



Validation of the Impact Index: can we measure disease effects on quality of life in patients with hip and knee osteoarthritis?

K. D. Valentine^{1,2} · S. Brodney¹ · Ha Vo¹ · Karen Sepucha^{1,2} · Floyd J. Fowler³ · Michael J. Barry^{1,2}

Accepted: 30 November 2020 / Published online: 2 January 2021

© The Author(s), under exclusive licence to Springer Nature Switzerland AG part of Springer Nature 2021

Abstract

Purpose To validate the Impact Index, a short, publicly available scale that measures the extent to which a respondent's health problem adversely impacts their quality of life.

Methods Secondary analysis of patients with hip or knee osteoarthritis surveyed after visiting a surgeon at baseline ($N=322$) and about 6 months after the visit ($N=283$). Patients responded to the Impact Index and previously validated questionnaires about overall health, pain, and function. The Impact Index includes four questions that ask how much the respondent is bothered, worried, limited, or in pain due to their health condition over the past 30 days. Total scores range from 0 to 12; higher scores indicate more deleterious impact.

Results Patients were mostly female (55%), majority white (95%), had an average age of 65 (SD=9), and most had surgery (64%). The baseline Impact Index score was 9.48 (SD=2.63); at follow up 4.75 (SD=3.54). Impact Index was related to overall health at baseline ($r=-0.49$). For knee patients at baseline, Impact Index was negatively related to their knee symptoms ($r=-0.49$) and knee pain ($r=-0.67$). For hip patients at baseline, Impact Index was negatively related to the Harris Hip score ($r=-0.62$). Scale directions varied; however, the signs of all correlations were as hypothesized. The Impact Index was predictive of surgical choice ($p<.001$, OR=1.45), however, overall health ($p=.88$) and comorbidity ($p=.24$) measures were not. Reliability was acceptable ($\alpha=0.85$). Responsiveness statistics suggested overall health, pain, function, and Impact Index measures reflected improvement patients experienced from surgery. The Impact Index had the largest effect sizes (>-3.4) and Guyatt Responsiveness Statistics (>-2.3).

Conclusions The Impact Index demonstrated strong evidence of validity, reliability, and responsiveness in hip or knee osteoarthritis patients.

Keywords Patient-reported outcomes · Reliability · Validity · Quality of life · Functional status · Hip and knee osteoarthritis

Plain English summary

This study reports on how well a scale called the Impact Index measures how much a health problem has a negative impact on a patient's quality of life. The scale is made up of four questions that ask patients about how much they

worried about their health because of their health problem, how much they were bothered by their health problem, how much their health problem limited what they were able to do, and how much pain their health problem caused them. Higher scores on the scale tell us that the impact of the health problem on the patient is larger. Patients with prostate problems have used this scale before and the scale was able to measure how much the disease had a negative impact on the patient's life. Here we show that this is also true for patients with hip or knee osteoarthritis. The Impact Index was compared to scales that measure overall health and pain and we saw that the Impact Index scores were higher for patients with worse health and higher for patients with more pain. We also saw that the scale was able to measure the improvement patients saw from having joint replacement

✉ Michael J. Barry
MBarry@mgh.harvard.edu

¹ Division of General Internal Medicine, Massachusetts General Hospital, 100 Cambridge St, 16th Floor, Boston, MA 02114, USA

² Harvard Medical School, Boston, MA, USA

³ Center for Survey Research, University of Massachusetts, Boston, Boston, MA, USA

surgery. With these results we believe the Impact Index is a short scale that can be used to measure how patients are impacted by their health problem over time.

Introduction

Disease processes often affect not only the quantity of people's lives, but also through bothersome symptoms and functional impairments, their quality of life as well. Similarly, medical treatment is aimed at mitigating the effects of illness through reducing symptoms and improving functioning defined broadly, including physical, mental, and social domains. Because only people themselves can rate their health in these areas, patient-reported outcome measures (PROMs) have been developed and are increasingly used in both research and clinical practice to describe the natural history of disease and the response to treatment in terms most relevant to patients [1, 2]. When patient-reported measures of their health status are repeated over time, particularly before and after treatment, they become outcome measures, putting the "O" in PROMs.

Black broadly classifies PROMs as disease specific and generic [1, 2]. Disease-specific measures focus on the effect of a particular disease on related aspects of a person's health, while generic measures focus on aspects of a person's health as may be affected by any (and all) disease processes. Disease-specific measures make more clinical sense to patients and clinicians, while generic measures allow comparisons of disease impact from condition to condition. Disease-specific measures are usually more sensitive to the effects of treatment than generic measures [3]. Disease-specific measures and generic measures are often combined to get a broader view of the impact of a disease and its treatment. However, when a disease-specific measure improves with treatment and a generic measure does not, it is often unclear whether the impact of the disease is just not so important in terms of general health, or whether the generic measures are just not sensitive enough to detect disease impacts that are indeed very important to patients.

In our previous qualitative work with men with prostate diseases, we found four main recurrent issues when patients affected by a health condition described how that condition affected them: discomfort, limiting what they could do, worry about their health, and a general feeling that they were bothered by having the condition. Using that information, we developed and validated a self-reported four-item disease-specific measure for men affected by benign prostatic hyperplasia (BPH). This scale, called the BPH Impact Index, showed good consistency, reliability, and validity [4]. Moreover, the measure was much more responsive than generic measures of health status for men with this condition who underwent treatment [4, 5].

We believe the domains of disease impact defined among men with BPH should be generalizable to other symptomatic conditions. The Impact Index has already been successfully adapted to assess the impact of androgen deprivation for men after radical prostatectomy and identified that androgen deprived men were more impacted by their cancer and its treatment [6]. Therefore, we have used simple replacement of the health condition described in the BPH Impact Index to allow this measure to be used in various health contexts with the same items simply asking about different conditions. If the Impact Index proves similarly reliable and valid across conditions, then the set of measures may provide the specificity and responsiveness of disease-specific measures with the comparability across conditions of generic measures. Such a demonstration of validity across conditions may suggest that the Impact Index can begin to bridge the divide between disease-specific and generic measures by creating a generalizable measure that can be tailored to be disease specific for a given disease. The purpose of this study is to examine the reliability, validity and responsiveness of the 4-item Impact Index modified for patients with hip and knee osteoarthritis.

Methods

Participant and procedures

This study is a secondary analysis of data from a longitudinal randomized trial. The DECIDE-OA study was a multi-site randomized controlled trial conducted at an academic medical center, community hospital, and specialty hospital that examined the impact of patient-directed and physician-directed decision support strategies on the quality of treatment decisions for hip and knee osteoarthritis [7, 8]. Eligible patients were 21 or over, read and spoke English or Spanish, had a diagnosis of hip or knee osteoarthritis, and attended a visit with a surgeon.

Participants completed surveys at three time points: shortly before the visit, about one week after the visit, and at a follow-up either 6 months after the visit (if they did not undergo surgery) or 6 months after surgery. The Impact Index was added to the post-visit and follow up surveys midway through the study and therefore only a subset of the original study's sample responded to the Impact Index. All patients completed a self-reported generic measure of overall health. Knee and hip surgery candidates were also asked to complete previously validated self-reported disease-specific measures of their knee and hip problems, respectively. Chart reviews were conducted 6 months after the visit to determine surgical status and to extract data to identify the degree of osteoarthritis (OA) radiographic severity as well as their score on the Charlson Comorbidity Index.

Partners Human Research Committee was the Central Institutional Review Board for the study. The DECIDE-OA study is registered on Clinicaltrials.gov (# NCT02729831).

Measures

Impact index

The Impact Index, which includes four questions, was collected from a subset of patients via self-report surveys both 1 week after the visit and at follow-up. The items ask about how much the person is bothered, worried, limited, or in pain from their health condition over the past 30 days (see Table 1). The response options included none = 0, a little = 1, some = 2, and a lot = 3. A total score (range 0–12), is calculated by summing scores on each of the four items. The higher the score, the greater impact the patient's condition is having on their quality of life [4].

Overall health

Patients self-reported their responses prior to the visit as well as at follow-up. The EQ-5D-3L is a widely used 5-item generic measure of overall health status with 3 response options for each item. The items are scored on a scale of 0–1 with 0 indicating death and 1 indicating full health [9]. A change of 0.1 points is considered the minimum important change [10].

OA Severity

OA radiographic severity was a 5-level variable (mild, mild–moderate, moderate, moderate–severe, severe) taken from radiologists' written reports from patient X-rays where they identified the radiographic severity.

Charlson Comorbidity Index (CCI)

The CCI is a method of categorizing prognostic comorbidity that uses administrative data in a patient's chart to identify their risk of death from comorbid diseases. Scores range from 0 (no disease burden) to 29 (maximal disease burden) [11].

KOOS pain/symptom/KOOS JR

The Knee Injury and Osteoarthritis Score (KOOS) was self-reported prior to the visit as well as at follow-up and is a measure of the patients' opinions about their knee problems. The five subscales include Pain, Symptoms, Function in Daily Living, Function in Sports and Recreation, and Knee-Related Quality of Life. Only the Pain (7 items) and Symptom (9 items) subscales were included. A normalized score for each scale was calculated ranging from 0 to 100 with 100 indicating no pain/symptoms and 0 indicating extreme pain/symptoms [12, 13]. The KOOS JR is a 7-item, validated, short form of the full 5-subscale KOOS score. This score uses 1 item from the Symptom subscale, 4 items from the Pain subscale, and 2 items from the Function in Daily Living subscale [14]. A normalized score was calculated ranging from 0 to 100 with 100 indicating perfect knee health [15].

Harris Hip Score

The Harris Hip Score was self-reported prior to the visit as well as at follow-up. The 10 items measure 4 domains covering pain, function, range of motion, and deformity for each hip. The measure is scored 0 to 100 and a higher score indicates less dysfunction [13, 16, 17].

Statistical analyses

First, we calculated the scores for each measure discussed above. Data were checked for the assumptions of correlation (i.e. outliers, normality, linearity, and homoscedasticity) and found to be acceptable.

Reliability: Cronbach's α and 95% confidence intervals (CIs) were used to assess inter-item consistency (targeting a coefficient > 0.7). Pearson's correlations were used to explore inter-item relationships as well as the relationship of each item to the overall Impact Index score. Here, we looked for correlations between 0.3 and 0.9, as those less than 0.3 would indicate that the item was assessing a different construct and those greater than 0.9 would indicate redundancy.

Validity we examined the relationships with overall health, patient-reported symptom severity for knees or hips, CCI, and radiographic severity using Pearson's correlations and CIs. Specifically, we tested whether Impact Index scores were negatively correlated with overall health, the KOOS-JR, KOOS Pain, and KOOS Symptom

Table 1 Impact Index items

-
- (1) Over the past 30 days, how much did you worry about your health because of your [hip/knee]?
 - (2) Over the past 30 days, how much were you bothered by your [hip/knee]?
 - (3) Over the past 30 days, how much did your [hip/knee] limit what you were able to do?
 - (4) Over the past 30 days, how much pain or physical discomfort did your [hip/knee] cause you?
-

scores, and Harris Hip scores (as higher values on these measures indicate perfect health). We tested whether the Impact Index was positively correlated with CCI as higher values on this measure indicate worse health. We also examined correlation between the X-ray severity and the Impact Index; however, we did not hypothesize a strong relationship between these as previous work has indicated that there is a limited relationship between patient-perceived functional impact and radiographic severity [18, 19]. Further, we tested the predictive validity of the Impact Index using hierarchical logistic regression models in three steps to predict participants' choice to undergo surgery. The first step included only the Impact Index predicting surgery choice; the second step added overall health with the EQ-5D measure; the third step additionally added CCI.

Responsiveness finally, we tested the responsiveness of the scale by calculating effect sizes and Guyatt's responsiveness statistics. To do this, first, raw changes in the measures were calculated by subtracting the score from baseline from the score from 6 months after the visit for those patients who chose to undergo surgery. The mean of these individual change scores was then divided by the standard deviation of the measure at baseline to create the standardized effect size. Next, Guyatt's responsiveness statistics were calculated for each measure by dividing the mean raw change by the standard deviation of the individual changes in scores of patients who chose not to undergo surgery. Generally, higher values (i.e. > 0.8) indicate substantial responsiveness in a scale [20]. For both

the effect sizes and Guyatt's responsiveness statistics 95% CIs were calculated [21].

Given that some patients will choose to undergo surgery and others will choose not to undergo surgery, we would also expect that those who undergo surgery will have a greater decrease in the impact from their symptoms than those who did not undergo surgery. Thus, the Guyatt responsiveness statistic indicates how much the group who underwent surgery decreased their symptoms or impact given the variability we see present in the group who did not undergo surgery.

Results

Of the subset of patients who were sent a survey containing the Impact Index one week after the visit, 322/383 completed the surveys; 283 also completed follow-up surveys. Participants at baseline had an average age of 64.8 (SD = 9.5), 55% ($N = 177$) were female, most were white (95.3%, $N = 307$), all spoke English, and most had knee osteoarthritis (66.0%, $N = 212$). Demographics for the 322 patients who completed the survey after the visit can be seen in Table 2.

For our sample, the average Impact Index score for all patients after the visit was 9.48 (SD = 2.63) and at follow-up was 4.75 (SD = 3.54). A total of 187 of the 283 follow up patients had surgery, while the remaining 96 patients had not undergone surgery by the 6-month follow-up. Scores on the Impact Index at baseline did not differ significantly between hip and knee patients ($p = 0.18$). Differences were

Table 2 Characteristics of 322 patients

Variable	Percent (N)
Female	55 (177)
White	95 (307)
Hip osteoarthritis	34 (110) ^a
Knee osteoarthritis	66 (212) ^b
Education	
8th grade or less	0 (1)
Some high school but not graduated	1 (4)
High school graduate or GED	11 (34)
Some college or 2-year degree	24 (75)
4-year college graduate	22 (71)
More than a 4-year college degree	42 (132)
How often do you have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy?	
Never	72 (231)
Rarely	16 (53)
Sometimes	9 (29)
Often	1 (4)
Always	2 (5)

^a84/110 had surgery over the follow-up interval

^b21/212 had surgery over the follow-up interval

seen by condition at follow up [$t(185) = -3.17, p = 0.002, d = -0.39$] as patients who had knee surgery had greater Impact Index scores ($M = 5.23, SD = 3.52$) compared to those who had hip surgery ($M = 3.87, SD = 3.42$). Given this difference, we report on responsiveness separately for hip and knee patients below.

Reliability Cronbach’s α for the baseline Impact Index scores was 0.85 (95% CI 0.83, 0.85) and for the follow-up scores was 0.91 (95% CI 0.89, 0.93). All four items in the Impact Index were highly correlated with one another (correlations range 0.5–0.8; see Table 3).

Validity

Descriptive information about each scale at baseline for all patients can be seen in Table 4. At baseline, no differences were found between hip and knee patients in correlations between Impact Index and Overall Health ($p = 0.833$), radiographic severity ($p = 0.26$), or CCI ($p = 0.72$) and so overall correlations are presented. Among knee patients, Impact Index was highly correlated with their KOOS Pain, KOOS JR, and KOOS symptom scores; as Impact Index increased indicating greater impact, KOOS scores decreased, indicating greater symptom burden. Among hip patients, Impact Index at baseline was highly correlated with their Harris Hip score; as Impact Index increased indicating greater impact, Harris Hip scores decreased, indicating lower quality of life due to hip symptoms. Impact Index was negatively correlated with overall health, indicating that patients who were more impacted by their symptoms had lower overall health scores. Impact Index was positively correlated with CCI ($r = 0.15$) indicating that those with greater comorbidity

burden were more impacted by their symptoms. Finally, Impact Index was positively correlated ($r = 0.26$) with the severity of radiographic osteoarthritis at baseline, albeit at a much lower level than the other correlations in the table; as Impact Index scores increased, so too did radiographic ratings of severity. Assessment of these relationships (sans CCI and radiographic severity which were only measured once) were assessed at follow-up where these relationships were maintained (or are even stronger) than at baseline (data not shown).

Assessment of predictive validity showed that in the first model Impact Index was predictive of surgical choice ($b = 0.37, p < 0.001, OR = 1.45$) indicating that those with greater Impact Index scores were more likely to choose to undergo surgery (see Table 5). When overall health was added into the model, the model is not improved ($p = 0.93$); thus overall health is not a significant predictor of choice, but Impact Index is still a significant predictor of choice. Finally, in the third step the model is still not improved ($p = 0.23$); thus, CCI and overall health are not significant predictors of choice, while Impact Index is still a significant predictor of choice.

Responsiveness

Responsiveness of the Impact Index measure, the 3 measures of knee symptom burden, the measure of hip symptom burden, and the measure of overall health were examined using data from the 283 patients who completed surveys both after the visit and at follow-up. Table 5 provides the average score after the visit, the mean raw change (follow-up score – baseline score), effect size, and Guyatt’s responsiveness statistic

Table 3 Correlations between overall Impact Index scores with individual Impact Index items and total score at baseline $N = 322$

Variable	Range	Pain	Worry	Bother	Limit
Pain	0 to 3	1			
Worry	0 to 3	0.5 (0.41, 0.57)	1		
Bother	0 to 3	0.8 (0.75, 0.83)	0.6 (0.53, 0.67)	1	
Limit	0 to 3	0.61 (0.54, 0.68)	0.56 (0.48, 0.63)	0.69 (0.63, 0.74)	1
Total score	0 to 12	0.83 (0.79, 0.86)	0.82 (0.78, 0.85)	0.89 (0.80, 0.91)	0.85 (0.81, 0.87)

Table 4 Descriptive statistics for measures and their correlations with Impact Index at baseline

Measure	N	Baseline range	Baseline Mean (SD)	Correlation with Impact Index	p -value
KOOS Pain	210	0 to 100	52.63 (18.9)	-0.67 (-0.74, -0.59)	<0.001
KOOS-JR	210	0 to 100	53.48 (15.27)	-0.62 (-.070, -.53)	<0.001
KOOS symptom	210	3.57 to 100	52.23 (19.30)	-0.49 (-.059, -0.38)	<0.001
Harris Hip	110	7 to 86	53.90 (16.85)	-0.62 (-.073, -.50)	<0.001
Overall Health	315	-0.04 to 1	0.65 (0.20)	-0.49 (-0.57, -0.40)	<0.001
CCI	321	0 to 4	0.44 (0.81)	0.15 (0.04, 0.25)	0.01
Radiographic Severity	207	1 to 5	3.92 (1.43)	0.26 (0.13, 0.38)	<0.001

Table 5 Model statistics for hierarchical logistic regression models predicting surgical choice

Variable	<i>b</i>	SD	<i>p</i>	OR (CI)
Model 1: Impact Index Baseline alone	0.37	0.06	<0.001	1.45 (1.29, 1.63)
Model 2: Impact Index Baseline	0.37	0.07	<0.001	1.45 (1.27, 1.66)
Model 2: Overall health added	−0.08	0.89	0.929	0.92 (0.15, 5.17)
Model 3: Impact Index Baseline	0.36	0.07	<0.001	1.43 (1.25, 1.64)
Model 3: Overall Health	−0.13	0.89	0.881	0.87 (0.14, 4.92)
Model 3: CCI added	0.24	0.2	0.241	1.27 (0.87, 1.92)

for each measure for those patients who underwent surgery. All of the measures presented in Table 6 reflected the improvement patients experienced as a result of surgery (all Guyatt values are greater than 0.80). However, the Impact Index measure consistently had the largest Effect Sizes and Guyatt statistics and the Impact Index measure and Overall Health measure were the only measures for which the confidence intervals for the effect sizes and Guyatt values did not cross zero.

Discussion

The Impact Index is a short measure that demonstrated high reliability and validity in this sample of patients with hip and knee osteoarthritis. Impact Index scores were strongly correlated with other well-established, patient-reported measures of quality of life and disease-specific symptoms. Further, the Impact Index was more responsive than other established measures for patients who underwent surgery. The items not only capture changes in the symptoms, which are the obvious target of the surgery, but also more broadly capture the fact that improving symptoms can also reduce the various adverse effects on patients' lives that those symptoms can have. These results provide new evidence that supports the generalizability of the measure outside of the previous context of BPH.

This measure has some key advantages over existing measures as it is both generic—in that it can measure broadly the impact on health from different disease processes—and disease-specific—in that it can focus in on the impact from a specific condition. Additionally, as the scale specifically indicates a set amount of time in which to consider the impact of the disease (i.e. “Over the past 30 days...”) this can be readily used as a measure of the change of impact over time which is a critical component of PROMs. Given the ease of translating the scale from the BPH context in which it was developed to the current hip and knee surgery context, we believe this scale is well poised to be a more generalizable measure acting as a bridge between generic and—with minimal modification—disease specific measure of disease impact that can be used across a wide variety of diseases to assess not only impact during an isolated point in time, but also changes in impact across time and with treatment.

Although we have evidence that the Impact Index is a valid and reliable measure of the impact of a disease in prior work among men with BPH [4], and in the current work among both men and women with hip and knee osteoarthritis, additional research is needed to identify if the Impact Index is valid and comparable across other conditions. Another yet unanswered question is if the Impact Index can discriminate between the impact of coexisting conditions in patients with multiple comorbidities. As health problems can adversely impact lives through more

Table 6 Responsiveness of the Impact Index for hip and knee replacement surgery

Measure	Baseline range	Baseline mean (SD)	Raw change (SD)	Effect size	Guyatt statistic
Hip Impact Index	5 to 12	10.13 (2.06)	−7.35 (2.97)	−3.57 (−4.03, −3.11)	−2.34 (−4.23, −2.91)
Knee Impact Index	3 to 12	10.40 (1.92)	−6.33 (3.37)	−3.3 (−3.66, −2.94)	−2.19 (−3.94, −2.66)
Knee Pain	0 to 88.89	47.59 (17.41)	34.76 (20.26)	2 (−0.51, 4.5)	1.87 (−0.92, 4.91)
KOOS-JR	0 to 84.6	49.90 (14.48)	26.41 (17.70)	1.82 (−0.26, 3.91)	1.65 (−0.72, 4.37)
Knee Symptom	3.57 to 100	47.30 (19.14)	27.64 (22.83)	1.44 (−1.31, 4.19)	1.46 (−1.84, 4.73)
Harris Hip	7 to 80	52.72 (15.93)	34.56 (15.47)	2.17 (−0.12, 4.46)	2 (−0.05, 4.39)
Overall Health	−0.04 to 0.83	0.62 (0.19)	0.24 (0.20)	1.25 (1.22, 1.28)	1.29 (1.22, 1.28)

For Impact Index larger scores represent worse outcomes, for Knee Pain, KOOS-JR, Knee Symptom, Harris Hip, and EQ-5D larger scores represent better outcomes. As change was calculated as the value at follow-up minus the value after the visit, decreases in Impact Index and increases in all other measures would be expected if patients were improving

than just symptoms, if a patient has multiple conditions that limit activities or cause pain, treating one condition may not actually improve overall quality of life. Therefore, it is important to not only measure changes in general health and quality of life, but also measure the change in impact for these individuals for their multiple diseases. Having a better understanding of how different diseases are impacting a single individual can provide insight into how to tailor treatment decisions to bring about the greatest improvements in quality of life across their diseases. For example, having a single individual respond to the Impact Index regarding both their knee pain and their BPH separately may better indicate where problems are arising or how problems are improving differently over time and condition. Measuring these changes will also be important to identifying how impact scores compare with more general measures of health and quality of life in these circumstances.

The results indicated that hip patients had better results 6 months after surgery than knee patients. Studies have similarly reported that recovery after a hip replacement is faster (and rehabilitation is easier) and that symptom scores are better as compared to knee surgery [22]. Both hip and knee patients who underwent surgery reported clinically meaningful improvements on the disease-specific measures, well in excess of the minimally important clinical difference of about 10–20 points for the KOOS [23]. As of yet the Impact Index does not have a defined minimally important clinical difference in hip and knee replacement patients, suggesting an avenue for future work.

There are a few study limitations that should be considered. First, the sample was highly educated and had very limited racial or ethnic diversity. Additional work would need to examine generalizability of these results to other patient populations within hip and knee osteoarthritis. Second, the follow-up assessment was conducted at 6 months which is fairly early in the recovery process for total joint replacement. Although most of the improvement has been achieved by then, patients still report additional benefits out to 2 years. Whether the relationship and responsiveness between the Impact Index and other established measures would be similar at 24 months is not known. This study was focused on only two conditions and additional work should be done to examine generalizability of these results to other common, symptomatic orthopedic conditions such as low back pain, shoulder (rotator cuff), tennis elbow, or carpal tunnel syndrome. Finally, it is not clear whether the Impact Index would also have strong reliability and validity of the impact index for other health conditions, including health conditions that may be asymptomatic for many patients, such as diabetes or high blood pressure.

Conclusions

The Impact Index has been shown to be a reliable, valid, and responsive measure of how impacted patients with hip or knee osteoarthritis are by their disease. This brief, 4-item, patient-reported measure outperformed commonly used generic and disease-specific scales in terms of responsiveness and can be readily adapted to other decision contexts.

Acknowledgements The study was funded in part through a Patient-Centered Outcomes Research Institute® (PCORI) Award (CDR#1503-28799). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Drs. Barry's and Brodney's time were funded in part through a grant from Healthwise, Incorporated, a nonprofit.

Author contributions Each author has contributed significantly to, and is willing to take public responsibility for, one or more aspects of the study. All authors have been actively involved in the drafting and critical revision of the manuscript, and each has provided final approval of the version submitted. Study design: HV, KS. Data acquisition: HV, KS. Analysis and interpretation: KDV, SB, HV, KS, FF, MJB. Drafting and critical revision of the manuscript: KDV, SB, HV, KS, FF, MJB.

Data availability Data can be accessed by contacting Drs. Barry or Sepucha and completing the necessary Institutional Review Board and data use agreements. Code availability All analytic code is provided in an R markdown document located in the Open Science Framework Repository here: osf.io/hrwp4.

Compliance with ethical standards

Conflict of interest With regards to potential conflicts of interest for this study. Drs. Fowler and Valentine have no conflicts of interest to disclose. Both Drs. Barry and Brodney receive Grant funding through Massachusetts General Hospital from Healthwise, a nonprofit.

Ethical approval Partners Human Research Committee was the Central Institutional Review Board for the study. The DECIDE-OA study is registered on Clinicaltrials.gov (# NCT02729831).

References

1. Black, N. (2013). Patient reported outcome measures may transform healthcare. *BMJ*, *346*(7896), 19–21.
2. Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. *BMJ*, *350*, 1–3.

3. Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care*, *27*, S217–S232.
4. Barry, M. J., et al. (1995). Measuring disease-specific health status in men with benign prostatic hyperplasia. *Medical Care*, *33*, AS145–AS155.
5. Fowler, F. J., Jr., & Barry, M. J. (1993). Quality of life assessment for evaluating benign prostatic hyperplasia treatments. *European Urology*, *24*, 24–27.
6. Fowler, F. J., Collins, M. M. N., Corkery, E. W., Elliott, D. B., & Barry, M. J. (2002). The impact of androgen deprivation on quality of life after radical prostatectomy for prostate carcinoma. *Cancer*, *95*(2), 287–295.
7. Sepucha, K., et al. (2019). Decision Support Strategies for hip and knee osteoarthritis: Less is more. A randomized comparative effectiveness trial (DECIDE-OA study). *Journal of Bone and Joint Surgery of America*, *101*(18), 1645–1653.
8. Mangla, M., et al. (2019). Protocol for a randomised trial evaluating the comparative effectiveness of strategies to promote shared decision making for hip and knee osteoarthritis (DECIDE-OA study). *British Medical Journal Open*, *9*(2), 1–12.
9. Rabin, R., & Charro, F. D. (2001). EQ-SD: A measure of health status from the EuroQoL Group. *Annals of Medicine*, *33*(5), 337–343.
10. Jansson, K. Å., & Granath, F. (2011). Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthopaedica*, *82*(1), 82–89.
11. Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, *40*(5), 373–383.
12. Roos, E. M., & Lohmander, L. S. (2003). The Knee injury and Osteoarthritis Outcome Score (KOOS): From joint injury to osteoarthritis. *Health and Quality of Life Outcomes*, *1*(64), 1–8.
13. Collins, N. J., & Roos, E. M. (2012). Patient-reported outcomes for total hip and knee arthroplasty: Commonly used instruments and attributes of a ‘good’ measure. *Clinics in Geriatric Medicine*, *28*(3), 367–394.
14. Lyman, S., Lee, Y. Y., Franklin, P. D., Li, W., Cross, M. B., & Padgett, D. E. (2016). Validation of the KOOS, JR: A Short-form Knee Arthroplasty Outcomes Survey. *Clinical Orthopaedics and Related Research*, *474*(6), 1461–1471.
15. KOOS, Jr Scoring Instructions. (2017). <https://www.hss.edu/files/KOOS-JR-Scoring-Instructions-2017.pdf>. Accessed 10 Dec 2020.
16. Nilsson, A., & Bremander, A. (2011). Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care and Research*, *63*(Suppl. 11), 200–207.
17. Söderman, P., & Malchau, H. (2001). Is the Harris Hip Score system useful to study the outcome of total hip replacement? *Clinical Orthopaedics and Related Research*, *384*, 189–197.
18. Bedson, J., & Croft, P. R. (2008). The discordance between clinical and radiographic knee osteoarthritis: A systematic search and summary of the literature. *BMC Musculoskeletal Disorders*, *9*, 1–11.
19. Bastick, A. N., et al. (2016). Defining knee pain trajectories in early symptomatic knee osteoarthritis in primary care: 5-Year results from a nationwide prospective cohort study (CHECK). *British Journal of General Practice*, *66*(642), e32–e39.
20. Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology*, *53*(5), 459–468.
21. De Morton, N. A., Davidson, M., & Keating, J. L. (2011). Reliability of the de Morton mobility index (DEMMI) in an older acute medical population. *Physiotherapy Research International*, *16*(3), 159–169.
22. Goodman, S. M., et al. (2020). Validation of the hip disability and osteoarthritis outcome score and knee injury and osteoarthritis outcome score pain and function subscales for use in total hip replacement and total knee replacement clinical trials. *Journal of Arthroplasty*, *35*(5), 1200–1207.e4.
23. Lyman, S., Lee, Y. Y., McLawhorn, A. S., Islam, W., & MacLean, C. H. (2018). What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clinical Orthopaedics and Related Research*, *476*(12), 2432–2441.

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