FROM PLATE TO PELVIS™ · FOR CLINICIANS

Outcome Measures Toolkit

Four validated pelvic floor and mental health outcome measures, with reported sensitivity/specificity, minimal clinically important difference (MCID) where established, cutoff scores, and interpretation guidance for both patients and clinicians.

Not every outcome measure has a published sensitivity/specificity against a diagnostic gold standard; some are severity or distress indices rather than case-finding screens. This toolkit notes clearly, for each measure, what has and has not been established in the literature, rather than filling in a number that doesn't exist.

PHQ-9 (Patient Health Questionnaire-9)

Depression screening and severity, 9 items, self-report

Property	Value
What it measures	Depressive symptom frequency over the past 2 weeks (each item 0-3; total 0-27)
Screening cutoff	≥10 is the most widely used threshold for a positive depression screen
Sensitivity / Specificity	Original validation: ~88% / ~88% at cutoff 10. Pooled meta-analyses report a wider range, roughly 78-85% sensitivity and 87-89% specificity at cutoff 10, varying by setting and population
MCID	Approximately 5 points is commonly cited, though estimates vary by study and population; treat as a general benchmark, not a fixed rule
Severity bands	0-4 minimal · 5-9 mild · 10-14 moderate · 15-19 moderately severe · 20-27 severe

Patient interpretation

A higher score reflects more frequent depressive symptoms over the last two weeks, not a diagnosis on its own. A drop of about 5 points after starting treatment is generally a meaningful improvement worth noticing.

Clinician interpretation

A score ≥10 warrants further clinical evaluation for major depressive disorder; it is a screen, not a diagnosis. Item 9 (thoughts of self-harm) should always be reviewed directly with the patient regardless of total score, and any endorsement should prompt immediate follow-up and, if needed, a safety plan or referral.

Sources: Kroenke et al. 2001 (original validation); Levis et al. 2019 pooled meta-analysis, BMJ/CMAJ.

GAD-7 (Generalized Anxiety Disorder-7)

Anxiety screening and severity, 7 items, self-report

Property	Value
What it measures	Anxiety symptom frequency over the past 2 weeks (each item 0-3; total 0-21)
Screening cutoff	≥10 is the standard threshold for a positive generalized anxiety screen
Sensitivity / Specificity	Original validation: ~89% / ~82% at cutoff 10 against a structured clinical interview
MCID	Approximately 4 points, based on anchor-based responsiveness studies

Severity bands	0-4 minimal · 5-9 mild · 10-14 moderate · 15-21 severe

Patient interpretation

A higher score reflects more frequent worry and anxiety symptoms recently, not a fixed label. A change of about 4 points is generally noticeable and meaningful.

Clinician interpretation

A score ≥10 warrants further evaluation for generalized anxiety disorder or another anxiety-spectrum condition. The GAD-7 also has reasonable sensitivity for panic disorder, social anxiety, and PTSD, so a positive screen with a negative anxiety workup should prompt consideration of those alternatives.

Sources: Spitzer et al. 2006 (original validation); Toussaint et al. 2020 MCID study.

ICIQ-UI SF

Urinary incontinence severity and impact, 4 items (3 scored), self-report

Property	Value
What it measures	Frequency, amount, and quality-of-life impact of urine leakage (scored items total 0-21)
Screening cutoff	Not designed as a diagnostic screen against a gold-standard reference; it is a severity and bother index, not a case-finding tool, so sensitivity/specificity are not typically reported
Sensitivity / Specificity	Not established in the traditional screening sense; used to grade severity and track change, not to rule incontinence in or out
MCID	Approximately 4 points is the most commonly cited threshold (anchor-based estimates have ranged roughly 3-5 points depending on follow-up interval and population)
Severity guide	1-5 slight · 6-12 moderate · 13-18 severe · 19-21 very severe (informal bands used clinically, not a formally validated cutoff set)

Patient interpretation

A higher score reflects more frequent leakage and a bigger impact on daily life. A drop of about 4 points after treatment is generally a change you'd actually notice.

Clinician interpretation

Use the ICIQ-UI SF to establish baseline severity and impact, and to track change over the course of pelvic floor PT or other treatment. It does not distinguish stress, urge, or mixed incontinence on its own; pair with a bladder diary or the ICIQ-UI SF's optional “type of leakage” item for that detail.

Sources: Avery et al. 2004 (original validation); Nystrom et al. 2015 and Sung et al. 2019 MCID studies.

PFDI-20 (Pelvic Floor Distress Inventory)

Pelvic floor symptom distress across 3 domains, 20 items, self-report

Property	Value
What it measures	Symptom-related distress across prolapse (POPDI-6), colorectal-anal (CRADI-8), and urinary (UDI-6) domains (total 0-300)
Screening cutoff	Not a diagnostic screen; it is a symptom-distress inventory, so there is no sensitivity/specificity against a gold-standard diagnosis
Sensitivity / Specificity	Not established in the traditional screening sense; it is used to quantify bother and track change, not to diagnose a specific condition
MCID	Reported estimates vary widely by population and treatment type, roughly 23-45 points; a commonly cited surgical benchmark is about 45 points (Barber et al.), while conservative-treatment and population-based studies have reported figures closer to 23-25 points. Use the estimate closest to your patient's treatment context
Reference point	A postoperative score of 60 or lower has been proposed as an “acceptable symptom state” in prolapse surgery populations, not a universal cutoff

Patient interpretation

A higher score reflects more bother from pelvic floor symptoms across bladder, bowel, and pelvic pressure domains. There is no single “pass/fail” score; what matters most is the change over time relative to where you started.

Clinician interpretation

Because the MCID estimates vary substantially by population and treatment type, use the PFDI-20 primarily as a within-patient tracking tool rather than a strict pass/fail threshold, and interpret any single MCID figure with that variability in mind.

Sources: Barber et al. 2009 MCID study; Wiegersma et al. 2017; 2021 PFDI-20/POPDI-6 MID and PASS study.

This toolkit summarizes publicly available psychometric literature for clinical reference. It is not a substitute for reading the original validation and MCID studies before applying a threshold to clinical decision-making, and does not replace clinical judgment or a full diagnostic workup.

Created by Dr. Kendra Garrett, PT, DPT, WCS, Cert. DN · Board-Certified Women's Clinical Specialist · Your Pelvic PT Auntie