Evaluation of total hip arthroplasty devices using a total joint replacement registry

Elizabeth W. Paxton1*, Christopher F. Ake1, Maria C.S. Inacio1, Monti Khatod2, Danica Marinac-Dabic3 and Art Sedrakyan3,4

1 Department of Surgical Outcomes and Analysis, Kaiser Permanente, San Diego, CA, USA
2 Department of Orthopedic Surgery, Kaiser Permanente, Baldwin Park, CA, USA
3 Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD, USA
4 Weill Medical College, Cornell University, New York, NY, USA

ABSTRACT

Purpose The purpose of this paper is to describe the infrastructure of the total joint replacement registry of a large integrated healthcare system’s and emphasize challenges associated with orthopedic device classification and evaluation.

Methods Using a large integrated healthcare system innovative infrastructure including electronic health record data, administrative data sources, and registry data collection, we evaluated device choice and outcomes of total hip arthroplasty (THA). Devices were classified into type of bearing surface (alternative versus traditional). Multiple imputation was used to accommodate missing data, and a logistic regression model was applied to assess the impact of patient and surgeon factors on choice of bearing surface. A Cox regression model was used to evaluate risk of aseptic revision while controlling for surgeon, site, and patient characteristics. Adjusted cumulative probability-of-event curves were created, comparing survival of alternative against traditional bearings of devices, with aseptic revision as the outcome of interest.

Results The study sample consisted of 25,377 primary THAs with an average follow-up of 2.7 years. Choice of bearing surface varied by surgeon and patient characteristics. After adjusting for patient, surgeon, and hospital covariates, results showed that the risk of aseptic revision associated with alternative bearings did not differ significantly from traditional bearing surfaces (hazard ratio = 1.33; 95% confidence interval: 0.90, 1.98).

Conclusions Clinically rich data from a registry with linkages to electronic health records and other administrative databases improve identification of exposures, outcomes, and patient subgroups in medical device evaluation. These various data sources facilitate refined adjustment for potential confounders such as hospital, surgeon, and patient factors and ensure comprehensive device performance evaluation within registries. Copyright © 2012 John Wiley & Sons, Ltd.

KEY WORDS—total joint replacement; arthroplasty; registry; selection model; bearing surfaces

Received 6 September 2011; Revised 11 January 2012; Accepted 11 January 2012

INTRODUCTION

Total hip arthroplasty (THA) is a widely successful procedure for reducing pain and stiffness associated with osteoarthritis. Approximately 270,000 THA surgeries are conducted each year in the USA.1 The volume and the cost of this procedure are projected to increase significantly over the next decade as a result of market penetration, increased incidence of osteoarthritis, aging demographics, and introduction of new procedures and technologies.2,3

As the volume and the cost of THA procedures increase, new technology is constantly introduced into the market with minimal evidence of clinical effectiveness. This practice has substantial implications for healthcare costs and the public health. Hence, the development of a comprehensive national surveillance system is critical to ensure timely evaluation of the safety and effectiveness of devices. In this context, total joint replacement registries provide a critical mechanism for identifying and evaluating selection, safety, and effectiveness of THA devices.
Total hip arthroplasty devices with alternative bearing surfaces are good examples of technology introduced into the market with little evidence of safety and effectiveness. Whereas traditional bearing surfaces such as metal on polyethylene have demonstrated 80.9% prosthesis survival at 25-year follow-up, alternative bearing surfaces such as metal on metal and ceramic on ceramic were introduced to reduce wear and improve durability, especially in younger patients. Although these alternative bearing surfaces have demonstrated initial 84.4% survival at 20-year follow-up, many recent concerns have been related to their clinical performance and cost effectiveness, particularly in the elderly. There are also concerns related to metal ion release in metal on metal and to fractures and squeaking in ceramic on ceramic alternative bearing surfaces that indicate the need for comprehensive evaluation of these THA technologies.

To evaluate the attributes of devices such as bearing surface type, device registries must adequately identify the devices and classify them into meaningful categories. In the absence of unique device identifiers for medical devices, total joint replacement registries typically identify devices by using manufacturer catalog and lot numbers. However, these numbers are not standardized, and the devices have no universal identification and classification system. In addition, various factors play important roles in making device choices, including patient and surgeon characteristics not universally available in the electronic data sources. Device factors, patient factors (age, gender, and co-morbid conditions), as well as surgical training and experience, must be assessed to accurately evaluate device performance while controlling for potential confounders.

The purpose of this study was to describe the infrastructure of a large integrated healthcare system registry that facilitates comprehensive data collection, by using the example of a THA device bearing surface study to emphasize the unique challenges of device classification and evaluation.

METHODS

Data sources

Using the Kaiser Permanente Total Joint Replacement Registry (KP TJRR), we identified all primary THAs performed between 04/01/2001 and 03/31/2009. The KP TJRR was designed as a longitudinal cohort study to answer specific clinical questions including (i) incidence of primary and revision arthroplasty in our system, (ii) identification of total joint arthroplasty revision and complication rates, (iii) assessment of small-area variation in clinical practices and implant selection, and (iv) identification of patient, implant, and surgical risk factors associated with revisions and complications.

Data collection procedures of the KP TJRR have been described in previous publications. In brief, three kinds of standardized forms are completed by surgeons and staff, one at each patient encounter (pre-operatively, intra-operatively, and at every post-operative visit). The KP TJRR is integrated into Kaiser Permanente HealthConnect, the institution’s electronic health record, capturing patient baseline characteristics, surgical techniques, implant characteristics, and patient outcomes. In addition to its own standardized documentation, the KP TJRR is linked to other administrative databases within the Kaiser Permanente system, such as mortality, membership status, and claims databases, to provide comprehensive post-market surveillance.

Study population

We included registry cases of primary elective THAs performed in the Kaiser Permanente system between 04/01/2001 to 03/31/2009. We excluded (i) cases performed in the mid-Atlantic region (n = 21) because of the small number of cases as well as (ii) bilateral cases (n = 308), which are not commonly indicated for THA patients and could affect the surgical outcome. We then excluded cases with a recorded bearing surface category other than the six categories of interest: metal on metal, ceramic on ceramic, ceramic on conventional polyethylene, ceramic on highly cross-linked polyethylene, metal on conventional polyethylene, and metal on highly cross-linked polyethylene (n = 118) and those with a missing bearing surface category (n = 622).

Data quality control and validation

The KP TJRR uses quality control queries and data entry restriction to identify out-of-range, inconsistent, and missing values. Outliers are manually reviewed and confirmed via chart review by using electronic medical health records. Discrepancies between administrative databases and the registry are adjudicated via electronic health record review. All revisions and complications are validated by chart review by using strict guidelines.

Total hip arthroplasty device classification

Total hip arthroplasty device information is captured in the registry by using the manufacturer catalog number.
Catalog numbers are captured either electronically using the institution’s electronic health record operative module or manually through centralized data entry based on manufacturer stickers containing catalog and lot numbers. Comprehensive lists of catalog numbers and implant descriptions are assembled at the KP TJRR data coordinating center, which contacts each implant manufacturer annually to request updated catalogs. Once these comprehensive look-up tables are created, implant design characteristics and attributes are then specified using information provided by the manufacturers, consultation of orthopedic surgeons, online research of more detailed implant characteristics, or by contacting companies directly. Catalog numbers obtained from the registry forms are linked to these look-up lists electronically and used to characterize implants and their attributes.

With the use of THA insert material and THA femoral component material determined by implant look-up tables, the bearings of a THA implant can be determined. Alternative bearings are metal on metal or ceramic on ceramic, whereas traditional bearings include metal on ceramic or polyethylene bearings (ceramic on conventional polyethylene, ceramic on highly cross-linked polyethylene, metal on conventional polyethylene, and metal on highly cross-linked polyethylene).

**Constructed and augmented variables**

Surgeon and site volume were calculated using primary and revision THAs (including all bilateral and all bearing surfaces) performed by a specific surgeon or site per year. Surgeon and site annual averages were calculated by summing cases for a given surgeon or site and dividing by total number of years from date of first case to date of last case, creating one time-independent volume measure.

Surgeon fellowship training status was ascertained via a survey of surgeons participating in the registry. All contributing surgeons were contacted and asked whether they had post-residency joint replacement/adult reconstruction fellowship training. This was categorized as a yes or no (Y/N) variable and added to the KP TJRR database.

**Geographically Enriched Member Sociodemographics**

Registry cases initially missing race/ethnicity (r/e) data were provided with values from the Kaiser Permanente Geographically Enriched Member Sociodemographics (GEMS) datamart, when available. Where possible, GEMS populates missing r/e data with imputed values. It standardizes r/e values into six mutually exclusive categories: white (non-Hispanic) only, black only, Hispanic (regardless of race), Asian/Pacific Islander only, Native American/Native Alaskan only, and multiracial.

For missing r/e data, GEMS uses Geographic Information System software licensed from ESRI to geocode a member’s last valid address data to assign the person to a census block group, thereby providing an estimate of a member’s r/e from the known distribution in their assigned group. Surname analysis employs a US Census Bureau list of approximately 150,000 last names containing probabilities that a given surname belongs to each of the six standard categories.

A Bayesian Surname and Geocoding algorithm developed at the Research and Development Corporation11 is used to combine prior information on a member’s r/e based on geocoding with updating from surname data to produce a posterior r/e likelihood distribution. R/e data are then imputed for an individual by assigning a probability to each of the six categories for that person. For such a case, in our regression analyses, we then used these six imputed probabilities as is, that is, we did not assign likeliest category or otherwise round off values.12 Elliott et al.13 tested the accuracy of this algorithm by using 2006 data for 1921 133 enrollees from a large national health plan (Aetna) in several ways, which included comparing individual predicted probabilities with self-reported classification. They found an aggregate correlation of 0.76 between individual predicted r/e and self-reported r/e.

**Body mass index.** The KP TJRR includes a variable for body mass index (BMI) measured within 3 months of the primary procedure. Because of the proportion of missing data for this measure (n = 5254, 20.71%), we created a new BMI variable with an expanded definition. If registry data for BMI were missing, a value outside the 3-month window but closest in time to the procedure from the electronic health record was used. For such electronic health record values used, the time between being recorded and the surgical procedure had a median of 26.7 months, with a minimum of 0.03 months and a maximum of 80.1 months (interquartile range: 14.9, 41.7 months). All our analyses used this expanded-definition BMI variable.

**Statistical analyses**

**Missing data.** The extent of missingness among variables used in our analyses was as follows: American Society of Anesthesiologists (ASA) score, 577 cases (2.3%); age, 4 cases (0.02%);
gender, 4 cases (0.02%); r/e (after inclusion of GEMS data), 3207 cases (12.6%); BMI (after inclusion of values outside the 3-month window), 1677 cases (6.6%); surgeon fellowship status, 64 cases (0.25%); surgical approach, 2490 cases (9.8%); fixation, 1752 cases (6.9%); and operative time (length), 5292 cases (20.9%).

For our analyses, we considered the assumption that data were missing at random or nearly so to be plausible, especially given the inclusion of a number each of patient, surgeon, site, procedure, and device covariates in the imputation model and the consequent density of observed data. The imputation model contained all the variables to be entered as covariates in the Cox regression models as well as a binary indicator of whether a given case had the outcome in question (aseptic revision), and also the Nelson–Aalen estimator of the cumulative baseline hazard. We produced 10 imputed data sets by using the Markov chain Monte Carlo method with a single chain having Jeffreys’ prior and initial values from Expectation Maximization (EM) with 400 burn-ins and 2000 iterations.

We treated binary variables with missing values as continuous in the imputation process and left imputed values unrounded in our analyses. For some analyses, any imputed values outside the [0, 1] interval were rounded up to 0 or down to 1. Any non-binary categorical variable with n categories with missing data was entered into the imputation model as a set of n − 1 binaries.

In the imputation model, we left any continuous variable, such as age, in its continuous form to maximize available information. In our Cox regression models, however, we dichotomized some continuous variables (e.g., femoral head size, surgeon volume, and site volume) if there was clinical interest in comparisons across a particular cut point.

Selection model. To assess the importance of factors that might influence the choice between an alternative bearing surface and a traditional one, we developed a selection model that included age; ASA score of 3, 4, or 5 vs. 1 or 2; gender; BMI 25–35 vs. <25 kg/m²; BMI ≥35 vs. <25 kg/m²; diabetes (Y/N); osteoarthritis (Y/N); black vs. white; Hispanic vs. white; Asian vs. white; other r/e (including Native American) vs. white; surgeon annual volume <30 vs. ≥30 cases/year; surgeon fellowship training status (Y/N); site annual volume <75 vs. ≥75 cases/year.

To adjust simultaneously for surgeon and site clustering, we used the GLIMMIX procedure in SAS (SAS Institute Inc., Cary, NC, USA) to run a logistic regression model in the form of a generalized linear mixed model, with random effects for both surgeon and site. Odds ratios (ORs) and 95% confidence intervals (CI) are provided. We did not adjust for clustering by patient, because we excluded bilateral cases from our analyses.

Survival analyses. Survival analyses were conducted with aseptic revision as the outcome event. Along with cases of death or membership termination, a primary THA that was revised because of infection was treated as a censored case (with censoring date = date of septic revision).

In Cox regression models, in addition to an indicator variable for alternative versus traditional bearing surface, the covariates were age; ASA score of 3, 4, or 5 vs. 1 or 2; gender; BMI 25–35 vs. <25 kg/m²; BMI ≥35 vs. <25 kg/m²; diabetes (Y/N); osteoarthritis (Y/N); black vs. white; Hispanic vs. white; Asian vs. white; other r/e (including Native American), vs. white; surgeon annual volume <30 vs. ≥30 cases/year; surgeon fellowship training status (Y/N); site annual volume <75 vs. ≥75 cases/year; operative time; three indicators for surgical approach: anterolateral surgical approach vs. posterior; direct lateral surgical approach vs. posterior; and other surgical approach vs. posterior; femoral head size <36 mm (Y/N); fixation (hybrid, uncemented, and cemented); and antibiotics in cement (Y/N). The sandwich covariance estimator was employed to accommodate clustering by surgeon, which had a stronger effect on results than site clustering.

Cumulative probability-of-event curves adjusted for the mean of covariates were generated for comparison of survival experience of the two bearing surface subgroups.

Sensitivity analyses. The impacts of dropout due to termination of health plan membership and (separately) due to mortality were studied via a form of sensitivity analysis. In these analyses, the Cox regression models used six bearing surface categories as distinct covariates.

In the best-case scenario for dropout due to membership termination, any such case was instead assumed to have survived until the end of the study period without an outcome event. Similarly, in the best-case...
scenario for mortality dropout, any death within the study period was instead assumed to have survived to the study period end date without an outcome event. In each such scenario, hazard ratio (HR) point estimates and 95%CI were calculated for covariates and compared with those from the non-scenario Cox regression model.

In addition, in a series of “worst-case” scenarios for dropout due to membership termination, randomly selected percentages of such cases were instead assumed to have an outcome event on the day after their recorded dropout date. The percentages used were 2%, 5%, and 10%, with HR estimates and 95%CIs also compared with those from the non-scenario model. The same method, using the same three percentages, was then used for mortality dropout as well.

To attempt to gauge the impact of missing values for the two variables with the most missingness (BMI and r/e) on the results of our survival analysis, we also re-ran the Cox regression models excluding those two variables. All analyses were initially conducted using SAS 9.1 with some subsequent confirmatory analyses performed in SAS 9.2.

RESULTS

Data supplementation

The addition of GEMS data reduced missingness for race and ethnicity values from 36.7% to 12.6%. Utilizing supplementary values as well as 3-month values reduced BMI missingness from 20.7% to 6.6%.

Bearing surface use and selection

Metal on highly cross-linked polyethylene bearing surfaces (an advanced form of traditional bearing) accounted for the largest proportion of the THA devices. THA bearing surface use differed across time (Figure 1). Whereas the use of alternative bearing surfaces such as metal on metal and ceramic on ceramic increased over the study period, the use of metal on conventional polyethylene bearing surfaces decreased.

Alternative bearing surfaces were used more frequently in patients who were younger (OR = 0.30, 95%CI: 0.28, 0.31 per 10-year increase in age), healthier (OR for diabetes = 0.50, 95%CI: 0.41, 0.59; OR for ASA ≥ 3 vs. 1 or 2 = 0.84, 95%CI: 0.75, 0.94), and male (OR for female vs. male = 0.47, 95%CI: 0.43, 0.52) (Tables 1 and 2). In addition, assessment of bearing surface use by individual surgeon suggested wide variation (Figure 2).

Table 1. Patient characteristics by total hip arthroplasty bearing type

<table>
<thead>
<tr>
<th>Variable</th>
<th>Alternative bearing (n = 4160)</th>
<th>Traditional bearing (n = 21217)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, STD)</td>
<td>56.4 (10.3)</td>
<td>67.5 (11.1)</td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>43.1</td>
<td>60.2</td>
</tr>
<tr>
<td>ASA score (median, [IQ])</td>
<td>2 [2, 2]</td>
<td>2 [2, 3]</td>
</tr>
<tr>
<td>Osteoarthritis (%)</td>
<td>86.4</td>
<td>91.4</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>10.1</td>
<td>15.4</td>
</tr>
<tr>
<td>BMI (mean, STD)</td>
<td>30.2 (6.2)</td>
<td>29.0 (5.8)</td>
</tr>
</tbody>
</table>

STD, standard deviation; IQ, interquartile range; BMI, body mass index; ASA, American Society of Anesthesiologists.

Table 2. Binary logistic regression results of patient factors associated with alternative total hip arthroplasty bearing selection

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95%CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (in 10-year increments)</td>
<td>0.30</td>
<td>0.28, 0.31</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>1.61</td>
<td>1.38, 1.88</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes</td>
<td>0.50</td>
<td>0.41, 0.59</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female vs. male</td>
<td>0.47</td>
<td>0.43, 0.52</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ASA ≥3 vs. 1 or 2</td>
<td>0.84</td>
<td>0.76, 0.93</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI: 25–35 kg/m² vs. &lt;25 kg/m²</td>
<td>0.84</td>
<td>0.74, 0.95</td>
<td>0.01</td>
</tr>
<tr>
<td>BMI: ≥35 kg/m² vs. &lt;25 kg/m²</td>
<td>1.18</td>
<td>1.03, 1.34</td>
<td>0.02</td>
</tr>
</tbody>
</table>

BMI, body mass index; OR, odds ratio; CI, confidence interval; ASA, American Society of Anesthesiologists.

Total hip arthroplasty bearing surface survival

We compared the survival of alternative bearing surface devices (ceramic on ceramic or metal on metal) versus all others (traditional). After adjusting for patient, surgeon, and hospital covariates, alternative bearing use did not differ significantly from the traditional with respect to risk of revision (HR = 1.33, 95%CI: 0.90, 1.98). A diagnosis of osteoarthritis was associated with a significantly lower risk than other
diagnoses (HR = 0.62, 95%CI: 0.44, 0.86), as was Hispanic versus white race (HR = 0.41, 95%CI: 0.19, 0.85). A femoral head size of less than 36 mm was associated with a significantly higher risk (HR = 1.85, 95%CI: 1.33, 2.56). An adjusted cumulative probability-of-event curve is shown in Figure 3.

Sensitivity analyses
Results from the sensitivity analyses indicated that parameter estimates and 95%CIs were stable across the best-case and the three worst-case scenarios, with the possible exception of those for race. Instability of estimates for the other versus white comparison was most likely due to the small number of cases and of outcome events in the "other" racial category.

In the Cox regression omitting the variables with greatest original missingness (BMI and r/e), the parameter estimates for the remaining covariates differed to a very minor degree from those in the full model: for alternative bearing surface, HR = 1.34, 95%CI: 0.91, 1.99 in the reduced model versus HR = 1.33, 95%CI: 0.90, 1.98 in the full model.

DISCUSSION
A total joint replacement registry from a large integrated healthcare system with linkage to existing administrative databases provided clinically rich data with which to evaluate THA device selection and performance. The KP TJRR provided critical information on patient demographics, surgical techniques, hospital, and implant characteristics for evaluation of THA device performance. Supplementation of registry data with administrative data reduced missingness of important patient characteristics such as BMI and race.

We found that both surgeon and patient factors were related to selection of THA bearing surface. Patients receiving alternative bearing surface devices were younger, had fewer co-morbidities, and were more likely to be male. Surgeon use of alternative bearing surface also varied; some surgeons used these alternative bearings in all procedures, whereas others never used alternative bearings at all. These findings suggest that patient and surgeon factors must be accounted for in assessment of devices.

After controlling for these known confounders, our findings did not support better performance by alternative bearing surfaces. Other national registries have reported higher risk of revision for alternative bearing surfaces such as metal on metal,16–18 prompting a recent worldwide recall of the Johnson & Johnson Articular Surface Replacement (DePuy Orthopaedics, Warsaw, IN, USA) metal-on-metal device.19 Although our results did not indicate the increased risk of revision for alternative bearing surfaces reported by other national registries, the use of alternative devices in
our healthcare system was less frequent, and our study had only short-term to midterm follow-up, which may explain this discrepancy. Further research is necessary to evaluate long-term outcomes of THA bearing surfaces. International collaborations amongst existing institutional, regional, and national total joint replacement registries, such as the International Consortium of Orthopedic Registries, may further elucidate THA bearing surface performance and improve post-market surveillance of these devices by allowing early detection of adverse events.

Conclusions

Clinically rich data from a registry with linkages to electronic health records and other administrative databases improve identification of exposures, outcomes, and patient subgroups in medical device evaluation. They facilitate more refined adjustment for potentially confounding hospital, surgeon, and patient factors, a critical analytical step to ensure adequate assessment of device performance by using registry data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The authors would like to thank all Kaiser Permanente orthopedic surgeons and the staff of the Department of Surgical Outcomes and Analysis who have contributed to the success of the National Total Joint Replacement Registry. We would also like to thank Alan L. Schepps, MS for his assistance with the data management for the study and creating the figures used.

REFERENCES