Surgical learning curves and operative efficiency: a cross-specialty observational study

Mahiben Maruthappu,1 Antoine Duclos,1,2 Stuart R Lipsitz,1 Dennis Orgill,1,3 Matthew J Carty1,3

ABSTRACT

Objectives: To evaluate the learning curves of three high-volume procedures, from distinct surgical specialties.

Setting: Tertiary care academic hospital.

Participants: A prospectively collected database comprising all medical records of patients undergoing isolated coronary artery bypass grafting (CABG), total knee replacement (TKR) and bilateral reduction mammoplasty (BRM) at the Brigham and Women’s Hospital, USA, 1996–2010. Multivariate generalised estimating equation (GEE) regression models were used to adjust for patient risk and clustering of procedures by surgeon.

Primary outcome measure: Operative efficiency.

Results: A total of 1052 BRMs, 3254 CABGs and 3325 TKRs performed by 30 surgeons were analysed. Median number of procedures per surgeon was 61 (range 11–502), 290 (52–973) and 99 (10–1871) for BRM, CABG and TKR, respectively. Mean operative times were 134.4 (SD 34.5), 180.9 (62.3) and 101.9 (30.3) minutes, respectively. For each procedure, attending surgeon experience was associated with significant reductions in operative time (p<0.05). After 15 years of experience, the learning curves of CABG and TKR were not statistically different, whereas BRM showed a more pronounced improvement. After 30 years of experience, the learning curves of TKR and BRM were not statistically different, whereas CABG continued to improve. The use of high-granularity data enabled the identification of performance differences between surgeons, which can be leveraged for quality improvement initiatives.

Conclusions: Common trends in surgical learning exist. Dependent on the procedure, experience can serve as a powerful driver of improvement or have clinically insignificant impacts on operative time.

INTRODUCTION

Performance measurement and the application of quality, efficiency and safety metrics in surgery has dramatically expanded over the past two decades. This has met with variable degrees of compliance and success. The operating room (OR) represents a high-risk environment, requiring the coordination of technology, competence and resources under time pressure by the operating team.1,2 When combined with the technical and dexterous nature of surgery, it is apparent that several factors must be considered and accounted for to make accurate, equitable and comparable measurements of performance between surgeons. One of these factors is the learning curve inherent to operative procedures.3–8

Until recently, studies have lacked adequate data volume and resolution to gain an appropriate understanding of procedural dynamics, or to robustly identify suitable performance metrics. Work thus far has predominantly focused on surgery at the institutional or departmental level, rather than that of the individual surgeon. Furthermore, no studies to date have compared learning curves across different surgical specialties using statistically uniform methods.8

We performed a cross-specialty, comparative study evaluating the learning curves of three high-volume and well-defined procedures: bilateral reduction mammoplasty (BRM), isolated coronary artery bypass grafting (CABG) and total knee replacement (TKR). Our intent was to define commonalities and discrepancies in the learning curves of these procedures, and thereby identify overarching themes that may inform future surgical performance improvement efforts.

METHODS

Design and population

Data for all CABG, TKR and BRM procedures performed at Brigham and Women’s Hospital, USA, 1996–2010.
Hospital (BWH), 2001–2010, 1996–2009 and 1995–2007, respectively, were culled from a combination of electronic medical records, an electronic operative time tracking application and physician employee databases. The data sets were filtered to exclude CABGs that were accompanied by concomitant procedures (eg, valve surgery, MAZE, etc); erroneously coded procedures including partial knee replacements, combined procedures and bilateral TKR procedures; gynaecomastia, resections and partial/total mastectomies; and incomplete or incorrect data entries (details of the exclusion process are included in the online supplementary appendix). Data sets were subsequently filtered to only include surgeons who conducted more than 10 procedures annually, focusing on the first 20 years of the learning curve.

For CABG, the primary dependent variable was operative time defined as the sum of cardiopulmonary bypass time (CPB) and cross clamp time. For TKR and BRM, operative time was defined as the time elapsed from skin incision to skin closure. The operative experience of the attending surgeon was calculated as the difference between the date of the procedure and that of the surgeon’s completion of training. This study was performed under Institutional Review Board approval (protocol 2006p000586).

Statistical analysis
The characteristics of patients and surgeons were described using absolute frequencies with percentages for categorical variables. The mean values with SDs and median values with minimum–maximum intervals were calculated for continuous variables. The expected learning curve of surgeons over time was generated based on a multivariate generalised estimating equations (GEE) regression model. A number of possible shapes of learning curves were tested in order to obtain the best fitting. Outcomes were adjusted for clustering of patients by surgeon, as well as patient case-mix factors (for CABG: patient age, sex, preoperative cardiogenic shock, diabetes mellitus and preoperative congestive heart failure; for TKR: patient age, sex and comorbidities including coronary artery disease, chronic obstructive pulmonary disease, diabetes mellitus, hypertension, obesity and smoking; for BRM: volume of breast reduction). Selection of covariates for generation of case-mix models was achieved using the approach described by Collett. Model estimates were obtained using the GENMOD procedures in SAS V.9.2 (SAS Institute Inc, Cary, North Carolina, USA). The expected reductions in operative time associated with attending surgeon experience were plotted. All tests were two-tailed, and p values <0.05 were considered significant.

RESULTS
Surgeon and cohort characteristics for each procedure are listed in tables 1 and 2. Figure 1 displays the learning curve of each of the procedures. The number of procedures per surgeon per year are displayed in the online supplementary appendix.

A total of 1052 BRMs were completed by eight surgeons. Mean attending experience was 11.3±4.5 years, with a median case volume of 61 (range: 11–502). Mean operative time was 134.4±34.5 min. Multivariate regression showed breast reduction volume to be the only case-mix factor to have a significant effect on operative time. After adjusting for patient characteristics and clustering, we found a significant association between operative time and attending surgeon experience entered both as a linear (p=0.0002) and a quadratic (p=0.0427) term in the model. Five, 10 and 15 years of experience were associated with 33.74, 57.02 and 69.82 min reductions in operative time, respectively, equating to 18.51%, 31.28% and 38.30% reductions (table 3).

A total of 3254 CABGs were completed by nine surgeons. Mean attending experience was 9.7±5.9 years, with a median case volume of 290 (range: 52–973). Mean operative time was 180.9±62.3 min. Multivariate regression showed the following case-mix factors to have a significant effect on operative time: patient age, sex, preoperative cardiogenic shock, diabetes mellitus and preoperative congestive heart failure. After adjusting for patient characteristics and clustering, we found a significant association between operative time and attending surgeon experience entered as a logarithmic term (p=0.0374) in the model. Five, 10 and 15 years of experience were associated with 5.85, 11.69 and 17.54 min reductions in operative time, respectively, equating to 2.61%, 5.22% and 7.83% reductions (table 3).

A total of 3325 TKRs were completed by 13 surgeons. Mean attending experience was 14.5±3.8 years, with a median case volume of 99 (range: 10–1871). Mean operative time was 101.9±30.3 min. Multivariate regression showed the following case-mix factors to have a significant effect on operative time: patient age, sex and comorbidities including coronary artery disease, chronic obstructive pulmonary disease, diabetes mellitus, hypertension, obesity and smoking. After adjusting for patient characteristics and clustering, we found a significant association between operative time and attending surgeon experience entered both as a linear (p<0.0001) and a quadratic (p=0.0002) term in the model. Five, 10 and 15 years of experience were associated with 38.20, 69.68 and 94.41 min reductions in operative time, respectively, equating to 19.60%, 35.75% and 48.44% reductions (table 3).

DISCUSSION
We performed a cross-specialty evaluation of learning curves in surgery, investigating over 7000 procedures performed by 30 attending surgeons. By comparing these learning curves using statistically uniform data and techniques, our study permits identification of important commonalities and differentiations, yielding several important findings.
First, in contrast to the notion that efficiency is optimised within a fairly narrow temporal window following the start of clinical practice, our data suggest that operative learning curves, for some procedures, exhibit ongoing improvement in efficiency over the course of a surgeon’s career, with time courses much longer than previously anticipated. This emphasises the necessity to draw equitable comparisons between surgeons at similar stages of the learning curve, and supports proposals for continual monitoring, training and behavioural interventions aimed to accelerate operative maturation.

Second, our results demonstrate the different learning curve dynamics that exist between procedures. BRM is typified by an initial phase of variability, followed by a period of rapid improvement, followed by a relative plateau phase. TKR and CABG, on the other hand, demonstrate a more linear improvement over time. Such findings suggest that certain procedures may demonstrate characteristic learning curves, with some achieving maturation more rapidly than others. The factors contributing to these characteristics should be the subject of further investigation.

Third, the magnitude of improvements in efficiency over time varies from procedure to procedure. In some cases, efficiency is markedly augmented with increasing surgeon experience (BRM and TKR); in others, however, the improvement is substantially more marginal, and possibly clinically insignificant (CABG; <10% reduction after 15 years of experience). This observation may be because cardiothoracic fellows begin their training than cardiothoracic surgeons, precluding this is the possibility that plastic and orthopaedic surgeons may learn a broader range of procedures during their training than cardiothoracic surgeons, precluding as much experience to be gained in a given operative technique—a difference of breadth and depth. This observation suggests that interventions aiming to accelerate surgical experience acquisition, such as simulation training, may be particularly effective for enhancing efficiency in TKR and BRM, but possibly less so in CABG, given the differential impact of individual surgeon experience across these procedures.

In addition to this, procedural learning curves were found to have differential sensitivities to case-mix factors. BRM was only affected by breast reduction volume; in contrast, the TKR learning curve was found to be influenced by a variety of patient-factors including age, sex and comorbidities. This differential is no doubt due, at least in part, to varying tolerances on the part of surgeons in each of the highlighted specialties to operate on patients with marginal health status; this must be taken into consideration when making cross-procedural comparisons.

Importantly, the development of experience-based learning curves, such as those elucidated in this study, has the potential to inform efforts to prospectively monitor individual surgical performance metrics on an ongoing basis. Through the application of iterative quality improvement systems, such as statistical process control (SPC), the depicted learning curves could be utilised as a baseline against which to chart the progress of surgical personnel—whether to meet administrative or educational goals. Such methodologies are designed to first identify and then address unwarranted variation in clinical processes, which generally leads to subsequent increases in experience. Superimposed on this is the possibility that plastic and orthopaedic surgeons may learn a broader range of procedures during their training than cardiothoracic surgeons, precluding as much experience to be gained in a given operative technique—a difference of breadth and depth. This observation suggests that interventions aiming to accelerate surgical experience acquisition, such as simulation training, may be particularly effective for enhancing efficiency in TKR and BRM, but possibly less so in CABG, given the differential impact of individual surgeon experience across these procedures.

In addition to this, procedural learning curves were found to have differential sensitivities to case-mix factors. BRM was only affected by breast reduction volume; in contrast, the TKR learning curve was found to be influenced by a variety of patient-factors including age, sex and comorbidities. This differential is no doubt due, at least in part, to varying tolerances on the part of surgeons in each of the highlighted specialties to operate on patients with marginal health status; this must be taken into consideration when making cross-procedural comparisons.

Importantly, the development of experience-based learning curves, such as those elucidated in this study, has the potential to inform efforts to prospectively monitor individual surgical performance metrics on an ongoing basis. Through the application of iterative quality improvement systems, such as statistical process control (SPC), the depicted learning curves could be utilised as a baseline against which to chart the progress of surgical personnel—whether to meet administrative or educational goals. Such methodologies are designed to first identify and then address unwarranted variation in clinical processes, which generally leads to subsequent increases in experience. Superimposed on this is the possibility that plastic and orthopaedic surgeons may learn a broader range of procedures during their training than cardiothoracic surgeons, precluding as much experience to be gained in a given operative technique—a difference of breadth and depth. This observation suggests that interventions aiming to accelerate surgical experience acquisition, such as simulation training, may be particularly effective for enhancing efficiency in TKR and BRM, but possibly less so in CABG, given the differential impact of individual surgeon experience across these procedures.

First, in contrast to the notion that efficiency is optimised within a fairly narrow temporal window following the start of clinical practice, our data suggest that operative learning curves, for some procedures, exhibit ongoing improvement in efficiency over the course of a surgeon’s career, with time courses much longer than previously anticipated. This emphasises the necessity to draw equitable comparisons between surgeons at similar stages of the learning curve, and supports proposals for continual monitoring, training and behavioural interventions aimed to accelerate operative maturation.

Second, our results demonstrate the different learning curve dynamics that exist between procedures. BRM is typified by an initial phase of variability, followed by a period of rapid improvement, followed by a relative plateau phase. TKR and CABG, on the other hand, demonstrate a more linear improvement over time. Such findings suggest that certain procedures may demonstrate characteristic learning curves, with some achieving maturation more rapidly than others. The factors contributing to these characteristics should be the subject of further investigation.

Third, the magnitude of improvements in efficiency over time varies from procedure to procedure. In some cases, efficiency is markedly augmented with increasing surgeon experience (BRM and TKR); in others, however, the improvement is substantially more marginal, and possibly clinically insignificant (CABG; <10% reduction after 15 years of experience). This observation may be because cardiothoracic fellows begin their training than cardiothoracic surgeons, precluding as much experience to be gained in a given operative technique—a difference of breadth and depth. This observation suggests that interventions aiming to accelerate surgical experience acquisition, such as simulation training, may be particularly effective for enhancing efficiency in TKR and BRM, but possibly less so in CABG, given the differential impact of individual surgeon experience across these procedures.

In addition to this, procedural learning curves were found to have differential sensitivities to case-mix factors. BRM was only affected by breast reduction volume; in contrast, the TKR learning curve was found to be influenced by a variety of patient-factors including age, sex and comorbidities. This differential is no doubt due, at least in part, to varying tolerances on the part of surgeons in each of the highlighted specialties to operate on patients with marginal health status; this must be taken into consideration when making cross-procedural comparisons.

Importantly, the development of experience-based learning curves, such as those elucidated in this study, has the potential to inform efforts to prospectively monitor individual surgical performance metrics on an ongoing basis. Through the application of iterative quality improvement systems, such as statistical process control (SPC), the depicted learning curves could be utilised as a baseline against which to chart the progress of surgical personnel—whether to meet administrative or educational goals. Such methodologies are designed to first identify and then address unwarranted variation in clinical processes, which generally leads to subsequent increases in experience. Superimposed on this is the possibility that plastic and orthopaedic surgeons may learn a broader range of procedures during their training than cardiothoracic surgeons, precluding as much experience to be gained in a given operative technique—a difference of breadth and depth. This observation suggests that interventions aiming to accelerate surgical experience acquisition, such as simulation training, may be particularly effective for enhancing efficiency in TKR and BRM, but possibly less so in CABG, given the differential impact of individual surgeon experience across these procedures.
substantial improvements in efficiency, safety and quality. While initially conceived for the manufacturing industry, they have recently been recognised as having increasing relevance to the healthcare sphere.\textsuperscript{19 20} Compared to more traditional approaches of quality control, they offer a more accurate means of assessing performance and guiding improvement initiatives, particularly when they are, as in the case of the data presented here, tailored to a specific operative procedure.\textsuperscript{11 12}

Implementation of SPC offers several potential benefits. Through improvement of surgical performance, these efforts have the potential to simultaneously improve surgical efficiency, safety, quality and cost at the individual as well as institutional levels. In the context of training, mapping of performance throughout a surgeon’s career will permit the evolution of young trainees to be monitored, giving rise to appraisals based on performance rather than career chronology alone, potentially ensuring progression only on acquisition of sufficient expertise.\textsuperscript{31 11 22} Deconstruction of the performance curve and comparison of curves between high and low performers will permit the elements contributing to training to be dissected. By doing so, the mechanics of surgical learning will be better understood, providing methods by which the learning curve can be accelerated. SPC methodology also presents a means of accurately assessing the efficacy of new training tools or simulation programmes at the individual level, both rapidly and sensitively. Finally, this methodology may have the capacity to identify deteriorations occurring towards the end of a surgeon’s career, serving as a safety monitor and supplementing the introduction of continuing education programmes.\textsuperscript{4 22}

These implications, however, must be considered in the context of this study’s limitations:

1. Our focus on operative efficiency as an outcome for CABG, TKR and BRM did not explicitly incorporate considerations related to surgical safety such as complication rates, which may be considered more patient

---

Table 2

<table>
<thead>
<tr>
<th>Surgical procedure</th>
<th>Factors included in patient case-mix adjustment</th>
<th>Time period data available</th>
<th>Surgeons (N)</th>
<th>Surgeon-years (N)</th>
<th>Procedures (N)</th>
<th>Median number of procedures per surgeon (min–max)</th>
<th>Mean operative time (SD)</th>
<th>Model</th>
<th>BRM, bilateral reduction mammoplasty; CABG, coronary artery bypass grafting; GEE, generalised estimating equation; TKR, total knee replacement.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CABG</td>
<td>Age, sex, preoperative congestive heart failure, chronic obstructive pulmonary disease, diabetes mellitus, hypertension, obesity, smoking</td>
<td>9 years (1/1/1995–31/12/2007)</td>
<td>9</td>
<td>9</td>
<td>3254</td>
<td>290 (62–973)</td>
<td>180.9 (62.3)</td>
<td>GEE (clustering and case-mix adjusted; logarithmic model used)</td>
<td></td>
</tr>
<tr>
<td>BRM</td>
<td>Age, sex, coronary artery disease, chronic obstructive pulmonary disease, diabetes mellitus</td>
<td>13 years (1/5/2002–30/9/2010)</td>
<td>8</td>
<td>48</td>
<td>3254</td>
<td>1052 (11–502)</td>
<td>134.4 (34.5)</td>
<td>GEE (clustering and case-mixture adjusted; linear and quadratic model used)</td>
<td></td>
</tr>
<tr>
<td>TKR</td>
<td>Age, sex, coronary artery disease, chronic obstructive pulmonary disease, diabetes mellitus, hypertension, obesity, smoking</td>
<td>14 years (28/7/1999–29/6/2009)</td>
<td>13</td>
<td>123</td>
<td>3325</td>
<td>98 (10–1871)</td>
<td>101.9 (30.2)</td>
<td>GEE (clustering and case-mix adjusted; linear and quadratic model used)</td>
<td></td>
</tr>
</tbody>
</table>

---

Figure 1

Learning curves showing bilateral reduction mammoplasty (BRM), coronary artery bypass grafting (CABG) and total knee replacement (TKR) operative times versus surgeon experience.
3. Our investigation utilised years of training and/or practice has been utilised as an acceptable substitute for surgical experience in prior published studies in the surgical literature.3 4 8 11 12 30

4. Additional controls could have been incorporated into the model, such as team familiarity, team experience, fellowship-status and the presence of a resident. Although cross-specialty analyses of such work have yet to be conducted, evidence exists at the procedure-specific level.16 30

This study quantitatively characterised the learning curves of three procedures from distinct surgical specialties. We identified common trends in surgical learning and demonstrated that, dependent on the procedure, surgical experience can serve as a powerful driver of improvement or have clinically insignificant impacts on efficiency. Appreciation of these findings may guide implementation of performance-tracking and quality-improvement strategies in surgery.

Table 3 Learning curve—reductions in operative time with surgical experience

<table>
<thead>
<tr>
<th>Experience</th>
<th>BRM</th>
<th>Total operative time (min)</th>
<th>Operative time reduction (min)</th>
<th>Operative time reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>193.3097</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>-7.5867</td>
<td>185.723</td>
<td>3.9</td>
<td>-1.1694099</td>
</tr>
<tr>
<td>2</td>
<td>-14.7546</td>
<td>178.5551</td>
<td>7.6</td>
<td>-2.3388198</td>
</tr>
<tr>
<td>3</td>
<td>-21.5037</td>
<td>171.8061</td>
<td>11.1</td>
<td>-3.5082297</td>
</tr>
<tr>
<td>4</td>
<td>-27.834</td>
<td>165.4757</td>
<td>14.4</td>
<td>-4.6776396</td>
</tr>
<tr>
<td>5</td>
<td>-33.7455</td>
<td>159.5642</td>
<td>17.5</td>
<td>-5.8470495</td>
</tr>
<tr>
<td>6</td>
<td>-39.2382</td>
<td>154.0715</td>
<td>20.3</td>
<td>-7.0164594</td>
</tr>
<tr>
<td>7</td>
<td>-44.3121</td>
<td>148.9976</td>
<td>22.9</td>
<td>-8.1858693</td>
</tr>
<tr>
<td>8</td>
<td>-48.9672</td>
<td>144.3425</td>
<td>25.3</td>
<td>-9.3552792</td>
</tr>
<tr>
<td>9</td>
<td>-53.2035</td>
<td>140.1062</td>
<td>27.5</td>
<td>-10.524689</td>
</tr>
<tr>
<td>10</td>
<td>-57.021</td>
<td>136.2887</td>
<td>29.5</td>
<td>-11.694099</td>
</tr>
<tr>
<td>11</td>
<td>-60.4197</td>
<td>132.89</td>
<td>31.3</td>
<td>-12.863509</td>
</tr>
<tr>
<td>12</td>
<td>-63.3996</td>
<td>129.9101</td>
<td>32.8</td>
<td>-14.032919</td>
</tr>
<tr>
<td>13</td>
<td>-65.9607</td>
<td>127.349</td>
<td>34.1</td>
<td>-15.203239</td>
</tr>
<tr>
<td>14</td>
<td>-68.103</td>
<td>125.2067</td>
<td>35.2</td>
<td>-16.371739</td>
</tr>
<tr>
<td>15</td>
<td>-69.8265</td>
<td>123.4832</td>
<td>36.1</td>
<td>-17.541148</td>
</tr>
<tr>
<td>16</td>
<td>-71.1312</td>
<td>122.1785</td>
<td>36.8</td>
<td>-18.710558</td>
</tr>
<tr>
<td>17</td>
<td>-72.0171</td>
<td>121.2926</td>
<td>37.3</td>
<td>-19.879968</td>
</tr>
<tr>
<td>18</td>
<td>-72.4842</td>
<td>120.8255</td>
<td>37.5</td>
<td>-20.19903</td>
</tr>
<tr>
<td>19</td>
<td>-72.5325</td>
<td>120.7772</td>
<td>37.5</td>
<td>-22.18788</td>
</tr>
</tbody>
</table>

BRM, bilateral reduction mammoplasty; CABG, coronary artery bypass grafting; TKR, total knee replacement.
REFERENCES