The relationship between provider volume and short
term complications after an elective total hip replace-
ment was studied on Belgian hospital discharge
administrative database from 2004. The analysis
included 11 856 patients. Hospitals were classified in
low-volume ($\leq 60$ interventions per year), medium
volume ($61-110$) or high volume ($> 110$). Surgeons
were labelled low-volume ($\leq 6$), medium volume ($7-
20$) or high volume ($> 20$). After adjustment for age,
sex, principal diagnosis and comorbidity, surgeon
volume was much more predictive of short term com-
plications than centre volume. Patients treated by
small volume surgeons (respectively medium volume
surgeons) had a 43% higher odds of complications
than patients operated by high volume surgeons
(respectively 37%). Despite some limitations, Belgian
administrative hospital discharge databases can be
used to assess the volume outcome relationship for
orthopaedic surgery. The study has emphasized the
need to closely monitor individual performance, for
hospitals and surgeons. Providers requiring further
auditing can be effectively identified with funnel plots
used routinely in quality control programs.

Keywords: total hip arthroplasty; provider volume;
short-term complication.

INTRODUCTION

Measuring and understanding the association
between outcome and volume of surgical proce-
dures has been the focus of much research since the
1980s. Many studies have shown that, for specific
diagnoses and procedures, patients admitted to low
volume hospitals or treated by low volume sur-
geons have a higher mortality rate or a higher com-
plications rate than patients admitted to high vol-
ume hospitals or treated by high volume surgeons
(2,6). This association was first demonstrated for
cardiac surgery and for cancer resection (2). For
orthopaedic surgery the literature is more scarce. A
systematic review published in 2005, based on 4
original studies in the United States, showed that
failure rate, dislocation within 90 days and deep
wound infection within 90 days after total hip
replacements (THR) were lower in high volume
hospitals compared to low volume hospitals (the
volume threshold was heterogeneous varying from
25 to 100 THR depending on the primary study)
THR is a frequent procedure, performed approximately 18,000 times in 2008 in Belgium (source: RIZIV/INAMI 2008). Arthroplasty may be performed urgently after a trauma such as a fracture but more frequently it is an elective procedure performed to relieve pain and improve function impaired by osteo-arthritis. Short term complications after THR are of two types: those associated with the surgical procedure itself (postoperative infection, mechanical malfunctions of the prosthesis and dislocation of the hip), or those caused by the immobilization (deep vein thrombosis, pulmonary embolism, pneumonia, muscular atrophy). Long term complications include the revision of joint replacement. The present article focuses on short term complications.

The Belgian Health Care Knowledge Centre (KCE), which provides health decision makers with recommendations based on scientific evidence, explored the feasibility to study the volume-outcome association on administrative databases, for a total of 12 surgical procedures. Most of these procedures were of cardiovascular and oncological nature, but due to the above mentioned evidence elective THR was also studied. The purpose of this study was thus to verify the hypothesis that patients with THR performed in high volume hospitals or by high volume surgeons had lower odds of developing complications on the short term than those operated in low volume hospitals or by low volume surgeons.

MATERIALS AND METHODS

Source data

Since 1991 the registration of the Hospital Clinical Data (HCD) is mandatory for every hospital in Belgium. This means that for each hospitalized patient, administrative information such as date of birth, gender, postal code of domicile, length of hospital stay have to be recorded, along with ICD-9-CM encoding of relevant diagnoses as well as diagnostic and therapeutic procedures performed. After data reception, the federal Ministry of Health runs the 3M APR-DRG-grouper software (All Patient refined Diagnostic Groups), version 15.0 to assign the APR-DRG and the Major Diagnostic Category (MDC) to each stay. Since 1997 a legally instituted ‘Technical Cell’ links the HCD records to the Hospital Billing Data (HBD) of the NHIHD (INAMI/RIVIZ) that include the reimbursement costs of each hospital stay. The HBD in particular, contain an identification of the surgeon for every surgical procedure performed during hospitalization.

The data analyzed in this study were selected in the HCD-HBD 2004-2005 (the most recent data available at the time of study): primary elective THR were selected in 2004, and 2005 was used to identify complications of patients operated late 2004.

Data selection

The target population consisted of patients with an elective primary total hip replacement in 2004. The selection of codes was based on a combination of ICD-9-CM procedure and NHIHD procedure codes, as used in previous studies. Patients were included if the ICD-9-CM procedure code 8151 ‘Total hip replacement’ or the Belgian NHIHD procedure code 289085 ‘Hip arthroplasty with a total prosthesis (acetabulum and femoral head)’ was recorded. Patients with bilateral procedures in APR-DRG 300 ‘Bilateral & multiple major joint procedures of lower extremity’ or with two primary THR during the whole study period were excluded. The other exclusion criteria were: THR for trauma (APR-DRG 301 ‘Major joint & limb reattachment procedure of lower extremity for trauma’), MDC different from MDC 8 musculoskeletal system in order to avoid very specific cases in another MDC and THR for complications as the focus was on elective surgery (principal diagnosis 996x ‘Complications peculiar to certain specified procedures’ or V54x ‘Other orthopaedic aftercare’).

Definition of outcomes

The outcome analyzed was the complications within 90 days after the intervention. This choice of outcome was based on previous studies. The 90 days period of time was originally chosen by Katz et al to maximize the likelihood of causality between the joint replacement and the adverse event as effect of the procedure.

The ICD-9-CM diagnostic and procedure codes used to identify complications were:
- 451.1x ‘Phlebitis and thrombophlebitis of deep vessels of lower extremities’,
- 415.1 ‘Pulmonary embolism and infarction’,
- 996.4 ‘Mechanical complication of internal orthopaedic device, implant, and graft’ (mechanical
complications involving: external (fixation) device utilizing internal screw(s), pin(s) or other methods of fixation grafts of bone, cartilage, muscle, or tendon internal (fixation) device such as nail, plate, rod, etc)

996.66 ‘Infection and inflammatory reaction due to internal joint prosthesis’,

996.77-79 ‘Other complications of internal prosthesis, implant, and graft, due to internal joint prosthesis or other internal orthopaedic or prothetic device, implant, and graft’,

835.x ‘Dislocation of hip’,

79.75 ‘Closed reduction of dislocation of hip’

79.85 ‘Open reduction of dislocation of hip’.

Complications within 90 days include patients having a complication during the index admission and patients readmitted within 90 days with one of the codes listed above. Complications occurring after discharge that did not require a hospitalization are not recorded.

Definition of volume

The hospital or surgeon volume was defined by the annual number of patients selected in the target population defined above (see data selection section). When the same surgeon operated in different hospitals, all interventions were taken into account. Tertiles (dividing the hospitals in three groups of the same size) were used to define low volume centers (less or equal to 60 THR per year), medium volume centres (between 60 and 110) and large volume centres (above 110 per year), low volume surgeons (less or equal to 6 per year), medium volume surgeons (at least 7 and less than or equal to 20) and high volume surgeons (more than 20 per year). These criteria were chosen as to maximize the power of the statistical tests (maximal on balanced sample sizes), but they have the drawback that they do not differentiate very high volume surgeons. Sensitivity analyses were thus performed on the surgeon volume using these additional criteria: (≤ 6, 7-12, 13-25, 26-50, 51-100 and above 100). The second cutoff of 12 interventions per year was defined as the cutoff that identifies low from high volume surgeons in the U.S. study of Losina et al on the impact of the surgeon volume on revisions (11).

Risk adjustment

Since some pre-existing patient factors such as disease severity and co-morbidities are clear determinants of outcome and might be distributed differently among low and high volume providers, it is necessary to account for differences in such factors to make valid comparisons between high and low volume providers (6). The following risk factors were taken into account in the analyses:

- patient’s age (as a continuous variable),
- patient’s gender,
- principal diagnosis of admission (osteoarthritis or other),
- Charlson score (co-morbidity). The Charlson score is a validated score based on patient’s co-morbidities, initially developed to predict 1-year mortality, but which has also been used in other settings (3,13). It is the sum of predefined weights attributed to a list of specific conditions (myocardial infarct, congestive heart failure, peripheral vascular disease, dementia, chronic pulmonary disease, connective tissue disease, ulcer disease, mild liver disease, hemiplegia, moderate or severe renal disease, diabetes, tumours, leukaemia, lymphoma, moderate or severe liver disease and metastatic solid tumour). The score is transformed into a five level scale 0, 1, 2, 3 and 4 respectively for scores 0, 1-2, 3-4, 5-6 and > 6, and is analyzed as a continuous variable, as suggested by D’Hoore (3). The information on co-morbidities is retrieved from the variable ‘secondary diagnosis’ which is encoded in the HCD.

Statistical models

Logistic regression models were used to assess the effect of the volume of the provider on the complications rate at 90 days. Three models were fitted: the first model included only tertile of hospital volume (unadjusted estimate). The second model includes tertiles of hospital volume and surgeon volume. In the third model patient case mix variables were added: gender, age, Charlson score and principal diagnosis (715 osteo-arthritis or other). The generalized estimating equations method was used to adjust for clustering of patients within centres. This is standard methodology in volume-outcome studies (15). The last model (Model 4 sensitivity analysis) included more categories of surgeon volume.

Funnel plots

Funnel plots, a common graphic in meta analysis to identify publication bias (4), have also been used as a tool for a first informal assessment of the volume-outcome relationship (17). These allow assessing visually which units were within the expected variability range.
(within the funnel) and which units were not (outside the funnel). Two ranges of control limits are constructed, such that the chance of exceeding these limits for a «in control» unit is 0.2% (99.8% control limits) or 5% (95% control limits).

All analyses were performed with SAS 9.1.3. (SAS Institute, Cary, North Carolina, USA), funnel plots were generated with the R 2.7.1. software (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

A total of 11,856 patients, operated by 522 surgeons in 115 hospitals, had an elective THA in 2004 and were included in the study cohort.

Patients’ characteristics

Patients were on average 67.6 years old, and 60% were female; 91.7% were operated for osteoarthritis, and 6.6% for other disorders of bone & cartilage (Table I).

Low THR volume hospitals had slightly more male and older patients with more co-morbidities (i.e. a higher Charlson index score). Similarly, surgeons who performed few THA’s had higher risk patients with a Charlson score of three or more, compared to other surgeons. High volume centres and surgeons operated more patients with a principal diagnosis of osteo-arthritis.

Complications within 90 days

The complications rate within 90 days was 3.33% (n = 395, Table II). The different reasons were: deep venous thrombosis (n = 16, 0.13%), pulmonary embolism (n = 45, 0.38%), deep wound infection (n = 21, 0.18%), mechanical complication (n = 232, 1.96%), dislocation of the hip (n = 200, 1.69%) and other complications (n = 56, 0.47%). Patients above 75 years old, patients operated for another diagnosis than osteoarthritis and patients with comorbidities had higher rates of short term
In high volume hospitals, the short term complications rate was 2.9%, compared to 4.1% in low volume hospitals. Similarly, high volume surgeons (> 20 THR per year) had a complications rate of 3.0%, compared to 5.0% for low volume surgeons (≤ 6). For very high volume surgeons (> 100), this rate was 2.2% (Table II, sensitivity analysis). Lowest complications rates are achieved by high volume surgeons in high volume hospitals (Fig. 1).

Unadjusted estimates from logistic regression (model 1, Table III) show that small volume hospitals have 40% higher odds of complications than high volume hospital (OR 1.40 95% CI 0.99, 1.97). The effect of hospital volume (small versus high) was reduced to 1.06 (0.72, 1.54), but estimates of surgeon volume were still numerically high: the odds of complications were 43% higher for small volume surgeons compared to high volume surgeons.

In the sensitivity analyses (Table III), results show a stronger effect of surgeon volume, with estimates above 70% increase in odds of complications for surgeons with 12 or less THR compared per year to very high volume surgeons (above 100).

### Table II. — Complications within 90 days by patients’ characteristics and volume of providers

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>n complications</th>
<th>N patients</th>
<th>% complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td></td>
<td>395</td>
<td>11856</td>
<td>3.3</td>
</tr>
<tr>
<td>Age</td>
<td>&lt; 75 years</td>
<td>240</td>
<td>8262</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>≥ 75 years</td>
<td>155</td>
<td>3594</td>
<td>4.3</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>155</td>
<td>4669</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>240</td>
<td>7187</td>
<td>3.3</td>
</tr>
<tr>
<td>Principal diagnosis</td>
<td>715 Osteo-arthrosis</td>
<td>345</td>
<td>10873</td>
<td>3.2</td>
</tr>
<tr>
<td></td>
<td>733 Other disorders of bone &amp; cartilage</td>
<td>40</td>
<td>778</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>10</td>
<td>205</td>
<td>4.9</td>
</tr>
<tr>
<td>Charlson score (comorbidity)</td>
<td>0</td>
<td>255</td>
<td>9109</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>99</td>
<td>2281</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>30</td>
<td>383</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>45</td>
<td>13.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>38</td>
<td>13.2</td>
</tr>
<tr>
<td>Hospital Volume</td>
<td>Tertiles Low (&lt; = 60)</td>
<td>54</td>
<td>1327</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Tertiles Medium (61-110)</td>
<td>133</td>
<td>3248</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Tertiles High (&gt; 110)</td>
<td>208</td>
<td>7281</td>
<td>2.9</td>
</tr>
<tr>
<td>Surgeon Volume</td>
<td>Tertiles &lt;= 6</td>
<td>30</td>
<td>604</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>7-20</td>
<td>89</td>
<td>1960</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>&gt; 20</td>
<td>276</td>
<td>9292</td>
<td>3.0</td>
</tr>
<tr>
<td>Other categories</td>
<td>(sensitivity analysis)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;= 6</td>
<td>30</td>
<td>604</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>7-12</td>
<td>41</td>
<td>735</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>13-25</td>
<td>74</td>
<td>2018</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>26-50</td>
<td>102</td>
<td>2598</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>51-100</td>
<td>87</td>
<td>3087</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>&gt; 100</td>
<td>61</td>
<td>2814</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Note: OR = Odds Ratio; CI = Confidence Interval.
1.68 for surgeons with 6 or less THR, OR 1.98 for surgeons between 7 and 12). There is however almost no difference in short term complications rates between surgeons operating between 50 and 100 THR and surgeons performing more than 100 operations per year (OR 1.05).

Funnel plots on complications rate within 90 days and THR hospital volume (Fig. 2) show that three hospitals are outliers, based on 99.8% limits of variability, and would deserve further scrutiny.

DISCUSSION

This study explored the association between the volume (both institutional and surgeon) of elective THR and the short term complications rate (within 90 days of the intervention). Results show that, after adjustment for age, sex, principal diagnosis and patient comorbidity, hospital volume is not a predictor of short term complications, but that surgeon volume is. The overall short term complications rate was 3.3% ; it was 5.0% for patients operated by surgeons who perform up to 6 operations per year and 3.0% for surgeons performing at least 20 operations per year. Post hoc sensitivity analyses on the surgeon cutoff revealed that surgeons performing at least 100 interventions per year reached a complications rate of 2.2%.

It is challenging to compare our results with those of previous studies. No author used a composite outcome as we did. Our partial rates after THR are lower than those of Katz et al: pulmonary embolism 0.38% versus 0.93%, deep wound infection 0.18% versus 0.23% and dislocation of the hip 1.69% versus 3.13% (8). This is possibly due to differences in outcome definitions based on ICD-9-CM codes. Katz et al defined complications using more codes than we do, such as V54.0 ‘Aftercare involving removal of fracture plate or internal fixation device’, 996.67 ‘Infection and inflammatory reaction due to other internal orthopaedic device, implant & graft’.

Low volume hospitals and low volume surgeons tend to treat older, higher risk patient than high volume hospitals or surgeons. Katz et al made similar observations on their population-based cohort study of patients with a THR in 1995 (9). An inverse relation between (hospital as well as surgeon) volume and hip dislocation was found by Katz et al while Solomon et al found that only the surgeon volume was a predictor of adverse events (8,16). In the 1999 French study by Or et al, readmission within 30 days was inversely related to the hospital volume (14).

This study also has several limitations. First, it is limited to short term complications and hence does not allow to conclude on the effect of volume on...
long term outcome, such as revisions. Nevertheless, a recent Belgian study on data from sickness funds analyzed revision rates with a 19 years follow-up, and included 54,200 patients who had an elective THR between 1990 and 2008 (1). This study showed that patients operated by small volume surgeons (up to 5 interventions per year) had 53% more chance of undergoing a revision than patients operated by high volume surgeons (> 20 patients/year).

This study inherits the usual pitfalls of administrative data. We cannot exclude that due to poor quality of data registration, THR for fractures would still be present in our cohort, despite the fact that we exclude stays with an APR-DRG 301 “THR for trauma”. Fracture cases represent a very different population with a much higher complications risk. This represents a potential source of bias that may explain some differences between high and low volume surgeons, if the latter treat proportionally more fractures than elective cases. In the same line, the fact that the ICD-9-CM principal diagnosis code 733 ‘Other disorders of bone and cartilage’ was more frequently coded in small hospitals could reveal either a different case-mix or could be due to less specific coding behaviours in low volume hospitals. We could not verify which hypothesis was true. In addition, the accurateness of the composite outcome relies on the quality of the hospital coding for complications, which is known to present flaws in Belgium (5). Hopefully, this is less of a problem with the composite outcome which combines different possible codes and which includes re-hospitalizations. Another limitation is the lack of
laterality indication in the data. We have excluded single-stage bilateral THR and bilateral THR during the study period, but we could not verify if all complications were univocally linked with the side operated. Another important disadvantage of the HCD is that clinically relevant outcome measures such as loss of dependence, loss of mobility or residual pain are not registered. The Charlson score which was used for risk-adjustment is probably less appropriate to predict morbidity and mortality than the Physiological and Operative Severity Score for enUmeration of Mortality and morbidity (POSSUM) which was adapted for orthopaedic patients (12). However, the orthopaedic POSSUM system includes a physiological assessment and an analysis of operative severity; this information is not available in the HCD. No hospital characteristic other than volume operated was integrated in the regression models. Finally, using data from more than 2 years would improve the estimates’ statistical precision.

Our study contributes in several ways to the current body of evidence on volume-outcome relationship in THR. The available data cover all Belgian hospitalizations. The composite outcome retaining the main postoperative complications gives the proportion of patients who did not experience any complication. Outcomes are analyzed at hospital as well as surgeon level, and small volume providers are not discarded. Outcomes are in hospital as well as after discharge. Finally, funnel plots represent a convenient tool to identify outlying centres.

These data show that Belgian administrative hospital discharge databases, despite some limitations, can be used to assess the volume outcome relationship for orthopaedic surgery. They have emphasized the need to closely monitor individual performance, for hospitals and surgeons. Funnel plots have proven effective in identifying centres requiring further auditing, and could be used routinely in quality control programs.

Acknowledgements

Grateful acknowledgment is made to Patrice Chalon, Knowledge Manager at the KCE for the editing and formatting support. The authors also thank the two anonymous referees for their helpful comments.

REFERENCES


