One-year outcomes of early-crossover patients in a cohort receiving nonoperative care for lumbar disc herniation

Anurekha Ramakrishnan, MS, K. Michael Webb, MD, and Matthew C. Cowperthwaite, PhD

OBJECTIVE The authors comprehensively studied the recovery course and 1-year outcomes of early-crossover patients who were randomized to the nonoperative care arm of the Leiden–The Hague Spine Intervention Prognostic Study. The primary goal was to gain insight into the differences in the recovery patterns of early-crossover patients and those treated nonoperatively; secondary goals were to identify predictors of good 1-year outcomes, and to understand when and why patients were likely to cross over.

METHODS Individual EuroQol-5D scores were obtained at baseline and at 2, 4, 8, 12, 26, 38, and 52 weeks for 142 patients. Early-crossover patients were defined as those electing to undergo surgery during the first 12 weeks of treatment. Crossover and noncrossover groups were compared using Kruskal-Wallis, Wilcoxon-Mann-Whitney, and chi-square tests. Linear mixed-effects models were used to examine the growth trajectories of crossover and noncrossover groups. Recursive partitioning trees were used to model crossover events and the timing of crossover decisions. Multivariable logistic regression models were used to identify predictors of good 1-year outcomes.

RESULTS Of the 142 patients randomized to receive prolonged nonoperative care, 136 were selected for the study. In this cohort, 43/136 (32%) opted for surgery, and 31/43 (72%) of crossover events occurred before the 12-week time point. Early-crossover patients had significantly greater functional impairment at Week 2 than noncrossover patients (p = 0.031), but experienced greater recovery by 26 weeks and better 1-year outcomes (p = 0.045). Patients who did not experience an improvement in their symptoms between 2 and 8 weeks were more likely to cross over (OR 3.5, 95% CI 1.2–10.1; p = 0.01). Recursive partitioning trees were able to identify crossover patients with 76% accuracy. Regression models suggested that better recovery at 26 weeks (p < 0.01) was predictive of good 1-year outcome; declining health status between Weeks 4 and 8 was negatively predictive of good outcome (p < 0.01).

CONCLUSIONS This study is the first to comprehensively analyze the recovery and outcomes of crossover patients, and compare them to nonoperatively treated patients. The results suggest that patients who have a low EuroQol-5D score during the early weeks of treatment and who do not respond to nonoperative care during the first few weeks of treatment are most likely to cross over. Early-crossover patients experience a greater rate of recovery and more frequently have a good 1-year outcome when compared with nonoperatively treated patients. The current results motivate a broader investigation into the timing of surgery and the identification of patient populations that will be most benefited by early surgical treatment for lumbar disc herniation.

https://thejns.org/doi/abs/10.3171/2017.2.SPINE16760

KEY WORDS early crossovers; outcomes; statistics; lumbar disc herniation; comparative effectiveness; degenerative

Lumbar disc herniation is one of the most common causes of low-back pain and radiculopathy. Treatment for patients with a herniated lumbar disc usually begins with nonoperative care such as analgesics, epidural steroid injections, and physical therapy. Surgery is usually reserved for patients with severe nerve root or cauda equina dysfunction, or if nonoperative therapy is unsuccessful in controlling the symptoms. Many studies have compared the effectiveness of nonoperative and surgical treatment protocols for treating herniated lumbar discs. However, interpretation of these studies is often challenged by high crossover rates...
(20%–50%), which confound the commonly used intent-to-treat (ITT) analysis approach. As a result, these studies have arrived at varying conclusions and, moreover, what has emerged as “fact” to the general public is that long-term outcomes (i.e., recovery) are equivalent for nonoperative and surgical treatment. We previously studied patients within a nonoperative cohort and showed that nearly half of the nonoperatively treated patients did not make a complete recovery after 1 year of nonoperative treatment.6 Thus, the following dilemma arises: although surgery should be avoided whenever a full recovery is reasonably expected to occur, when nonoperative treatment fails, patients must decide between tolerating a poor quality of life (QOL) for an extended and indefinite time period or opting for surgery that is associated with its own risks and uncertainties regarding outcomes. Medical professionals who care for these patients are similarly challenged because, in the absence of severe nerve root dysfunction, there is no clear guidance on whether and when to consider surgery as a treatment option to improve pain- and mobility-related QOL in their patients.

We present an analysis of the crossover subjects among a cohort of patients randomized to the prolonged nonoperative treatment arm of the Leiden–The Hague Spine Intervention Prognostic Study. Our goals were to understand the characteristics of crossover patients and to determine the patterns in their recovery that are associated with opting out of prolonged conservative therapy.

**Methods**

**Patient Cohort**

The cohort contained 142 patients randomized to a protocol of prolonged nonoperative care as part of the Leiden–The Hague Spine Intervention Prognostic Study. Eligible patients were those 18–65 years of age with a radiographically confirmed disc herniation and lumbosacral radicular symptoms starting 6–12 weeks prior; the cohort’s demographics and baseline characteristics have been previously described.9,10,13 Conservative-care management was supervised by each patient’s general practitioner and included analgesics and physical therapy as determined by the treating physician.

The EuroQol–5 Dimensions (EQ-5D) instrument was used to measure a patient’s health status (utility) at baseline and at 2, 4, 8, 12, 26, 38, and 52 weeks after enrollment into the study; the average duration of sciatica prior to enrollment was 9.5 weeks. Utility is measured on a scale anchored at 0 (as bad as dead) and 1 (perfect health). To estimate utility, the EQ-5D assesses a patient’s functional impairment in 5 domains: mobility, self-care, usual activities, pain, and anxiety. For each domain, patients self-report the scores of 1 (no problems), 2 (some problems), or 3 (extreme problems). The EQ-5D scores were computed using the US valuation model,11 which can clearly distinguish subjects reporting no health problems (EQ-5D = 1) from those reporting at least some health problems (EQ-5D ≤ 0.86). Our results are independent of the particular valuation model (not shown). Follow-up completeness of EQ-5D measures ranged from 98% at 2 weeks to 90% at 38 weeks.

**Statistical Analysis**

All statistical analyses were conducted in Microsoft R Open 3.2 (http://mran.microsoft.com/open). Continuous variables are presented as the mean ± 1 SD, and were compared using 2-tailed Student t-tests or Wilcoxon-Mann-Whitney tests. Categorical variables are presented as contingency tables, and proportions were compared between groups using chi-square and Fisher exact tests. Significance was assessed at the α < 0.05 level, unless otherwise indicated.

Because the data were obtained from a longitudinal study, a linear mixed-effects model was used to study the growth trajectories of crossover and nonoperatively treated patients. The change in EQ-5D score starting from Week 2 in the 2 groups was modeled using the baseline EQ-5D score, observation week (Week 2, 4, 8, 12, 26, 38, or 56), and treatment (surgery or nonoperative) as predictors. The model was fitted using the “nlme” package.

Decision tree and random forest models were used to model crossover events and to study which patients were likely to cross over and when. These models included EQ-5D scores up to 12 weeks as predictors, with the response variable being a binary indicator (surgery or no surgery) for the crossover event; no additional clinical or demographic covariates were included in the models. Decision tree models were fitted using the “part” package. Precision (positive predictive value) and recall (sensitivity) rates were used to assess the performance of the model, and the precision-recall area under the receiver-operating characteristic curve (AUC) was calculated using the “PRROC” package. Random forest models are a powerful alternative to decision trees that often give improved predictive accuracy on relatively complex multivariate decision problems. Random forest models were fitted using the “randomForest” package. The mean decrease in the Gini impurity statistic, a common measure used by Classification and Regression Tree (CART) algorithms to determine the best way to split the records, was used to assess the importance of each predictor variable.

Multivariable logistic regression models were used to determine patient factors associated with good outcomes (complete recovery) at the 1-year observation point. The dependent variable was 1-year outcome (full recovery, defined as EQ-5D = 1, or incomplete recovery, defined as EQ-5D ≤ 0.86), and independent variables included demographics (age, sex, marital status, and so on), patient history (education, days of missed work at baseline, and so on), and recovery course (current EQ-5D, periodic changes in EQ-5D, and so on). Crossover status was included in the models and was constrained to be included in the final model to control for this important variable of interest. Predictor variables were selected for inclusion into the models by using a 2-step process: 1) chi-square tests of association were used to independently assess the association of predictor variables with outcomes; and 2) bidirectional stepwise selection of predictors with a Bayesian Information Criterion (BIC) was used to select the final set of terms to be included in the model. The prediction accuracy of the model was assessed using the AUC, and prediction consistency was assessed using leave-one-out cross-validation to compute a mean-squared error (MSE) estimate of prediction error.
Results

Patient Data Set

The data set contained 142 patients randomized to a protocol of prolonged nonoperative care as part of the Leiden–The Hague Spine Intervention Prognostic Study. Of those, 136 patients were used in the present study; 5 patients were removed because most (> 50%) of their EQ-5D scores were missing, and 1 patient who had 2 surgery events during the observation period was also removed. The EQ-5D measures from baseline to 1 year were included in subsequent analyses; the completeness of follow-up measurements ranged from 90% to 98% at the follow-up time points. In this cohort, 43/136 (32%) patients elected to have surgery before the end of the 1st year; 31/43 (72%) of that subgroup of patients elected to have surgery on or before the 12-week time point and are hereafter referred to as “early-crossover patients.” We considered the 12 (28%) patients who had surgery after Week 12 as “late-crossover patients.” However, there were too few patients in the late-crossover group to have sufficient statistical power to compare this subgroup with early-crossover or nonoperatively treated patients (discussed below). Table 1 summarizes the baseline demographics of the study cohort.

Recovery Patterns

Studies comparing operative and nonoperative treatment for lumbar disc disease (LDD) have typically used an ITT data analysis approach, in which the cohorts are analyzed according to their initial randomization. In the ITT model, crossover patients would be included with their original randomization cohort; hence, crossover patients would be treated as if they had undergone nonoperative treatment for the duration of the study period. We therefore sought to compare the subcohort of early-crossover patients to the subcohort of nonoperatively treated (i.e., noncrossover) patients to determine if their recovery patterns differed significantly.

Early-crossover patients had significantly lower EQ-5D scores during the early weeks of treatment. We used multilevel models to estimate the EQ-5D scores of the 2 subcohorts at each observation week. The average score at Week 2 for nonoperatively treated patients was 0.697 and the average score of crossover patients was 0.615. The EQ-5D scores varied by an average of 0.06 ± 0.0016 (SD) between the crossover and noncrossover patients during the first 12 weeks of treatment (multilevel model, p = 0.031). Figure 1 shows the mean EQ-5D scores of the 2 groups at each observation week, and the trajectories of late-crossover patients. Despite greater functional impairment during the early weeks of treatment, the early-crossover patients experienced higher rates of recovery. The average rate of recovery for a nonoperatively treated patient was 0.004 ± 0.0004 (SD) at each week, and the average rate for early-crossover patients was 0.006 ± 0.0012 (SD) at each week (multilevel model, p = 0.009). Additionally, early-crossover patients experienced 3 times faster recovery from Week 8 to Week 26: the EQ-5D rate of change was 0.006 for the nonoperative group versus 0.019 for the early-crossover group (Student t-test, p < 0.01). This was associated with a greater recovery at the 6-month observation period (0.81 for nonoperatively treated vs 0.86 for crossover patients, p > 0.1). Over the first 6 months, nonoperatively treated patients recovered by an average of 103% of their baseline status, whereas crossover patients recovered an average of 157% (Fig. 2).

| TABLE 1. Demographic data and baseline descriptors of the study cohort |
|---------------------------------|-----------------|----------------|
| **Baseline Descriptor**         | **Early Crossover** | **Nonoperative** | **Late Crossover** |
| Age in yrs                      | 42.5 (40)        | 42.5 (42)       | 46.1 (47)         |
| Sex, % male                     | 71.8             | 68.0            | 75                |
| Race, % Caucasian               | 93.8             | 94.3            | 100               |
| BMI                             | 26.0 (26.6)      | 25.5 (25.1)     | 27.27 (28)        |
| No. of children                 | 1.7 (2)          | 1.5 (2)         | 2 (2)             |
| Domestic partner, %             | 84.3             | 81.9            | 91.6              |
| Live w/ partner, %              | 75.0             | 79.0            | 91.6              |
| Employment, % FT                | 75.0             | 70.5            | 83.3              |
| Higher education, %*            | 84.4             | 95.2            | 91.6              |
| Missed work, % >4 wks           | 37.5             | 41.9            | 25                |
| Disability support, %†          | 68.8             | 64.8            | 58.3              |
| Work type, % moderate/heavy     | 43.8             | 41.9            | 58.3              |
| Intellectually demanding, % high| 62.5             | 65.7            | 66.7              |
| Physically demanding, % high    | 25.0             | 37.1            | 50                |
| EQ-5D score                     | 0.507 (0.562)    | 0.567 (0.689)   | 0.539 (0.608)     |

BMI = body mass index; FT = full time.
The subcohorts did not differ significantly at baseline with respect to any demographic category. Values are expressed as the mean (with median in parentheses), or as percentages.
* Dutch education system; percent with secondary or higher education (college, trade school, etc.).
† Percent with any disability support.
Early surgery was associated with a high probability of a “good” outcome, which we previously defined as no reported health problems (EQ-5D = 1) after 1 year (Fig. 3). The probability of a good outcome (Fig. 3, gray bars) appeared to be associated with surgical timing. The lowest probability (55.6%) of a good outcome occurred with surgery at 4 weeks, whereas the highest probability (71.4%) of a good outcome occurred with surgery at 12 weeks. The average EQ-5D scores of patients at crossover time (Fig. 3, dashed line) was not strongly associated with the likelihood of a good 1-year outcome. Most crossovers occurred early during nonoperative treatment (Fig. 3, solid line), which probably corresponds to patients with the highest levels of discomfort.

Decision tree and random forest models were used to predict surgery events and to identify the timing of crossover decisions. Patients whose health declined between 2 weeks and 8 weeks were more likely to cross over (OR 3.5, 95% CI 1.2–10.1; Fisher exact test, p = 0.01). Figure 4 shows the most frequent rules used to generate trees in the random forest algorithm. The model selection was done using a 3-fold cross-validation and the precision-recall AUC was 76%.

**Posttreatment Outcomes**

We previously showed that nonoperatively treated patients within this cohort segregated into 2 groups: a “low-utility” group in which patients reported some health problems at 1 year, and a “high-utility” group in which patients reported no health problems at 1 year. We therefore studied 1-year outcomes to determine whether early-crossover patients experienced better long-term outcomes relative to the nonoperatively treated patients within the cohort.

At the 6-month time point, the proportion of nonoperatively treated (16; 28%) and early-crossover (12; 38%) patients in the high-utility group was nearly equivalent (Fisher exact test, p > 0.10). However, at 1 year, early-crossover patients (20/32; 63%) were 2.6 times more likely to completely recover than nonoperatively treated patients (22/57; 39%) at the end of the 1-year study period (Fisher exact test, p < 0.05). Four (7%) nonoperatively treated patients reported complete recovery at 6 months, but then reported health problems at 1 year; in contrast, no early-crossover patients followed this pattern.

Multivariable logistic regression models were used to identify demographic and clinical variables that were predictive of good outcomes at 1 year. Just over half of the candidate predictor variables (29/57; 50.8%) were not independently associated with 1-year outcomes. The step-wise model selection process selected 3 variables from the remaining 28 candidate predictors for inclusion in the final model: 26-week utility score, crossover status, and Week 4–8 decline in utility score. The subject’s EQ-5D score at 26 weeks was positively predictive of good outcomes ($\beta = 9.2$, p < 0.01), whereas a decline between Weeks 4 and 8 ($\beta = -2.3$, p < 0.01) was negatively predictive of a good outcome. Crossover status ($\beta = 0.6$, p > 0.1) was not significantly predictive of a good 1-year outcome. The final model was both predictive of good outcomes (AUC 0.82) and reasonably robust (MSE 0.185).

**Late-Crossover Patients**

We investigated whether late-crossover patients followed a different recovery pattern and outcomes than early-crossover patients. Given the small number (n = 12) of late crossovers, the analysis was hindered, but we note the following observations. We did not observe any significant differences in the baseline EQ-5D scores or the recovery rates between early- and late-crossover patients, or between
late-crossover and nonoperatively treated patients. Figure 1 shows the mean utility score of late-crossover along with early-crossover and nonoperatively treated patients at each observation point, and suggests that late-crossover patients follow a very similar trajectory to nonoperatively treated ones. Early-crossover patients were more likely to experience complete recovery at 52 weeks than late-crossover ones (Fisher exact test, \( p = 0.066 \)); however, this result is confounded by the shorter follow-up time of the late-crossover patients.

**Discussion**

Several studies have sought to compare the effectiveness of operative and nonoperative treatment for the care of LDD.\(^1\)\(^,\)\(^2\)\(^,\)\(^8\)\(^,\)\(^13\)\(^,\)\(^16\) However, interpretation of these studies is often challenged by high crossover rates (20%–50%) that confound the use of the ITT analysis approaches used, and additionally, the outcomes of crossover patients have not been separately analyzed. In an ITT analysis, crossover patients are statistically treated as belonging to their original randomization cohort; for example, early-surgery crossover patients would be treated as if they had gone through the complete course of nonoperative therapy. With positive clinical outcomes of early-surgery crossover patients and generally high crossover rates, ITT analysis could potentially bias the interpretation of the outcomes of nonoperative treatment of LDD toward the outcomes of the surgery crossover patients.

The present study provides, to our knowledge, the first comprehensive look at the recovery and outcomes of crossover patients, and compares them to nonoperative patients. The following observations are particularly interesting. First, early-crossover patients had lower utility scores than nonoperative patients over the first 12 weeks of observation. The lower utility scores were probably the result of incomplete recovery during this time frame, although we could not conclusively confirm this prediction in our relatively small cohort. For clinicians, this should make sense; patients who are reporting particularly severe pain or other impairments during an office visit are more frequently inclined toward surgery because of the severity of their pain.

However, despite lower baseline utility scores, early-crossover patients reported significantly higher utility scores than nonoperative patients at the 12-week observation point. Additionally, the recovery of early-crossover patients appeared more stable (no relapses to health problems) than in nonoperative patients. Although far from a conclusive point, it is interesting that neither postoperative discomfort nor postoperative complications had a significant effect on the reported QOL after surgery, suggesting that more studies are needed to determine the effect of surgery on the reported QOL during the recovery period.

We also studied the rate at which individual patients recovered their utility status. Our previous work suggested that early recovery was strongly associated with better 1-year outcomes in patients treated with nonoperative care.\(^6\) We found that, over the first 6 months of treatment, early-crossover patients reported improvement that

![FIG. 3. Graph showing the proportion of crossover events (solid line) and average EQ-5D score at crossover time (dashed line) at each time point in comparison with the probability (Prob.) of a good outcome (gray bars). Good outcome is defined as reporting no health problems (EQ-5D = 1) at the 1-year observation point.](image)

![FIG. 4. Chart showing the most frequently used rules by recursive partitioning tree algorithms to identify crossover events. Patients who experience decline in utility scores from 2 to 4 weeks and from 4 to 8 weeks were more likely to cross over.](image)
was almost twice that of nonoperative patients, which may explain why early-crossover patients were able to sustain higher utility values over the last 6 months of treatment.

A limitation of this study lies in the use of the EQ-5D as the primary outcome measure. Although the EQ-5D is valid, and is able to consistently measure differences in QOL, it does not specifically report utility values related to low-back pain and radiculopathy. We have assumed that changes and differences in utility scores are due to changes in the symptoms of LDD, but there remains the possibility that factors other than LDD could be responsible for the changes in utility scores observed.

Ultimately, this study provides a starting point for quantitatively evaluating what surgeons who deal with lumbar disc herniation have intuitively grasped. Patients with a very low initial utility score and patients who desire a quicker recovery tend to favor surgery and generally have good outcomes associated with it. Patients with a higher utility score or for whom time is not a factor should continue with nonoperative treatment. A more complete study is ultimately needed to determine which of these groups, if any, do better over the long term.

Conclusions

In a cohort of 142 patients receiving prolonged nonoperative care, we found that 32% opted for surgery and that most (72%) of the crossover decisions occurred before the 12-week time point. Our findings suggest that patients who report low initial utility status and who do not respond to nonoperative care during the first 8 weeks of treatment are most likely to opt for early surgery. Although utility status was initially low, crossover patients reported greater recovery by 6 months and more often experienced complete recovery by 1 year. Our study motivates a broader investigation into the characteristics and patterns associated with early-surgery decisions.

Acknowledgments

This work was supported in part by a charitable grant from the St. David's Foundation Impact Fund. The authors are solely responsible for the results presented. We thank Dr. W. B. Van Den Hout for sharing the data.

References

5. Breiman L, Friedman JH, Olshen RA, Stone CJ: Classifica-

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Acquisition of data: Cowperthwaite. Analysis and interpretation of data: Ramakrishnan. Drafting the article: Cowperthwaite, Ramakrishnan. Critically revising the article: Webb. Reviewed submitted version of manuscript: Ramakrishnan. Statistical analysis: Cowperthwaite, Ramakrishnan. Study supervision: Cowperthwaite.

Correspondence

Matthew C. Cowperthwaite, St. David’s Medical Center, 1015 East 32nd St., Ste. 414, Austin, TX 78705. email: matthew.cowperthwaite@stdavids.com.