ARTICLE IN PRESS

VALUE IN HEALTH ∎ (2016) ∎∎∎–∎∎∎

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Using Patient-Reported Outcomes for Economic Evaluation: Getting the Timing Right

Chris Schilling, MSc[1,2,*], *Michelle M. Dowsey, PhD*[2,3], *Philip M. Clarke, PhD*[1], *Peter F. Choong, MBBS, MD, FRACS*[2,3]

[1]*Centre for Health Policy, School of Population and Global Health, the University of Melbourne, Carlton, Victoria, Australia*
[2]*The University of Melbourne Department of Surgery, St. Vincent's Hospital, Fitzroy, Victoria, Australia;* [3]*Department of Orthopaedics, St. Vincent's Hospital Melbourne, Melbourne, Victoria, Australia*

ABSTRACT

Background: Patient-reported outcome measures (PROMs) are becoming increasingly popular in orthopedic surgery. Preoperative and postoperative follow-up often elicit PROMs in the form of generic quality-of-life instruments (e.g., Short Form health survey SF-12 [SF-12]) that can be used in economic evaluation to estimate quality-adjusted life-years (QALYs). However, the timing of postoperative measurement is still under debate. **Objectives:** To explore the timing of postoperative PROMs collection and the implications for bias in QALY estimation for economic evaluation. **Methods:** We compared the accuracy of QALY estimation on the basis of utilities derived from the SF-12 at one of 6 weeks, 3 months, 6 months, and 12 months after total knee arthroplasty, under different methods of interpolation between points. Five years of follow-up data were extracted from the St. Vincent's Melbourne Arthroplasty Outcomes (SMART) registry (n = 484). The SMART registry collects follow-up PROMs annually and obtained more frequent outcomes on subset of patients (n = 133). **Results:** Postoperative PROM collection at 6 weeks, 6 months, or 12 months biased the estimation of QALY gain from total knee arthroplasty by −41% (95% confidence interval [CI] −59% to −22%), 18% (95% CI 4%–32%), and −8% (95% CI −18% to −2%), respectively. This bias was minimized by collecting PROMs at 3 months postoperatively (6% error; 95% CI −9% to 21%). **Conclusions:** The timing of PROM collection and the interpolation assumptions between measurements can bias economic evaluation. In the case of total knee arthroplasty, we recommend a postoperative measurement at 3 months with linear interpolation between preoperative and postoperative measures. The design of economic evaluations should consider timing and interpolation issues.

Keywords: PROMs, economic evaluation, timing, total knee arthroplasty.

Introduction

Patient-reported outcome measures (PROMs) are advocated by a growing range of authors to help measure the value and quality of orthopedic surgery [1–4]. Rolfson and Malchau [4] suggest that "the debate is not primarily why or if we should measure PROMs, but rather how, when and what to measure, and how to interpret the results." Although the literature is increasingly investigating the *what* [5,6] and *how* [7,8], there is less discussion of *when* to efficiently measure PROMs. This is perhaps because clinical outcomes are generally evaluated at a point in time after the patient outcome has stabilized, with the arthroplasty deemed a success if the change in outcome is greater than some minimum important difference [4,9]. However, PROMs can also be used for economic evaluation [10]. Generic instruments such as the EuroQol five-dimensional questionnaire and Short Form health survey (SF-12) measure health-related quality-of-life (HRQOL) utility scores, which are used to estimate quality-adjusted life-years (QALYs) for economic evaluation. QALYs weight any improvement in HRQOL by duration, and thus measure outcomes *over time* rather than at a single point in time. *When* PROMs are observed is therefore particularly important for economic evaluation.

To calculate QALY gains from arthroplasty, health economists will use the HRQOL scores recorded at discrete times to estimate a continuous HRQOL curve. Assumptions are required about how the patient's HRQOL varies between follow-up measurements and into the future after the final follow-up measurement [11,12]. Such assumptions could bias cost-utility analysis. The methodological literature notes that when considering uncertainty, researchers need to consider not only sampling error but also the error associated with interpolation assumptions [11,12]. In practice, HRQOL curves can be based on as little as two measurements. Jenkins et al. [13], for example, use one preoperative and one postoperative measurement to evaluate the cost-utility of total hip and knee replacement. As Devlin and Appleby [10] note, "inferring the benefit of treatment from just two observations of

Conflict of interest: Each author certifies that he or she has no commercial associations that might pose a conflict of interest in connection with the submitted article.

* *Address correspondence to*: C. Schilling, Centre for Health Policy, the University of Melbourne, 207 Bouverie Street, Carlton 3051, Australia.

E-mail: chris.schilling@unimelb.edu.au.

1098-3015$36.00 – see front matter Copyright © 2016, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

http://dx.doi.org/10.1016/j.jval.2016.05.014

PROMs makes the timing of the second observation crucial. For example, collecting PROMs data six months after hip surgery might miss the time when patients first get back to their usual activities, as well as giving no real indication of the longer-term outcomes and durability."

As the use of PROMs for economic evaluation increases, it is paramount that they are collected in a manner that maximizes the information that can be derived from them [14]. This article examined the timing of PROMs data for economic evaluation. Specifically, we aimed to show how the timing of HRQOL measurement can lead to significant bias in QALY estimation, and to highlight how this bias can be reduced by repositioning follow-up measurement. We investigate this issue using the example of total knee arthroplasty (TKA).

Methods

Data Collection

Data for this study were derived from the St. Vincent's Melbourne Arthroplasty Outcomes (SMART) registry, which collects clinical and PROMS data in all patients who undergo elective lower limb arthroplasty at the institution. Baseline data are prospectively collected and include patients' demographic characteristics, diagnoses, and self-reported comorbidities. Follow-up captures an extensive range of outcomes, including surgery and prosthesis-related variables. Patients complete a condition-specific questionnaire and general health questionnaire (SF-12) [15] 12 weeks before surgery and annually postoperatively. The SF-12 and its longer form the 36-item short form health survey are recognized tools that are used widely for the economic evaluation of TKA (see, e.g., Fordham et al. [16], Liebs et al. [17], and Losina et al. [18]), and are widely recorded in arthroplasty registries around the world [19]. A systematic review by Jones and Pohar [20] into HRQOL after orthopedic surgery confirmed the use and construct validity of the SF-12 for economic evaluation of TKA. Data entry and questionnaire follow-up are completed by a dedicated registry coordinator. Mortality data are checked against data from the Registrars of Births, Deaths and Marriages via the Australian Orthopaedic Association National Joint Replacement Registry [21]. The SMART registry has been approved by the Human Research Ethics Committee of St. Vincent's Hospital Melbourne (HREC-A 100/14), and informed consent is obtained before entry onto the registry.

Data

Preoperative and annual postoperative SF-12 scores out to 5 years were available in 484 patients who underwent primary elective TKA between January 2006 and December 2007. Within the same time frame, additional SF-12 scores were collected in a subset of patients ($n = 133$) at 6 weeks, 3 months, and 6 months postsurgery. These additional follow-ups provided measured data to more accurately describe the HRQOL path after surgery. This was used to test the accuracy of QALY gains estimated using just one follow-up point at 12 months. They were available because of an otherwise unassociated randomized controlled trial conducted at the site. Selection into the trial was based on allocation to participating surgeons and use of a standard prosthesis type routinely used at the site [22]. The prosthesis offered no improvement in outcomes relative to other prostheses used at the site [23]. We tested for differences in the characteristics of patients in the subset and the full cohort.

QALY Estimation

Results from the SF-12 assessments at each measurement point were converted into utility scores using the established Brazier algorithm [24]. An average utility curve, which measures the mean HRQOL trajectory for the given cohort, is derived by interpolating between the follow-up measurement points, and extrapolating to 5 years. The QALY gains from surgery were then calculated as the "area under the curve" between the estimated average utility curve and the baseline set at the average preoperative utility score. QALY gains after 1 year postsurgery were discounted at 3% per annum [25].

Generalizing Findings

Bootstrapping was performed to account for potential variation between the true population average utility curves and the sample average utility curve calculated from the SMART registry data. This involved calculating 1000 average utility curves by sampling at the individual level, and calculating 95% confidence intervals (CIs) on the bootstrapped results. We used nonparametric bootstrapping specifically to evaluate the statistical significance of the population bias within the subset and the full cohort.

Interpolation Methods

Two typical methods for interpolation *between* measurement points were used on the basis of the available TKA literature [13,16] and wider methodological cost-effectiveness literature [11,12]:

1. Linear: Assume that the trajectory between any two HRQOL measurement points is linear (Fig. 1A).
2. Immediate: Assume that the HRQOL measurement applies constantly to the time period before the measurement point; for example, the HRQOL score measured at 1-year follow-up is assumed to have accrued immediately after surgery (Fig. 1B).

Figure 1 shows how each interpolation method would apply to the same set of HRQOL points. The QALY gain measured by each method is the shaded area under the curve. It varies dramatically depending on the interpolation method.

Extrapolation Methods

In addition to interpolation between measurement points, extrapolation assumptions may be required to estimate HRQOL into the future if the follow-up measurements do not sufficiently cover the intended study time horizon. In arthroplasty, this extrapolation is usually relatively simplistic: the final follow-up measurement will be carried forward into the future (e.g., see Jenkins et al. [13]), sometimes known as the last-observation-carried-forward extrapolation [26]. A combination of approaches can be used within the same evaluation: linear or immediate interpolation between measurement points, and the last-observation-carried-forward extrapolation for outcomes past the follow-up period.

Scenario Analysis

We completed an investigation of the accuracy of the 5-year QALY gain from TKA under different follow-up timing schedules and interpolation assumptions. First, we calculated the "true" QALY gain from TKA using all the HRQOL measurements at 6 weeks, 3 months, 6 months, 1 year, 2 year, 3 year, 4 year, and 5 year with linear interpolation between measurement points. This provided the base case on which to evaluate potential bias. We then tested the accuracy of QALY estimation using a single