CORRESPONDENCE

Open Access

Classifications of good versus poor outcome following knee arthroplasty should not be defined using arbitrary criteria

Daniel L. Riddle[1*] and Levent Dumenci[2]

Abstract

A recently published paper by te Molder and colleagues in BMC Musculoskeletal Disorders confirmed prior reports indicating that definitions of good versus poor outcome cutoff scores for relevant knee arthroplasty outcomes including pain and function are heterogeneous and that this heterogeneity prevents generalizable inferences. In this Correspondence, we highlight an additional and, in our view, a more important problem with the substantial literature on this topic. There also is high homogeneity in that all studies relied on arbitrarily defined cutoff scores to differentiate good versus poor outcome. We discuss this problem and propose a method to avoid repeating the same problem in future studies designed to group patients into those with good versus those with poor outcome following knee arthroplasty.

Keywords: Knee, arthroplasty, outcome, pain, function

Main text

The systematic review by te Molder and colleagues [1] summarized various methods used by investigators to dichotomize outcomes of patients with knee arthroplasty (KA) as either good or poor. There are important reasons for wanting to know if a patient's KA outcome is good or poor. For example, interventions to improve outcome can be specifically designed and targeted to patients fitting the poor outcome phenotype. The dilemma with categorizing outcome, as te Molder et al. and others [2, 3] have noted, is that definitions of good versus poor outcome vary substantially across the many studies that have attempted to categorize outcomes following KA.

Variation precludes consensus and prevents meaningful comparisons across study cohorts. We noted an additional problem with evidence classifying outcome as good or poor [4]. Definitions of good versus poor outcome are grounded in the use of arbitrary cutoff values, whether based on final outcome score, percent or absolute change from baseline or the Minimal Clinically Important Difference (MCID) family of change indicators.

The main conclusion of the study by te Molder and colleagues was that there was substantial heterogeneity in the 47 definitions of good versus poor KA outcomes. In our view, te Molder et al. should also have focused on implications related to the homogeneity of these 47 definitions. All studies in the review used the cutoff method to determine good versus poor outcome. Cutoff scores are, by definition, arbitrary. Supplemental file 3 in the study by te Molder et al.

* Correspondence: dlriddle@vcu.edu
[1]Departments of Physical Therapy, Orthopaedic Surgery and Rheumatology, Virginia Commonwealth University, Richmond, VA, USA
[2]Department of Epidemiology and Biostatistics, 1301 Cecil B. Moore, Ave., Ritter Annex, Room 939, Temple University, Philadelphia, PA 19122, USA

[1] provides a partial list of definitions used to establish arbitrary cutoff scores (including two of our prior studies [5, 6]). For example, Brander and colleagues indicated that a 0 (no pain) to 100 (worst pain imaginable) visual analogue pain scale of > 40 indicated a poor pain outcome [7]. This cutoff is arbitrary.

Over three decades ago, researchers and clinicians were warned about the arbitrary nature of the cutoff method for clinical decision making and proposed latent class analysis as a scientifically defensible alternative [8]. Recent methodological developments also have been extensively documented [9]. In 2011, we further elaborated on why the cutoff method should not be used to determine patient groupings in scientific research, developed methods originating from discrete latent variable modeling approaches to circumvent problems associated with the arbitrary cutoff method, and provided multiple examples using real-life data to illustrate how new methods could be used to answer scientific questions [10]. In 2019, we used methods originating from a longitudinal discrete latent variable modeling framework to define poor versus good outcomes in KA [4]. For reasons that were unclear to us, given that it met inclusion criteria by te Molder and colleagues, our 2019 study [11] was not included in the review. This latent variable modeling method does not rely on biased good versus poor cutoffs but rather on statistical modeling that is free of arbitrary decision-making.

The cutoff method is an impediment to scientific progress. If we continue to overlook homogeneity, and don't acknowledge that this evidence relies on arbitrary cutoff scores, we will keep using arbitrary cutoff scores to define poor outcome in KA. Going down this road would lead to even more studies that rely on arbitrary cutoffs and we'll have made no progress. In our view, the answer to the lack-of-consensus problem posed by te Molder et al. for defining good versus poor outcome in KA is not to continue relying on arbitrary cutoff scores. Instead, we should rely on a non-biased statistical model-based approach to categorizing good versus poor outcome [11].

Once the cutoff method is replaced with model-based approaches, we suggest the following strategy: Researchers focus on factors that matter most as the sources of outcome variability. For example, what constitutes the KA outcome (e.g., self-reported knee pain, function, health-related quality of life)? Whose perspective(s) should be captured (e.g., patients, relatives, surgeons, or a combination)? What are the optimal time point(s) for measuring outcome (e.g., 2 weeks before and after KA, and four additional times over subsequent 2 years)? What are the key predictors of good versus poor outcome classes? We contend that a coordinated consensus-based strategy like the one described above is needed to shift the paradigm of this type of work and advance the science of good versus poor outcome identification in KA.

Abbreviations
KA: Knee arthroplasty; MCID: Minimal Clinically Important Difference

Author's response
Malou E. M. te Molder[3,4], José M. H. Smolders[5], Petra J. C. Heesterbeek[4] and Cornelia H. M. van den Ende[6,7]

[3]Department of Orthopedic Research, Sint Maartenskliniek, Nijmegen, The Netherlands
[4]Department of Orthopaedic Surgery, Radboud University Medical Center, Nijmegen, The Netherlands
[5]Department of Orthopedic Surgery, Sint Maartenskliniek, Nijmegen, The Netherlands
[6]Department of Rheumatology Research, Sint Maartenskliniek, Nijmegen, The Netherlands
[7]Department of Rheumatology, Radboud University Medical Center, Nijmegen, The Netherlands

Thank you for giving us the opportunity to write a response to the correspondence "Classification of good versus poor outcome following knee arthroplasty should not be defined using arbitrary criteria".

We thank Riddle *et al* for their interest and critical assessment of our inventory review in which we summarized definitions of poor response to total knee arthroplasty (TKA). Riddle *et al* suggest that we should have focused on implications related to the arbitrary and homogeneous use of cutoff points. Instead, Riddle *et al* strongly recommend to rely on a model-based approach to define poor response to TKA. Several model-based approaches are available to identify subgroups with different growth curves. We acknowledge the value of those models. However, a major limitation of these types of models is that membership of poor and good outcome classes can only be determined afterwards and that results with regard to membership of classes cannot be transferred to other study populations.

We fully agree that a drawback of dichotomizing data is data reduction and that a continuous measure is more sensitive to change, and, therefore, more useful on individual level and in clinical decision making. Mixture models can provide more in-depth insight in the course of outcome over time and its determinants. However, to allow comparisons of the prevalence of poor responders to TKA across hospitals, countries, and over time, a strict definition is necessary with clearly defined criteria and thresholds. For this purpose, a dichotomous outcome is more appropriate while the use of mixture models is preferred if the purpose is to gain insight in factors underlying outcomes over time.

The second remark relates to the reason why the 2019 study by Dumenci *et al* [4] was not included in the review. As the authors indicated in their correspondence, inclusion criteria for our inventory review focused on predefined dichotomized cutoffs to define poor outcome and, therefore, studies including model-based approaches (such as the study by Dumenci *et al*) were not included.

Nevertheless, we thank Riddle *et al* for their suggested strategy to focus on factors, perspective(s) and optimal time point(s) for measuring good versus poor outcome in TKA. The intended strategy of our project is exactly what Riddle *et al* proposed. We use the results of our inventory review and an ongoing qualitative study in patients as well as health care providers to focus on relevant concepts underlying a poor response to TKA. Once the relevant concepts have been identified, we can start the discussion among panelists of a subsequent Delphi study. The ultimate aim of our project is to reach consensus on a definition of poor response to TKA after which we, hopefully, can properly compare the prevalence of poor responders across hospitals and countries.

References

1. te Molder ME, Smolders JM, Hesterbeek PJ, Van Den Ende CH. Definitions of poor outcome after total knee arthroplasty: an inventory review. BMC Musculoskelet Disord. 2020;21:378.
2. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients BMJ Open. 2012;2:e000435.
3. Wylde V, Dieppe P, Hewlett S, Learmonth ID. Total knee replacement: is it really an effective procedure for all? Knee. 2007;14:417–23.
4. Dumenci L, Perera R, Keefe F, Ang D, Slover J, Jensen M, et al. Model-based pain and function outcome trajectory types for patients undergoing knee arthroplasty: a secondary analysis from a randomized clinical trial. Osteoarthr Cartil. 2019;27:878–84.
5. Riddle DL, Wade JB, Jiranek WA, Kong X. Preoperative pain catastrophizing predicts pain outcome after knee arthroplasty. Clin Orthop Relat Res. 2010; 468:798–806.
6. Riddle DL, Golladay GJ, Jiranek WA, Perera RA. External validation of a prognostic model for predicting nonresponse following knee Arthroplasty. J Arthroplast. 2017;32:1153–8.
7. Brander V, Gondek S, Martin E, Stulberg SD. Pain and depression influence outcome 5 years after knee replacement surgery. Clin Orthop Relat Res. 2007;464:21–6.
8. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. Stat Med. 1986;5:21–7.
9. Rupp AA, Templin J, Henson RA. Diagnostic measurement: theory, method, and application. New York: The Guilford Press; 2010.
10. Dumenci L. The psychometric latent agreement model (PLAM) for discrete latent variables measured by multiple items. Organ Res Methods. 2011;14: 91–115.
11. Riddle DL, Dumenci L. Modeling longitudinal osteoarthritis data to identify homogeneous subgroups: opportunities and challenges in a burgeoning literature. Osteoarthr Cartil. 2015;23:1035–7.