The impact of early recovery on long-term outcomes in a cohort of patients undergoing prolonged nonoperative treatment for lumbar disc herniation

Clinical article

MATTHEW C. COWPERTHWAITE, PH.D.,1,2 WILBERT B. VAN DEN HOUT, PH.D.,3 AND K. MICHAEL WEBB, M.D.1

1NeuroTexas Institute, St. David’s Healthcare, Austin; 2Center for Systems and Synthetic Biology, The University of Texas at Austin, Texas; and 3Department of Medical Decision Making, Leiden University Medical Center, Leiden, The Netherlands

Object. The authors comprehensively studied the recovery of individual patients undergoing treatment for lumbar disc herniation. The primary goal was to gain insight into the variability of individual patient utility scores within a treatment cohort. The secondary goal was to determine how the rates and variability of patient recovery over time, represented by improvement in utility scores, affected long-term patient outcomes.

Methods. EuroQol Group–5 Dimension (EQ-5D) scores were obtained at baseline and at 2, 4, 8, 12, 26, 38, and 52 weeks for 93 patients treated under a prolonged conservative care protocol for lumbar disc herniation. Gaussian kernel densities were used to estimate the distribution of utility scores at each time point. Logistic regression and multistate Markov models were used to characterize individual patient improvement over time. Fisher exact tests were used to compare the distribution of EQ-5D domain scores.

Results. The distribution of utility scores was bimodal at 1 year and effectively sorted patients into a “higher” utility group (EQ-5D = 1; 43% of cohort) and a “lower” utility group (EQ-5D ≤ 0.86; 57% of cohort). Fisher exact tests revealed that pain/discomfort, mobility, and usual activities significantly differed between the 2 utility groups (p << 0.001). The utility groups emerged at 8 weeks and were stable for the remainder of the treatment period. Using utility scores from 8 weeks, regression models predicted 1-year outcomes with 62% accuracy.

Conclusions. This study is the first to comprehensively consider the utility recovery of individual patients within a treatment cohort for lumbar disc herniation. The results suggest that most utility is recovered during the early treatment period. Moreover, the findings suggest that initial improvement is critical to a patient’s long-term outcome: patients who do not experience significant initial recovery appear unlikely to do so at a later time under the same treatment protocol.

Key words • outcome • statistics • lumbar disc herniation • comparative effectiveness • evidence-based medicine • EuroQol Group–5 Dimension

An Abbreviations used in this paper: AUC = area under a receiver-operating curve; EQ-5D = EuroQol Group–5 Dimension.

Lumbar disc herniation is one of the most common causes of low-back pain and radiculopathy.1 Treatment for patients with a herniated lumbar disc usually begins with conservative care such as analgesics, epidural steroid injections, and physical therapy,1,3 with surgery reserved for patients with severe nerve root or cauda equina dysfunction or if conservative therapy is unsuccessful in controlling the symptoms.

Several recent studies have compared the effectiveness of conservative care and surgical treatment protocols for treating herniated lumbar discs, and have arrived at varying conclusions.2,3,9,10,15–18 However, these studies have generally considered outcomes over a period of years, which is a significant length of time for patients who are waiting for their quality of life to improve. In clinical practice, this often leads to the following dilemma: most patients, particularly those with moderate symptoms, would prefer to avoid surgery, but are unwilling to wait an indefinite period of time for their symptoms to resolve. Unsurprisingly, lumbar discectomy is the most frequently performed surgical procedure in the US.17,18 Moreover, the above-mentioned studies have typically compared the average difference between treatment
groups, without regard for individual recovery within the cohort. Additionally, this approach assumes that recovery in the protocols being compared proceeded similarly between observation intervals. To better understand the treatment responses of individual patients and the time frames of their responses, we comprehensively analyzed a cohort of patients undergoing a prolonged conservative care treatment protocol to gain insight into the dynamics of individual patient recovery over time, and whether these recovery dynamics influence long-term outcomes.

Methods

Study Data Set

The data set contained 142 patients randomized to a protocol of prolonged conservative care as part of the Leiden–The Hague Spine Intervention Prognostic Study.10,12 The Sciatica Trial was reviewed and approved by the Medical Ethics Committee of Leiden University Medical Center.13 Patients were enrolled into the Sciatica Trial entirely in the Netherlands.

In the Sciatica Trial, all patients aged 18 to 65 years, with persistent radicular pain in the L-4, L-5, or S-1 dermatome (with or without mild neurological deficit), severe disabling leg pain (lumbosacral radicular syndrome) lasting 6–12 weeks, and radiologically (MRI) confirmed disc herniation were considered eligible to enroll in the trial. Cauda equina syndrome or severe paresis, prior complaints of lumbosacral radicular syndrome in the previous 12 months, history of same-level unilateral disc surgery, spinal canal stenosis, and degenerative or lytic spondylolisthesis were all exclusion criteria. Cohort demographics and baseline characteristics were previously described; all patients reported both back and leg pain, but leg pain was generally more severe (mean leg pain 67.2 ± 27.7 vs back pain 33.8 ± 29.6, measured on a 100-point, horizontal visual analog scale).15

The Sciatica Trial used a pragmatic study design: conservative-care management was influenced as little as possible and was supervised by each patient’s general practitioner. Use of analgesics and physical therapy was determined by the treating physician. In this cohort, 46 patients (32%) elected to have surgery before the end of the 1st year; the mean timing of surgery was 12.6 weeks after the start of treatment. The surgical patients and 3 crossover patients will be discussed in a separate study (data not shown).

To determine whether specific EQ-5D domains differed between utility groups, Fisher-exact tests were conducted on contingency tables of the number of patients in each utility group that reported scores of 1, 2, or 3. Significance was assessed using a Bonferroni-corrected p value of 0.01.

Two-state, continuous-time Markov models were used to study the patterns and probabilities of patients transitioning between a “lower” utility (EQ-5D ≤ 0.86) and a “higher” utility group (EQ-5D = 1). The threshold utility value defining the groups remained fixed over time and was used to assign each patient to a utility group at each observation time. The models were fitted using the “msm” package14 with piecewise-constant transition intensity matrices (Q) estimated for each time interval between the points t = 0, 4, 8, 12, 26, 38, 52 (t = 2 was omitted because there were insufficient transitions to yield a robust model). Transition intensities were permitted to change between subsequent observation intervals, but remained homogeneous within each observation interval. The starting transition intensities were based on the observed frequencies of transitions in the data set and were calculated using the formula

\[ q_{ij} = n_{ij} / (n_i T) \]

in which \( n_{ij} \) is the observed number of transitions from Group i to Group j over the duration of the study period (T), and \( n_i \) is the initial number of patients in Group j. The fitted models were robust to the choice of starting transition intensities and yielded qualitatively similar pa-

Statistical Analysis

All statistical analyses were conducted using the R statistical environment (version 2.9.2; http://www.r-project.org/) with the additional “msm,”14 “ROCR,”14 and “rms” packages (all freely available from http://cran.r-project.org). Continuous variables are presented as means (± SEM) and were compared using 2-tailed Student t-tests. Significance was assessed at an \( \alpha \leq 0.05 \) significance level, unless otherwise indicated. Missing EQ-5D measures were imputed using the mean of the measures at adjacent time points; our results are qualitatively similar under forward or backward imputation schemes (not shown).

Gaussian kernel density estimates were computed to estimate the distribution of utility scores. The kernel density estimates were estimated using a Silverman’s “rule-of-thumb” bandwidth and a Gaussian smoothing kernel.13 The left- and right-most points were set to the theoretical minimum and maximum EQ-5D values, respectively, so that the area under the density curve summed to 1.

The fitted models were robust to the choice of starting transition intensities and yielded qualitatively similar pa-
parameter estimates over a range of starting parameters (not shown). The likelihood function was maximized using a Nelder-Mead algorithm, and convergence was visually verified and typically occurred well short of the maximum number of iterations.

Logistic regression models were used to test whether utility measurements from earlier time points could predict long-term outcomes. These models only included utility values up to a particular time point as predictors, with the response variable being the patient’s 1-year outcome (higher or lower utility group) modeled as a dichotomous variable; no additional clinical or demographic covariates were included in the models. The models were fitted using the "rms" package7 and the fit was assessed using chi-square tests (\( \alpha \leq 0.05 \)). Separate regression models were created for all utility measurements up to and including those for 2, 4, 8, 12, and 26 weeks; for example, the 8-week model would include utility measurements at 0, 2, 4, and 8 weeks. The AUC statistic was used to assess the performance of the models and was calculated using the ROCR package.14

Results

Delineation of Higher and Lower Utility Groups

The distributions of patient utility scores markedly changed over the course of 1 year of conservative care (Fig. 1). At baseline, the majority of patients reported a relatively poor quality of life; the mean EQ-5D score was 0.55 (median 0.60). Two distinct utility groups were found to be present at baseline: a “lower” utility group (EQ-5D \( \leq 0.86 \)) and a “higher” utility group (EQ-5D = 1). At 6 months, the lower utility group (n = 62, 67%) was larger than the higher utility group (n = 31, 33%); at 1 year, the lower utility group (n = 53, 57%) had declined, but remained larger than the higher utility group (n = 40, 43%).

EQ-5D Domain Scores Between Groups

The average scores in each domain of the EQ-5D (Table 1) suggested that the pain/discomfort (low score = 1.9, high score = 1.0), mobility (low score = 1.4, high score = 1.0), and usual activities (low score = 1.5, high score = 1.0) domains differed most significantly between the high and low utility groups (p << 0.001). The anxiety (low score = 1.2, high score = 1.0) and self-care (low score = 1.1, high score = 1.0) domains differed much less between the 2 utility groups, although they were also significant (p < 0.01).

Trajectory of Patient Utility Over Time

The series of patient utility scores measured over the study period are referred to as utility “trajectories,” which were studied to understand how patients recovered over the study period. In the study cohort, all patients experienced improvement during at least 1 observation period; only 19.3% (n = 18) never experienced a decline during their recovery. Recovery was variable: 49.5% of the patients (n = 46) experienced at least 2 reversals, which were defined as improvements (declines) immediately followed by declines (improvements) at the next observation. Furthermore, only 29% of patients (n = 27) had stable trajectories with no reversals. Overall, increases in utility were 4 times more common than decreases in utility.

The utility of the entire cohort increased by 0.296 (51.8% above baseline; p << 0.001, Wilcoxon Mann-Whitney test) over the year (Fig. 2), but was markedly faster during the first 2 months (0.022/week) than the final 3 months (0.005/week). Over the same time frames, utility scores improved by 0.178 (35.2% above the baseline average) over the first 2 months and by 0.063 (1.3% above the
9-month average) during the final 3 months. The mean utility scores significantly differed between the 2 final utility groups at 8 weeks and remained significant for the rest of the year (p < 0.01, Student t-test; Fig. 2).

Modeling Patient Recovery

Given that 2 utility groups were present over the study period, Markov models were used to study the robustness of these groups by estimating the likelihood of patients switching between the groups. The models suggested that the average probability of a patient remaining within their utility group was 97.9% and 97.6% for patients currently in the low and high utility groups, respectively (Fig. 3). The probability of a patient transitioning from the low to the high utility group was 2.1%; the corresponding probability for transitions from the high to the low utility group was 2.3%.

The models also suggested that the likelihood of a patient transitioning to another utility group declined over the study period. During the first 8 weeks, 2.8% and 3.5% of patients experienced low-to-high and high-to-low transitions, respectively; over the last 3 months, 1.6% and 1.3% of patients experienced low-to-high and high-to-low group transitions, respectively.

Predicting Individual Patient Outcome

At 8 weeks, logistic regression models could predict a patient’s outcome (final utility group) with modest accuracy (AUC = 0.62, or 62%). The accuracy of the models steadily increased as data from later time points were included; the 26-week model performance was good with an AUC of 0.78 (Fig. 4). The amount of improvement in utility scores from baseline to 8 weeks was also investigated as a predictor of good outcome (higher utility group). Patients with EQ-5D scores that improved by at least 0.30 during the first 8 weeks of treatment were 60% more likely to have a good outcome.

![Graph of mean patient utilities at each measurement time point. Error bars represent 95% CIs about the mean. High and low utility group refers to the final group in which the patient belongs at the 1-year time point.](image)

![Graphs of the Markov transition probabilities (per week) for transitions within (lower) and between (upper) utility groups. Each point is centered at the middle of each time interval and represents the maximum-likelihood estimate of the per-week transition probability during the entire interval. Error bars (mean width of the 95% CI was 1.8) were omitted for clarity because the differences were not significant.](image)

![Graph of the Markov transition probabilities (per week) for transitions within (lower) and between (upper) utility groups. Each point is centered at the middle of each time interval and represents the maximum-likelihood estimate of the per-week transition probability during the entire interval. Error bars (mean width of the 95% CI was 1.8) were omitted for clarity because the differences were not significant.](image)

### TABLE 1: Distribution of scores in each EQ-5D domain at 1 year

<table>
<thead>
<tr>
<th>Score</th>
<th>Mobility (EQ1)</th>
<th>Self-Care (EQ2)</th>
<th>Usual Activity (EQ3)</th>
<th>Pain/Discomfort (EQ4)</th>
<th>Anxiety/Depression (EQ5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>40</td>
<td>42</td>
<td>40</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

*p value* <<0.001 <0.01 <<0.001 <<0.001 <0.01

* Significance was assessed using chi-square tests with Bonferroni-adjusted p values.
Early recovery and long-term outcomes in lumbar disc herniation

Several studies have sought to compare the relative effectiveness of surgery and conservative care for treatment of a lumbar disc herniation. Generally, these studies have compared “average” differences between the study cohorts, while the individual trajectories by which patient utility changes over time have received less attention. To our knowledge, this study provides the first comprehensive statistical analysis of individual patient-level utility data from a large cohort of patients randomized to a prolonged conservative-care treatment protocol for lumbar disc herniation. The decision to proceed with surgery is straightforward in patients with severe, disabling symptoms or neurological deficits. Likewise, the decision to continue conservative care is simple for patients with mild symptoms or those who are content to live with their symptoms indefinitely. However, patients with moderate symptoms often present a greater challenge because most patients would prefer to avoid surgery if possible, but are also not content to wait indefinitely for their pain to resolve. These patients often ask for more than just the overall probability they will improve eventually; they usually want to know when they will recover. Moreover, they are usually interested in whether their current symptoms and progress affect the probability and extent of their future improvement.

For patients with moderate symptoms, the following observations from our study may be useful. First, the utility scores for individual patients diverged sharply at 8 weeks and were thereafter easily classified as either those reporting no health problems (higher utility, EQ-5D = 1) or those reporting at least some health problems (lower utility, EQ-5D ≤ 0.86). Among the lower utility group, the “pain/discomfort,” “mobility,” and “usual activities” domains of the EQ-5D differed most significantly from the higher utility group, which could potentially represent incompletely treated radiculopathy. Second, most improvement occurred early: almost one-third of the overall improvement in utility came in the first 2 months, while only 1% occurred in the last 3 months. Third, recovery is variable, with most patients (80%) experiencing at least 1 interval of deterioration and only 19% continuously improving without any setbacks. This may provide some reassurance to patients with generally good recovery to “stay the course” without resorting to more invasive measures such as surgery simply because of what may be a brief transient decrease in quality of life. Lastly, the probability of moving into another group was quite low (2%), which may be considered when counseling a patient who is not improving with his or her current treatment regimen.

We note the following limitations inherent in this cohort study. First, this is an observational study, and therefore we cannot infer causality for the emergence of the 2 utility groups, and because the individual treatment plans were unknown to us, we cannot comment on any specific type of conservative therapy. However, even if one considers the patients in the low utility group as nonresponders to conservative therapy (which is likely at least partly incorrect), the study does not imply that surgery would necessarily be beneficial in these patients. Second, the EQ-5D scores a patient’s overall health, and therefore unknown comorbid conditions likely account for at least some of the patients residing in the lower utility group and for part of the utility fluctuations. However, in the clinical setting, it should be obvious as to whether a patient’s symptoms are resulting from unresolved radiculopathy or from preexisting comorbidities. Lastly, we excluded crossover patients from our analysis. Crossover patients are likely those with the most severe symptoms and thus our results may be limited to patients with mild to moderate symptoms. However, we believe this exclusion is appropriate because, as mentioned above, the decision to operate is fairly straightforward when a patient has severe symptoms. From a clinical standpoint, patients with moderate symptoms and without neurological deficits after 8 weeks need the most information about the potential time course and extent of their nonoperative recovery to make an informed treatment decision.

The focus of the present study is individual utility recovery within a patient cohort rather than comparing average response to different treatment protocols. The goal was to gain insight into the dynamics of utility recovery among individual patients treated conservatively, but our approach could be applied to almost any treatment protocol. Studies of the changes (improvements or declines) in individual utility over time are useful because they may provide insights into a patient’s perception of their current treatment protocol (for example, patients in the low utility group would likely report a poor response to treatment), and also to identify a point at which continuing the same treatment is unlikely to improve a patient’s quality of life. Patients entering a conservative-care treatment protocol are likely to experience an initial period of rapid recovery, followed by a longer phase of more modest recovery. Our results suggest that, once the long-term recovery phase begins, patients are unlikely to spontaneously change their recovery for better or worse under the same treatment protocol. Lastly, patient utility scores early in the treatment process were reasonable predictors of long-term outcomes.

**Discussion**

Fig. 4. Graph showing the accuracy of classifiers based on patient utilities. The horizontal line is drawn at 0.50, above which models would perform better than randomly assigning patients to utility groups.
This study is a comprehensive characterization of individual patients’ recovery of health utility from a lumbar disc herniation, and provides a unique picture for clinicians taking care of these patients. Our findings suggest that most recovery occurs early during treatment, and this early recovery period is important to long-term outcomes.

Conclusions

In a cohort of patients undergoing prolonged conservative care for treatment of lumbar disc herniation, 57% of the patients had lingering health problems at 1 year. Utility was recovered most rapidly early in the treatment process, and the majority of utility was also recovered in the initial treatment period. After the initial recovery period, we could identify with reasonable accuracy those patients who would fully recover and those who would not. Over the course of the year, recovery was observed to be highly variable, although most fluctuations were relatively small and only transient. These findings suggest that patients not initially responding to their treatment protocol should consider other options because they are unlikely to respond at a later time. However, patients and clinicians should also be mindful of transient decreases in quality of life, and carefully consider any changes in their treatment plan.

Disclosure

This work was partially supported by a charitable grant from the St. David’s Foundation Impact Fund to Dr. Cowperthwaite, and does not necessarily represent the views of the Impact Fund or the St. David’s Foundation.

Author contributions to the study and manuscript preparation include the following. Conception and design: all authors. Acquisition of data: Cowperthwaite, van den Hout. Analysis and interpretation of data: all authors. Drafting the article: Cowperthwaite. Critical revision of the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Cowperthwaite. Statistical analysis: Cowperthwaite, van den Hout. Administrative/technical/material support: Cowperthwaite. Study supervision: Cowperthwaite.

References


Portions of this work were presented as a podium presentation at the 4th Annual Meeting of the Lumbar Spine Research Society, Chicago, Illinois, on May 6, 2011.

Please include this information when citing this paper: published online June 28, 2013; DOI: 10.3171/2013.5.SPINE12992.

Address correspondence to: Matthew C. Cowperthwaite, Ph.D., The University of Texas at Austin, Texas Advanced Computing Center, J.J. Pickle Research Campus, ROC 1.101, 10100 Burnet Rd., Austin, TX 78758. email: mctcowp@mac.com.