Stratum-Specific Likelihood Ratio Analysis: An Evidence-Based and Pragmatic Approach to Meaningful Thresholds in Lower Extremity Arthroplasty

Sergio M Navarro, BS  
Medical Student  
Department of Orthopaedic Surgery  
Baylor College of Medicine  
Houston, Texas  
Said Business School  
University of Oxford  
Oxford, United Kingdom

Michael A. Mont, MD  
Director of Joint Arthroplasty  
Department of Orthopaedic Surgery  
Lenox Hill Hospital  
Northwell Health  
New York, New York

Viktor Krebs, MD  
Orthopaedic Surgeon  
Department of Orthopedic Surgery  
Cleveland Clinic  
Cleveland, Ohio

Heather S Haederle, BS  
Medical Student  
Department of Orthopaedic Surgery  
Baylor College of Medicine  
Houston, Texas

Prem N. Ramkumar, MD, MBA  
Resident Physician, PGY-3  
Department of Orthopedic Surgery  
Cleveland Clinic  
Cleveland, Ohio

ABSTRACT

Background: With the transition toward a value-based care delivery model, an evidence-based approach to quantify the effect of procedural volume on outcomes and cost presents an opportunity to understand and optimize the delivery of lower extremity arthroplasty. Stratum-specific likelihood ratio (SSLR) analysis has been recently applied to define benchmarks which confer a significant advantage in value at the hospital or surgeon level.

Materials and Methods: In this report, the role, statistical technique, and future applications of SSLR analysis are described with an example outlined for total hip arthroplasty (THA).

Results: SSLR analysis provides multiple significant value-based thresholds, providing an advantage over previous methods used to describe the effects of surgeon and hospital volume. These benchmarks have been developed for THA, total knee arthroplasty (TKA), hip fracture, and several other orthopaedic procedures.
Recent value-based changes and policy pressures in the delivery of lower extremity arthroplasty have led to increased emphasis on volume-related effects. Focus has increasingly shifted from providing volume-based care to value-based care that improves outcomes while simultaneously limiting costs.1–3 The recent initiative of the Comprehensive Care for Joint Replacement (CJR) model seeks to support better and more efficient care for patients undergoing hip and knee arthroplasties via a strategic bundled payment model.4 Further, private payers, including United HealthCare and other large employers, have recently announced plans to shift compensation to bundled payments, a burgeoning value-based care model for hip and knee arthroplasty.5,6

As the search for outcome optimization in healthcare systems continues, studies have demonstrated that high-volume surgeons and hospitals achieve better outcomes for patients at a lower cost of care.7–11 The notion of “practice makes perfect” is intuitive in understanding why higher-volume surgeons may produce better outcomes at lower costs. However, the exact number of procedures required before a surgeon or hospital system is truly proficient is ill-defined and often arbitrary. With increasing evidence mounting in favor of higher-volume hospitals and surgeons translating to higher quality care, both patients and payers seek them out for value-based care.12–15 Value is defined by both parties, in many cases in volume-based models, and is defined as the ratio of the outcome over the specific cost of the intervention.1

To date, no consensus exists as to what the volume–value curve looks like, or what threshold differentiates surgeons and hospitals as “high” or “low” volume. Classically, the stratification of hospital and surgeon volume from continuous variables into discrete categories (e.g., low, medium, high volume) has been utilized for ease of data interpretation and the implementation of interventions. Thus, volume-outcome studies have often relied on establishing arbitrary cutoffs or splitting patients into quantiles for analysis without considering surgeon experience or systemic infrastructure, thus producing studies limited by inconsistency in established benchmarks and thresholds.9 Without evidence-based volume thresholds, standardized policy metrics of quality, patient safety, and reimbursement cannot be generalized across institutions to improve operational efficiency and value.9,16 Several hospitals have already established volume standards for certain procedures requiring surgeons to reach a specific level of repetitions before performing particular cases, and hospital transfers are commonplace for patients requiring tertiary levels of care.17 Thus, determining appropriate volume-based performance thresholds that confer value is a necessary first step in communicating with administrators, payers, and policy makers.

Several studies have reported on favorable patient outcomes and decreased costs for surgeons and hospitals with higher volumes than those with lower volume.18–22 However, practical experience informs us that learning and improvement is not a linear process, but rather stepwise, and there may also exist a point of diminishing returns. Is a surgeon who does 1,000 primary total hip arthroplasties (THAs) truly different than one who does 850 per year? It is likely that both fall into a spectrum that are better than surgeons who perform 10 to 20 per year in terms of quality and cost-effectiveness. SSLR analysis enables us to more objectively answer these questions.

SSLR analysis is a methodology originally pioneered by Peirce and Cornell.21 Statistically, SSLR analysis applies receiver operating characteristic

---

**Figure 1.** Comparison of AUC- versus SSLR-based groupings for threshold determination.
(ROC) curves to identify volume thresholds with differing risks between strata. Within the realm of orthopaedic surgery, Wilson first described the volume-outcomes relationships in the context of total knee arthroplasty, based on determining the threshold where the complication risk curve dropped precipitously. As an example, a resident physician at one institution who successfully performed 300 primary THAs over a five-year training program is unlikely to be less equipped than a resident physician at another institution who performed 150. However, compared to a resident physician who replaced 10 hips, there is likely to be a difference in comfort level and skill. Where the threshold exists is unclear, but it is attainable using stratum-specific likelihood ratio (SSLR).

**How Does SSLR Work?**

Once a data set is prepared in terms of the desired time period, the first step of SSLR involves splitting the cohort into many different groupings. Using our example of a group of resident physicians who performed THAs involving creating many groups consisting of the total unique number of THAs they performed over their residency at varying increasing increments—ranging from groups of one to two THAs per five-year program to groups of one to 10 and one to 15 per year, incrementally increasing. Next, a risk ratio is calculated for each group created, defined as the ratio of sensitivity to 1 − specificity. The sensitivity (i.e., true positive rate) and 1 − specificity (i.e., false positive rate) are based on an established cut-off, generally the mean or first standard deviation for the outcome measure of interest. For evaluating a Medicare population, one outcome measure could be a length of stay (LOS) greater than three days or an overall cost of greater than $25,000 per inpatient stay. This outcome measure selection determines the point at which observed hospital cost and LOS were meaningfully different from an outcome perspective, which allows for statistically significant groupings to be created based on changes to the ROC curve.

With these groupings and associated risk ratios, statistical significance is computed. These groups are then merged into progressively larger groups until a significant difference in risk ratios is observed between adjacent groups. This allows the conversion of the risk groups into numerous discrete groups. Such an approach allows a more objective statistical approach to stratification and maximizes the number of thresholds that may be created. Once the statistically significant groupings are created, odds ratios and confidence intervals are calculated to compare grouping thresholds, computing a point estimate along with confidence coefficient and standard error, with the highest threshold grouping used as the reference group. These thresholds can be computed and aggregated at the level of the surgeon and hospital level.

**SSLR versus Linear Stratification: A Stepped Approach to Growth**

SSLR analysis provides a stepwise approach to targeting improved performance and communicating growth at the surgeon and hospital level for technical and systemic proficiency, respectively. Individual performance outcomes can be tracked with respect to the aforementioned stratified groupings as benchmarks. More importantly, SSLR analysis allows us to establish thresholds that are evidence-based, practical, and meaningfully related to value. SSLR extends beyond attempts made using the routine area under the curve (AUC) analysis of ROC curves, which do not provide statistically dissimilar threshold groupings within a comprehensive data set. While AUC analysis is a rigorous statistical approach, it does not provide sufficient information for multi-strata threshold determination. SSLR analysis improves upon AUC analysis and provides a rigorous statistical approach to volume stratification by identifying numerous thresholds through the preservation of more information from the ROC curve (see Fig. 1).

**What Have We Learned From SSLR Analysis?**

SSLR can be used to define thresholds for additional outcome metrics, including patient-reported outcome measures (PROMs), inpatient complications, readmission rates, and revision rates. Thus, established thresholds may change if we were to examine annual orthopaedic surgeon and hospital volume using a different patient cohort or outcome. Additionally, specific details of each procedure may be required to perform more relevant risk-adjusted analyses. In recent applications to orthopaedic procedures, SSLR has been used to analyze total knee arthroplasty, shoulder arthroplasty, anterior cruciate ligament (ACL) reconstruction, adolescent idiopathic scoliosis procedures, and total hip arthroplasty. These studies have demonstrated that an increased volume has been associated with reduced LOS and reduced cost at the hospital and surgeon level. Lyman et al. demonstrated that the learning curve for hip arthroscopy was unexpectedly demanding, as cases performed by surgeons with career volumes greater than 500 had significantly lower risks of subsequent hip surgery than those performed by lower-volume surgeons—finding a *p*-value less than 0.0001 between the groups. This could similarly apply to the total arthroplasty learning curve over a five-year training program or as part of a fellowship program. The findings of previous SSLR studies have been outlined in Table 1 and

---

### Table I

<table>
<thead>
<tr>
<th>Independent Strata</th>
<th>TKA</th>
<th>TSA</th>
<th>THA</th>
<th>Hip Fracture</th>
<th>Spinal Fusion</th>
<th>Laminctomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon LOS</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Surgeon Cost</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Hospital LOS</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Hospital Cost</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>
### Table II
Average reduction of LOS and cost by strata

<table>
<thead>
<tr>
<th>Avg Reduction by Strata</th>
<th>TKA</th>
<th>TSA</th>
<th>THA</th>
<th>Hip Fracture</th>
<th>Spinal Fusion</th>
<th>Laminectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon LOS (d)</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.9</td>
<td>1.1</td>
</tr>
<tr>
<td>Surgeon Cost ($)</td>
<td>$767</td>
<td>$1,233</td>
<td>$167</td>
<td>$800</td>
<td>$2,100</td>
<td>$3,200</td>
</tr>
<tr>
<td>Hospital LOS (d)</td>
<td>0.4</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td>Hospital Cost ($)</td>
<td>$867</td>
<td>$1,133</td>
<td>$1,033</td>
<td>$1,400</td>
<td>$1,450</td>
<td>$1,300</td>
</tr>
</tbody>
</table>

### Table III
Surgeon volume vs. length of stay (LOS) strata

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TKA</th>
<th>TSA</th>
<th>THA</th>
<th>Hip Fracture</th>
<th>Spinal Fusion</th>
<th>Laminectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>&lt;13</td>
<td>&lt;5</td>
<td>&lt;70</td>
<td>&lt;31</td>
<td>&lt;58</td>
<td>&lt;17</td>
</tr>
<tr>
<td>Medium</td>
<td>13–64</td>
<td>5–14</td>
<td>70–120</td>
<td>N/A</td>
<td>58–90</td>
<td>17–40</td>
</tr>
<tr>
<td>High</td>
<td>65–124</td>
<td>&gt;14</td>
<td>&gt;120</td>
<td>31+</td>
<td>91–124</td>
<td>41–71</td>
</tr>
<tr>
<td>Very High</td>
<td>&gt;124</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>&gt;71</td>
</tr>
</tbody>
</table>

### Table IV
Surgeon volume vs. cost strata

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TKA</th>
<th>TSA</th>
<th>THA</th>
<th>Hip Fracture</th>
<th>Spinal Fusion</th>
<th>Laminectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>&lt;14</td>
<td>&lt;5</td>
<td>&lt;74</td>
<td>&lt;31</td>
<td>&lt;44</td>
<td>&lt;17</td>
</tr>
<tr>
<td>Medium</td>
<td>14–77</td>
<td>5–14</td>
<td>74–123</td>
<td>N/A</td>
<td>N/A</td>
<td>17–33</td>
</tr>
<tr>
<td>High</td>
<td>&gt;77</td>
<td>&gt;14</td>
<td>&gt;123</td>
<td>31+</td>
<td>44+</td>
<td>34–86</td>
</tr>
<tr>
<td>Very High</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>&gt;86</td>
</tr>
</tbody>
</table>

### Table V
Hospital volume vs. length of stay (LOS) strata

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TKA</th>
<th>TSA</th>
<th>THA</th>
<th>Hip Fracture</th>
<th>Spinal Fusion</th>
<th>Laminectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Low</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>&lt;43</td>
</tr>
<tr>
<td>Low</td>
<td>&lt;51</td>
<td>&lt;4</td>
<td>&lt;121</td>
<td>&lt;60</td>
<td>&lt;81</td>
<td>43–96</td>
</tr>
<tr>
<td>High</td>
<td>89–204</td>
<td>&gt;15</td>
<td>&gt;357</td>
<td>&gt;146</td>
<td>&gt;471</td>
<td>148–172</td>
</tr>
<tr>
<td>Very High</td>
<td>&gt;204</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>&gt;172</td>
</tr>
</tbody>
</table>

### Table VI
Hospital volume vs. cost strata

<table>
<thead>
<tr>
<th>Threshold</th>
<th>TKA</th>
<th>TSA</th>
<th>THA</th>
<th>Hip Fracture</th>
<th>Spinal Fusion</th>
<th>Laminectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Low</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>&lt;43</td>
</tr>
<tr>
<td>Low</td>
<td>&lt;194</td>
<td>&lt;4</td>
<td>&lt;122</td>
<td>&lt;126</td>
<td>&lt;95</td>
<td>43–82</td>
</tr>
<tr>
<td>Medium</td>
<td>194–260</td>
<td>4–15</td>
<td>122–309</td>
<td>N/A</td>
<td>N/A</td>
<td>83–115</td>
</tr>
<tr>
<td>High</td>
<td>&gt;260</td>
<td>&gt;15</td>
<td>&gt;15</td>
<td>126+</td>
<td>95+</td>
<td>116–169</td>
</tr>
<tr>
<td>Very High</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>&gt;169</td>
</tr>
</tbody>
</table>
Applying Performance Thresholds at the Surgeon-Level

Our studies suggest that low-volume surgeons may benefit from performing additional procedures to achieve a certain volume threshold. In the case of arthroplasty, achieving the minimum threshold is necessary to yield improved outcomes, with further thresholds possible at high and very high volumes. In many cases, the thresholds to becoming a medium- or even high-volume surgeon that result in improved value are actually quite low and, in some cases, would take as few as 70 to 80 cases a year for a knee arthroplasty surgeon to be considered high-volume. Whereas, our threshold analysis oversimplifies the complexity of surgical experience, these population-level data do not apply to the competence or expertise of individual surgeons. In addition, volume is just one of a number of different considerations that should be taken into account in choosing a surgeon or hospital. That said, if 70 to 80 procedures per year is the necessary “experience” to providing quality care from a strictly LOS and cost perspective, this threshold has meaningful consequences that may have relevance to numerous applications, from administrative targets to trainee benchmarks.

Limitations of SSLR Analysis

There exists specific limitations with SSLR-based analysis. The database being utilized may not be representative of a hospital or surgeon pool that thresholds are applied to, and different databases may result in various strata and threshold levels. As such, values established represent a guideline and may not be generalizable. By running SSLR across larger state and national databases with the specifics of each procedure better quantified, more broadly applicable cutoffs for growth and performance development may emerge. In some cases, goals and milestones may need to be set at a more specific level, including sub-stratification by type of surgical fellowship training and years of experience.

What Next?

Adoption of performance threshold analysis at the hospital and residency-training level can further enable favorable outcomes from the volume-value relationships in orthopaedic surgery. The breadth of studies examining performance thresholds in orthopaedics has been limited to retrospective analyses. Furthermore, parameters examined can be expanded to incorporate PROMs and other outcome measures. Machine learning data approaches may enable multiple outcomes to be examined simultaneously, with a generalized outcome threshold generated. SSLR is able to define baseline thresholds for key patient outcome metrics, as well as to establish training benchmarks where surgeons and hospitals alike are able to improve performance and outcomes. Furthermore, with SSLR, individual surgeon and hospital performance may be more objectively stratified and evaluated on a periodic basis in these groupings at an administrative level.

CONCLUSION

The role of volume-based optimization of value in orthopaedic surgery allows the dual ability to improve care and create efficiencies, an area of extreme scrutiny in lower extremity arthroplasty. SSLR is a preferred evidence-based technique that can define multiple performance strata rooted in evidence that expand on previous attempts to define the value-volume relationships at both the hospital and surgeon level. It can also be used to define thresholds for additional outcome metrics, as well as establish standardized training guidelines for orthopaedic surgeons, as has been shown for total joint arthroplasty. Finally, the advanced data approach SSLR offers may further enable threshold analysis to establish volume-based performance benchmarks in orthopaedic surgery relevant to administration and policy makers.

AUTHORS’ DISCLOSURES

Dr. Mont is a consultant for, or has received institutional or research support from, the following companies: CyMedica Orthopedics, Inc., Performance Orthopaedics, Inc., Perfor...
Stratum-Specific Likelihood Ratio Analysis: An Evidence-Based and Pragmatic Approach to Meaningful Thresholds in Lower Extremity Arthroplasty

NAVARRO/HAEBERLE/MONT/KREBS/RAMKUMAR


Dr. Krebs is a consultant for Stryker Corporation.

All other authors have no conflicts of interest to disclose.

REFERENCES