Real-world Performance of On-line Models to Predict Success after Lumbar Fusion

Leah Yacat Carreon[1], Morgan E. Brown[2], Christy Daniels, Mladen Djurasovic, Steven D Glassman[3]
[1]Norton Leatherman Spine Center, [2]Norton Healthcare, [3]Norton Leatherman Spine

INTRODUCTION:

There is a push to use modelling to predict patient outcomes after spine surgery. These predictive models were developed and tested on similar patient samples but have not been tested using a different patient sample. Recent systematic reviews identified only one easily accessible on-line model (Lumbar Fusion Calculator, LFC). "Dialogue Support" (DS) has recently been made available on the EuroSpine Website. The purpose of this study is to evaluate the real-world performance of on-line models used predict patient outcomes after lumbar fusion for spinal stenosis (LSS) or disc herniation (LDH).

METHODS:

Patients enrolled in the Quality Outcomes Database (QOD) [25-27] from a single, multi-surgeon site with complete baseline data to complete the risk assessment tools and 12-month post-operative data were identified. Although some institutions enrolled every surgical patient, our institution followed the original protocol of enrolling only six cases a week. Data on each individual patient was entered into each risk assessment tool and the probability of success was collected. As the LFC specifically states "Here are your predicted outcomes if you choose to have a lumbar fusion" only patients who had a lumbar fusion were included in the analysis. In addition, since only the diagnostic options of disc herniation and spinal stenosis are available in both LFC and DS, only patients with these two diagnoses were included in the study.

Predicted probability of a good outcome between patients who actually achieved successful outcome versus those who did not were compared using independent t-tests. A threshold p-value of 0.01 was considered statistically significant given the multiple concurrent analysis and relatively small sample size. The ability of the models to identify patients who achieve a successful outcome was determined using Receiver-Operating Characteristic Curve Analysis. Area Under the Curve, threshold values, sensitivity and specificity were calculated. An AUC of 0.5 suggests no discrimination 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding [31]. The ratio of expected to observed results which describes the overall calibration of the prediction model were also calculated, with 1.0 considered to be optimal. Binary logistic analysis was also performed to determine associations between the predicted probability of a good outcome and the achievement of success. A p-value of >0.3 was considered to indicate no lack of fit.

RESULTS:

Of 1362 cases enrolled at a single site in QOD, 957 (70%) had complete 12-month data available for analysis, 252 cases with Stenosis and 92 cases with Disc Herniation. This follow-up rate was similar to that reported for the LFC (66-70%) and better than that reported for DS (40%).

Of the 252 cases with Stenosis enrolled locally in the QOD, 188 cases had a concomitant fusion with a mean age of 61.0 ± 11.7 years, mean BMI of 31.4 ± 6.5 kg/m^2 and mean 1.6 ± 0.8 surgical levels fused. There were 97 males, 42 smokers, 66 had private insurance, 21 were revision cases, and most (135, 72%) were ASA grade 3. The age distribution were similar across the current study, the LFC and DS cohorts. The sample in the current study has more smokers, less revisions and worse baseline ODI scores compared to the development samples of the LFC and DS. There were no statistically significant difference ($p>0.01$) in the predicted probabilities of a good outcome between the patients who actually achieved success and those who did not, except for having a 12-month ODI ≤ 22 ($p<0.000$), 12-month BP ≤3 ($p<0.000$) and patient satisfaction ($p=0.003$).

ROC analysis (Figure 2) showed that the LFC model had an acceptable ability to predict the probability of a patient achieving a 12-month ODI ≤ 22 (AUC=0.855) and 12-month BP ≤ 3 (AUC=0.700). However, only the model's ability to predict BP ≤ 3 had a Hosmer-Lemeshow p-value greater than 0.200, denoting an acceptable fit. The expected to observed ratios in the current study (0.52 – 0.77) are smaller than that reported in the original article describing LFC which had expected to observed ratios of 0.92 to 1.02.

Of the 92 cases with Disc Herniation enrolled locally in the database, 46 cases had a concomitant fusion with a mean age of 48.0 ± 14.5 years, mean BMI of 31.2 ± 6.02kg/m^2 and mean 1.2 ± 0.5 surgical levels fused. There were 25 males, 13 smokers, 2 had private insurance, 22 were revision cases and around half (24, 52%) were ASA 3. The age and sex distribution were similar between the current study cohort and the Dialogue Support cohort. There were more smokers and more revision patients in the current study compared to the DS cohort. Except for BP Change ≥ 2, there were no statistically significant difference ($p>0.01$) in the predicted probabilities of a good outcome between the patients who actually achieved a successful outcome and those who did not (Table 5).

ROC analysis (Figure 4) showed that the LFC model had an acceptable ability to predict the probability of a patient achieving 12-month ODI ≤ 22 (AUC=0.857), BP change ≥ 2 (AUC=0.792) and 12-month BP ≤ 3 (AUC=0.735). Dialogue

Support had an ROC of 0.940 to predict a patient having 12-month LP=0. These four measures also had Hosmer-Lemeshow p-values greater than 0.200, denoting an acceptable fit.

DISCUSSION AND CONCLUSION: Existing on-line models to predict success after lumbar fusion may overestimate the probability of success in patients undergoing lumbar fusion for disc herniation or stenosis. The Lumbar Fusion Calculator showed acceptable identification of lumbar spinal stenosis patients who will have minimal back pain and lumbar disc herniation patients who will notice an improvement in their back pain one year after fusion surgery. Future studies to identify ways to improve the accuracy and reliability of these models are needed.