

3 Βαθμολογίες και μετρήσεις: εγκυρότητα, αξιοπιστία και ευαισθησία

Περίληψη

Σε αυτό το κεφάλαιο θα διερευνήσουμε ιδιότητες οι οποίες είναι κοινές για όλες τις μορφές μετρήσεων. Σε αυτές συμπεριλαμβάνονται οι μετρήσεις μεμονωμένων ερωτημάτων, όπως η απάντηση σε μια μόνο γενική ερώτηση, και οι αθροιστικές βαθμολογίες που προέρχονται από κλίμακες πολλαπλών ερωτημάτων, όπως οι βαθμολογίες των αθροιστικών κλιμάκων που χρησιμοποιούνται σε πολλά εργαλεία μέτρησης Ποιότητας Ζωής. Σε αυτές τις ιδιότητες συμπεριλαμβάνονται η εγκυρότητα, η αξιοπιστία, η ευαισθησία και η ανταποκρισιμότητα. Αυτό το κεφάλαιο εστιάζεται στις πτυχές των ιδιοτήτων που ισχύουν στις βαθμολογίες των μεμονωμένων ερωτημάτων και των αθροιστικών κλιμάκων. Το Κεφάλαιο 4 διαπραγματεύεται τις σχετικές τεχνικές που ισχύουν για τις κλίμακες πολλαπλών ερωτημάτων, όταν μπορούν να εξεταστούν οι σχέσεις μέσα στις κλίμακες και μεταξύ των ερωτημάτων.

3.1 ΕΙΣΑΓΩΓΗ

Όλες οι μετρήσεις, από την πίεση του αίματος έως τις αξιολογήσεις της Ποιότητας Ζωής, πρέπει να ικανοποιούν κάποιες βασικές ιδιότητες, αν πρόκειται να έχουν κλινική χρησιμότητα. Αυτές είναι: εγκυρότητα, αξιοπιστία, επαναληψιμότητα, ευαισθησία και ανταποκρισιμότητα.

Η πιστοποίηση εγκυρότητας των εργαλείων μέτρησης είναι η διαδικασία, κατά την οποία διευκρινίζεται το κατά πόσον ευσταθεί να πιστέψουμε ότι ένα εργαλείο μετρά αυτό που σκοπεύει να μετρήσει και ότι είναι χρήσιμο για τον προτεινόμενο σκοπό. Για παράδειγμα, σε ποιο βαθμό είναι λογικό να υποστηρίξουμε ότι ένα ερωτηματολόγιο «ποιότητας ζωής» πραγματικά αξιολογεί την ΠΖ; Από τη στιγμή που προσπαθούμε να μετρήσουμε μια μεταβλητή, η οποία δεν έχει οριστεί με σαφήνεια και δεν μπορούμε άμεσα να την παρατηρήσουμε, αφού είναι

λανθάνουσα (η ΠΖ), μπορούμε μόνο να συμπεράνουμε ότι ένα εργαλείο μέτρησης είναι έγκυρο στο βαθμό που συσχετίζεται με άλλη παρατηρήσιμη συμπεριφορά. Αυτή η διαδικασία πιστοποίησης της εγκυρότητας αποτελείται από μια σειρά σταδίων, κατά τα οποία ελπίζουμε να συλλέξουμε πειστικές αποδείξεις ότι το εργαλείο μέτρησης αξιολογεί τις έννοιες που σκοπεύει, και παράγει χρήσιμες μετρήσεις που απεικονίζουν την ΠΖ των ασθενών. Η εγκυρότητα μπορεί να υποδιαιρεθεί σε τρεις κύριες πτυχές.

Η *εγκυρότητα περιεχομένου* αφορά το βαθμό στον οποίο τα ερωτήματα είναι εύλογα και απεικονίζουν τους τομείς ενδιαφέροντος που σκόπευαν. Η *εγκυρότητα κριτηρίου* αφορά το κατά πόσον η κλίμακα συνδέεται εμπειρικά με εξωτερικά κριτήρια, όπως άλλα καθιερωμένα εργαλεία μέτρησης. Η *εγκυρότητα δομής* ερευνά τη θεωρητική σχέση των ερωτημάτων αναμεταξύ τους και με τις υποτιθέμενες κλίμακες. Από αυτά τα τρία είδη εγκυρότητας, η δομική εγκυρότητα προσφέρεται περισσότερο από όλες για διερεύνηση μέσω αριθμητικής ανάλυσης. Δύο πτυχές της δομικής εγκυρότητας είναι η *συγκλίνουσα εγκυρότητα* και η *διακριτική εγκυρότητα*. Κάποια ερωτήματα ή κλίμακες, όπως το άγχος και η κατάθλιψη, αναμένεται να έχουν υψηλή συσχέτιση ή να συγκλίνουν. Άλλα ίσως αναμένεται να είναι σχετικά ασύνδετα ή να αποκλίνουν και να έχουν *διακριτική εγκυρότητα*. Αν συμπεριληφθεί μια ομάδα ασθενών με ένα μεγάλο φάσμα διαγνώσεων και θεραπειών, η πολύ υψηλή συσχέτιση μεταξύ των κλιμάκων μπορεί να υποδεικνύει χαμηλή *διακριτική* εγκυρότητα και μπορεί να δείχνει ότι οι δύο κλίμακες μετρούν όμοια πράγματα. Αν οι συσχετίσεις μεταξύ των κλιμάκων δεν ανταποκρίνονται σε γενικές γραμμές σε αυτό που ήταν αναμενόμενο, οι προτεινόμενες σχέσεις μεταξύ των κλιμάκων τίθενται υπό αμφισβήτηση.

Η *αξιόπιστία* και η *επαναληψιμότητα* αφορούν την τυχαία μεταβλητότητα που συνδέεται με τις μετρήσεις. Ιδανικά, οι ασθενείς των οποίων η Ποιότητα Ζωής δεν έχει αλλάξει, πρέπει να δίνουν πολύ όμοιες ή επαναλαμβανόμενες αντιδράσεις κάθε φορά που αξιολογούνται. Εάν υπάρχει σημαντική τυχαία μεταβλητότητα, οι μετρήσεις είναι αναξιόπιστες. Είναι δύσκολο να γνωρίζουμε πώς να ερμηνεύσουμε τα αποτελέσματα των μεμονωμένων ασθενών αν οι μετρήσεις δεν είναι αξιόπιστες. Η χαμηλή αξιοπιστία μπορεί κάποιες φορές να αποτελεί προειδοποίηση ότι η εγκυρότητα ίσως είναι ύποπτη και ότι η μέτρηση εντοπίζει κάτι διαφορετικό από αυτό που σκοπεύουμε να μετρήσουμε.

Η *ευαισθησία* είναι η ικανότητα των μετρήσεων να ανιχνεύουν διαφορές ανάμεσα σε ασθενείς ή σε ομάδες ασθενών. Αν μπορούμε να επιδείξουμε ότι μια μέτρηση είναι ευαίσθητη και ανιχνεύει διαφορές που πιστεύεται ότι υπάρχουν μεταξύ ομάδων ασθενών, όπως π.χ. οι διαφορές μεταξύ ασθενών με κακή και καλή πρόγνωση, θα είμαστε πιο σίγουροι ότι είναι έγκυρη και μετρά αυτό που πιστεύουμε ότι μετρά. Η ευαισθησία είναι, επίσης, σημαντική στις κλινικές δοκιμές, αφού η μέτρηση δεν είναι ιδιαίτερα χρήσιμη, αν δεν μπορεί να

εντοπίζει τις διαφορές στην ΠΖ που τυχόν υπάρχουν μεταξύ τυχαιοποιημένων ομάδων.

Η ανταποκρισιμότητα μοιάζει με την ευαισθησία, αλλά συνδέεται με την ικανότητα να ανιχνεύονται μεταβολές, όταν ένας ασθενής βελτιώνεται ή χειροτερεύει. Μία μέτρηση έχει περιορισμένη χρήση στην παρακολούθηση του ασθενή, αν δεν αντανακλά τις μεταβολές της κατάστασής του. Μια μέτρηση με ευαισθησία θα ανταποκρίνεται συνήθως, αλλά όχι απαραίτητα, στις αλλαγές.

Η εγκυρότητα, η αξιοπιστία, η ευαισθησία και η ανταποκρισιμότητα είναι αλληλένδετες, αλλά η κάθε μία είναι σημαντική από μόνη της. Η αξιολόγηση της εγκυρότητας, συγκεκριμένα, είναι μια σύνθετη και ατέρμονη διαδικασία. Στην έρευνα της ΠΖ ποτέ δεν μπορεί να αποδειχθεί η εγκυρότητα μιας κλίμακας. Αντί αυτού, όμως, η διαδικασία της πιστοποίησης συμπεριλαμβάνει τη συλλογή όλο και περισσότερων αποδείξεων ότι οι κλίμακες είναι εύλογες και συμπεριφέρονται με τον αναμενόμενο τρόπο.

Για τη μελέτη της στατιστικής σημαντικότητας και των τιμών p , που αναφέρονται σε αυτό το κεφάλαιο, δείτε τα Κεφάλαια 4 και 14.

3.2 ΕΓΚΥΡΟΤΗΤΑ ΠΕΡΙΕΧΟΜΕΝΟΥ

Η εγκυρότητα περιεχομένου σχετίζεται με την επάρκεια του περιεχομένου ενός εργαλείου μέτρησης και αναφέρεται στον αριθμό και το εύρος των μεμονωμένων ερωτήσεων που περιλαμβάνει. Χρησιμοποιεί τους ορισμούς των εννοιών που αξιολογούνται και περιλαμβάνει την επανεξέταση του εργαλείου, για να διασφαλιστεί ότι αυτό φαίνεται λογικό και καλύπτει όλα τα σχετικά ζητήματα. Συνεπώς, η πιστοποίηση εγκυρότητας του περιεχομένου απαιτεί την κριτική εξέταση της βασικής δομής του εργαλείου, την ανασκόπηση της διαδικασίας που ακολουθήθηκε για την ανάπτυξη του ερωτηματολογίου και, επίσης, τη μελέτη των προοπτικών εφαρμογής στο θέμα που προτίθεται να ερευνήσει. Για να μπορέσουμε να ισχυριστούμε ότι ένα εργαλείο μέτρησης έχει εγκυρότητα περιεχομένου, πρέπει ο σχεδιασμός και η ανάπτυξη του να ακολουθούν αυστηρά καθορισμένες διαδικασίες.

ΕΥΡΟΣ ΚΑΛΥΨΗΣ ΚΑΙ ΣΥΝΑΦΕΙΑ ΤΩΝ ΕΡΩΤΗΜΑΤΩΝ

Μια από τις σημαντικές πτυχές της εγκυρότητας περιεχομένου είναι η συνολική κάλυψη. Επίσης, πρέπει να καλύπτεται από το εργαλείο μέτρησης το πλήρες εύρος των σχετικών ζητημάτων. Για παράδειγμα, ένα εργαλείο μέτρησης, που στοχεύει στην αξιολόγηση της συμπτωματολογίας, θα πρέπει να συμπεριλάβει ερωτήματα που αναφέρονται σε όλα τα σχετικά κύρια συμπτώματα. Ειδάλλως, ενδέχεται να υπάρχουν διαφορές μεταξύ ομάδων ασθενών, οι οποίες δεν εντο-

πίζονται. Σε ακραίες περιπτώσεις, μπορεί να μην ανιχνευτούν και να μην αναφερθούν σημαντικές παρενέργειες. Αν και αυτές οι παρενέργειες ίσως έχουν ουσιαστική επίδραση στην ΠΖ, μία μεμονωμένη γενική ερώτηση που αφορά τη συνολική Ποιότητα Ζωής ενδέχεται να μην έχει την εξειδίκευση και την ευαισθησία για να ανιχνεύσει διαφορές μεταξύ ομάδων.

Ο βαθμός στον οποίο τα ερωτήματα παρέχουν επαρκή κάλυψη δεν είναι εύκολο να ελεγχθεί με επίσημες στατιστικές δοκιμές και εξαρτάται κυρίως από τη διασφάλιση της ανάπτυξης του εργαλείου σύμφωνα με αυστηρά προκαθορισμένες μεθόδους. Η διαδικασία με την οποία σχηματίζονται τα ερωτήματα πρέπει να περιλαμβάνει πληροφόρηση από ειδικούς στον τομέα της ασθένειας, την ανασκόπηση δημοσιευμένων δεδομένων και βιβλιογραφίας, καθώς και συνεντεύξεις με ασθενείς που πάσχουν από την ασθένεια. Οι αποδείξεις ότι έχουν ακολουθηθεί επίσημες τεκμηριωμένες διαδικασίες θα υποστηρίξουν τους ισχυρισμούς περί εγκυρότητας του περιεχομένου ενός εργαλείου.

Ταυτόχρονα, όλα τα ερωτήματα που συμπεριλαμβάνονται πρέπει να σχετίζονται με την αξιολογούμενη έννοια και συγχρόνως πρέπει να αποκλείονται όλα τα άσχετα ερωτήματα. Η σχετικότητα των ερωτημάτων συχνά αντιμετωπίζεται, από τη μια, χρησιμοποιώντας μια ομάδα ειδικών για την αξιολόγηση της καταλληλότητας επιμέρους ερωτημάτων, τα οποία αναφέρονται στην έννοια που διερευνάται, και από την άλλη, ζητώντας την άποψη ασθενών ως προς τη συνάφεια των ερωτήσεων. Επίσης, οι μέθοδοι πιστοποίησης της εγκυρότητας μπορούν να προσδιορίσουν τα ερωτήματα που συμπεριφέρονται διαφορετικά από τα υπόλοιπα σε μια κλίμακα (βλ. Κεφάλαιο 4). Αυτά τα ερωτήματα μπορεί να αναθεωρηθούν κριτικά, ώστε να αποφασιστεί εάν συνδέονται πραγματικά ή όχι με την έννοια που αξιολογείται. Μπορεί, επίσης, να αποκλειστούν ερωτήματα ως περιττά, αν επικαλύπτουν ή επαναλαμβάνουν πληροφορίες που περιέχονται σε άλλα ερωτήματα.

ΕΓΚΥΡΟΤΗΤΑ ΜΟΡΦΗΣ

Η *εγκυρότητα μορφής* αφορά τον έλεγχο του κατά πόσο τα ερωτήματα ενός εργαλείου μέτρησης δείχνουν «στη μορφή τους» να καλύπτουν τα θέματα, στα οποία αποσκοπούν καθαρά και αναμφισβήτητα. Η εγκυρότητα μορφής συνδέεται στενά με την εγκυρότητα περιεχομένου, ενώ συχνά θεωρείται μια διάστασή της. Η κύρια διαφορά είναι ότι η εγκυρότητα μορφής αφορά την κριτική αναθεώρηση ενός εργαλείου μέτρησης *μετά* τη δημιουργία του, ενώ το μεγαλύτερο μέρος της πιστοποίησης εγκυρότητας του περιεχομένου διασφαλίζει ότι ακολουθήθηκαν και τεκμηριώθηκαν συνολικές και ενδεδειγμένες διαδικασίες.

Η εγκυρότητα περιεχομένου βελτιστοποιείται, όταν στη διαδικασία δημιουργίας μετέχουν πολλοί διαφορετικοί άνθρωποι και με παρόμοιο τρόπο μεγιστοποιείται και η εγκυρότητα μορφής. Συνεπώς, όταν επιβεβαιώνεται η εγκυρότητα μορφής, πρέπει να αναζητείται η γνώμη των ειδικών (όπως ιατρών, νοσο-

κόμων και κοινωνικών επιστημόνων), και να ερωτώνται οι ασθενείς εάν έχει λογική το εργαλείο μέτρησης. Παρότι οι ερευνητές, που περιγράφουν την πιστοποίηση εγκυρότητας των εργαλείων μέτρησης, ισχυρίζονται συχνά ότι αναζητήθηκε η συμφωνία απόψεων και ότι το εργαλείο μέτρησης θεωρείται ότι έχει καλή εγκυρότητα μορφής, συχνά δεν αναφέρονται σαφείς λεπτομέρειες. Είναι, όμως, σημαντικό να περιγράφεται η σύνθεση και η λειτουργία των απόμων που συμμετείχαν στις διαδικασίες ανάπτυξης και πιστοποίησης της εγκυρότητας.

Παράδειγμα από τη βιβλιογραφία

Ο Langfitt (1995) συνέκρινε τρία εργαλεία μέτρησης της Ποιότητας Ζωής για ασθενείς με ανίατη επιληψία. Χρησιμοποιήθηκαν το Washington Psychosocial Seizure Inventory (WPSI), το Epilepsy Surgery Inventory 55 (ESI-55) και το SIP. Δυο ανεξάρτητοι ψυχολόγοι ταξινομήσαν κάθε ερώτημα των τριών ερωτηματολογίων σύμφωνα με τη συνάφειά του σε διάφορους τομείς της ΠΖ. Ένας τρίτος ανεξάρτητος βαθμολογητής έλυσε τις τυχόν διαφωνίες. Έτσι, για παράδειγμα, το SIP και το WPSI περιελάμβαναν 26 ερωτήσεις το καθένα, οι οποίες αναφέρονταν στην κοινωνική λειτουργικότητα, ενώ το ESI-55 μόνο δύο (Πίνακας 3.1)

Ο συγγραφέας συμπέρανε ότι το ESI-55 και το SIP έχουν πιο κατάλληλο περιεχόμενο για την αξιολόγηση των γενικών επιπτώσεων της επιληψίας στην Ποιότητα Ζωής από το εργαλείο WPSI, το οποίο επικεντρώνεται στην ψυχολογική και κοινωνική προσαρμογή. Επίσης ο Langfitt, υποστηρίζοντας τις υποδείξεις του, παρουσίασε πληροφορίες για την αξιοπιστία και την εγκυρότητά τους.

Πίνακας 3.1 Αριθμός ερωτημάτων των ESI-55, SIP και WPSI τα οποία σχετίζονται με ασθενείς με επιληψία, κατά τομέα Ποιότητας Ζωής (Βασισμένο στο Langfitt, 1995)

Τομείς Ποιότητας Ζωής		ESI-55	SIP	WPSI
Σωματικός	Συμπτώματα	2	5	2
	Λειτουργική κατάσταση	10	55	2
	Ενέργεια/ύπνος και ξεκούραση	4	8	4
Ψυχολογικός	Κοινωνική λειτουργικότητα	2	26	26
	Συναισθηματική κατάσταση	6	7	37
	Τρόπος αντίληψης των πραγμάτων	5	16	7
	Αντιλήψεις υγείας	7	2	3
	Γενική ικανοποίηση από τη ζωή	3	1	10
Δραστηριότητες ρόλων		15	16	16
Εκτός κατηγοριών		1	0	25
Σύνολο		55	136	132

3.3 ΕΓΚΥΡΟΤΗΤΑ ΚΡΙΤΗΡΙΟΥ

Η *εγκυρότητα κριτηρίου* αναφέρεται στην αξιολόγηση ενός εργαλείου μέτρησης συγκριτικά με μια πραγματική τιμή, ή συγκριτικά με κάποια άλλη σταθερά που θεωρείται ότι προσφέρει μια ένδειξη των πραγματικών τιμών που μετρούνται. Διακρίνεται σε *συγχρονική εγκυρότητα* και *προγνωστική εγκυρότητα*.

ΣΥΓΧΡΟΝΙΚΗ ΕΓΚΥΡΟΤΗΤΑ

Συγχρονική εγκυρότητα σημαίνει συμφωνία με την πραγματική τιμή. Τέτοια «χρυσή σταθερά» δεν υπάρχει για τα εργαλεία μέτρησης της Ποιότητας Ζωής, αφού αυτά μετρούν υποθετικές έννοιες που είναι πειραματικές και υποκειμενικές. Επομένως, η πιο συχνή προσέγγιση περιλαμβάνει τη σύγκριση νέων ερωτηματολογίων με ένα ή περισσότερα καθιερωμένα εργαλεία. Αυτό ίσως είναι λογικό, εάν ο στόχος της δημιουργίας ενός νέου εργαλείου είναι η κατασκευή ενός συντομότερου ή απλούστερου ερωτηματολογίου, οπότε το πιο λεπτομερές, εδραιωμένο ερωτηματολόγιο μπορεί να θεωρηθεί ότι θέτει το πρότυπο στο οποίο στοχεύουμε. Πιο συχνά, η λογική της δημιουργίας ενός νέου ερωτηματολογίου είναι το ότι οι ερευνητές πιστεύουν πως τα ήδη υπάρχοντα είναι κατώτερης ποιότητας. Σε αυτή την περίπτωση, η σύγκριση του νέου με το καθιερωμένο είναι περιορισμένης αξίας, αφού το τελευταίο ουσιαστικά ήδη έχει απορριφθεί ως «χρυσή σταθερά». Μια άλλη προσέγγιση αποτελεί η χρήση έμμεσων μεθόδων σύγκρισης. Μια λεπτομερής συνέντευξη, όπου χρησιμοποιείται προσωπικό εκπαιδευμένο σε τεχνικές συνέντευξης, μπορεί να αποφέρει εκτιμήσεις των εννοιών που θεωρούνται πως πιθανώς προσεγγίζουν την πραγματική τιμή.

Παράδειγμα

Το άγχος και η κατάθλιψη είναι ψυχολογικές έννοιες, οι οποίες έχουν αξιολογηθεί παραδοσιακά με τη χρήση συνεντεύξεων σε βάθος για την αξιολόγηση της δριμύτητάς τους και για τον εντοπισμό «περιστατικών» ασθενών που χρειάζονται ψυχιατρική παρέμβαση. Εάν η ψυχιατρική αξιολόγηση θεωρείται ότι προσεγγίζει το πραγματικό επίπεδο αυτών των καταστάσεων, μπορεί να χρησιμοποιηθεί ως κριτήριο με το οποίο να συγκριθούν τα αυτό-συμπληρούμενα ερωτηματολόγια του ασθενούς. Το άγχος και η κατάθλιψη ίσως διαφέρουν από τις περισσότερες κλίμακες Ποιότητας Ζωής στο ότι υπάρχει (με αμφιβολίες) ένας πιο σαφής ορισμός και μια καλύτερη ομοφωνία μεταξύ των ψυχιάτρων για τη σημασία αυτών των όρων. Βάσει αυτής της παραδοχής, θα φαινόταν λογικό ένα σύντομο ερωτηματολόγιο, το οποίο απευθύνεται στον ασθενή και συμπληρώνεται σε λίγα μόνο λεπτά, να θεωρηθεί ως μια εύχρηστη μέθοδος για τον υπολογισμό των «πραγματικών» τιμών της λεπτομερούς συνέντευξης.

Όπως ήδη αναφέρθηκε, ένα καινούργιο εργαλείο μέτρησης συνήθως συγκρίνεται με τις τιμές άλλων καθιερωμένων ή περισσότερο εκτενών εργαλείων, όπως λ.χ. μία συνέντευξη σε βάθος ή η αξιολόγηση ενός παρατηρητή. Εάν η συμφωνία μεταξύ των δύο μεθόδων θεωρηθεί φτωχή, η συγχρονική εγκυρότητα είναι χαμηλή. Ίσως είναι δύσκολο να καθοριστεί με βεβαιότητα αν η μία ή και οι δύο μέθοδοι έχουν χαμηλή εγκυρότητα, αλλά το χαμηλό επίπεδο συμφωνίας λειτούργει ως ένας δείκτης ότι κάτι δεν πάει καλά.

Παράδειγμα από τη βιβλιογραφία

Οι Zigmond και Snaith (1983) ζήτησαν από 100 εξωτερικούς ασθενείς μιας κλινικής γενικής ιατρικής να συμπληρώσουν το ερωτηματολόγιο HADS. Στη συνέχεια χρησιμοποίησαν μια 20-λεπτη ψυχιατρική συνέντευξη για την αξιολόγηση του άγχους και της κατάθλιψης. Σύνοψη των αποτελεσμάτων παρουσιάζεται στον Πίνακα 3.2, όπου οι ασθενείς ομαδοποιήθηκαν σε τρεις κατηγορίες, ανάλογα με το αν ήταν ψυχιατρικά περιστατικά, ήταν αμφίβολα περιστατικά ή δεν ήταν περιστατικά άγχους και κατάθλιψης.

Για τη διάγνωση ψυχιατρικών περιστατικών, η κλίμακα κατάθλιψης κατέταξε 1% λανθασμένα θετικά και 1% λανθασμένα αρνητικά, ενώ η κλίμακα άγχους 5% λανθασμένα θετικά και 1% λανθασμένα αρνητικά περιστατικά.

Πίνακας 3.2 Το ερωτηματολόγιο HADS συμπληρωμένο από 100 εξωτερικούς ασθενείς μιας κλινικής γενικής ιατρικής (Βασισμένο στους Zigmond και Snaith, 1983).

Βαθμολογία HADS	Κατάθλιψη			Άγχος		
	Μη περιστατικά	Αμφίβολα περιστατικά	Περιστατικά	Μη περιστατικά	Αμφίβολα περιστατικά	Περιστατικά
0-7	57	11	I	41	4	I
8-10	8	7	3	10	9	1
11-21	I	4	8	5	15	14

Τα λανθασμένα θετικά και τα λανθασμένα αρνητικά σημειώνονται με έντονη πλάγια γραφή

ΠΡΟΓΝΩΣΤΙΚΗ ΕΓΚΥΡΟΤΗΤΑ

Η *προγνωστική εγκυρότητα* αφορά την ικανότητα ενός εργαλείου μέτρησης να προβλέπει τη μελλοντική κατάσταση υγείας, τα μελλοντικά γεγονότα ή τα μελλοντικά αποτελέσματα των δοκιμασιών. Για παράδειγμα, συχνά αναφέρεται ότι η βαθμολογία της συνολικής ΠΖ σε δοκιμές καρκίνου μπορεί να προβλέψει τον επακόλουθο χρόνο επιβίωσης και ότι η αξιολόγηση της ΠΖ παρέχει πρόσθετες

προγνωστικές πληροφορίες που συμπληρώνουν τις πιο αντικειμενικές μετρήσεις, όπως είναι το στάδιο του όγκου και η έκταση της ασθένειας. Έχοντας ως συνέπεια, η μελλοντική κατάσταση υγείας να μπορεί να λειτουργεί ως κριτήριο με το οποίο να συγκρίνεται ένα εργαλείο μέτρησης. Επίσης, όσον αφορά την εγκυρότητα κριτηρίου, η μελλοντική κατάσταση θεωρείται καλύτερος δείκτης της τρέχουσας πραγματικής τιμής της λανθάνουσας μεταβλητής, σε σύγκριση με τις παρατηρούμενες απαντήσεις των ασθενών από το εργαλείο μέτρησης που δημιουργείται. Για να προβεί ο ερευνητής σε μια τέτοια παραδοχή, πρέπει να δημιουργήσει ένα εννοιολογικό μοντέλο του αξιολογούμενου μεγέθους και της σχέσης του μεγέθους με τη μελλοντική έκβαση. Συνεπώς, η προγνωστική εγκυρότητα μπορεί να μελετηθεί πιο εύκολα, αν τη σκεφτούμε ως μια πτυχή της εγκυρότητας δομής.

3.4 ΕΓΚΥΡΟΤΗΤΑ ΔΟΜΗΣ

Η *εγκυρότητα δομής* είναι ένα από τα πιο σημαντικά χαρακτηριστικά ενός εργαλείου μέτρησης. Αξιολογεί το βαθμό στον οποίο ένα εργαλείο μετρά την έννοια που σχεδιάστηκε να εκτιμήσει. Το θέμα της εγκυρότητας δομής είναι δύσκολο και επίμαχο. Αφορά, κατ' αρχήν, τη δημιουργία ενός υποθετικού μοντέλου, το οποίο περιγράφει τις έννοιες που αξιολογούνται και επίσης προτείνει τις σχέσεις τους. Στη συνέχεια συλλέγονται δεδομένα και γίνεται αξιολόγηση του βαθμού στον οποίο επιβεβαιώνονται αυτές οι σχέσεις. Εάν τα αποτελέσματα επιβεβαιώνουν τις προηγούμενες προσδοκίες που αναφέρονται στις έννοιες, έπεται ότι το εργαλείο μέτρησης *ίσως* να είναι έγκυρο και να μπορούμε να το χρησιμοποιούμε για να εξάγουμε συμπεράσματα για τους ασθενείς.

Η δήλωση «*ίσως* να είναι έγκυρο» αποσκοπεί να δώσει έμφαση στην πτυχή της εγκυρότητας δομής, για την οποία υπάρχει αρκετή διαμάχη. Η δυσκολία έγκειται στο ότι το κριτήριο και η έννοια δεν είναι άμεσα μετρήσιμα. Ως εκ τούτου, δεν μπορεί να διεξαχθεί μια κανονική στατιστική δοκιμή. Αφού η αξιολόγηση της εγκυρότητας δομής βασίζεται στην έκφραση απόψεων, που αφορούν τις αναμενόμενες σχέσεις μεταξύ των εννοιών, και την επιβεβαίωση ότι οι παρατηρούμενες μετρήσεις συμπεριφέρονται όπως αναμενόταν, δεν μπορούμε να *αποδείξουμε* ότι τα ερωτήματα του ερωτηματολογίου είναι έγκυρες μετρήσεις αυτών των εννοιών, ούτε ότι οι έννοιες είναι έγκυρες αναπαραστάσεις συμπεριφοράς. Το μόνο που μπορούμε να κάνουμε είναι να συλλέξουμε όλο και περισσότερες αποδείξεις ότι οι μετρήσεις φαίνονται λογικές, ότι οι υποτιθέμενες έννοιες συμπεριφέρονται όπως αναμενόταν, και ότι δεν υπάρχουν λόγοι να τις απορρίψουμε. Όσες περισσότερες αποδείξεις έχουμε για αυτό, τόσο μεγαλύτερη εμπιστοσύνη μπορούμε να έχουμε ότι το μοντέλο μας αποτελεί μια ικανοποιη-

τική απεικόνιση των εννοιών τις οποίες ονομάζουμε Ποιότητα Ζωής.

Ακολουθώντας όλους και τους κανόνες και τις διαδικασίες, η εγκυρότητα δομής περιλαμβάνει μια ποικιλία τεχνικών που όλες στοχεύουν να αξιολογήσουν δύο πράγματα: πρώτον, εάν η υποτιθέμενη θεωρητική έννοια φαίνεται να αποτελεί ένα κατάλληλο μοντέλο και, δεύτερον, εάν η κλίμακα μέτρησης φαίνεται να αντιστοιχεί σε αυτή την υποτιθέμενη έννοια. Μεταφράζοντας τα παραπάνω σε πρακτική εφαρμογή, η εγκυρότητα δομής ασχολείται κυρίως με τον έλεγχο της:

1. *Διαστασιμότητας*: Σχετίζονται όλα τα ερωτήματα μιας υποκλίμακας με μία μοναδική λανθάνουσα μεταβλητή ή υπάρχουν ενδείξεις ότι είναι απαραίτητες περισσότερες μεταβλητές για να ερμηνευθεί η παρατηρούμενη μεταβλητότητα;
2. *Ομοιογένειας*: Φαίνεται να συνδέονται εξίσου ισχυρά όλα τα ερωτήματα της υποκλίμακας με την ίδια λανθάνουσα μεταβλητή;
3. *Επικάλυψης μεταξύ λανθανουσών μεταβλητών*: Συσχετίζονται κάποια ερωτήματα μιας υποκλίμακας με άλλες λανθάνουσες μεταβλητές;

Η πιστοποίηση της εγκυρότητας πρόκειται για μια μακρά και συνεχή διαδικασία που περιλαμβάνει το να μάθουμε περισσότερα για τη δομή, να κάνουμε καινούριες προβλέψεις και να τις δοκιμάσουμε. Κάθε μελέτη που υποστηρίζει τη θεωρητική δομή εξυπηρετεί την ενδυνάμωση της θεωρίας, ενώ ένα μόνο αρνητικό εύρημα μπορεί να θέσει σε αμφισβήτηση ολόκληρη τη δομή.

Κατά κύριο λόγο, η αξιολόγηση της εγκυρότητας δομής χρησιμοποιεί συσχετίσεις, μεταβολές κατά το πέρασμα του χρόνου και διαφορές μεταξύ ομάδων ασθενών. Περιλαμβάνει την κατασκευή και τη δοκιμασία εννοιολογικών μοντέλων, τα οποία εκφράζουν την υποτιθέμενη σχέση μεταξύ των υποθετικών περιοχών της ΠΖ και των κλιμάκων, οι οποίες δημιουργούνται για να μετρήσουν τις περιοχές αυτές.

ΠΙΣΤΟΠΟΙΗΣΗ ΕΓΚΥΡΟΤΗΤΑΣ ΓΝΩΣΤΩΝ ΟΜΑΔΩΝ

Μια από τις πιο απλές μορφές της πιστοποίησης μιας δομής είναι η *εγκυρότητα των γνωστών ομάδων*. Αυτή στηρίζεται στην αρχή, ότι κάποιες συγκεκριμένες ομάδες ασθενών αναμένεται να έχουν διαφορετική βαθμολογία από άλλες, και το εργαλείο μέτρησης πρέπει να είναι ευαίσθητο στις διαφορές αυτές. Για παράδειγμα, ασθενείς με προχωρημένο καρκίνο ίσως αναμένεται να έχουν χαμηλότερη Ποιότητα Ζωής από εκείνους στις αρχικές φάσεις της ασθένειας. Μια έγκυρη κλίμακα πρέπει να εντοπίζει διαφορές, προς την αναμενόμενη κατεύθυνση, μεταξύ αυτών των ομάδων. Συνεπώς, οι συγκρίσεις γνωστών ομάδων αποτελούν συνδυασμό της δοκιμασίας εγκυρότητας και μιας μορφής αξιολόγησης ευαισθησίας ή ανταποκρισιμότητας. Μια κλίμακα που δεν μπορεί να διακρίνει αυτές

τις διαφορές επιτυχώς, είτε λόγω έλλειψης ευαισθησίας, είτε επειδή αποφέρει αποτελέσματα αντίθετα από τις προσδοκίες, είναι μάλλον απίθανο να έχει αξία για πολλές χρήσεις.

Οι ερευνητές συχνά επιλέγουν ασθενείς στους οποίους θα αναμένονταν μεγάλες διαφορές μεταξύ των ομάδων. Αυτό σημαίνει ότι ακόμα και μια έρευνα πολύ μικρού δείγματος μπορεί να προσφέρει αρκετές αποδείξεις, ώστε να επιβεβαιώσει πως οι παρατηρούμενες διαφορές είναι απίθανο να οφείλονται στην τύχη. Αυτό που έχει μεγαλύτερη σημασία είναι το μέγεθος των διαφορών και όχι οι τιμές p . Αν και οι δοκιμασίες στατιστικής σημαντικότητας δεν προσφέρουν ιδιαίτερες πληροφορίες και δεν αξίζει να αναφερόμαστε σε αυτές τις περιπτώσεις, συχνά βλέπουμε δημοσιεύσεις να περιγράφουν όλες τις διαφορές ως στατιστικά υψηλά σημαντικές με τιμές p μικρότερες από 0.0001.

Παράδειγμα από τη βιβλιογραφία

Οι Schag και συν. (1992) αξιολόγησαν το HIV Overview of Problems–Evaluation System (HOPES), συγκρίνοντάς το με το MOS–HIV και άλλα εργαλεία σε ασθενείς με τον ιό HIV. Οι ασθενείς διαχωρίστηκαν σε «γνωστές ομάδες»: Ασυμπτωματικών, πασχόντων από σύνδρομο συνδεδεμένο με το AIDS (ARC), AIDS και AIDS σε συνδυασμό με καρκίνο (AIDS+C). Η ανάλυση της εγκυρότητας γνωστών ομάδων στον Πίνακα 3.3 δείχνει ότι υψηλές βαθμολογίες στο HOPES ή χαμηλές στο MOS–HIV δηλώνουν χαμηλή Ποιότητα Ζωής και χειρότερη λειτουργικότητα.

Σε όλες τις κλίμακες η ασυμπτωματική ομάδα έδειξε σαφείς διαφορές από τις υπόλοιπες τρεις ομάδες. Οι διαφορές μεταξύ των ARC, AIDS και AIDS με καρκίνο υπήρξαν πολύ μικρότερες. Με μία εξαίρεση, η ANOVA έδωσε υψηλά σημαντικές τιμές p μικρότερες από 0.0001 – κυρίως λόγω των ακραίων διαφορών μεταξύ των ασυμπτωματικών ασθενών και των υπολοίπων.

Πίνακας 3.3 Εγκυρότητα γνωστών ομάδων των εργαλείων μέτρησης HOPES και MOS–HIV σε ασθενείς με HIV (Βασισμένο στους Schag και συν., 1992)

	Ασυμπτωματικοί $n=119$	ARC $n=65$	AIDS $n=78$	AIDS + C $n=56$	ANOVA τιμή p
<i>HOPES</i>					
Γενική βαθμολογία	0.75	1.27	1.18	1.24	0.0001
Σωματική	0.44	1.03	1.09	1.08	0.0001
Ψυχοκοινωνική	1.02	1.49	1.24	1.42	0.0001
Ιατρική αλληλεπίδραση	0.30	0.70	0.62	0.55	0.0001
Σεξουαλική	1.29	1.54	1.67	1.98	0.0001

Πίνακας 3.3 συνέχεια

	Ασυμπτωματικοί <i>n</i> =119	ARC <i>n</i> =65	AIDS <i>n</i> =78	AIDS + C <i>n</i> =56	ANOVA τιμή <i>p</i>
<i>MOS-HIV</i>					
Σωματική	81.78	53.07	40.98	46.72	0.0001
Ψυχική υγεία	70.25	59.61	66.54	64.20	0.01
Λειτουργικότητα ρόλων	82.62	44.62	30.77	39.29	0.0001
Ποιότητα Ζωής	67.82	55.38	56.73	52.68	0.0001

ΣΥΓΚΛΙΝΟΥΣΑ ΕΓΚΥΡΟΤΗΤΑ

Η συγκλίνουσα εγκυρότητα είναι μια ακόμη σημαντική πλευρά της εγκυρότητας δομής και έγκειται στην απόδειξη ότι μια θεωρούμενη διάσταση της ΠΖ συσχετίζεται αξιωματικά με όλες τις άλλες διαστάσεις με τις οποίες, βάσει της θεωρίας, θα έπρεπε να έχει σχέση. Αυτό σημαίνει, ότι ίσως πιστεύουμε πως κάποιες διαστάσεις της Ποιότητας Ζωής σχετίζονται και, συνεπώς, αναμένουμε οι παρατηρούμενες μετρήσεις να συσχετίζονται. Για παράδειγμα, κάποιος ίσως να αναμένει οι ασθενείς με ισχυρό πόνο να εμφανίζουν κατάθλιψη και, επομένως, θα έπρεπε να υπάρχει συσχέτιση μεταξύ των βαθμολογιών του πόνου και των μετρήσεων της κατάθλιψης.

Πολλές από τις διαστάσεις της ΠΖ ενδοσυσχετίζονται. Ασθενείς που είναι βαριά άρρωστοι συχνά υποφέρουν από ποικιλία συμπτωμάτων και δίνουν υψηλές βαθμολογίες σε μεγάλο εύρος ψυχολογικών διαστάσεων. Πολλές, και κάποιες φορές σχεδόν όλες οι διαστάσεις της Ποιότητας Ζωής συσχετίζονται μεταξύ τους. Συνεπώς, η αξιολόγηση της συγκλίνουσας εγκυρότητας συνίσταται στην πρόβλεψη των ισχυρότερων και ασθενέστερων συσχετίσεων και στην επιβεβαίωση ότι οι επακόλουθες παρατηρούμενες τιμές συμφωνούν με τις προβλέψεις. Η ανάλυση συνίσταται στον υπολογισμό όλων των συντελεστών συσχέτισης κατά ζεύγη μεταξύ των βαθμολογιών των διαφόρων κλιμάκων Ποιότητας Ζωής.

Μια πολύ υψηλή συσχέτιση μεταξύ δύο κλιμάκων εγείρει την ερώτηση εάν και οι δύο κλίμακες μετρούν τον ίδιο παράγοντα και εάν θα μπορούσαν να συνδυαστούν σε μια ενιαία κλίμακα χωρίς απώλεια πληροφορίας. Η απόφαση της συγχώνευσης κλιμάκων πρέπει να λάβει υπόψη τη σύσταση των επιμέρους κλιμάκων και την ύπαρξη κλινικών, ψυχολογικών ή άλλων επιχειρημάτων, που θα έδειχναν ότι η εγκυρότητα μορφής θα διακυβευόταν, οπότε θα ήταν καλύτερο να διατηρηθούν οι χωριστές κλίμακες. Εναλλακτικά, μία πολύ υψηλή συσχέτιση μπορεί να υπονοεί ότι μία από τις κλίμακες είναι περιττή και μπορεί να αφαιρεθεί από το εργαλείο. Η συγκλίνουσα εγκυρότητα συνήθως υπολογίζεται μαζί με τη διακριτική εγκυρότητα.