)

Χρησιμοποιούνται διάφορα είδη αξιοπιστίας ενός οργάνου μέτρησης. Τα πλέον συνήθη είδη αξιοπιστίας είναι των επαναληπτικών μετρήσεων, των εναλλακτικών τύπων, των ημίκλαστων ή ημίσεων, της εσωτερικής συνοχής ή συνάφειας, των μετρήσεων του ίδιου παρατηρητή/βαθμολογητή και μεταξύ παρατηρητών/βαθμολογητών.

2.1 Αξιοπιστία επαναληπτικών μετρήσεων

Για να εξακριβωθεί εάν ένα εργαλείο μέτρησης διαθέτει σταθερότητα αναφορικά με τις μετρήσεις, χρησιμοποιείται η αξιοπιστία επαναληπτικών μετρήσεων ή ο έλεγχος-επανελέγχος (test-retest reliability).^{9,10} Συγκεκριμένα, ένα εργαλείο μέτρησης χρησιμοποιείται σε ένα δείγμα (έλεγχος, test) που έχει επιλεγεί και σε ένα διάστημα χρόνου που μεσολαβεί επαναχορηγείται στο ίδιο ακριβώς δείγμα (επανελέγχος, retest) και κάτω από τις ίδιες συνθήκες (εικόνα 1). Ο βαθμός αξιοπιστίας είναι ο συντελεστής r , ο οποίος προκύπτει από τη συσχέτιση των δύο βαθμολογιών (scores) που προέρχονται από τις δύο μετρήσεις.⁹ Επισημαίνεται ότι η χρονική απόσταση από τη μία μέτρηση στην άλλη (test-retest) δε θα πρέπει να είναι μεγάλη. Διαφορετικά, δε θα εκτιμηθεί η σταθερότητα του εργαλείου μέτρησης, καθώς και των απαντήσεων. Υπάρχει η άποψη ότι όταν μετράται η επαναληπτική αξιοπιστία (test-retest), τα άτομα που συμμετέχουν σ' αυτή εξοικειώνονται με το περιεχόμενο του εργαλείου μέτρησης και βασιζόμενοι στη μνήμη τους δίνουν τις ίδιες απαντήσεις που είχαν



Εικόνα 1. Αξιοπιστία επαναληπτικών μετρήσεων.

δώσει στην αρχική μέτρηση (test). Αυτό ονομάζεται επίδραση πράξης (practice effect/ memory effect) και είναι ένα θέμα, το οποίο οι ερευνητές λαμβάνουν υπόψη τους αναφορικά με το χρονικό διάστημα που θα μεσολαβήσει μεταξύ του ελέγχου και του επανελέγχου.⁹

Παράδειγμα: Σε έρευνα που πραγματοποιήθηκε για να μετρήσει το μετεγχειρητικό πόνο σε ένα δείγμα 12 ανδρών ($n=12$) οι οποίοι υποβλήθηκαν σε εγχείρηση κήλης, δημιουργήθηκε μια κλίμακα με 6 προτάσεις (items), καθεμιά από τις οποίες είχε διαβάθμιση του πόνου από 1–10. Το άθροισμα των απαντήσεων στις 6 προτάσεις της κλίμακας, έδινε και τη συνολική βαθμολογία που εκτιμούσε το μετεγχειρητικό πόνο. Η κλίμακα συμπληρώθηκε από το δείγμα της έρευνας τη 2η μετεγχειρητική ώρα (χρόνος 1) και την 4η μετεγχειρητική ώρα (χρόνος 2). Η επιλογή του μεσοδιαστήματος χρόνου (2 ώρες) μεταξύ των μετρήσεων ήταν κατάλληλη ώστε το επίπεδο του πόνου να είναι παρόμοιο. Μετά από τις δύο μετρήσεις του πόνου των ανδρών του δείγματος, η συνολική βαθμολογία της πρώτης μέτρησης του πόνου (χρόνος 1) συσχετίστηκε με τη συνολική βαθμολογία της δεύτερης μέτρησης (χρόνος 2) και βρέθηκε ότι ήταν $r=0.78$. Ο συντελεστής συσχέτισης $r=0.78$ δείχνει ότι η κλίμακα εκτίμησης του πόνου που δημιουργήθηκε είχε καλή αξιοπιστία στον επαναληπτικό έλεγχο (test-retest reliability) εκτίμησης του άμεσου μετεγχειρητικού πόνου.

2.2 Αξιοπιστία εναλλακτικών τύπων ή αξιοπιστία της εναλλακτικής φόρμας

Η αξιοπιστία εναλλακτικών τύπων ή αξιοπιστία της εναλλακτικής φόρμας (alternate form reliability ή parallel-forms reliability) παρέχει τη δυνατότητα να αποφευχθεί η επίδραση πράξης (practice effect), η οποία χαρακτηρίζει συχνά τις επαναλαμβανόμενες μετρήσεις, δηλαδή όταν οι ερωτώμενοι δίνουν τις ίδιες απαντήσεις επειδή θυμούνται αυτές που είχαν δώσει σε προηγούμενη μέτρηση.^{2,9} Ο πιο

συχνός τρόπος αξιοπιστίας εναλλακτικών τύπων είναι η αλλαγή της σειράς επιλογής των απαντήσεων σε κάθε πρόταση του εργαλείου μέτρησης. Εναλλακτικά, οι προτάσεις διατυπώνονται διαφορετικά, σε μια ωστόσο λειτουργικά ισοδύναμη μορφή, χωρίς να αλλάξει το χαρακτηριστικό που μετράται. Επισημαίνεται, ότι οι προτάσεις του εργαλείου μέτρησης θα πρέπει να διαφέρουν μόνο ως προς τη διατύπωση και όχι ως προς το νοηματικό περιεχόμενο και αυτό επιτυγχάνεται με τη χρήση παρόμοιου λεξιλογίου, με ίδιο νόημα και επίπεδο δυσκολίας.² Με τον τρόπο αυτόν δημιουργείται ένα εργαλείο μέτρησης παρόμοιο με εκείνο που χρησιμοποιήθηκε αρχικά, το οποίο μετράει την ίδια μεταβλητή χωρίς ωστόσο να είναι πανομοιότυπο. Τα δύο εργαλεία μέτρησης χορηγούνται στα ίδια άτομα σε διαφορετικό χρόνο, ενώ η συνολική βαθμολογία της κάθε κλίμακας συσχετίζεται με τη συνολική βαθμολογία της άλλης.⁹ Εάν ο συντελεστής συσχέτισης που προκύπτει είναι υψηλός, το εργαλείο της έρευνας θεωρείται ότι έχει αξιοπιστία εναλλακτικών τύπων.⁹

Παράδειγμα: Σε έρευνα που πραγματοποιήθηκε για να μετρήσει το stress σε ένα δείγμα νοσηλευτών, χορηγήθηκε σταθμισμένη κλίμακα για τη συλλογή των στοιχείων. Για την εκτίμηση της επαναληπτικής αξιοπιστίας (retest) της κλίμακας και για να αποφευχθεί η επίδραση πράξης (practice effect), η εξοικείωση δηλαδή των ερωτώμενων με το περιεχόμενο της κλίμακας, έγινε αλλαγή στη σειρά των επιλογών στις απαντήσεις κάθε πρότασης και χορηγήθηκε το ερωτηματολόγιο στο ίδιο δείγμα 2 εβδομάδες μετά από την αρχική μέτρηση. Ενδεικτικά, παρατίθεται μία πρόταση της αρχικής κλίμακας που χρησιμοποιήθηκε στην 1η μέτρηση (πίνακας 1) και η τροποποιημένη σειρά στην επιλογή των απαντήσεων, όπως χρησιμοποιήθηκε μετά από 2 εβδομάδες στον επανέλεγχο της κλίμακας, στη 2η μέτρηση (πίνακας 2).

2.3 Αξιοπιστία των ημίκλαστων ή ημίσεων

Στην αξιοπιστία των ημίκλαστων (split-half reliability) χωρίζονται όλες οι προτάσεις (items) ενός εργαλείου μέτρησης που θεωρείται ότι μετρούν την ίδια μεταβλητή,

σε δύο ομάδες δεδομένων.^{10,11} Το εργαλείο μέτρησης χορηγείται στο ίδιο δείγμα και η συνολική βαθμολογία κάθε συμμετέχοντα στην έρευνα διαιρείται τυχαία σε δύο ομάδες. Συνήθως, ακολουθείται ως πρακτική, η διαίρεση των ερωτήσεων σε μονές-ζυγές.⁹ Ο συντελεστής συσχέτισης που εφαρμόζεται ανάμεσα στη βαθμολογία (score) των δύο ομάδων είναι η αξιοπιστία των ημίκλαστων (split-half reliability). Εάν ο συντελεστής συσχέτισης εμφανίζει παρόμοια τιμή στα δύο ημίκλαστα/ημίσεια, το εργαλείο μέτρησης θεωρείται αξιόπιστο.¹⁰

Παράδειγμα: Σε έρευνα που πραγματοποιήθηκε για να εκτιμηθεί ο πόνος και η αντιμετώπισή του σε 30 ασθενείς ($n=30$) με μεταστατικό καρκίνο, οι ερευνητές για να ελέγξουν την αξιοπιστία της κλίμακας που μετράει την εμπειρία του πόνου, διαίρεσαν τις καταχωρημένες απαντήσεις των ερωτώμενων στο στατιστικό πρόγραμμα, σε μονές-ζυγές και υπολόγισαν το συντελεστή αξιοπιστίας στις δύο ομάδες δεδομένων. Από τον υπολογισμό του συντελεστή συσχέτισης βρέθηκε ότι τα δύο ημίκλαστα εμφάνισαν τιμές 0.85 και 0.84, επικυρώνοντας την αξιοπιστία ημίσεων της κλίμακας.

2.4 Αξιοπιστία εσωτερικής συνοχής ή συνάφειας

Χρησιμοποιείται για να μετρήσει την αξιοπιστία σε εργαλεία μέτρησης. Μπορεί να εφαρμοστεί όχι μόνο σε μία πρόταση (item) ή σε μία υποκλίμακα, αλλά και στο σύνολο του εργαλείου που μετράει την ίδια έννοια (μεταβλητή).^{9,12} Η αξιοπιστία εσωτερικής συνοχής (internal consistency) είναι ένας δείκτης που φανερώνει κατά πόσο διαφορετικές προτάσεις (items) μετρούν την ίδια έννοια (μεταβλητή).

Η αξιοπιστία εσωτερικής συνοχής τόσο σε μια υποκλίμακα, όσο και σε ολόκληρη κλίμακα εκτιμάται με το συντελεστή Cronbach's alpha που δείχνει την ομοιογένεια μιας κλίμακας.⁹ Για να θεωρείται αποδεκτή η τιμή του Cronbach's alpha θα πρέπει να είναι >0.7 .¹² Όσο μεγαλύτερη είναι η τιμή του συντελεστή Cronbach's alpha τόσο μεγαλύτερη είναι η αξιοπιστία εσωτερικής

Πίνακας 1. Πρώτη μέτρηση (έλεγχος, test) εκτίμησης του stress των νοσηλευτών.

Πρόταση ερωτηματολογίου εκτίμησης του stress	Ποτέ	Μερικές ημέρες	Κάθε ημέρα
Κατά τη διάρκεια των προηγούμενων 4 εβδομάδων αισθανόμουν πιεσμένος στο χώρο της εργασίας μου	0	1	2

Πίνακας 2. Δεύτερη μέτρηση (επανέλεγχος, retest) εκτίμησης του stress των νοσηλευτών.

Πρόταση ερωτηματολογίου εκτίμησης του stress	Κάθε ημέρα	Μερικές ημέρες	Ποτέ
Κατά τη διάρκεια των προηγούμενων 4 εβδομάδων αισθανόμουν πιεσμένος στο χώρο της εργασίας μου	2	1	0

συνοχής.⁹ Εάν μια κλίμακα εμφανίζει μικρού βαθμού εσωτερική συνοχή, ενδεχομένως μπορεί να βελτιωθεί με την προσθήκη προτάσεων (items) ή με την επανεξέταση της σαφήνειας των προτάσεων της. Όταν ένα εργαλείο μέτρησης εφαρμόζεται για πρώτη φορά σε συγκεκριμένο πληθυσμό, με διαφορετικά πολιτισμικά χαρακτηριστικά και γλώσσα, τότε επιβάλλεται ο έλεγχος της αξιοπιστίας εσωτερικής συνοχής.⁹

Παράδειγμα: Σε κλίμακα εκτίμησης του επαγγελματικού stress, η οποία περιλαμβάνει 8 υποκλίμακες, γίνεται η εκτίμηση της εσωτερικής συνοχής/συνάφειας (internal consistency) προκειμένου να υπολογιστεί ο συντελεστής Cronbach's alpha, τόσο για κάθε υποκλίμακα χωριστά όσο και για το σύνολο της κλίμακας. Ο συντελεστής Cronbach's alpha θα πρέπει να είναι ≥ 0.7 ώστε η κλίμακα να διαθέτει ικανοποιητικό επίπεδο εσωτερικής συνοχής.

2.5 Αξιοπιστία μετρήσεων του ίδιου παρατηρητή/βαθμολογητή και αξιοπιστία μετρήσεων μεταξύ παρατηρητών/βαθμολογητών

Μερικά όργανα μέτρησης δεν είναι αυτοσυμπληρούμενα, αλλά εργαλεία μέτρησης μιας παρατηρούμενης συμπεριφοράς ή φαινομένου. Τα εργαλεία που βασίζουν τη μέτρησή τους στην παρατήρηση, χρειάζεται να ελέγχονται για την αξιοπιστία μεταξύ βαθμολογητών. Σε αυτή την περίπτωση πραγματοποιείται επαναληπτική μέτρηση από τον παρατηρητή/βαθμολογητή μετά από μικρό χρονικό διάστημα (2–3 εβδομάδες). Από τις δύο μετρήσεις του ίδιου παρατηρητή προκύπτουν δύο ομάδες δεδομένων που συγκρίνονται μεταξύ τους. Η συσχέτιση που προκύπτει μεταξύ των δύο ομάδων δεδομένων, ονομάζεται επανέλεγχος αξιοπιστίας του ίδιου παρατηρητή/βαθμολογητή.¹²

Η αξιοπιστία μεταξύ των διαφορετικών βαθμολογητών/παρατηρητών (inter-rater reliability, inter-observer reliability) αφορά στο μεταξύ τους έλεγχο αξιοπιστίας. Συγκεκριμένα, αναφέρεται στο βαθμό που συμφωνούν οι μετρήσεις ή οι παρατηρήσεις τους αναφορικά με το προς μελέτη φαινόμενο¹³ (εικόνα 2).

Παράδειγμα: Στην προσπάθεια δημιουργίας μιας κλίμακας για την πρόληψη κινδύνου εμφάνισης ελκών κατάκλισης σε νοσηλεύομενους ασθενείς, διεξήχθησαν δύο έρευνες. Σε κάθε έρευνα συμπλήρωσαν το εργαλείο μέτρησης διαφορετικοί βοηθοί ερευνητών. Με την ολοκλήρωση των ερευνών συμπληρώθηκαν 88 ζεύγη παρατηρήσεων. Χρησιμοποιήθηκε ο συντελεστής Pearson για να συσχετίσει τις μετρήσεις των παρατηρήσεων μεταξύ των βαθμολογητών. Από τη συσχέτιση των παρατηρήσεων μεταξύ βαθμολογητών βρέθηκε ο



Εικόνα 2. Αξιοπιστία μετρήσεων μεταξύ δύο παρατηρητών/βαθμολογητών.

συντελεστής συσχέτισης $r=0.99$. Οι ερευνητές κατέληξαν ότι το εργαλείο πρόληψης κινδύνου εμφάνισης ελκών κατάκλισης διαθέτει αξιοπιστία μεταξύ διαφορετικών παρατηρητών και επομένως θεωρείται ότι έχει αποδεδειγμένη συνέπεια και σταθερότητα σε επαναλαμβανόμενες μετρήσεις ώστε να θεωρείται αξιόπιστο.

3. Εγκυρότητα

Επειδή ένα εργαλείο μέτρησης μπορεί να είναι αξιόπιστο αλλά όχι έγκυρο, παράλληλα με την εκτίμηση της αξιοπιστίας μιας κλίμακας ή ενός οργάνου μέτρησης θα πρέπει να ελεγχθεί και η εγκυρότητά του (validity).¹⁴ Συγκεκριμένα, όταν ένα εργαλείο μέτρησης είναι πράγματι έγκυρο, «αντανακλά» την έννοια (τη μεταβλητή) που προτίθεται ότι μετράει. Δηλαδή, ένα εργαλείο μέτρησης που δηλώνεται ότι μετράει την ανησυχία, δεν μπορεί να μετράει ταυτόχρονα και το stress.

Ένα εργαλείο μέτρησης θεωρείται έγκυρο όταν έχει χρησιμοποιηθεί επανειλημμένα με επιτυχία σε πληθυσμό για τον οποίο έχει σχεδιαστεί ερευνητικά.^{13–15}

Για τη διασφάλιση της εγκυρότητας στις έρευνες με ποσοτικά δεδομένα υπάρχουν οι ακόλουθες προσεγγίσεις: Η εγκυρότητα περιεχομένου, η φαινομενική εγκυρότητα, η εγκυρότητα κριτηρίου που περιλαμβάνει τη συντρέχουσα και την προβλεπτική εγκυρότητα και η εγκυρότητα εννοιολογικής κατασκευής ή δομική εγκυρότητα που περιλαμβάνει την παραγοντική, τη μέθοδο των γνωστών ομάδων, τη συγκλίνουσα, καθώς και την αποκλίνουσα. Στην αξιολόγηση ενός ερευνητικού άρθρου εκτιμάται η επάρκεια ενδείξεων για να διαπιστωθεί εάν ο τύπος εγκυρότητας που είχε επιλεγεί ήταν ο κατάλληλος για το εργαλείο μέτρησης.¹⁴

3.1 Εγκυρότητα περιεχομένου

Η διασφάλιση της εγκυρότητας περιεχομένου (content validity) ενός εργαλείου μέτρησης πρέπει να προηγείται της συλλογής των δεδομένων μιας έρευνας.¹⁶ Η εγκυρότητα περιεχομένου αναφέρεται στο κατά πόσο ένα εργαλείο μέτρησης καλύπτει εννοιολογικά το εύρος της μεταβλητής που μετράει. Αρχικά, ορίζεται η έννοια η οποία θα μετρηθεί και γίνεται προσπάθεια εντοπισμού των διαστάσεων που συνθέτουν την προς μέτρηση μεταβλητή, ώστε να περιλαμβάνονται στις προτάσεις (items) του εργαλείου μέτρησης που θα κατασκευαστεί.¹⁴ Όταν ολοκληρωθεί η εν λόγω διαδικασία, το εργαλείο μέτρησης υποβάλλεται για αξιολόγηση σε μια ομάδα ατόμων που θεωρούνται ειδικοί, αναφορικά με τη μεταβλητή που μετράται.² Η ομάδα των ειδικών εξετάζει κατά πόσο το σύνολο και το περιεχόμενο των προτάσεων σχετίζεται άμεσα με την έννοια που ερευνάται.¹⁶

Η εγκυρότητα περιεχομένου είναι σημαντική σε όργανα που μετρούν σύνθετα ατομικά και ψυχοκοινωνικά χαρακτηριστικά.⁶ Επισημαίνεται ότι, όταν σχεδιάζονται κλίμακες είναι σημαντικό να ερωτάται η ομάδα πληθυσμού που θα μελετηθεί για την καταλληλότητα περιεχομένου της κλίμακας.⁹

Παράδειγμα: Για τη μέτρηση της αποτελεσματικότητας της δίαιτας και της άσκησης, σε σχέση με την εμφάνιση καρδιολογικού προβλήματος αναπτύχθηκαν δύο ερευνητικά εργαλεία.¹⁷ Οι ερευνητές μετά από ανασκόπηση της βιβλιογραφίας διαμόρφωσαν ερωτηματολόγιο με 30 προτάσεις (items) που αναφέρονταν στην αποτελεσματικότητα της κατάλληλης δίαιτας και ένα άλλο με 30 προτάσεις που σχετίζονταν με την αποτελεσματικότητα της άσκησης αναφορικά με την εμφάνιση ή μη καρδιολογικού νοσήματος. Τα δύο όργανα μέτρησης υποβλήθηκαν σε ομάδα ειδικών, οι οποίοι βαθμολόγησαν κάθε πρόταση (item) σε μια πεντάβαθμη κλίμακα (από 1=μη σχετική έως 5=απόλυτα σχετική). Η βαθμολόγηση των ειδικών υπολογίστηκε για κάθε πρόταση του ερωτηματολογίου και παρέμειναν στο όργανο μέτρησης όσες προτάσεις εμφάνισαν τιμή τουλάχιστον 3. Οι 19 προτάσεις της κλίμακας εκτίμησης της αποτελεσματικότητας της δίαιτας και οι 18 προτάσεις της κλίμακας εκτίμησης της άσκησης στην πρόληψη καρδιολογικών νοσημάτων αποτέλεσαν τα τελικά εργαλεία μέτρησης.

3.2 Φαινομενική εγκυρότητα ή εγκυρότητα όψης

Ένας τύπος της εγκυρότητας περιεχομένου είναι η φαινομενική εγκυρότητα (face validity), η οποία αποτελεί μια στοιχειώδη μορφή εγκυρότητας που δείχνει ότι ένα εργαλείο μέτρησης φαίνεται ότι εκτιμά την έννοια (μετα-

βλητή) που αναφέρεται ότι μετράει.¹⁴ Στη φαινομενική εγκυρότητα, ο ερευνητής κάνοντας μια αδρή εκτίμηση του εργαλείου μέτρησης αναφορικά με την έννοια που μετράει, προσδιορίζει «με μια πρώτη ματιά» εάν το περιεχόμενο των ερωτήσεων-προτάσεων είναι εννοιολογικά σχετικό με ό,τι προτίθεται να μετρήσει. Όταν ένα ερωτηματολόγιο μετρά αντικειμενικά προφανή έννοια, είναι εύκολα κατανοητό και σχετίζεται άμεσα με το σκοπό της μελέτης.¹⁵ Συχνά, συγχέεται με την εγκυρότητα περιεχομένου. Ωστόσο, η φαινομενική εγκυρότητα είναι περισσότερο «επιφανειακή» λόγω της εμπλοκής σε μεγαλύτερο βαθμό της υποκειμενικής κρίσης, παρά της αντικειμενικής ανάλυσης.

Για την εκτίμησή της μπορεί να ζητηθεί από ειδικούς που γνωρίζουν ή ασχολούνται επαγγελματικά με την προς μέτρηση έννοια, να αξιολογηθεί εάν το εργαλείο μέτρησης στοχεύει σε ό,τι υποστηρίζει ότι μετράει.¹¹ Επισημαίνεται ότι πολλοί ερευνητές δεν τη λαμβάνουν υπόψη τους, επειδή πρόκειται για μια αδρή εκτίμηση της εγκυρότητας ενός εργαλείου μέτρησης.⁹

Παράδειγμα: Σε έρευνα που πρόκειται να χρησιμοποιηθεί κλίμακα για την εκτίμηση του επαγγελματικού stress των νοσηλευτών, η φαινομενική εγκυρότητα μπορεί να διαπιστωθεί από τον ίδιο τον ερευνητή. Συγκεκριμένα, με μια αδρή επισκόπηση μπορεί να διαπιστώσει εάν η κλίμακα περιέχει στο σύνολό της προτάσεις που αναφέρονται σε συνθήκες και καταστάσεις που ενδεχομένως προκαλούν stress στους νοσηλευτές στο χώρο εργασίας τους. Η φαινομενική εγκυρότητα της συγκεκριμένης κλίμακας βελτιώνεται με την εκτίμηση της εγκυρότητας της κλίμακας από ειδικούς που έχουν ασχοληθεί με το επαγγελματικό stress των νοσηλευτών ή χορηγώντας το σε εργαζόμενους νοσηλευτές, προκειμένου να αποφανθούν αν το περιεχόμενό του αφορά σε στρεσογόνες καταστάσεις που βιώνουν στο χώρο εργασίας τους.

3.3 Εγκυρότητα κριτηρίου

Στην εγκυρότητα κριτηρίου (criterion validity) λαμβάνεται υπόψη ένα κριτήριο βάσει του οποίου θα αποδειχθεί ότι ένα εργαλείο μέτρησης μετράει την έννοια-μεταβλητή που προτίθεται να μετρήσει. Το κριτήριο είναι συνήθως ένα δεύτερο εργαλείο μέτρησης που εκτιμά την ίδια μεταβλητή. Οι δύο υπότυποι εγκυρότητας κριτηρίου είναι η συντρέχουσα εγκυρότητα (concurrent validity) και η προβλεπτική εγκυρότητα (predictive validity).¹³

3.3.1 Συντρέχουσα ή ταυτόχρονη εγκυρότητα

Η συντρέχουσα εγκυρότητα (concurrent validity) φανερώνει το βαθμό συμφωνίας στο αποτέλεσμα, μετά από τη χρήση στην ίδια έρευνα και στην ίδια χρονική

στιγμή δύο διαφορετικών εργαλείων μέτρησης της ίδιας μεταβλητής, το ένα από τα οποία έχει ήδη σταθμιστεί και αποδεδειγμένα διαθέτει αξιοπιστία και εγκυρότητα. Η συμφωνία ως προς το αποτέλεσμα αναφέρεται στατιστικά στο βαθμό συσχέτισης των δύο εργαλείων μέτρησης της ίδιας μεταβλητής και εκφράζεται με το συντελεστή συσχέτισης r . Η εμφάνιση ισχυρού συντελεστή συσχέτισης (*correlation coefficient*) δείχνει ότι το κύριο εργαλείο μέτρησης που έχει χρησιμοποιηθεί είναι έγκυρο.¹⁸ Όταν ελέγχεται η συντρέχουσα εγκυρότητα, είναι σημαντικό να επιλέγονται ισοδύναμα εργαλεία μέτρησης που έχουν επιδείξει παρόμοιες ιδιότητες.^{9,19}

Παράδειγμα: Σε έρευνα εκτίμησης του επαγγελματικού stress σε νοσηλευτικό προσωπικό χρησιμοποιήθηκε ταυτόχρονα κλίμακα εκτίμησης του stress των νοσηλευτών, καθώς και μια άλλη σταθμισμένη κλίμακα εκτίμησης του stress των επαγγελματιών υγείας. Τα αποτελέσματα της έρευνας έδειξαν ότι το νοσηλευτικό προσωπικό και στις δύο κλίμακες εμφάνισε μέτριο βαθμό επαγγελματικού stress. Όταν συσχετίστηκαν οι βαθμολογίες (*scores*) των δύο κλιμάκων του stress, ο υψηλός συντελεστής συσχέτισης που εμφανίστηκε ($r=0.90$), επιβεβαίωσε ότι και οι δύο κλίμακες που χρησιμοποιήθηκαν διέθεταν συντρέχουσα εγκυρότητα, επιβεβαιώνοντας τα αποτελέσματα της μελέτης.

3.3.2 Προβλεπτική εγκυρότητα

Η προβλεπτική εγκυρότητα (*predictive validity*) δείχνει τη δυνατότητα ενός εργαλείου μέτρησης να προβλέψει μελλοντικά γεγονότα, συμπεριφορές, στάσεις ή αποτελέσματα. Η προβλεπτική εγκυρότητα αναφέρεται στο βαθμό συσχέτισης, από τη μέτρηση μιας μεταβλητής, με άλλες μελλοντικές μετρήσεις της ίδιας εννοιολογικά μεταβλητής.²⁰

Η προβλεπτική εγκυρότητα παρουσιάζει δυσκολίες κατά την εφαρμογή της σε προοπτικές έρευνες στις επισημικές υγείας, καθώς οι μετρήσεις χρονικά εκτείνονται σε μεγάλο χρονικό διάστημα, επειδή προκύπτουν μεταβολές στην πορεία της νόσου και στην κατάσταση του ασθενούς.⁹

Παράδειγμα: Σε 24 γυναίκες ($n=24$) οι οποίες επρόκειτο να υποβληθούν σε υστερεκτομή διενεργήθηκε έρευνα προεγχειρητικά για να ελεγχθεί η προβλεπτική εγκυρότητα του εργαλείου μέτρησης «Δείκτης Ανοχής στον Πόνο». Η μεγαλύτερη βαθμολογία στην κλίμακα αντιπροσώπευε τη μεγαλύτερη ανοχή στον πόνο. Μετά από την πραγματοποίηση της υστερεκτομής στις 24 ασθενείς, οι ερευνητές μελέτησαν τον ιατρικό φάκελο των ασθενών και εκτίμησαν τις δόσεις των ισχυρών αναλγητικών φαρμάκων που χορηγήθηκαν για την αντιμετώπιση του μετεγχειρητικού πόνου. Στη συνέχεια, αναζητήθηκε η συσχέτιση ανάμεσα στη συνολική βαθμο-

λογία της κλίμακας εκτίμησης της ανοχής στον πόνο και τη συνολική βαθμολογία των δόσεων των ισχυρών αναλγητικών που χορηγήθηκαν στο δείγμα των ασθενών. Ο συντελεστής συσχέτισης εμφάνισε ισχυρή αρνητική συσχέτιση ($r=-0.84$). Όσο δηλαδή αυξανόταν η ανοχή στον πόνο τόσο μειώνονταν οι δόσεις των αναλγητικών που χορηγήθηκαν. Συνεπώς, το εργαλείο μέτρησης «Δείκτης Ανοχής στον Πόνο» παρουσιάζει προβλεπτική εγκυρότητα.

3.4 Εγκυρότητα εννοιολογικής κατασκευής ή δομική εγκυρότητα

Η εγκυρότητα εννοιολογικής κατασκευής (*construct validity*) είναι πολύτιμη αλλά, ταυτόχρονα, και η πλέον δύσκολη στην εκτίμηση των εργαλείων μέτρησης. Συχνά, αποδίδεται σε ένα εργαλείο μέτρησης μετά από τη χρησιμοποίηση και τη βελτίωσή του σε πολλές έρευνες, σε διαφορετικό πληθυσμό και σε διάστημα κάποιων χρόνων.²¹ Επίσης, εκφράζει το βαθμό που ένα εργαλείο μέτρησης αποτυπώνει τις ιδέες ενός θεωρητικού πλαισίου ή μιας θεωρίας.⁸

Η αποτίμηση της εγκυρότητας εννοιολογικής κατασκευής ή αλλιώς της δομικής εγκυρότητας είναι μια διαδικασία που εφαρμόζεται κυρίως όταν δημιουργούνται όργανα ψυχομετρικών χαρακτηριστικών, αντιλήψεων και πεποιθήσεων. Η εγκυρότητα εννοιολογικής κατασκευής δεν είναι πάντα απευθείας παρατηρήσιμη αλλά χρειάζεται να ανιχνευτεί και γι' αυτό είναι δύσκολο να μετρηθεί. Συνήθως, είναι αναγκαίο να πραγματοποιηθεί παραγοντική ανάλυση (*factor analysis*) και να αναζητηθούν σε ένα εργαλείο μέτρησης, ομάδες προτάσεων (*items*) που σχετίζονται εννοιολογικά αλλά και στατιστικά μεταξύ τους. Στο πλαίσιο αναζήτησης της εγκυρότητας εννοιολογικής κατασκευής χρησιμοποιούνται διάφορες προσεγγίσεις: Η παραγοντική εγκυρότητα, η μέθοδος των γνωστών ομάδων, η συγκλίνουσα εγκυρότητα και η αποκλίνουσα εγκυρότητα.

Παράδειγμα: Το ερωτηματολόγιο που εκτιμά τις απόψεις για την ψυχική ασθένεια (*Opinion of Mental Illness*) έχει χρησιμοποιηθεί σε ένα μεγάλο αριθμό ερευνών, έχει υποστεί επαναληπτικά παραγοντική ανάλυση σε διαφορετικά πολιτισμικά πλαίσια και χώρους, καθώς και σε διαφορετικές πληθυσμιακές ομάδες (νοσηλευτές, πολίτες, ιατρούς, φοιτητές Νοσηλευτικής κ.λπ.). Οι παραπάνω διαδικασίες έχουν βελτιώσει το ερωτηματολόγιο και του έχουν αποδώσει εγκυρότητα εννοιολογικής κατασκευής ή δομική εγκυρότητα.

3.4.1 Παραγοντική εγκυρότητα

Η παραγοντική εγκυρότητα (*factorial validity*) ελέγχεται με την παραγοντική ανάλυση, μια σύνθετη στατιστική διαδικασία που χρησιμοποιείται για να αναζητήσει τις

υποκείμενες διαστάσεις ενός εργαλείου μέτρησης.² Με την παραγοντική ανάλυση ελέγχεται εάν οι προτάσεις ανήκουν στατιστικά σε ένα όργανο μέτρησης. Εάν η παραγοντική ανάλυση καταδεικνύει τις διαστάσεις που συμφωνούν με τον εννοιολογικό προσδιορισμό της κλίμακας, τότε θεωρείται ότι υπάρχουν ενδείξεις εγκυρότητας εννοιολογικής κατασκευής ή δομικής εγκυρότητας της κλίμακας.⁹

Παράδειγμα: Σε ένα δείγμα 388 ($n=388$) ατόμων του γενικού πληθυσμού, αφού χορηγήθηκε ερωτηματολόγιο αναφορικά με τις «Απόψεις για την Ψυχική Ασθένεια» πραγματοποιήθηκε παραγοντική ανάλυση, με σκοπό να ευρεθούν οι παράγοντες που συνθέτουν τις απόψεις των ερωτώμενων. Οι παράγοντες που αναδύθηκαν από την παραγοντική ανάλυση και οι οποίοι είχαν άμεση σχέση με τις αντιλήψεις και τη στάση του γενικού πληθυσμού απέναντι στην ψυχική νόσο ήταν πέντε: Η κοινωνική διάκριση, ο κοινωνικός περιορισμός, η κοινωνική φροντίδα, η κοινωνική ενσωμάτωση και η αιτιολογία ψυχικής νόσου. Οι παράγοντες συνδέονται και ερμηνεύουν τη στάση του γενικού πληθυσμού προς την ψυχική ασθένεια και επιβεβαιώνουν την εγκυρότητα εννοιολογικής κατασκευής του ερωτηματολογίου.

3.4.2 Μέθοδος των γνωστών ομάδων

Με τη μέθοδο των γνωστών ομάδων (known groups method) χορηγείται ένα όργανο μέτρησης σε δύο ομάδες, οι οποίες είναι γνωστό ότι διαφέρουν αναφορικά με το προς μέτρηση χαρακτηριστικό. Εάν από τα αποτελέσματα των μετρήσεων αποδειχθεί ότι οι δύο ομάδες εμφανίζουν διαφορά στην εκτίμηση του χαρακτηριστικού που μετράται, όπως αρχικά ήταν γνωστό, τότε θεωρείται ότι το όργανο μέτρησης έχει δομική εγκυρότητα.²⁰

Παράδειγμα: Από μια ομάδα ατόμων που έχουν διαγνωσμένο άσθμα και από μια ομάδα που δεν έχει διαγνωστεί άσθμα, συμπληρώνεται «η Κλίμακα Σοβαρότητας του Άσθματος». Από τα αποτελέσματα αναμένεται, τα άτομα με διάγνωση άσθματος να εμφανίζουν μεγάλη βαθμολογία (score) στην κλίμακα, ενώ, αντίθετα, τα άτομα που δεν έχουν άσθμα θα εμφανίσουν πολύ χαμηλή βαθμολογία. Με τη μέθοδο των «γνωστών ομάδων» επιβεβαιώνεται η δομική εγκυρότητα στην κλίμακα «Εκτίμηση της Σοβαρότητας του Άσθματος».

3.4.3 Συγκλίνουσα εγκυρότητα

Όταν οι μετρήσεις της ίδιας έννοιας (μεταβλητής) από δύο διαφορετικά εργαλεία μέτρησης που εκτιμούν εννοιολογικά την ίδια μεταβλητή, αποφέρουν τα ίδια αποτελέσματα, τότε στοιχειοθετείται συγκλίνουσα εγκυρότητα

(convergent validity). Η θετική συσχέτιση των βαθμολογιών μεταξύ των εργαλείων μέτρησης επιβεβαιώνει τη συγκλίνουσα εγκυρότητα και εξασφαλίζει ότι το όργανο μέτρησης που θα χρησιμοποιηθεί έχει εγκυρότητα εννοιολογικής κατασκευής ή δομική εγκυρότητα.²¹

Παράδειγμα: Σε έρευνα που πραγματοποιήθηκε σε 72 ασθενείς ($n=72$), η συνολική βαθμολογία (score) μιας πρόσφατα δομημένης κλίμακας που εκτιμά τη σοβαρότητα του άσθματος, εμφανίζει παρόμοια βαθμολογία (score) με μια «Κλίμακα Διαβάθμισης του Άσθματος», της οποίας έχει ελεγχθεί η αξιοπιστία και η εγκυρότητα. Στη συγκεκριμένη περίπτωση θεωρείται ότι η νέα αυτή κλίμακα εμφανίζει συγκλίνουσα εγκυρότητα.

3.4.4 Αποκλίνουσα ή διακρίνουσα εγκυρότητα

Η εγκυρότητα εννοιολογικής κατασκευής ή δομική εγκυρότητα μπορεί επίσης να διασφαλιστεί όταν το κύριο εργαλείο μέτρησης μιας έννοιας χρησιμοποιείται παράλληλα με ένα άλλο, το οποίο μετράει μια εννοιολογικά αντίθετη μεταβλητή.^{21,22}

Παράδειγμα: Για να εκτιμηθεί η αποκλίνουσα εγκυρότητα της κλίμακας «Ελπίδα» του Miller,¹⁴ σε ένα δείγμα 250 ατόμων ($n=250$) χρησιμοποιήθηκε παράλληλα και η κλίμακα «Απελπισία». Η συνολική βαθμολογία (score) των δύο κλιμάκων συσχετίστηκε και η αρνητική συσχέτιση που προέκυψε $r=-0.54$ έδειξε ότι η κλίμακα της «Ελπίδας» διαθέτει αποκλίνουσα εγκυρότητα.

4. Συμπεράσματα

Η αξιοπιστία και η εγκυρότητα των οργάνων μέτρησης είναι καθοριστική για την εξασφάλιση ορθών αποτελεσμάτων από τη διεξαγωγή μιας έρευνας. Όταν χρησιμοποιείται ένα εργαλείο μέτρησης, ανεξάρτητα από το εάν έχει ήδη σταθμιστεί ή έχει δημιουργηθεί για τις ανάγκες μιας έρευνας, θα πρέπει να ελεγχθεί η αξιοπιστία και η εγκυρότητά του. Επίσης, χρειάζεται να αναφερθούν τα είδη αξιοπιστίας και οι τύποι εγκυρότητας που έχουν υποστηρίξει το εργαλείο μέτρησης σε περίπτωση ανακοίνωσης ή δημοσίευσης της ερευνητικής μελέτης. Τα αποτελέσματα τέτοιων ερευνών αποτελούν σοβαρές ενδείξεις που θεωρούνται τεκμηριωμένες και μπορούν να αξιοποιηθούν κατάλληλα. Όταν δεν αναφέρεται η αξιοπιστία και η εγκυρότητα των οργάνων μέτρησης, τα αποτελέσματα μιας έρευνας είναι αποδεκτά με επιφύλαξη και η μελέτη παρουσιάζει πολλούς περιορισμούς ως προς την ορθότητα των αποτελεσμάτων της.

ABSTRACT

Validity and Reliability of Measurement Instruments in Quantitative Studies

Christina Ouzouni, Konstantinos Nakakis

Lecturers in Nursing, Department of Nursing, Technological Educational Institute of Lamia, Greece

In order to draw valid conclusions from empirical studies it is necessary to use assessment instruments which have previously been tested for their validity and reliability. Reliability refers to the repeatability and reproducibility of a set of measurements in yielding a consistent result, the homogeneity of an instrument and the degree to which it is free from random error. The assessment of reliability is performed by test-retest, alternate form, split-half, internal consistency, inter-observer and inter-rater reliability testing. Validity refers to whether an instrument measures what it aims to measure, and it is related to systematic error. Depending on the approach chosen to assess the validity of a measurement tool, various different types of validity need to be addressed, such as content, face, criterion (concurrent-predictive) and construct (factorial, known groups method, convergent and divergent or discriminant validity). The purpose of this continuing education article is to present, with examples, the major types of reliability and validity of measurement tools used in quantitative studies. **NOSILEFTIKI 2011, 50 (2): 231–239.**

Key-words: *measurement instruments, measurement tools, reliability, scales, validity*



Corresponding Author: Christina Ouzouni, Department of Nursing, Technological Educational Institute of Lamia, 3rd km Old National Road Lamia-Athens, GR-351 00 Lamia, Greece, tel.: +30 2231 060 178, e-mail: christinaouzouni@yahoo.com

Βιβλιογραφία

1. Νακάκης Κ, Ουζούνη Χ. Νοσηλευτική βασισμένη σε ενδείξεις. Μια νέα προοπτική για την ελληνική Νοσηλευτική. *Το Βήμα του Ασκληπιού* 2008, 7:208–220
2. Polit DF, Beck CT. *Nursing research: Principles and methods*. 8th ed. Lippincott Williams & Wilkins, Philadelphia, 2008
3. Gibbon B. Validity and reliability of assessment tools. *Nurs Res* 1995, 2:48–55
4. Oppenheim AN. *Questionnaire design and attitude measurement*. Heinemann, London, 1984
5. Polit DF, Hungler BP. *Nursing research: Principles and methods*. 6th ed. Lippincott Williams & Wilkins, New York, 1999
6. Σαχίνη-Καρδάση Α. Μεθοδολογία έρευνας. Εφαρμογές στο χώρο της υγείας. Εκδόσεις Βήτα, Αθήνα, 1997:155–161
7. Gilbert N. Research theory and method. In: Gilbert N (ed) *Researching social life*. Sage, London, 1993
8. Jacobson SF. Evaluating instruments for use in clinical nursing research. In: Frank-Stromborg M (ed) *Instruments for clinical research*. Jones and Bartlett Publishers, Boston, MA, 1992
9. Litwin MS. *How to measure survey reliability and validity*. Sage Publications, London, 1995
10. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. *Am J Health Syst Pharm* 2008, 65:2276–2284
11. Babbie E. *The practice of social research*. 9th ed. Wadsworth, USA, 2000
12. Houser J. Precision, reliability, and validity: Essential elements of measurement in nursing research. *J Spec Pediatr Nurs* 2008, 13:297–299
13. Bowling A. *Research methods in health*. 2nd ed. Open University Press, New York, 2002:150–156
14. LoBiondo-Wood G, Haber J. *Nursing research*. 3rd ed. Mosby, London, 1994
15. Υφαντόπουλος Ι. Μέτρηση της ποιότητας ζωής και το ευρωπαϊκό υγειονομικό μοντέλο. *Αρχ Ελλ Ιατρ* 2007, 24(Συμπλ 1):6–18
16. Berk RA. Importance of expert judgment in content-related validity evidence. *West J Nurs Res* 1990, 12:659–671
17. Hickey ML, Owen SV, Froman RD. Instrument development: Cardiac diet and exercise self-efficacy. *Nurs Res* 1992, 41:347–351
18. Oman KS, Krugman ME, Fink RM. *Nursing research secrets*. Hanley & Belfus Inc, Philadelphia, 2003
19. Streiner DL, Norman GR. *Health measurement scales*. 3rd ed. Oxford University Press, USA, 2005
20. Norwood SL. *Research strategies for advanced practice nurses*. Prentice Hall Health, New Jersey, 2000
21. Burns N, Grove SK. *The practice of nursing research: Conduct, critique and utilization*. 5th ed. WB Saunders Co, Philadelphia, 2005
22. Schwarz N, Oyserman D. Asking questions about behavior: Cognition, communication and questionnaire construction. *Am J Eval* 2001, 22:127–160