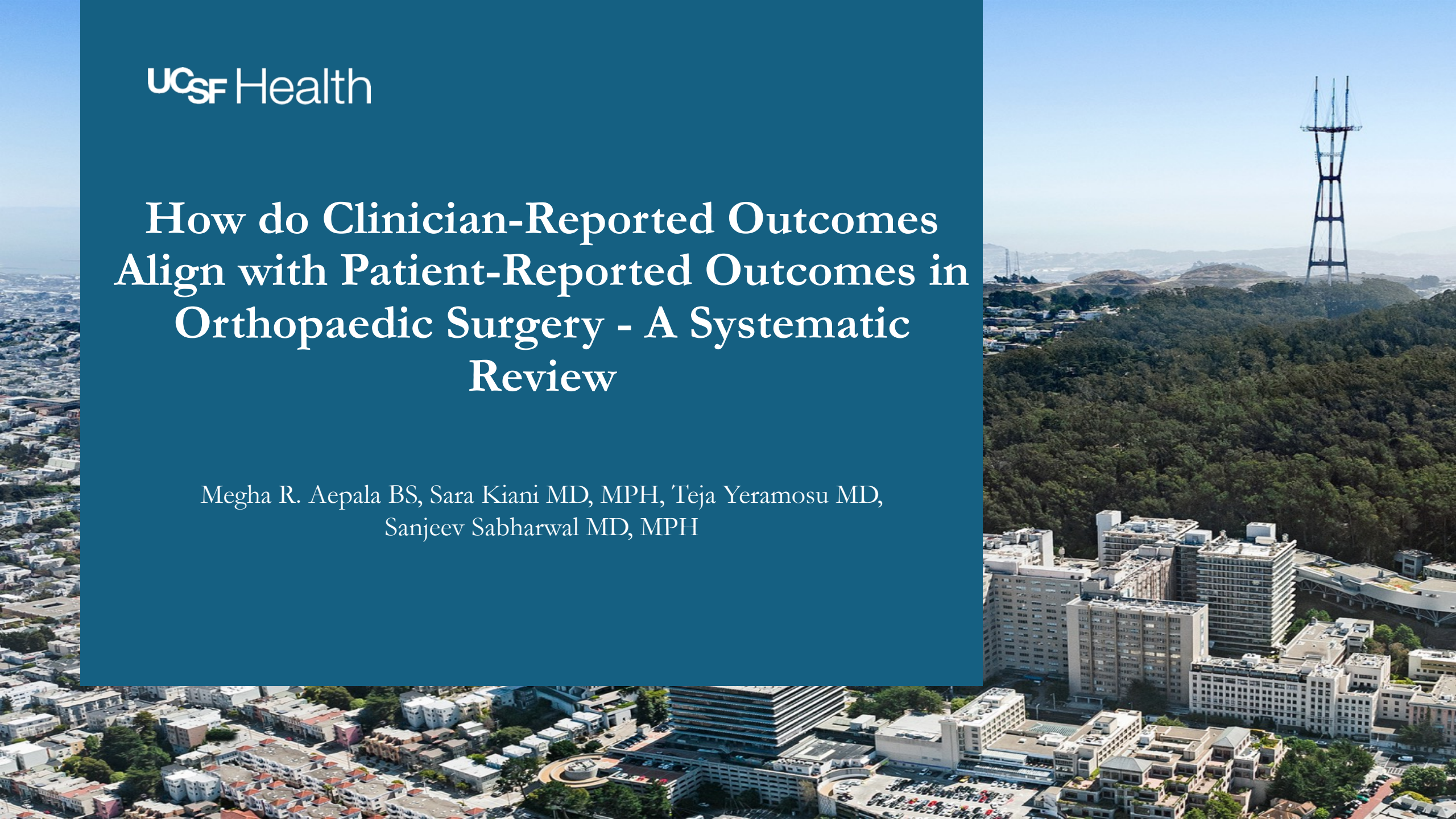
UCSF Health

How do Clinician-Reported Outcomes Align with Patient-Reported Outcomes in Orthopaedic Surgery - A Systematic Review

Megha R. Aepala BS, Sara Kiani MD, MPH, Teja Yeramosu MD,
Sanjeev Sabharwal MD, MPH



Disclosures

No Disclosures

Background

- PROMs capture the patient's perspective on pain, function, and quality of life.
- CROMs assess disease severity and functional status from the clinician's perspective.
- Both are widely used to evaluate outcomes in orthopaedic surgery.

Knowledge Gap

- The relationship between PROMs and CROMs remains unclear and varies across orthopaedic subspecialties.
- Understanding their agreement and discordance is important for interpreting outcomes.

AIMS

1) Do PROMs show distinct correlation with Hard CROMs versus Soft CROMs?

2) Do CROMs correlate differently with condition-specific PROMs compared with general health PROMs?

3) How do PROMs and CROMs compare in their ability to detect change over time as measured by responsiveness indices?

Methods

A literature search was performed using PubMed, Embase, and Web of Science. Title/abstract/full-text screening were performed independently by two reviewers using Covidence.

Inclusion Criteria:

- Orthopaedic conditions
- Examined the correlation, agreement, or responsiveness between PROMs and CROMs
- Reported PROM-CROM comparison as a primary outcome;
- Reported composite PROM scores (with validated subscales used only when total scores were unavailable)
- Assessed outcomes at a final postoperative or stable nonoperative time point, unless specifically evaluating responsiveness.

Correlation Analysis: Spearman/Pearson correlations were extracted and categorized by strength and instrument type (PROM vs CROM; Hard vs Soft)

Risk of Bias (COSMIN): An adapted COSMIN framework assessed study quality; most studies were low risk or had some concerns due to incomplete paired data

Categorization of Outcome Measures

- **PROMs** (Patient-Reported Outcome Measures):
  - Condition-specific PROMs: designed to assess outcomes related to a specific disease, condition, or body system (*e.g.,Western Ontario and McMaster Universities Osteoarthritis Index)*
  - General Health PROMs: measure overall health status and quality of life *(e.g.,Short Form (36) Health Survey)*

- **CROMs** (Clinician-Reported Outcome Measures)
  - Soft CROMs: Clinician-derived measures that incorporate patient input *(e.g., Harris Hip Score)*
  - Hard CROMs: Fully objective measures such as performance based tests *(e.g.,Timed up-and-go, grip strength, range of motion measured by goniometer)*

Responsiveness Measures

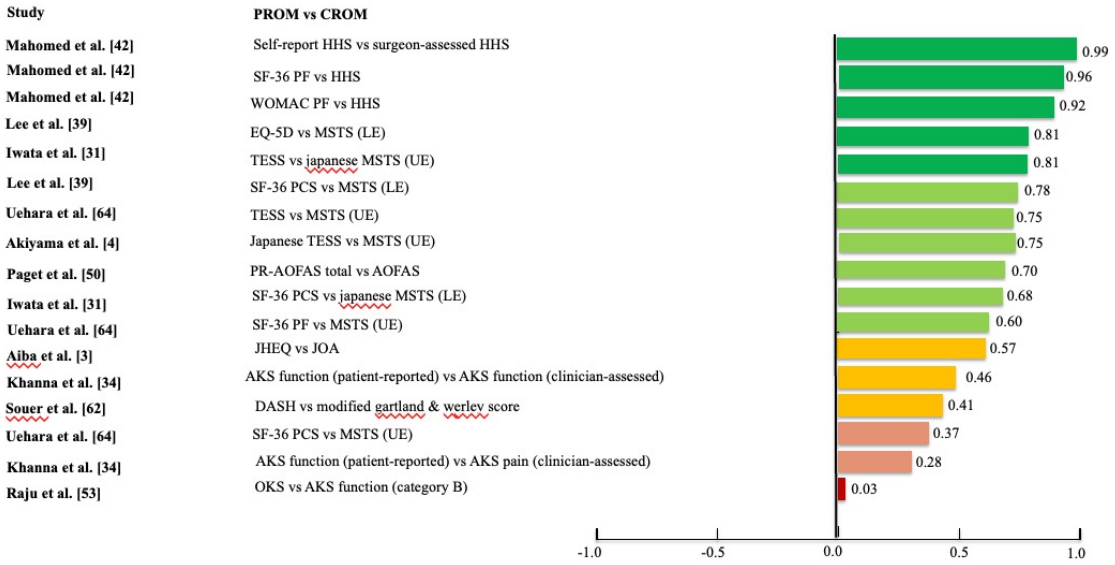
Responsiveness Measure	What it Represents	How is it Calculated?	Interpretation
Effect Size (ES)	Magnitude of change relative to baseline variability	Mean change ÷ SD of baseline scores	Larger values = greater detected change
Responsiveness Ratio (RR)	Ability to distinguish improvement from stability (signal-to-noise)	Mean change in improved ÷ SD of change in stable patients	Larger values = better detection of true change
Standardized Response Mean (SRM)	Consistency of change relative to variability in responses	Mean change ÷ SD of change scores	Larger values = greater change detected
Area Under the Curve (AUC)	Ability to discriminate improved vs stable patients	ROC analysis (sensitivity vs specificity)	Closer to 1.0 = better discrimination
Minimally Important Change (MIC)	Smallest change perceived as clinically meaningful	Anchor- or distribution-based methods	Larger values = larger change needed for clinical significance

Results

- 30 studies included spanning 1996-2025 Total of 27 PROM-CROM correlations; 9 studies assessed responsiveness
- Sample sizes ranged 20 to 717 (mean 143)
- Populations included: Arthroplasty/osteoarthritis (n=11), Foot & ankle (n=7), Musculoskeletal tumors (n=4), Distal radius fractures (n=3), Pediatric limb deformity (n=3), Spine (n=1), trauma (n=1)
- Majority conducted in North America, Europe, and Asia Most studies rated low risk or some concerns, primarily due to incomplete paired PROM-CROM data

PROM vs Soft CROMs

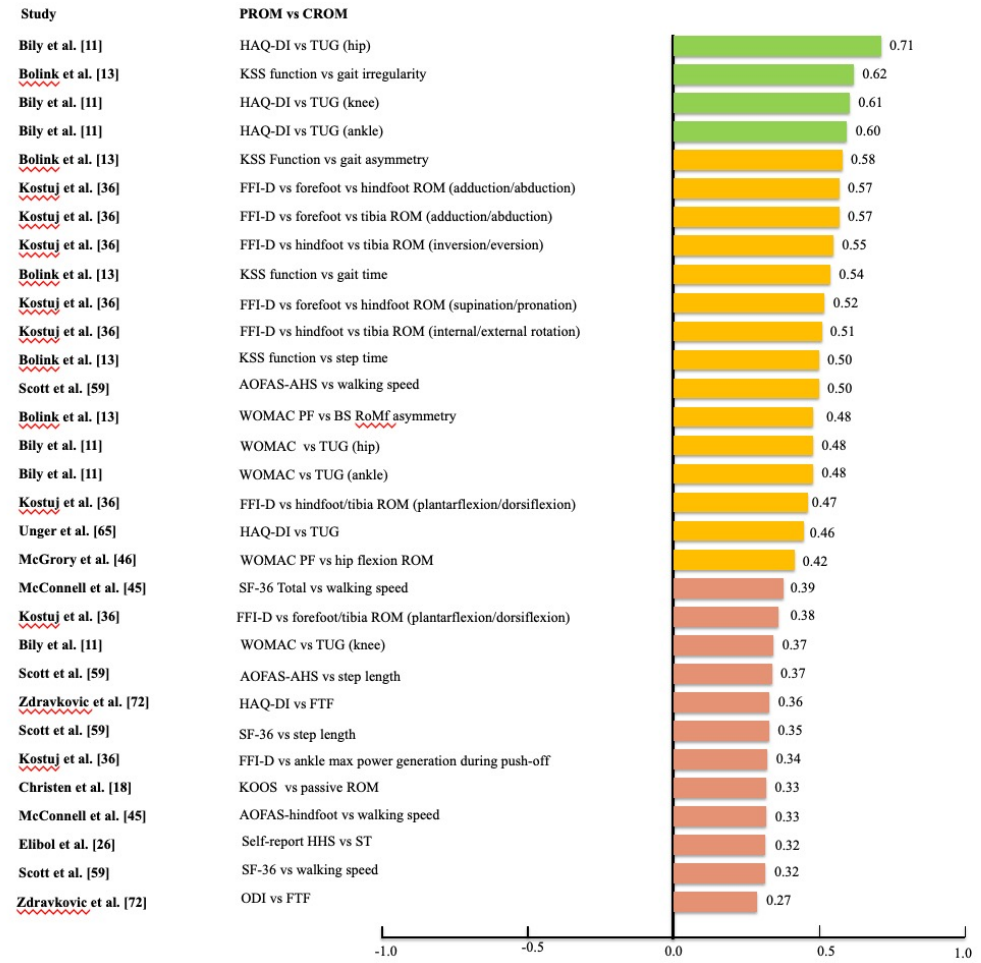
Representative subset of PROM vs Soft CROM correlations



- 52% strong/very strong correlations (r ≈ 0.60 to 0.99; -0.64 to 0.70)
- Strongest associations: HHS (patient vs clinician): r = 0.99, TESS vs MSTS: r = 0.75–0.81, AOFAS (patient vs clinician): r = 0.70
- Some variability in upper extremity measures (e.g., wrist scores: r = -0.07 to -0.70)

PROM vs Hard CROMs

Representative subset of PROM vs Hard CROM correlations



- 57% weak/very weak (range: r = 0.086 to 0.39; -0.001 to -0.36)
- 35% moderate (range: r = 0.42 to 0.58; -0.40 to -0.59)
- 8% strong (range: r = 0.60 to 0.71; -0.60 to -0.61)
- Weakest associations seen with: ROM in distal radius fractures: r = -0.01 to -0.03, Four Square Step Test: r = –0.19 to 0.09
- Soft CROMs show stronger associations with PROMs compared to Hard CROMs – Soft CROMs reflect greater conceptual overlap with PROMs, as both incorporate patient-reported domains

Condition-Specific vs General Health PROMs

- Correlations were highly variable (r = -0.70 to 0.99)
- Condition-specific PROMs showed stronger alignment:
  - TESS-MSTS: r = 0.75–0.81
  - AOFAS (patient vs clinician): r = 0.70
- General health PROMs showed weaker/inconsistent correlations:
  - EQ-5D: r = -0.29 to 0.13
  - Upper extremity PROMs: r = -0.046 to 0.41
  - ODI (spine): r = 0.27

Responsiveness Results

- Condition-specific PROMs showed stronger ability to detect change for postoperative recovery (e.g., WOMAC ES = 2.54)
- Soft CROMs demonstrated similarly high responsiveness (e.g., AOFAS SRM = 1.11–2.98; HHS RR = 1.70)
- Hard CROMs detected larger magnitude of change in early rehabilitation (e.g., ROM/TUG ES up to 1.41)
- General PROMs (SF-36, EQ-5D) showed lower ability to detect change (ES 0.09–0.77)
- Overall, responsiveness is instrument-specific, driven by item content, scoring range, and clinical context.

Conclusions

- Condition-specific PROMs offer better construct alignment, and Soft CROMs correlate more strongly with PROMs than Hard CROMs.
- Hard CROMs remain useful for short-term rehabilitation monitoring, while condition-specific PROMs and Soft CROMs capture patient-centered outcomes not reflected in objective measures.
- A protocol that emphasizes early Hard CROMs followed by later PROMs/Soft CROMs may best capture the full recovery trajectory.
- Orthopaedic studies and registries should include at least one condition-specific PROM (or Soft CROM) and one Hard CROM to ensure thorough assessment.
- Clinicians and researchers should rely on instrument-specific responsiveness data rather than generalizing across PROM or CROM types.
- Future research should directly compare specific PROM-CROM pairings within musculoskeletal domains to guide outcome assessment across conditions and populations.