The Journal of Arthroplasty xxx (2018) 1–6

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Comparison of 3 Types of Readmission Rates for Measuring Hospital and Surgeon Performance After Primary Total Hip and Knee Arthroplasty

Alex Bottle, PhD [a, *], Mark D. Loeffler, MBBS, FRCS (Tr & Orth) [b], Paul Aylin, MBChB, FFPH [a], Adam M. Ali, BMBCh, MRCS (Eng) [c, d]

[a] Dr Foster Unit, Department of Primary Care and Public Health, Imperial College, London, UK
[b] Colchester General Hospital, Colchester, UK
[c] St Mary's Hospital, London, UK
[d] Harvard University, Cambridge, Massachusetts

ARTICLE INFO

Article history:
Received 12 January 2018
Received in revised form
13 February 2018
Accepted 15 February 2018
Available online xxx

Keywords:
performance measures
readmission
total hip arthroplasty
total knee arthroplasty
administrative data
multilevel modelling

ABSTRACT

Background: All-cause 30-day hospital readmission is in widespread use for monitoring and incentivizing hospital performance for patients undergoing total hip arthroplasty (THA) and total knee arthroplasty (TKA). However, little is known on the extent to which all-cause readmission is influenced by hospital or surgeon performance and whether alternative measures may be more valid.

Methods: This is an observational study using multilevel modeling on English administrative data to determine the interhospital and intersurgeon variation for 3 readmission metrics: all-cause, surgical, and return-to-theater. Power calculations estimated the likelihood of identifying whether the readmission rate for a surgeon or hospital differed from the national average by a factor of 1.25, 1.5, 2, or 3 times, for both average and high-volume providers.

Results: About 259,980 THAs and 311,033 TKAs were analyzed. Variations by both surgeons and hospitals were smaller for the all-cause measure than for the surgical or return-to-theater metrics, although statistical power to detect differences was higher. Statistical power to detect surgeon-level rates of 1.25 or 1.5 times the average was consistently low. However, at the hospital level, the surgical readmission measure showed more variation by hospital while maintaining excellent power to detect differences in rates between hospitals performing the average number of THA or TKA cases per year in England. In practice, more outliers than expected from purely random variation were found for all-cause and surgical readmissions, especially at hospital level.

Conclusion: The 30-day surgical readmission rate should be considered as an adjunctive measure to 30-day all-cause readmission rate when assessing hospital performance.

Since 2012, the Centers for Medicare and Medicaid Services (CMS) have penalized hospitals with high risk-adjusted all-cause 30-day readmission rates for specified conditions as part of the Hospitals Readmissions Reduction Program [1]. It has included elective hip and knee arthroplasty since 2015. Given that the all-cause measure only partially reflects hospital performance [2–4], alternative metrics assessing readmissions resulting from "surgical" causes, and the subset of these requiring return to theater, may more accurately reflect preventability [5,6]. A recent Report to Congress from the US Department of Health and Human Services stated that there is "clearly a need for more research in this area" [7].

With such increasing focus on performance and with rates of hip and knee arthroplasty set to increase dramatically over the coming

One or more of the authors of this paper have disclosed potential or pertinent conflicts of interest, which may include receipt of payment, either direct or indirect, institutional support, or association with an entity in the biomedical field which may be perceived to have potential conflict of interest with this work. For full disclosure statements refer to https://doi.org/10.1016/j.arth.2018.02.064.

Ethical approval: The authors have approval from the Secretary of State and the Health Research Authority under Regulation 5 of the Health Service (Control of Patient Information) Regulations 2002 to hold confidential data and analyze them for research purposes (CAG ref 15/CAG/0005). They have approval to use them for research and measuring quality of delivery of healthcare, from the London–South East Ethics Committee (REC ref 15/LO/0824).

* Reprint requests: Alex Bottle, PhD, Dr Foster Unit, Department of Primary Care and Public Health, Imperial College, 3 Dorset Rise, London EC4Y 8EN, UK.

https://doi.org/10.1016/j.arth.2018.02.064
0883-5403/

decades [8], it is important to establish how well readmission measures perform on key statistical measures. First, ideal measures of surgical quality have good ability (statistical power) to detect differences in performance between hospitals and between surgeons, but this is unknown for readmission-type measures. Second, an ideal readmission measure is one which is strongly influenced by the hospital or surgeon rather than patient factors (ie, it has a large surgeon or hospital "footprint"). Third, ideal measures have a high power to detect small but important deviations from the average or other benchmark rate, even for a low-volume provider. Comparisons of alternative readmission measures with the all-cause version regarding these 3 key characteristics would inform policy.

Multilevel models allow us to see how much variation is attributed to each level, that is, patient, surgeon, and hospital [6]. Using national data, we compared the "footprint" for surgeons and hospitals in England for readmission after primary hip and knee arthroplasty, with power calculations run for typical caseloads and levels of relative performance. We assessed the statistical performance of 3 types of 30-day readmission indicator: all-cause readmission, surgical readmission, and readmission resulting in return to theater.

Materials and Methods

England's national administrative hospital database, Hospital Episode Statistics [9] covers National Health Service (NHS) (public) hospitals in England, including private patients treated in them and NHS patients treated in private hospitals [8]. Up to 20 diagnosis fields are coded using International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) [10], and 24 procedure fields coded using England's own Office of Population Censuses and Surveys procedure code system. For discharges between April 2010 and March 2015, we extracted all elective (planned) total hip and total knee arthroplasty (THKA) procedures at NHS hospitals with the Office of Population Censuses and Surveys codes related to primary THKA, including primary resurfacing arthroplasty if accompanied by an auxiliary code for the relevant joint (Appendix). A set of comorbidities were defined from our previous work using secondary diagnosis codes during the index admission and any admissions in the previous year [4].

Neighborhood socioeconomic status was taken from the small geographical area Carstairs deprivation score, converted into national population-weighted fifths; patients with unknown deprivation were excluded (0.1% of the total). The Carstairs score is derived from national census data on factors such as unemployment and household overcrowding. Each record has a field with the General Medical Council code for the consultant (senior surgeon) with responsibility for the patient; although this is not necessarily the person who performed the operation, we will refer to this as "surgeon" throughout as shorthand for surgical team.

Patients who died during the index admission were excluded. Three types of readmission indicator were defined within 30 days of index discharge: all-cause readmissions; surgical readmissions (where the primary ICD-10 diagnostic code related to the surgical site as previously defined [5], for example, "mechanical complication of internal joint prosthesis," "infection and inflammatory reaction due to internal joint prosthesis," "disruption of operation wound" and so forth); return-to-theater readmissions (determined from inspecting the list of postindex procedures actually recorded as previously defined) [6]. Return-to-theatre (RTTs) were a subset of surgical readmissions.

Analysis

Comorbidity, age group, sex, deprivation, procedure subtype, and the number of emergency admissions for any cause in the previous year were included as patient factors [6]. To ensure model convergence, surgeons with fewer than 50 operations recorded during the 5 years were combined into 1 pseudosurgeon. The same sets of models were built for each measure and joint: for models including patient factors, all candidate factors were included as predictors with no removal of nonsignificant ones. One-, 2-, and 3-level models were built with combinations of patient factors, surgeon and hospital, and the intraclass correlation coefficient (ICC) calculated for each model [11]. The ICC estimates the proportion of the total variation in the outcome that is accounted for by each level. For example, in a 2-level model with patient factors and surgeons, a small ICC implies that the influence of surgeon is small compared with the influence of patient factors. Higher ICC values are therefore desirable: the larger the ICC, the bigger the surgeon or hospital "footprint."

Median odds ratios (MORs) were also calculated from the multilevel models [12,13]. The MOR indicates the median value of the odds ratios obtained when comparing the odds of readmission in a patient from a randomly selected cluster with another patient with identical covariates but randomly selected from a different cluster. Cluster in our case refers to either the surgeon or the hospital. Each odds ratio from which we take the median has the cluster with the bigger odds in the numerator and the cluster with the smaller odds in the denominator, so the MOR is always at least 1. It can be thought of as the median increase in the odds of readmission that would arise when a patient moves from a lower-risk cluster to a higher-risk cluster. Unlike the ICC, which compares the variation at one level with the total variation, the MOR is purely a measure of how much variation in the odds of readmission exists between surgeons or hospitals. As with ICC, higher values for the MOR are desirable in an indicator.

In randomized controlled trials, it is hoped that some new treatment performs better than the control treatment by at least a certain specified amount (the minimum clinically important difference), and patients are recruited in sufficient numbers to be able to detect this clinically important difference with high probability (the statistical power, often set at 80% or 90%). Similarly, we can estimate with what probability (statistical power) each type of readmission rate can distinguish between the performance of a given hospital or surgeon and the benchmark rate when in fact that hospital or surgeon has, for example, half the benchmark readmission rate. Usually, however, we are interested in poorer than average performance. We therefore performed two-sided power calculations based on one proportion to estimate the likelihood of identifying whether a surgeon's or hospital's outcome rate differed from the national average by a factor of 1.25, 1.5, 2, or 3, that is, 1.25, 1.5, 2, or 3 times the average. In other words, what proportion of surgeons or hospitals with, for example, rates of 1.25 times the average are identified as having high rates for a given caseload? For surgeons, we considered an annual caseload of 50 and 100 procedures, that is, 250 or 500 procedures over 5 years. For hospitals, we considered 5-year totals of 1200 and 1600.

It is increasingly common to use funnel plots to classify units as statistical outliers—that is, to detect whether their performance as measured here by the readmission rate is unusually high or low. These plot each unit's readmission rate on the Y axis against the number of procedures performed on the X axis, with superimposed lines (control limits, shaped like funnels on their sides) that represent rates that are approximately 2 or 3 standard deviations from the national average rate. We report the number and percentage of surgeons and hospitals who were classified as high or low outliers based on 99.8% control limits. With these control limits, if units' readmission rates showed purely random variation, we would expect 1 in 1000 units to be labeled high outliers and 1 in 1000 units to be labeled low outliers, that is, to have 0.2% outliers in total. In screening terminology, such units are false positives