



# Patient-reported outcome measures: it is time for authors, reviewers, journal editors and health care strategists to take sufficient responsibility

Michael Rindom Krogsgaard<sup>1</sup> · Christian Fugl Hansen<sup>1</sup>

Received: 5 August 2022 / Accepted: 19 August 2022 / Published online: 1 September 2022

© The Author(s) under exclusive licence to European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2022

**Keywords** Patient-reported outcome scores · PROM · Content validity · Construct validity · Type 2 errors

The misuse and abuse of patient-reported outcome measures (PROMs) in musculoskeletal research is leading to increasingly irresponsible healthcare recommendations. PROMs are progressively used as primary outcomes in clinical studies of musculoskeletal conditions, and there are multiple examples of the marked impact of such studies on recommended healthcare strategies. Typically, these studies find no difference in patient-reported outcome scores when comparing surgical treatment with other interventions (in most cases physiotherapy), or find differences believed not to be clinically relevant. A number of these studies have been published in high-ranking journals, and consequently, many national healthcare strategies do not recommend surgical treatment, at least not as the first choice. For essentially all these high-impact studies, it is well-documented that the PROMs are of questionable quality.

Strangely, there is no universal standard for documenting the quality of PROMs as outcomes. It is essential that clinical studies involving humans follow rules related to ethical issues according to the Helsinki declaration, to the handling of data according to the General Data Protection Regulation, to authorship according to criteria of the International Committee of Medical Journal Editors, and to the validity of clinical measurement methods (i.e., clinical biochemical analyses or physical measurements), and also that data are structurally meticulously prepared and presented, for example according to the Consolidated Standards of Reporting Trials (CONSORT) guidelines. All these issues must be

thoroughly reported and documented (e.g., by including a completed CONSORT checklist or by presenting a copy of the ethical permission), when a manuscript is submitted for publication. For clinical measurement methods to be trustworthy, proof of their accuracy, sensitivity, specificity, and repeatability is documented. This has not been the case for the PROMs used in the high-impact studies mentioned above.

An example of this is a randomized study of treatment of acute anterior cruciate ligament (ACL) rupture published in 2010 in the *New England Journal of Medicine* (NEJM) [9]. Five-year follow-up results were published in the *British Medical Journal* (BMJ) in 2013 [10], and an additional sub-study in the *British Journal of Sports Medicine* (BJSM) in 2017 [8]. The primary outcome in all published articles was the Knee injury and Osteoarthritis Outcome Score 4, KOOS4, which is a sum-score derived from 4 of the 5 domains of the original KOOS [27]. However, it was demonstrated already in 2008, using Rasch analysis, that KOOS is invalid for patients with ACL injury [3]. Nor has the validity of KOOS4 ever been documented; and hence, even the KOOS homepage ([www.koos.nu](http://www.koos.nu)) advises against its use. Also, KOOS has insufficient content validity (i.e., relevance and coverage of items) for patients with ACL injury [13, 31], which is not surprising, since 3 of 5 domains were taken directly from the Western Ontario McMaster Universities Osteoarthritis Index, a PROM developed in the mid-1980s for elderly persons with end-stage osteoarthritis of the knee or hip. The authors, reviewers, and editors of the three high-ranking journals that published those studies should have been well-aware of the problems with applying KOOS as the primary outcome. Nevertheless, the manuscripts were accepted and published. Since KOOS4 scores were not significantly different between the treatment arms, the study concluded in favor of a strategy of rehabilitation

✉ Michael Rindom Krogsgaard  
Michael.Rindom.Krogsgaard@RegionH.dk

<sup>1</sup> Section for Sports Traumatology M51 (a part of IOC Research Center Copenhagen), Bispebjerg and Frederiksberg University Hospital, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark

plus optional delayed ACL reconstruction over primary reconstruction, even though there was significantly greater clinical laxity of the knee in the rehabilitation group. The study has had a substantial impact on the treatment of ACL injury despite robust evidence that KOOS is an inadequate outcome measure, and this controversy is not unique. Other studies with a substantial impact that have used inadequate PROMs include several studies on the treatment of meniscal problems published in the *BMJ* (primary outcome was KOOS4 [18]), the *Journal of the American Medical Association (JAMA)* (primary outcome International Knee Documentation Committee Subjective Knee Form (IKDC)) [29], and the *NEJM Evidence* (primary outcome KOOS4) [28]; a study on femoroacetabular impingement in *BMJ* (primary outcome Hip Outcome Score) [24]; a study on treatment of proximal humeral fractures in *JAMA* (primary outcome Oxford Shoulder Score (OSS)) [26]; and two studies on sub-acromial impingement published in *The Lancet* (primary outcome OSS) [1] and *BJSM* (primary outcome Constant-Murley Score) [11], respectively. Finally, a study on ankle sprain published in *The Lancet* used the Foot and Ankle Score as an outcome [22]. None of these PROMs were adequately developed for the patient groups they were used to evaluate [14], but this was not mentioned or discussed in any of these studies, and the conclusions were not modified considering the suboptimal measurement methods.

It is reasonable to expect that at least studies published in high-ranked journals use valid measurement methods, but results based on PROMs without robust evidence of validity are published as often in high- as in lower-ranked journals [12].

## The hierarchy of measurement properties

It is universally accepted that PROMs when used as primary outcomes in clinical studies, must be valid. Even though articles reporting PROM data usually state that the questionnaire(s) used in their study are “valid and reliable”, and authors commonly cite specific references, the assurance of validity can in most cases be characterized as nonsense. Valid PROMs possess a series of measurement properties, all of which are relevant, but not equally important. There is a hierarchical order in PROM validation [25]:

1. Content validity
2. Structural validity, internal consistency, cross-cultural validity
3. Reliability, criterion validity, hypotheses testing construct validity, responsiveness

While adequate content validity ensures what is measured, the properties in points 2 and 3 ensure how well it

is measured. If content validity is inadequate, some items of the PROM might be irrelevant to the patients, and key concepts of the condition as the patient sees it might be missing [15, 31]. This leads to an increased risk of a type 2 measurement error (i.e., falsely claiming no difference between groups), because the PROM has too little sensitivity and specificity [32]. Next, when the structural validity of a PROM is assessed by classical test theory models (as are most PROMs) and not through a statistical measurement model developed to validate item responses (i.e., an item response theory (IRT) model), then a number of implicit scale properties related to the sum-score are simply assumed. Without assessment of these properties using an IRT model, it is not known if the items in a domain (e.g., a set of pain items) measure the same underlying construct (i.e., that the scale is unidimensional). It is also not known if scale scores reflect a value on an interval or ordinal scale level, meaning whether a difference in scores means the same throughout the scale (e.g., that an increase in pain from 2 to 4 has the same magnitude as an increase from 6 to 8). Dimensionality and scale level are both imperative when the scale scores are interpreted, and other classically reported properties such as floor/ceiling effects, test–retest reliability, Cronbach's alpha, MCID, responsiveness, etc., are all meaningless if the scale is not valid, because in such case the output—the scale score—is not valid! The arguments for using IRT models and not classical test theory models for validation of PROMs are clearly described in the literature [2].

Less than 10% of musculoskeletal PROMs possess adequate content validity [14]. In addition, many PROMs have only been subjected to the lowest ranking validity analyses in the hierarchy. Still, many PROMs with no evidence of content and construct validity are repeatedly characterized as ‘valid and reliable’. Remarkably, many studies conclude that a PROM is valid and reliable even though the reported analyses show the exact opposite. There are even cases where the most relevant and robust analyses reveal inadequate structural validity of a PROM, but where these results are only reported in the supplementary material and are not mentioned in the main text, and the conclusion is that the PROM has “good reliability and validity” [30].

## The consequences of continued use of inadequate PROMs

Does it make a difference to use inadequate PROMs? PROMs that are adequate are more accurate than inadequate PROMs. Practically all PROM scores will improve after treatment, even if a PROM for shoulder conditions has been used to evaluate patients with an ACL injury. This is less important when a single group of patients is assessed, but when between-group differences are evaluated, it is

of utmost importance that a condition-specific and valid PROM is used. Randomized Controlled Trials (RCTs) that use adequate PROMs show more than twice as often a significant between-group difference compared with RCTs that use inadequate PROMs [12], as they are more responsive [7]. This means that there is more than a 50% risk that the conclusion of the RCT mentioned above for the treatment of ACL injury is incorrect.

It is time for editors and reviewers to accept this. Likewise, it is time that healthcare strategists realize that local or national strategies based on studies that used inadequate PROMs can carry more than a 50% risk that healthcare funding is expended on less effective treatments.

It is first the responsibility of the author to ensure that valid outcome measures are used. Clinicians and researchers may find it difficult to understand the theoretical background of PROMs, how to apply a PROM in the clinical setting, and difficult to evaluate whether articles that conclude that particular PROMs are valid and reliable are actually valid. Therefore, with the aim to make all relevant information about PROMs available and understandable for clinicians and researchers without any particular statistical education, we and a group of colleagues produced a 10-article series, which was recently published in the *Scandinavian Journal of Medicine and Science in Sports* [2, 4–6, 12, 14, 17, 19–21]. This also included an analysis of the validity of 61 musculoskeletal PROMs [14].

It appears to be an uphill battle to convince authors to use adequate and valid PROMs. What about reviewers and journal editors? No better. COSMIN guidelines for the validity of PROMs [23, 25] are rigid and complex, but they define strictly what it takes to design a valid PROM. We performed an assessment of five relevant PROMs for knee and hip conditions using the COSMIN guidelines [23, 25]. We also invited the developers of all five PROMs to supply additional information, which they did. The analytic process was reported in a completely transparent manner and based on the highest scientific standards. We found the content validity of three widely-used knee and hip PROMs: KOOS, IKDC, and the modified Harris Hip Score to be inadequate. Our manuscript was rejected by three highly ranked musculoskeletal journals. They were all high-relevance journals for such a study, as they all have published and still publish manuscripts that use these PROMs as outcomes. The paper was finally accepted by *Knee Surgery, Sports Traumatology, Arthroscopy* [13], and following its publication the American Orthopedic Society for Sports Medicine and the American Board of Orthopedic Surgery decided to fund a retrospective content analysis of IKDC [16]. This is a positive development, but it should also be remembered that content validity is secured through patient engagement in the developmental process. Secondary content validation of an already existing PROM, which did not involve patients in its

development should include concept elicitation techniques and open discussions with target patients to confirm content relevance and coverage of items themes until data saturation is achieved. This will most likely result in a substantially modified version of the PROM with different items and perhaps even new domains [15].

But it is worse than that. The problem is not just ignorance from reviewers, editors and journals. The greater problem is, there are already myriads of published studies that wrongly describe specific PROMs as valid and reliable. It can for instance be studies assessing the validity of a newly developed PROM or a local translation of a well-established PROM. These articles are used as references to state that PROMs used in clinical studies are “valid and reliable”. Through this large pool of articles, many journals support that clinical studies use invalid PROMs.

Ultimately, it is the responsibility of editors and journals only to publish validity studies of PROMs that are based on sufficient analytic methods. If this is not done, then journals are feeding readers with false information on invalid PROMs. A PROM cannot be considered valid if it does not have sufficient evidence of content validity, if only superficial measurement properties of the PROM have been assessed, or if the structure has not been assessed by IRT models [2]. Following these simple principles would reduce the number of articles that erroneously conclude a PROM to be valid and reliable by more than 90%. Journals should also consider adding a note of warning on already published articles that lack a scientific foundation for a statement of validity of a PROM.

We appeal to authors, reviewers, and particularly journal editors to meet their responsibility and publish only clinical studies that apply valid and adequate PROMs. The level of content validity and evidence of fit to a measurement model is the very least of what authors should state and editors demand - an example of an informative statement is given in [20]. Such studies will be the foundation for decisions on healthcare strategies and will help ensure that resources are not wasted on inferior treatments. This appeal also includes studies on the validity of PROMs.

What if no valid PROM for a particular study exists, for instance if the only PROM that has been developed through patient involvement is old and a secondary content validation has shown that it is outdated [15]? The best solution is of course to develop a new PROM—which is not realistic for most researchers. The next best solution is to use the ‘least inadequate’ PROM. This can be justified, but the resultant downgrade of the validity of the results should be discussed and be reflected in the discussion and conclusion. This should also be considered when the study is included as the basis for healthcare strategies. However, other clinical outcomes with more robust measurement properties (e.g., range of motion, stability, nights with disrupted sleep,

return-to-work/sport) will most likely be more appropriate as primary outcomes.

The idea that PROM scores are the most important outcomes fails in cases where an inadequate PROM has been used. In such cases, healthcare strategies are best if they are based on other outcomes.

**Author contributions** Both authors planned and completed the manuscript.

**Funding** No particular funding was provided for this research.

## Declarations

**Conflict of interest** Both authors declare no conflicts of interest.

**Ethical approval** No approval required.

**Informed consent** Not relevant.

## References

1. Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N et al (2018) Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. *Lancet* 391:329–338
2. Christensen KB, Comins JD, Krogsgaard MR, Brodersen J, Jensen J, Hansen CF et al (2021) Psychometric validation of PROM instruments. *Scand J Med Sci Sports* 31:1225–1238
3. Comins J, Brodersen J, Krogsgaard M, Beyer N (2008) Rasch analysis of the knee injury and osteoarthritis outcome score (KOOS): a statistical re-evaluation. *Scand J Med Sci Sports* 18:336–345
4. Comins JD, Brodersen J, Christensen KB, Jensen J, Hansen CF, Krogsgaard MR (2021) Responsiveness, minimal important difference, minimal relevant difference, and optimal number of patients for a study. *Scand J Med Sci Sports* 31:1239–1248
5. Comins JD, Brodersen J, Siersma V, Jensen J, Hansen CF, Krogsgaard MR (2021) Choosing the most appropriate PROM for clinical studies in sports medicine. *Scand J Med Sci Sports* 31:1209–1215
6. Comins JD, Brodersen J, Siersma V, Jensen J, Hansen CF, Krogsgaard MR (2021) How to develop a condition-specific PROM. *Scand J Med Sci Sports* 31:1216–1224
7. Comins JD, Siersma VD, Lind M, Jakobsen BW, Krogsgaard MR (2018) KNEES-ACL has superior responsiveness compared to the most commonly used patient-reported outcome measures for anterior cruciate ligament injury. *Knee Surg Sports Traumatol Arthrosc* 26:2438–2446
8. Filbay SR, Roos EM, Frobell RB, Roemer F, Ranstam J, Lohmander LS (2017) Delaying ACL reconstruction and treating with exercise therapy alone may alter prognostic factors for 5-year outcome: an exploratory analysis of the KANON trial. *Br J Sports Med* 51:1622–1629
9. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS (2010) A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med* 363:331–342
10. Frobell RB, Roos HP, Roos EM, Roemer FW, Ranstam J, Lohmander LS (2013) Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. *BMJ* 346:f232
11. Hallgren HC, Holmgren T, Oberg B, Johansson K, Adolfsson LE (2014) A specific exercise strategy reduced the need for surgery in subacromial pain patients. *Br J Sports Med* 48:1431–1436
12. Hansen CF, Jensen J, Brodersen J, Siersma V, Comins JD, Krogsgaard MR (2021) Are adequate PROMs used as outcomes in randomized controlled trials? an analysis of 54 trials. *Scand J Med Sci Sports* 31:972–981
13. Hansen CF, Jensen J, Odgaard A, Siersma V, Comins JD, Brodersen J et al (2021) Four of five frequently used orthopedic PROMs possess inadequate content validity: a COSMIN evaluation of the mHHS, HAGOS, IKDC-SKF, KOOS and KNEES-ACL. *Knee Surg Sports Traumatol Arthrosc*. <https://doi.org/10.1007/s00167-021-06761-y>
14. Hansen CF, Jensen J, Siersma V, Brodersen J, Comins JD, Krogsgaard MR (2021) A catalogue of PROMs in sports science - quality assessment of PROM development and validation. Article 7 in a series of 10. *Scand J Med Sci Sports*. <https://doi.org/10.1111/sms.13923>
15. Holmenlund C, Overgaard S, Bilberg R, Varnum C (2021) Evaluation of the Oxford hip score: does it still have content validity? Interviews of total hip arthroplasty patients. *Health Qual Life Outcomes* 19:237
16. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P et al (2001) Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med* 29:600–613
17. Jensen J, Hansen CF, Brodersen J, Comins JD, Krogsgaard MR (2021) Are PROMs used adequately in sports research? An analysis of 54 randomized controlled trials with PROMs as endpoint. *Scand J Med Sci Sports* 31:982–990
18. Kise NJ, Risberg MA, Stensrud S, Ranstam J, Engebretsen L, Roos EM (2016) Exercise therapy versus arthroscopic partial meniscectomy for degenerative meniscal tear in middle aged patients: randomised controlled trial with two year follow-up. *BMJ* 354:i3740
19. Krogsgaard MR, Brodersen J, Christensen KB, Siersma V, Jensen J, Hansen CF et al (2021) How to translate and locally adapt a PROM. Assessment of cross-cultural differential item functioning. *Scand J Med Sci Sports* 31:999–1008
20. Krogsgaard MR, Brodersen J, Christensen KB, Siersma V, Kreiner S, Jensen J et al (2021) What is a PROM and why do we need it? *Scand J Med Sci Sports* 31:967–971
21. Krogsgaard MR, Brodersen J, Jensen J, Hansen CF, Comins JD (2021) Potential problems in the use of patient reported outcome measures (PROMs) and reporting of PROM data in sports science. *Scand J Med Sci Sports* 31:1249–1258
22. Lamb SE, Marsh JL, Hutton JL, Nakash R, Cooke MW, Collaborative Ankle Support T (2009) Mechanical supports for acute, severe ankle sprain: a pragmatic, multicentre, randomised controlled trial. *Lancet* 373:575–581
23. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM et al (2018) COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27:1171–1179
24. Palmer AJR, Ayyar Gupta V, Fernquest S, Rombach I, Dutton SJ, Mansour R et al (2019) Arthroscopic hip surgery compared with physiotherapy and activity modification for the treatment of symptomatic femoroacetabular impingement: multicentre randomised controlled trial. *BMJ* 364:i185
25. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW et al (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27:1147–1157

26. Rangan A, Handoll H, Brealey S, Jefferson L, Keding A, Martin BC et al (2015) Surgical vs nonsurgical treatment of adults with displaced fractures of the proximal humerus: the PROFHER randomized clinical trial. *JAMA* 313:1037–1047
27. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD (1998) Knee injury and osteoarthritis outcome score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 28:88–96
28. Skou ST, Hölmich P, Lind M, Jensen HP, Jensen C, Garval M et al (2022) Early surgery or exercise and education for meniscal tears in young adults. *NEJM Evidence* 1:EVIDoa2100038
29. van de Graaf VA, Noorduyn JCA, Willigenburg NW, Butter IK, de Gast A, Mol BW et al (2018) Effect of early surgery vs physical therapy on knee function among patients with nonobstructive meniscal tears: the ESCAPE randomized clinical trial. *JAMA* 320:1328–1337
30. van der Linde JA, van Kampen DA, van Beers LW, van Deurzen DF, Terwee CB, Willems WJ (2015) The Oxford shoulder Instability score; validation in dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res* 10:146
31. van Meer BL, Meuffels DE, Vissers MM, Bierma-Zeinstra SM, Verhaar JA, Terwee CB et al (2013) Knee injury and osteoarthritis outcome score or international knee documentation committee subjective knee form: which questionnaire is most useful to monitor patients with an anterior cruciate ligament rupture in the short term? *Arthroscopy* 29:701–715
32. Zsidai B, Narup E, Senorski EH, Lind M, Spalding T, Musahl V et al (2022) The knee injury and osteoarthritis outcome score: shortcomings in evaluating knee function in persons undergoing ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc.* <https://doi.org/10.1007/s00167-022-06990-9>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.