Systematic Review

The Learning Curve for Hip Arthroscopy: A Systematic Review


Purpose: The learning curve for hip arthroscopy is consistently characterized as “steep.” The purpose of this systematic review was to (1) identify the various learning curves reported in the literature, (2) examine the evidence supporting these curves, and (3) determine whether this evidence supports an accepted number of cases needed to achieve proficiency. Methods: The electronic databases Embase and Medline were screened for any clinical studies reporting learning curves in hip arthroscopy. Two reviewers conducted a full-text review of eligible studies and a hand search of conference proceedings and reference sections of the included articles. Inclusion/exclusion criteria were applied, and a quality assessment was completed for each included article. Descriptive statistics were compiled. Results: We identified 6 studies with a total of 1,063 patients. Studies grouped surgical cases into “early” versus “late” in a surgeon’s experience, with 30 cases being the most common cutoff used. Most of these studies used descriptive statistics and operative time and complication rates as measures of competence. Five of 6 studies showed improvement in these measures between early and late experience, but only one study proposed a bona fide curve. Conclusions: This review shows that when 30 cases was used as the cutoff point to differentiate between early and late cases in a surgeon’s experience, there were significant reductions in operative time and complication rates. However, there was insufficient evidence to quantify the learning curve and validate 30, or any number of cases, as the point at which the learning curve plateaus. As a result, this number should be interpreted with caution. Level of Evidence: Level IV, systematic review of Level IV studies.

Hip arthroscopy is currently one of the fastest developing fields in orthopaedic surgery.¹ The recent focus on the diagnosis and treatment of pathologic intra-articular hip conditions such as femoroacetabular impingement and labral tears has led to a rapid increase in both the number of surgeons performing hip arthroscopy and the number of hip arthroscopies being performed. Over a 5-year period from 2006 to 2010, the number of reported hip arthroscopies performed by American Board of Orthopaedic Surgery Part II examinees increased by more than 600%, from 83 to 636.² Additionally, this growth was reflected in a parallel increase of approximately 500% in the corresponding literature (22 citations in 2005, increasing to 101 in 2010), as discussed by Ayeni et al.³

Hip arthroscopy is a technically demanding procedure and is almost universally described as having a difficult or “steep” learning curve.⁴,⁵ This may be because of the thickness of the surrounding structures limiting maneuverability, depth of the joint, distance of hands from the manipulating instruments, high level of congruency between the acetabulum and femoral head, and the need for specialized instrumentation. Although a learning curve is an intuitive concept, reflecting an individual’s improvement in skill or competence over time, there is no universally accepted definition.⁶ Rather, there appears to be considerable ambiguity regarding the use of this term. There is, nonetheless, the need to understand the various factors that influence the acquisition of skills and performance to safely and effectively perform a procedure such as hip arthroscopy.

The purpose of this review was to systematically evaluate the literature to identify the various learning curves reported for hip arthroscopy, examine the
evidence on which these learning curves are based, and ultimately determine whether there is a reasonably accepted number of cases needed to achieve proficiency in this area.

**Methods**

**Search Strategy**

Medline (1946 to July 2013) and Embase (1980 to July 2013) were searched by 2 reviewers for any clinical studies or abstracts involving learning curves in hip arthroscopy. The search strategy used combinations of the following MeSH terms: hip, arthroscopy, injury, femoroacetabular impingement, learning, learning curve, clinical competence, treatment outcome, experience, assessment, complication, and result (the Appendix provides the full search strategy). A broad search strategy was used to capture articles that described a learning curve but did not explicitly use this term.

In addition, these searches were supplemented with hand searches by both reviewers of a number of conference proceedings from 2010 to 2013, including the American Academy of Orthopaedic Surgeons, the Arthroscopy Association of North America, the American Orthopaedic Society for Sports Medicine, the American Association of Hip and Knee Surgeons, the International Society for Hip Arthroscopy, and the International Society of Arthroscopy, Knee Surgery, and Orthopaedic Sports Medicine. In addition, we hand searched the reference sections of all included articles.

**Inclusion and Exclusion Criteria**

Articles meeting the following eligibility criteria were included: (1) articles in any language; (2) Level of evidence of I, II, III, or IV; (3) studies involving human patients; (4) surgical treatment involving hip arthroscopy; (5) a formal discussion of the learning curve by table, graph, or statistical technique, or even descriptive comparison between time points on a learning curve; and (6) evaluation of any clinical outcome in relation to the learning curve, including but not limited to the following: traction time, operative time, complication rate, conversion rate to an open procedure, or conversion to a total hip arthroplasty. Exclusion criteria included the following: (1) studies that did not compare outcomes based on surgeons' experience, (2) studies conducted in a simulation laboratory, or (3) review articles, technique papers, letters, comments, or editorials.

**Identification of Studies**

Title, abstract, and full-text screening were conducted in duplicate in a stepwise manner. Any discordant opinion at any stage of the review process was resolved by discussion among all authors. Cohen’s kappa statistic was calculated at each stage of the abstract and full-text screening was used to measure inter-rater agreement7; if either reviewer felt a study was appropriate after the title screening, it was included.

**Assessment of Methodological Quality**

All articles identified for full-text review were graded for level of evidence by the 2 reviewers using the criteria discussed by Wright et al.8 Nonrandomized studies were graded for quality using the Methodological Index for Non-Randomized Studies (MINORS), which has been shown to have high test-retest reliability, high internal consistency, and good inter-reviewer agreement.9 If available, randomized controlled trials were assessed with the Consolidated Standards of Reporting Trials checklist, a guideline developed to improve the quality of reporting in randomized controlled trials.10

**Data Collection**

Baseline data were abstracted from all included articles by both reviewers and included authors, date of publication, journal, study design, sample size, patient characteristics, and mean age. Type of learning curve described, outcomes, complications, and mean operative time were also recorded.

**Data Analysis**

Descriptive statistics from the selective studies were tabulated and reviewed. Further statistical analysis was not possible because of heterogeneity in the study design and outcome measures.

**Results**

**Study Identification**

The literature search yielded 1,023 articles after removal of duplicates. Four studies from the database search (431 patients)11-14 and 2 abstracts (632 patients)15,16 from the search of conference proceedings met the inclusion criteria and were selected for data abstraction (Fig 1). Cohen’s kappa statistic measuring inter-rater agreement was 0.40 (95% confidence interval [CI], 0.31-0.50) after the title screening, 0.66 (95% CI, 0.43-0.90) after the abstract screening, and 1.0 after the full text screening, indicating moderate, high, and perfect agreement, respectively, at each successive stage.

**Study Characteristics**

Table 1 lists descriptive characteristics of the included studies. All studies were Level IV evidence, consisting of case series of hip arthroscopies performed by a single surgeon. Four of the studies were prospective, and the remaining 2 were retrospective. Only 2 of the articles described the previous experience of the surgeon.11,12 The mean number of cases per study was 177 (range, 40 to 400). All studies were of low to moderate methodological quality, with an average MINORS quality
assessment of 9.7 (range, 9 to 11) of 16. In general, studies had an ambiguous explanation of the inclusion and exclusion criteria, data collection practices, and assessment of outcomes.

**Description of Individual Studies**

Although all the articles that were included in the review mentioned the words “learning curve,” all, with the exception of Konan et al., grouped the cases into “early” versus “late” in terms of experience and compared outcomes between the 2 groups rather than provide data on every consecutive case, which would translate into a bona fide “curve.” Although Lee et al. grouped patients into early and late cases, they additionally performed a cumulative sum (CUSUM) analysis to determine how many cases were required for the cumulative failure rate not to differ significantly from an acceptable rate.

There was no uniformity in the criteria used to define “early” and “late” stages in a surgeon’s career. The most
common cutoff point reported was 30 cases, which was used in 4 of the studies. Lee et al.\textsuperscript{12} compared the first 20 cases with the second 20 cases (but also performed a CUSUM as noted), Souza et al.\textsuperscript{13} used consecutive groups of 30 cases, Vilchez et al.\textsuperscript{14} compared the first 30 cases with the next 67 cases, Sobau et al.\textsuperscript{16} compared the first 100 cases with the last 300 cases, whereas Comba et al.\textsuperscript{15} took the first 30 cases versus the last 202 cases. Konan et al.\textsuperscript{11} did attempt to provide a curve by using blocks of 10 for the first 100 cases and comparing the first 30 cases with the remaining 70 cases.

The previous experience of the surgeon was reported in only 2 studies\textsuperscript{11,12}. All authors were contacted and asked to provide more information concerning the hip arthroscopy training of the surgeons involved. Four studies chose to include data beginning with the surgeon’s initial hip arthroscopy performed in practice, although their training varied. Only in Lee et al.\textsuperscript{12} did the surgeon complete a hip surgery fellowship, assisting on 117 hip arthroscopies, whereas the remaining 3 surgeons had no formal training. The surgeon in the study of Konan et al.\textsuperscript{11} attended instructional course lectures on hip arthroscopy, and the surgeon in the study of Vilchez et al.\textsuperscript{14} attended 2 instructional courses and a course using cadavers and observed 15 hip arthroscopies during a 2-week visit to an academic center. Finally, Comba et al.\textsuperscript{15} participated in a course using cadavers and observed 25 hip arthroscopies during a month-long observership at 2 institutions in the United States. As a result, because their experiences and training differed, the surgeons may have begun at different points on the learning curve, although the data represents the outcomes from their initial hip arthroscopy.

The specific surgical indications were not given in any of the studies, although Konan et al.\textsuperscript{11} noted that the types of skills performed increased in difficulty as cases accumulated.

### Outcome Measures Used

Outcomes used to assess skill or learning included rate of complications, operative time, and a heterogeneous assortment of patient-related outcomes. The most common outcome measure used was rate of complications, which was reported in 5 of the studies. One study found a significant decrease in complications, and a second showed no significant difference in rates stratified by groups. The remaining 3 studies did not use comparative statistics, but we were able to compute a $P$ value and CIs for the relative risk reduction (Table 2). The most common type of complication reported was neurologic injury, but in general the studies did not provide detailed information on the type or severity of complications. At the time of these publications, however, there was no universally accepted method of grading complications or adverse events associated with hip arthroscopy.
Operative time was used to measure learning in 4 studies, all of which showed a substantial decrease between early and late cases, with 2 studies reporting a significant difference, whereas the remaining 2 studies provided only descriptive information (Table 3).

Lee et al.12 attempted to use patient-oriented outcomes to determine the learning curve. They measured patient outcomes using modified Harris hip scores (mHHS) 6 months postoperatively and performed a CUSUM analysis with these scores to compare cumulative failure rates as the number of cases increased. They defined failure as a mHHS less than 80 and assumed an acceptable failure rate to be 20%. They found that after 21 cases, the failure rate began to decrease consistently, and after 30 cases, the failure rate was equal to the defined acceptable failure rate. From this, they estimated that at least 20 cases are required to achieve a satisfactory outcome.

Of the 6 articles, only 2 empirically established the number of cases required for the learning curve to plateau. Konan et al.11 reported a value of 30 cases, and Lee et al.12 reported a range of 20 to 30 cases for this to occur. A third study, Comba et al.,15 used a value of 30 cases based on Konan et al.11 and obtained similar results, and so concluded that 30 cases is an appropriate value for the learning curve.

## Discussion

Our review found that when 30 cases was used as the cutoff number to differentiate between early and late cases in a surgeon’s career, there were significant reductions in operative time and complication rates. Three of the 6 articles reported statistical significance, whereas an additional 2 studies showed trends supporting 30 cases as the number needed but did not perform any statistical analysis. However, there was insufficient evidence to quantify the learning curve and validate that 30 cases is the point at which the learning curve plateaus. For example, the parameters used and goals of defining various learning curves for arthroscopic hip procedures are variable and have not been well defined. In addition, inherent surgical skill, manual dexterity, and learning ability have not always been carefully identified. Therefore, this value should be interpreted with extreme caution.

One of the purposes of our review was to understand the different types of learning curves that have been reported for hip arthroscopy. Conceptually, a learning curve should be identified as a graph in which a consecutive number of cases are presented on the horizontal axis and some measure of proficiency or learning is presented on the vertical axis.18 One may assume that the most learning with regard to the parameters being evaluated has occurred at the point at which the learning curve plateaus. Unfortunately, with one exception,12 this is not how the term was used in the studies included in this review. Instead, a learning curve was replaced by a dichotomous comparison of early versus late stages in a surgeon’s operative experience. In addition, the word “steep” has been used to describe the learning curve for hip arthroscopy. In common usage, steep (as in climbing a steep hill) connotes difficulty, but as a curve, steep means a rapid increase in learning over a relatively small number of cases, which is the opposite intent of the term. In the end, it is clear that hip arthroscopy is rapidly evolving and relatively challenging, yet the learning curve has yet to be defined with precision as steep or gradual.

Three articles reported 30 cases a priori as the boundary to define early versus late in a surgeon’s experience. As pointed out in Perez-Carro and Tey,19 this number appears to have first been suggested as an appropriate number of cases to define a learning curve for hip arthroscopy.

### Table 2. Summary of Outcome Data

<table>
<thead>
<tr>
<th>Study</th>
<th>Grouping of Cases</th>
<th>Early Group</th>
<th>Late Group</th>
<th>Relative Risk Reduction (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Souza et al.13</td>
<td>Consecutive groups of 30 cases</td>
<td>3.3%</td>
<td>7.3%</td>
<td>-1.195 (-15.261-0.704)</td>
<td>.771</td>
</tr>
<tr>
<td>Vilchez et al.14</td>
<td>First 30 v 31-97</td>
<td>17%</td>
<td>3%</td>
<td>0.821 (0.128-0.963)</td>
<td>.041*</td>
</tr>
<tr>
<td>Konan et al.11</td>
<td>First 30 v 31-100</td>
<td>26.7%</td>
<td>2.9%</td>
<td>0.750 (-0.081-0.942)</td>
<td>.080*</td>
</tr>
<tr>
<td>Sobau et al.16</td>
<td>First 100 v 101-400</td>
<td>11%</td>
<td>2%</td>
<td>0.818 (0.521-0.931)</td>
<td>.005*</td>
</tr>
<tr>
<td>Lee et al.12</td>
<td>First 20 v 21-40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Comba et al.15</td>
<td>First 30 v 31-232</td>
<td>23%</td>
<td>5%</td>
<td>0.809 (0.526-0.923)</td>
<td>.002</td>
</tr>
</tbody>
</table>

NOTE. Boldface indicates that the P values are significant (P < .05).

*These 3 P values were not provided in the corresponding articles but were calculated for this review.

### Table 3. Mean Operative Time Outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Grouping of Cases</th>
<th>Mean Operative Time (min)</th>
<th>Early Group</th>
<th>Late Group</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Souza et al.13</td>
<td>Consecutive groups of 30 cases</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Vilchez et al.14</td>
<td>First 30 v 31-97</td>
<td>150</td>
<td>90</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Konan et al.11</td>
<td>First 30 v 31-100</td>
<td>75</td>
<td>45</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Sobau et al.16</td>
<td>First 100 v 101-400</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lee et al.12</td>
<td>First 20 v 21-40</td>
<td>135</td>
<td>89</td>
<td>-</td>
<td>.012</td>
</tr>
<tr>
<td>Comba et al.15</td>
<td>First 30 v 31-232</td>
<td>137</td>
<td>110</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

NOTE. Boldface indicates that the P values are significant (P < .05).
curve in 2006 by a panel of experts at the 2nd International Arthroscopy Hip Meeting in Homburg, Germany. However, by not establishing a curve empirically, it is impossible to determine where 30 cases sits on the learning curve.

There is also concern with the outcome measures used to infer learning. Typically, improvement is assessed indirectly by 2 types of variables: either measures of patient outcome (complication rate, functional scores, or survival), or measures of task efficiency (time, number of hand movements, and so on).6 Five of the articles reviewed used operative time as an outcome, i.e., a measure of task efficiency. Operative time, from incision to closure, may provide an indirect measure of skill, because one would presume the length of a procedure would decrease as a surgeon’s skill and competence increases. However, many uncontrollable variables affect the length of an operation, and the total operative time includes not only the surgical procedure but also the time required for anesthesia induction and patient positioning. Because none of the studies indicated exactly what was measured, 30 cases may well reflect the entire operating team’s comfort with the procedure and equipment rather than the surgeon’s skill. Also, many of the authors of the included articles stated that as the surgeon’s comfort with the procedure increased, the complexity of the cases did as well. One may expect that a surgeon performing more complex procedures or a higher percentage of revision procedures may result in a longer mean operative time in comparison with a surgeon performing more basic procedures. For these reasons, the use of operating time as a variable to define a learning curve has been called both “crude and probably unacceptable”20 and a “weak proxy for learning.”6

Additionally, operating time may not be an accurate indicator of patients’ clinical outcomes. Aside from Lee et al.12 who used the mHHS scores, none of the articles looked at patient-important outcomes, such as patient satisfaction, validated functional scores, return to sport, or revision rates. Further, the mHHS score was originally validated for patients undergoing hip arthroplasty and may not be sensitive enough in a younger, more active population because of its ceiling effect.21 Recently, Aprato et al.22 reported that the mHHS had limited predictive value in their study regarding patient satisfaction for patients who underwent arthroscopic hip procedures.

Complication rate was another common outcome measure used, but rarely were chondral injuries reported, which may be a better measure of how a surgeon can maneuver within the joint. Another concern was that there was no distinction in any of the articles regarding the type of surgery performed. It is unknown whether the learning curve for basic removal of loose bodies or debridement differs measurably from the learning curve for more technically demanding procedures such as labral repair, labral reconstruction, capsular repair/plication, or femoroacetabular impingement corrective procedures. Hence, if the specific procedures and level of complexity of the surgical procedures are not controlled, the data obtained might not apply to all surgeons and might represent different points on the hip arthroscopy learning curve.

Finally, the surgeon’s baseline skill was not quantified. Previous experience, fellowship training, and number of cases performed before inclusion in the various studies generally was not stated. It is unknown at what point the individual investigators started on the learning curve for the various studies, and therefore it is difficult to extrapolate data and make firm conclusions regarding this topic.

The strengths of this study include use of a comprehensive literature search involving 2 major databases, hand searches of all relevant conference proceedings, and a sufficiently broad search strategy including articles in languages other than English. Moreover, multiple reviewers screened and evaluated articles.

**Limitations**

The findings in our review are limited by the lack of high-quality studies. Despite our broad search strategy, only case series (noncomparative) data were available, and each study involved the variable experience and ability of a single surgeon, making generalizations to a larger group of surgeons difficult. Additionally, the articles included used varying lengths of follow-up and heterogeneous outcomes.

**Controversies**

This study was motivated by the desire to understand at what stage a surgeon may be considered ready to expect reasonable outcomes and reliable rates of success when performing arthroscopic hip procedures. This is a contentious question because hip arthroscopy is still in its youth, and there are no guidelines or specific criteria describing who should be performing hip arthroscopy.23 For instance, after a residency with limited exposure to hip arthroscopy, should a resident be expected to perform the procedure? Is attending a weekend instructional course adequate for a surgeon to begin carrying out hip arthroscopies? Even after a fellowship in sports medicine, some surgeons may have assisted in only a handful of hip arthroscopies, which may not provide enough exposure to surgical technique or decision making. Peters et al.24 in a panel discussion by hip preservation experts, reported that: “Diagnosis and management of young adult hip deformities frequently are not formally covered in sports medicine or adult reconstruction curricula.”
If we had a better understanding of the learning curve for hip arthroscopy, it would then be possible to establish guidelines concerning the number and type of arthroscopic hip procedures one should observe, assist, or perform and the amount of experience regarding the surgical and nonsurgical decision-making process that is required before being considered competent. This could lead to the establishment of mentorship or certification programs or the creation of more formalized hip preservation fellowships so that surgeons could consider themselves subspecialists in hip arthroscopy.

Future Directions

One of the critical questions derived from this review is how to best define and evaluate the learning curve. Hip arthroscopy is simply an approach that allows for various surgical procedures to be performed. This approach can be used to treat various patterns of hip impingement, subtle dysplastic variants, combined impingement and instability, loose bodies, snapping hip, traumatic instability, and other disorders in the native hip and after hip arthroplasty. If the procedures performed and the level of experience (which determines a surgeon’s baseline skill) before conducting the study are not well defined, it will not be possible for other surgeons to see how specific reported learning curves apply to their individual practices. There are a number of factors regarding arthroscopic hip procedures that might have their own associated learning curves, and future studies might evaluate and attempt to define each of these. Complication rates and operative time might not necessarily correlate with functional outcomes and, therefore, learning curves for each of these parameters might need to be defined separately. Similarly, there might be different cutoff points for when surgical time or complication rates become consistently low, or when one can perform a hip arthroscopy without significant chondral injury. Examining these learning curves individually might help to better differentiate when a surgeon can perform safely and efficiently and when the surgeon develops appropriate surgical indications and proficiency with arthroscopic hip surgical techniques.

Future research requires a prospective formalized analysis of the outcomes of consecutive hip arthroscopies by individual surgeons, stratified by similarity of procedure and using one or more outcome measures that truly reflect learning. This might ultimately allow us to determine the point at which a plateau in the learning curve occurs.

Conclusions

This review shows that when 30 cases was used as the cutoff number to differentiate between early and late cases in a surgeon’s experience, there were significant reductions in operative time and complication rates. However, there was insufficient evidence to quantify the learning curve and validate 30 or any other number of cases as the point at which the learning curve plateaus. As a result, this number should be interpreted with caution.

References


Appendix. Full Database Search Strategy

**MEDLINE Search Strategy**

(exp Hip Fractures/ or exp Hip Dislocation/ or exp Hip Injuries/ or hip/mp or exp Hip Contracture/ or exp Hip Joint/ or exp Hip Dislocation, Congenital/ or exp Hip/ or ext Osteoarthritis, Hip/ or exp Hip Dysplasia or exp Acetabulum/ or exp Femoroacetabular Impingement/ or exp Femur Head) AND (arthroscopy.mp or exp Arthroscopy/) AND exp Learning Curve/ or exp Learning/ or Clinical Competence.mp or exp Clinical Competence or exp Treatment Outcome/ or exp outcome.mp or exp Motor Skills/ or outcome/mp or exp Fatal Outcome/ or exp “Outcome Assessment (Health Care)”/ or exp Treatment Outcome or exp Intraoperative Complications/ or exp Postoperative Complications/ or compiliation.mp)

**EMBASE Search Strategy**

(exp hip disease/ or exp hip osteoarthritis/ or exp hip injury/ or exp hip osteotomy/ or exp congenital hip dislocation/ or exp hip pain/ or exp hip distractor/ or exp hip radiography/ or exp hip arthroscopy/ or exp hip surgery/ or exp “Hip Disability and Osteoarthritis Outcome Score”/ or exp hip dislocation. Or exp hip fracture/ or exp hip dysplasia/ or exp hip contracture/ or exp hip/ or exp hip malformation/ or exp hip arthroplasty/ or exp femoroacetabular impingement/ or exp femur head/ or femur/ or exp femur malformation) AND (exp arthroscopy/ or exp hip arthroscopy/ or arthroscopy.mp.) AND (exp learning curve/ or exp learning/ or exp clinical competence/ or competence or exp experience or exp adverse outcome or exp treatment outcome/ or exp motor performance/ or exp complication/ or exp postoperative complication/)

---

**ARTHROSCOPY TECHNIQUES**

Have you been to www.arthroscopytechniques.org lately? Check it out; we now publish three video tech notes a week!