Learning curves in robot-assisted spine surgery: a systematic review and proposal of application to residency curricula

Zach Pennington, MD,1 Brendan F. Judy, MD,2 Hesham M. Zakaria, MD,3 Nikita Lakomkin, MD,1 Anthony L. Mikula, MD,1 Benjamin D. Elder, MD, PhD,1 and Nicholas Theodore, MD2

1Department of Neurosurgery, Mayo Clinic, Rochester, Minnesota; 2Department of Neurosurgery, Johns Hopkins University School of Medicine, Baltimore, Maryland; and 3Department of Neurosurgery, California Pacific Medical Center, Sutter Health, San Francisco, California

OBJECTIVE Spine robots have seen increased utilization over the past half decade with the introduction of multiple new systems. Market research expects this expansion to continue over the next half decade at an annual rate of 20%. However, because of the novelty of these devices, there is limited literature on their learning curves and how they should be integrated into residency curricula. With the present review, the authors aimed to address these two points.

METHODS A systematic review of the published English-language literature on PubMed, Ovid, Scopus, and Web of Science was conducted to identify studies describing the learning curve in spine robotics. Included articles described clinical results in patients using one of the following endpoints: operative time, screw placement time, fluoroscopy usage, and instrumentation accuracy. Systems examined included the Mazor series, the ExcelsiusGPS, and the TiRobot. Learning curves were reported in a qualitative synthesis, given as the mean improvement in the endpoint per case performed or screw placed where possible. All studies were level IV case series with a high risk of reporting bias.

RESULTS Of 1579 unique articles, 97 underwent full-text review and 21 met the inclusion and exclusion criteria; 62 articles were excluded for not presenting primary data for one of the above-described endpoints. Of the 21 articles, 18 noted the presence of a learning curve in spine robots, which ranged from 3 to 30 cases or 15 to 62 screws. Only 12 articles performed regressions of one of the endpoints (most commonly operative time) as a function of screws placed or cases performed. Among these, increasing experience was associated with a 0.24- to 4.6-minute decrease in operative time per case performed. All but one series described the experience of attending surgeons, not residents.

CONCLUSIONS Most studies of learning curves with spine robots have found them to be present, with the most common threshold being 20 to 30 cases performed. Unfortunately, all available evidence is level IV data, limited to case series. Given the ability of residency to allow trainees to safely perform these cases under the supervision of experienced senior surgeons, it is argued that a curriculum should be developed for senior-level residents specializing in spine comprising a minimum of 30 performed cases.

©AANS 2022, except where prohibited by US copyright law

ABBREVIATIONS ACGME = Accreditation Council for Graduate Medical Education; PGY = postgraduate year; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.


INCLUDE WHEN CITING DOI: 10.3171/2021.10.FOCUS21496.

KEYWORDS spine robot; learning curve; resident education; residency curricula
surgery, including several meta-analyses and one small randomized, unblinded prospective trial. The identified benefits have included increased instrument accuracy, reduced radiation dosage, reduced blood loss, and a shorter hospital length of stay. However, the heterogeneity of the published series makes it unclear as to whether the results of robot-assisted spine surgery are dependent on the experience level of the user. It is known from the orthopedic arthroplasty and general surgery literature that there exists a learning curve, a certain minimum number of cases that must be performed for the user to become proficient. Systematic reviews of the robotic laparoscopy and arthroplasty literature have suggested that the number to achieve baseline proficiency is 15 to 35 cases. To this end, many authors describing their experience with spine robots have similarly documented the presence of a learning curve. This is important, as the rates of complication may be higher early on in the learning curve and may actually make robotic surgery less safe than more conventional means as surgeons and trainees master these assistive devices.

Given the potential for robotic systems to be less efficient and less effective, at least early in the experience, there exists the question as to whether or not the teaching of robotic systems should be integrated within neurosurgery residency curricula. The objectives of the present review are to summarize the literature on learning curves in spine robotics and to propose a means by which the teaching of spinal robotics may be integrated within current residency curricula.

### Methods

To identify the current evidence for learning curves in spine robotic surgery, a systematic review of the literature was performed on July 22, 2021. Databases queried were PubMed/MDLILINE, Ovid MEDLINE, Web of Science, and Scopus. The bibliographies of included articles were queried for additional relevant references. Search queries are listed in Table 1. Articles were included if they described primary data from a cohort, trial, or series examining spine surgical robots used to place pedicle screws or other spine instrumentation in patients. To be included, articles must have presented data on one of the following endpoints as a surrogate of improvement in proficiency with the device: pedicle screw accuracy (graded or rate of misplacement), pedicle screw placement time, robot registration time, fluoroscopy utilization (time or radiation delivered), or operative time. Studies using any of the commercially approved robots (ExcelsiusGPS, Mazor SpineAssist, Mazor Renaissance, Mazor X, Mazor X Stealth, TiRobot, ROSA ONE Spine, and Cirq) were included. Articles were excluded if they did not present primary data (i.e., were case reports, reviews, commentaries, letters to the editor, or methods descriptions), if they described the use of a non-spine robot (e.g., DaVinci, Intuitive Surgical), if they studied learning outside the clinical setting (e.g., in cadavers, polyurethane models, or animals), or if they studied robot skills learning in non-spine surgery.

Articles were screened independently by two reviewers for inclusion and exclusion criteria according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Fig. 1). In cases of disagreement, a third reviewer was recruited to resolve the disagreement. Data extraction was then performed by the two reviewers with additional review by the third, independent reviewer. As the primary endpoint of the study was the presence or absence of a learning curve (i.e., a difference in the examined outcomes as a function of screws placed or cases performed) and none of the series provided patient-level data, a pooled analysis was not possible. Consequently, only a qualitative analysis of the gathered data is presented. Gathered endpoints were summarized as ranges and the number of studies reporting each outcome. All included studies were classified as level IV therapeutic studies according to the North American Spine Society levels of evidence guidelines, and were at high risk of reporting bias.

### Results

Our systematic review identified 1579 unique articles, of which 97 met criteria for full-text review. On full-text review, 76 articles were excluded, most commonly because they did not present learning curve data (n = 62) or did not present primary data (n = 7). This left 21 articles that met inclusion and exclusion criteria (Table 2). All were case series (level IV evidence) at high risk of reporting bias. Of the included articles, 2 described experiences with the Mazor SpineAssist, 8 with the Mazor Renaissance, 2 with the Mazor X, 2 with the Mazor X Stealth, 3 with the ExcelsiusGPS, and 1 with the TiRobot. Fourteen series described the experience of a single attending surgeon, described the pooled experience of multiple attending surgeons, and described the experiences of supervised residents or spine fellows, and described the experiences of a single attending surgeon and two spine fellows.
Series varied in size from 13 patients\textsuperscript{32} to 258 patients\textsuperscript{19} only 9 series considered > 50 treated patients\textsuperscript{12–16,19,23,27,28} with 2 having examined experience with the Mazor Spine-Assist\textsuperscript{19,27} 2 with the Mazor Renaissance\textsuperscript{12,28} 1 with the TiRobot\textsuperscript{15} 2 with the Mazor X\textsuperscript{14,16} 1 with the ExcelsiusGPS\textsuperscript{23} and 1 with the Mazor X Stealth\textsuperscript{13}. For endpoints, 8 articles used pedicle screw accuracy as a measure of the learning curve\textsuperscript{12,17,19,23,27,28,30,31} 10 used total operative time\textsuperscript{9,13–16,18,24,25,31,32} 8 used screw placement/insertion time\textsuperscript{13,17,18,20,22,24,28,29} 7 used fluoroscopy time per screw\textsuperscript{14,18,20,21,28,29,31} and 2 used system registration time\textsuperscript{13,17}.

**Learning Curve**

Of the 21 included articles, 18 observed the presence of a learning curve using ≥ 1 of the endpoints analyzed\textsuperscript{9,12–25,28,31,32} including 4 of the 5 series considering the experiences of surgeons who had employed the robot in ≥ 50 cases\textsuperscript{12,13,15,16} Among the 8 articles using screw accuracy as a measure of the learning curve\textsuperscript{12,17,19,23,27,28,30,31} 4 observed the presence of a learning curve, ranging from 10 to 30 cases\textsuperscript{12,17,19,23,27,28,30,31} or 30 screws, as in the analysis of Siddiqui et al\textsuperscript{23} When considering total operative time, 8 of the 10 articles reported the presence of a learning curve, ranging from 20 to 30 cases\textsuperscript{14,16,24,25,32} or 62 screws, as in the analysis of Onen et al\textsuperscript{18} Four of the 8 articles using screw placement/insertion time as a marker of proficiency observed the presence of a learning curve, ranging from 3 to 15 cases\textsuperscript{17,20,22} or 62 screws, again, as in Onen et al\textsuperscript{18} Three of the 7 articles using fluoroscopy time per placed screw observed the presence of a learning curve, ranging from 8 to 10 cases\textsuperscript{27,29} or 62 screws (Onen et al\textsuperscript{18}) One of the 2 articles considering system registration time observed a decrease with increased proficiency\textsuperscript{13}.

The majority (n = 13) of included studies performed dichotomous analyses with a predetermined case or screw threshold\textsuperscript{12,13,16–23,27,29,31} For those performing this analysis, the described cut-points were 3 to 30 cases or 15 to 62 screws Ten articles performed linear regressions of the examined endpoint as a function of the number of cases performed or screws placed\textsuperscript{9,14,16,17,20,24,25,28,30,32} For those performing said linear regressions, the improvement in screw placement time was estimated at 0.10 minutes per screw for each additional case performed\textsuperscript{20} and the improvement in operative time ranged from 0.24 to 4.6 minutes per case for each additional case performed\textsuperscript{9,16,24,25,32} Two studies, those of Avrumova et al\textsuperscript{13} and Chen et al\textsuperscript{15} performed logistic regressions Avrumova et al\textsuperscript{15} performed a regression for screw placement time as a function of screws placed in addition to a dichotomous analysis In this analysis they found that the learning curve was rela-
<table>
<thead>
<tr>
<th>Authors &amp; Year</th>
<th>System</th>
<th>Provider Level</th>
<th>Study Population</th>
<th>Outcome</th>
<th>Learning Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hu &amp; Lieberman, 2014</td>
<td>Mazor Renaissance</td>
<td>1 attending</td>
<td>• 150 pts underwent elective thoracolumbar fusion (12 robot failures; 19 MIS; 9 mixed open/MIS; 1699 screws)</td>
<td>• Screw accuracy (scale not given)</td>
<td>↑ screw placement accuracy from 1st 30 to subsequent cases (82 vs 93%; p &lt; 0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Stratified by quintiles</td>
<td>• Rate of conversion to manual placement</td>
<td>↓ rate of conversion from robot to manual comparing 1st 30 cases w/ subsequent 120 (17 vs 5%; p &lt; 0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Mean age 51 yrs; 29% M</td>
<td>• Screw malposition</td>
<td>NS ↓ screw malposition rate (0.8 vs 0.7%; p = 0.72)</td>
</tr>
<tr>
<td>Onen et al., 2014</td>
<td>Mazor Renaissance</td>
<td>1 attending</td>
<td>• 27 pts underwent thoracolumbar fusion for fracture or degenerative disease (136 screws; 60% PPS)</td>
<td>• Op time</td>
<td>Comparison of 1st 62 screws w/ next 74 showed ↓ op time (73.2 vs 46.1 min; p &lt; 0.05), ↓ screw placement time (15.5 vs 8.6 sec; p &lt; 0.05), &amp; ↓ fluoro time (1.8 vs 0.9 sec; p &lt; 0.05)</td>
</tr>
<tr>
<td>Schatlo et al., 2015</td>
<td>Mazor Spine-Assist</td>
<td>13 attendings</td>
<td>• 258 pts underwent thoracolumbar instrumentation (1265 screws)</td>
<td>• Screw accuracy on GR scale</td>
<td>Improvement nonlinear; ↑ misplacement from cases 11–15, peak btwn 16 &amp; 20, then improvement</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Pt demographics not given</td>
<td>• % total cases w/ screw malposition ↓ w/ surgeons w/ &gt;15 cases performed</td>
<td></td>
</tr>
<tr>
<td>van Dijk et al., 2015</td>
<td>Mazor Spine-Assist</td>
<td>2 attendings</td>
<td>• 112 pts underwent PPS lumbar fusion (494 screws)</td>
<td>• Screw accuracy on GR scale</td>
<td>NS difference in screw accuracy (p &gt; 0.13) or screw deviation (p &gt; 0.15) when comparing 1st 30 placed screws w/ subsequent screws for each operator</td>
</tr>
<tr>
<td>Hyun et al., 2017</td>
<td>Mazor Renaissance</td>
<td>1 attending</td>
<td>• 30 pts underwent 1- or 2-level MIS fusion (130 screws)</td>
<td>• Instrumentation time/screw, Fluoro time/screw</td>
<td>↓ time/screws for 2nd 15 vs 1st 15 (4.0 vs 5.5 mins; p = 0.23); curve: time (mins) = −0.0952 (case no.) + 6.1845 (r² = 0.20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Mean age 67 yrs; 30% M</td>
<td>• NS ↓ insertion time for “spine-dedicated” vs “not dedicated” residents (4.5 vs 3.84 min/screw)</td>
<td></td>
</tr>
<tr>
<td>Kim et al., 2017</td>
<td>Mazor Renaissance</td>
<td>1 attending</td>
<td>• 37 pts underwent 1- to 2-level PLIF for lumbar stenosis (158 screws)</td>
<td>• Screw accuracy on GR scale</td>
<td>1st 8 vs subsequent 29 cases showed ↓ fluoro time in platform attachment (7.75 ± 3.86 vs 2.9 ± 1.44 sec; p &lt; 0.001), ↓ pedicle screw insertion time (14.86 ± 3.3 vs 9.03 ± 3.85 sec; p &lt; 0.001) &amp; total fluoro time (27.5 ± 4.87 vs 18.45 ± 3.90 sec; p &lt; 0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Mean age 65 ± 10 yrs; 50% M</td>
<td>• Total fluoro time decreased mildly btwn cases 1 &amp; 9; stable thereafter</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>↓ total fluoro time came from ↓ fluoro time for robot platform attachment &amp; screw placement</td>
<td></td>
</tr>
<tr>
<td>Urakov et al., 2017</td>
<td>Mazor Renaissance</td>
<td>Junior residents (PGY1–5), senior residents (PGY6 &amp; 7), fellows (PGY8)</td>
<td>• 30 pts underwent short-segment fusion of thoracolumbar spine: T4–pelvis (306 screws; 6 cases MIS)</td>
<td>• Pedicle screw insertion time</td>
<td>↓ insertion time/screw w/ increasing case no. (no statistics); 1–3 cases vs 4–7 cases (4.6 vs 3.6; p = 0.057)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Mean age 69 yrs; 50% M</td>
<td>• NS ↓ insertion time for “spine-dedicated” vs “not dedicated” residents (4.5 vs 3.84 min/screw)</td>
<td></td>
</tr>
<tr>
<td>Kam et al., 2019</td>
<td>Mazor Renaissance</td>
<td>3 attendings</td>
<td>• 73 pts underwent PPS instrumentation (352 screws)</td>
<td>• Screw accuracy on GR scale</td>
<td>Plot of screw time vs case no. flat (r² = 2.6 × 10⁻¹)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Mean age 67 ± 11 yrs; 37% M</td>
<td>• Screw placement time</td>
<td>No difference in median screw placement time (p = 0.61) or screw accuracy (p = 0.31) w/ increased experience</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Fluoro time/screw</td>
<td>No trend in mean fluoro time/screw across cases</td>
</tr>
</tbody>
</table>

CONTINUED ON PAGE 5 »
TABLE 2. Summary of studies examining learning curves for spinal robotic systems

<table>
<thead>
<tr>
<th>Authors &amp; Year</th>
<th>System</th>
<th>Provider Level</th>
<th>Study Population</th>
<th>Outcome</th>
<th>Learning Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khan et al., 2019&lt;sup&gt;29&lt;/sup&gt;</td>
<td>Mazor X</td>
<td>1 attending</td>
<td>20 pts underwent 1-level lumbar fusion (75 screws)</td>
<td>• Screw placement time</td>
<td>Comparing 1st 10 w/ 2nd 10 cases showed no significant differences, though fluoro time (15.8 ± 21.8 vs 10.5 ± 4.8 sec) &amp; radiation dose (35.6 ± 44.4 vs 24.2 ± 14.1 mGy) were lower</td>
</tr>
<tr>
<td>Siddiqui et al., 2019&lt;sup&gt;23&lt;/sup&gt;</td>
<td>Excelsius-GPS</td>
<td>1 attending, 2 fellows</td>
<td>120 pts underwent thoracolumbar fusion (95% lumbar), mostly for degenerative disease &amp; adjacent-segment disease (665 screws)</td>
<td>• Screw tip deviation, tail deviation, &amp; angular deviation from plotted trajectory</td>
<td>Attending &amp; 1 fellow showed ↑ screw tip accuracy after 30 screws placed (p = 0.01 &amp; p = 0.002, respectively)</td>
</tr>
<tr>
<td>Bäcker et al., 2020&lt;sup&gt;24&lt;/sup&gt;</td>
<td>Mazor Renaissance</td>
<td>1 attending</td>
<td>46 pts underwent multilevel lumbar fusion for degenerative disease (281 screws); 16 cases excluded because of robot system setup failure</td>
<td>• Op time</td>
<td>NS ↓ robot usage time (slope −0.69 mins/op ($r^2$ = 0.07; p = 0.094), but 2 extreme outliers present</td>
</tr>
<tr>
<td>Fayed et al., 2020&lt;sup&gt;30&lt;/sup&gt;</td>
<td>Excelsius-GPS</td>
<td>2 attendings</td>
<td>20 pts underwent PPS placement for lumbar interbody fusion (103 screws)</td>
<td>• Screw accuracy on GR scale</td>
<td>Minor trend toward ↓ screw breach rate in 2nd half vs 1st half of screws (p = 0.151)</td>
</tr>
<tr>
<td>Jiang et al., 2020&lt;sup&gt;9&lt;/sup&gt;</td>
<td>Excelsius-GPS</td>
<td>1 attending</td>
<td>28 pts underwent 1- to 2-level lumbar fusion for degenerative disease (113 screws; 23 open cases)</td>
<td>• Screw accuracy on GR scale</td>
<td>Significant ↓ op time over 1st 30 cases; curve: time (mins) = −4.6 (case no.) + 275.5; $r^2$ = −0.26 (p = 0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Op time</td>
<td>Greatest ↓ in op time occurred in 1st 10 cases</td>
</tr>
<tr>
<td>Khan et al., 2020&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Mazor X</td>
<td>1 attending</td>
<td>22 pts underwent 1- to 2-level PLIF degenerative disease (92 screws)</td>
<td>• Screw accuracy on Ravi scale</td>
<td>Cases 1–11 associated w/ ↑ fluoro time vs cases 12–22 but difference NS</td>
</tr>
<tr>
<td>Mao et al., 2020&lt;sup&gt;25&lt;/sup&gt;</td>
<td>Mazor X</td>
<td>1 attending</td>
<td>39 pts underwent 1–4+ level thoracolumbar instrumentation for multiple indications (318 screws)</td>
<td>• Screw time learning curve: time (min) = −1.064 ln (screw no.) + 8.5814; $r^2$ = 0.1458; curve flat beyond screw 50</td>
<td>No difference in op time, radiation dose, or screw accuracy btwn groups</td>
</tr>
<tr>
<td>Avrumova et al., 2021&lt;sup&gt;13&lt;/sup&gt;</td>
<td>Mazor X</td>
<td>1 attending</td>
<td>65 pts underwent PLIF (311 screws; 53 converted to freehand)</td>
<td>• Screw time learning curve: time (min) = −1.064 ln (screw no.) + 8.5814; $r^2$ = 0.1458; curve flat beyond screw 50</td>
<td>Significant ↓ op time from cases 1 to 2; relatively flat after case 2</td>
</tr>
<tr>
<td></td>
<td>Stealth</td>
<td></td>
<td></td>
<td>• Screw time learning curve: time (min) = −1.064 ln (screw no.) + 8.5814; $r^2$ = 0.1458; curve flat beyond screw 50</td>
<td>↓ registration time 1st 10 vs last 10 cases (9.1 ± 2.0 vs 6.1 ± 2.5 min; p = 0.006)</td>
</tr>
</tbody>
</table>
tively flat beyond 50 screws placed. Chen et al.,\textsuperscript{15} however, performed a regression of operative time and robot usage time as a function of case number and reported a relative plateau after approximately 20 cases.

**Discussion**

In the present systematic review, 21 articles were identified that investigated the presence of a learning curve in spinal robotics, of which 18 provided evidence for the existence of a learning curve, ranging from 3 to 30 cases or 15 to 62 screws, depending on the endpoint examined.\textsuperscript{9,12–25,28,31,32} Importantly, of those series with the greatest experiences described—5 series in which the surgeon had used the robot in ≥ 50 cases—described very small experiences of 13 and 20 cases.\textsuperscript{30,32} Given that a significant plurality of the series documented a learning curve of 20 to 30 cases to achieve proficiency, it is likely that these series were too small to determine the presence or absence of a learning curve. Similarly, the third article tested for the presence of a learning curve using a threshold of 30 screws,\textsuperscript{32} which is far below the number of screws that would be placed in 20 single-level fusions. Consequently, the data would argue for the presence of a learning curve given that a sufficient number of cases are examined.

Despite the evidence gathered supporting the existence of a learning curve in spinal robotics, the body of evidence is of low quality, comprising single-surgeon series. A pooled analysis of prior experiences would help to generate high-level evidence and enable more powerful conclusions to be drawn. Additionally, such a pooled analysis would help to control for baseline surgeon experience, as greater familiarity with spine surgery, in general, would be

## TABLE 2. Summary of studies examining learning curves for spinal robotic systems

<table>
<thead>
<tr>
<th>Authors &amp; Year</th>
<th>System</th>
<th>Provider Level</th>
<th>Study Population</th>
<th>Outcome</th>
<th>Learning Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bydon et al., 2021\textsuperscript{14}</td>
<td>Mazor X</td>
<td>Multiple attendings</td>
<td>77 pts treated for degenerative disease; 53 MIS TLIF, 18 PLIF (5 MIS), 6 PPS fusion (402 screws)</td>
<td>• Op time</td>
<td>↓ op time (R = −0.39; p &lt; 0.001); 6-min reduction/mo</td>
</tr>
<tr>
<td>Chen et al., 2021\textsuperscript{15}</td>
<td>TiRobot</td>
<td>1 attending</td>
<td>52 pts underwent 1-level MIS TLIF (208 screws)</td>
<td>• Screw accuracy on GR scale</td>
<td>Relative plateau after ~20 cases</td>
</tr>
<tr>
<td>Lee et al., 2021\textsuperscript{16}</td>
<td>Mazor X</td>
<td>1 attending</td>
<td>51 pts underwent 1- to 2-level lumbar fusion for degenerative disease (240 screws)</td>
<td>• Op time</td>
<td>↓ op time (40 mins) after 1st 20 cases (193 ± 30 vs 153 ± 55 min; p = 0.019)</td>
</tr>
<tr>
<td>Morse et al., 2021\textsuperscript{17}</td>
<td>Mazor X</td>
<td>1 attending</td>
<td>19 pts underwent pediatric scoliosis surgery (194 screws planned; 168 placed)</td>
<td>• Pedicle screw insertion time</td>
<td>↓ screw insertion time w/ ↑ screw no. (r = −0.34; p &lt; 0.001) &amp; ↑ case no. (r = −0.34; p &lt; 0.001)</td>
</tr>
<tr>
<td>Wang et al., 2021\textsuperscript{32}</td>
<td>Mazor Renaissance</td>
<td>1 attending</td>
<td>13 pts underwent MIS TLIF (15 interbodies placed)</td>
<td>• Op time</td>
<td>Using fluoro-guided MIS TLIF as baseline, no trend in op time for robot-assisted cases; time (mins) = 0.24 (case no.) + 40.56 (r² = 0.001)</td>
</tr>
</tbody>
</table>

\textsuperscript{Fluoro = fluoroscopy; GR = Gertzbein-Robbins; ln = natural log; LOS = length of stay; MIS = minimally invasive surgery; NS = not significant; PLIF = posterior lumbar interbody fusion; PPS = percutaneous pedicle screw; pts = patients; TLIF = transforaminal lumbar interbody fusion; ↑ = increase in; ↓ decrease in.}

Mean values are often presented as mean ± SD.
expected to serve as a sounder knowledge base onto which the intricacies of robotic surgery could be overlaid. Only 2 of the identified studies examined surgeon experience or level of training as a factor in the learning curve for robotic systems. Urakov et al. saw no difference in the time of screw insertion for junior residents (postgraduate years [PGYs] 1–5) compared with senior residents (PGYs 6–7) or fellows using the Mazor Renaissance. However, the number of cases performed by each resident was quite small, and the interaction of learning curve with level of training was not reported. Similarly, Siddiqui et al. suggested that there was minimal interaction between level of training and learning curve in a three-surgeon series composed of a single attending and two postgraduate neurosurgical spine fellows using the ExcelsiusGPS. All three surgeons experienced significant improvement in screw tail placement accuracy over the first 30 screws placed. Only one of the fellows failed to show improvement in screw tip placement; however, their baseline performance was already superior to that of the attending and other fellow. The reason for this is unclear; however, it may be that all three surgeons had the minimum knowledge base required to take advantage of the robot, whereas a more junior resident might not.

Influence of Surgical Robots on Residency Curricula

Although novel for the field of neurosurgery, the issue of whether and how best to integrate surgical robotics within residency curricula is not unique. A similar issue was faced by the general surgery, obstetrics/gynecology, and urology fields in the past decade as they were forced to determine how best to integrate the increased utilization of surgical robots across their surgical disciplines. From a top-down perspective, all three fields have elected to consider robot-assisted procedures differently. Both obstetrics/gynecology and general surgery treat robot-assisted cases as a subtype of laparoscopic surgery, which had already been integrated into the case minimums outlined by the Accreditation Council for Graduate Medical Education (ACGME), the accrediting body for US residency programs. By contrast, effective as of 2021, ACGME mandates that graduating urology residents complete a minimum of 80 robot-assisted cases. The latter move has been suggested to, at least in part, been driven by the increasing volume of robot-assisted urological procedures, which comprised nearly half of all major urological surgeries during the 2016 to 2017 year. Consequently, reworking case minimums to better equip residents for the reality of clinical practice was a consideration to improve job prospects.

Penetration of surgical robotics within general surgery and obstetrics/gynecology is far lower, however, accounting for only 29% of uterine procedures and 8% of colorectal procedures between 2001 and 2015 (compared with 77% of prostate procedures). Additionally, many general surgery and obstetrics/gynecology programs do not regularly expose trainees to robotic cases. Consequently, imposing case log minimums for these procedures would place an unreasonable burden on trainees. Additionally, it has been noted that increased use of surgical robots is associated with decreased trainee participation in surgical cases and decreased surgical volume in other tracked categories.

Application of Spinal Robotics to Neurosurgery Residency Curricula

Unlike robotics in general surgery, urology, and obstetrics/gynecology, spine surgery robots are relatively new. Therefore, few programs have these devices available, thereby precluding their inclusion in ACGME case minimums. Additionally, their indications are relatively limited, as their main purpose is to facilitate the accurate placement of pedicle screw instrumentation. As a result, the proportion of neurosurgical cases in which such a device could be used comprises the minority of cases that will be seen by or require logging by neurosurgical residents. Nevertheless, formalization of a spine robot curriculum for residents pursuing specialization in spine surgery can help to ensure these residents experience their robot learning curves during residency, and, thereby, shift the poorer outcomes from the time during which they are practicing independently to one in which they are supervised by a senior surgeon adept at the use of the robot. Additionally, acquisition of these skills may increase the marketability of residency graduates, as the majority of the public perceives surgical robots to be associated with better results and fewer complications, irrespective of the available data.

Based on the evidence identified in the literature review here, we propose a curriculum in which senior-level residents (≥ PGY5) can obtain a certificate verifying competence with spine surgery robots. Under this curriculum, residents must perform a minimum of 30 surgical cases with one of the currently approved and marketed spine robots for which learning curves have been previously described (i.e., Mazor X, Mazor X Stealth, and ExcelsiusGPS). In addition, these residents should also complete any industry-sponsored training programs that are offered to surgeons who will be using these technologies, as these may highlight unique features of the varied systems, some of which may not be present at the residents’ home institutions. To participate, residents should have met their ACGME-mandated minimums for spine cases (300 in the latest ACGME guidelines). Having met these minimums helps to ensure that residents have acquired the basic anatomical knowledge necessary to verify the fidelity of robotic trajectories. It also helps to ensure that residents can troubleshoot cases in which the robot fails to register patient anatomy (up to 8% of cases) or otherwise forces screws to be placed manually after registration (up to 9% in one large series). The ability of surgeons to fall back on more conventional placement techniques (image guided, freehand, or fluoroscopy guided) is essential and ensures that robots are only used as tools to facilitate superior outcomes for spine surgeons as opposed to being employed by those who are incapable of performing the primary procedure without assistive technology.

In truth, no curriculum can ensure that surgeons are competent with robotics or any procedure for that matter, but akin to residency training in general, it can improve the probability that users are competent with the device. An alternative to the aforementioned structure would require a minimum number of screws (vs cases) be placed to achieve certification. However, some of the articles identified in this review, such as that of Chen et al., found...
that the major driver of decreased operative time was a decrease in overall robot usage time (vs screw instrumentation time alone). Consequently, the learning curve (and, therefore, assessment of a trainee’s progression along said curve) should be determined by both the number of screws placed and the number of times the trainee has gone through the workflow of setting up the robot and plotting screw trajectories. Accomplishing both occurs at the case level; therefore, it is argued that ensuring a minimum number of robotic cases performed helps to better ensure a minimum level of competency.

Limitations

The present review has several limitations, foremost of which is the low level of evidence. All included studies were level IV case series, the majority of which were single-surgeon series. Ideally, individual, patient-level data could be included in a future analysis to reach more generalizable results, as well as to investigate the relative influence of user training level and robotic platform on the learning curve. Additionally, the source data are all retrospective, and many series are nonconsecutive, meaning that included cases could have been selected for robot assistance preoperatively based on surgeon-identified factors that would facilitate a better outcome. It is impossible to know the degree to which this affected the present results, and future studies that prospectively investigate and document surgeon results with surgical robots are required. Furthermore, the evidence in the present review is drawn from both current general robots (ExcelsiusGPS, Mazor X Stealth, Mazor X, and TiRobot) and prior versions (Mazor SpineAssist and Mazor Renaissance). The latter employed more fragile platforms and less-refined imaging software to plot screw trajectories. Consequently, they may have been associated with longer learning curves as surgeons adjusted to their varying results.

Another limitation of the present review is that a diverse set of endpoints are reported to describe the learning curves associated with spine robot use. This may help to account for the variability in the identified minimum number of cases required to achieve proficiency. However, it must be noted that few studies assessed a total robot usage time, with most looking at screw placement time, screw accuracy, or screw fluoroscopy time. It has previously been noted that robot system registration time represents a nontrivial time cost associated with spine robot system utilization. Time expended with robot system positioning in the operating room, reference frame placement and registration, and reregistration of the system in the case of long-segment constructs represent a greater proportion of the total case time than does actual screw placement. Consequently, the learning curve associated with robot system usage may prove to be different when focusing on these steps. Further investigation of robotic learning curves that examines both changes in screw placement time and system setup and registration times is necessary to better appreciate the true learning curve of these devices.

Lastly, the proposed development of a certificate program for residency curricula is contingent upon residency programs having a sufficient robotic case volume. As no good data yet exist for the percentage of spine surgeries being performed robotically, or the percentage of training centers actively using robots for their spine cases, it is unclear that such a certificate program is feasible at this time. Nevertheless, given the increasing spine robot case volumes reported by market research, and the propensity for these advanced technologies to be found at large academic centers with sufficient financial infrastructure and case volume to support their use, it seems likely that in the near future a sufficient number of training programs will see volume sufficient enough to render the proposed certificate program feasible. Consequently, we advocate for earlier discussion of such curriculum changes in anticipation that they will be needed in the near future. Such discussions will also broach topics of import, such as whether a generalized curriculum is sufficient, or whether platform-specific curricula will be necessary. This decision will likely benefit from better data on the learning curves associated with spine robot workflow steps other than screw placement.

Conclusions

In the present systematic review, we found that the majority of published series demonstrate the presence of a learning curve in spinal robotics. Descriptions of the length of the curve and the metrics used to define the curve are highly heterogeneous. However, multiple studies have described that surgeons experienced continued improvement in outcomes up to a nadir of 20 to 30 cases, which coincides with our previously described experience. Based on this, we propose that an optional robotic spine surgery certificate program could be developed for senior-level residents pursuing spine specialization who have already completed the ACGME minimum requirements for spine surgery cases. However, additional investigation in prospective series using standardized outcomes (i.e., screw placement time, operative time, and instrumentation placement accuracy) on currently employed systems is required to account for potential reporting bias present in the extant series.

References


Disclosures

Dr. Theodore is the inventor of a device (ExcelsiusGPS) manufactured by Globus Medical Inc., which was used in the study discussed in this publication. He is entitled to royalty payments on future sales of the device and is a paid consultant for and direct stock owner in Globus Medical Inc. He also receives royalties from DePuy Synthes. Dr. Elder is a direct stock owner in Injectsense; consultant for Johnson & Johnson; and received research support for the study described from Stryker and SI Bone.

Most devices mentioned in this article—Mazor SpineAssist, Mazor Renaissance, Mazor X, Mazor X Stealth, and Globus ExcelsiusGPS—have been FDA approved for use in the placement of thoracolumbar spine instrumentation and are discussed only for approved indications. One device, the Tinavi TiRobot, is not currently FDA approved, but is approved for use in China and has been previously described in the peer-reviewed literature.

Author Contributions

Conception and design: Pennington, Zakaria. Acquisition of data: Pennington, Lakomkin, Mikula. Analysis and interpretation of data: Pennington, Judy, Zakaria, Lakomkin, Mikula. Drafting the article: Pennington. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Pennington.

Correspondence

Zach Pennington: Mayo Clinic, Rochester, MN. pennington.zachary@mayo.edu.