Orthopedics Articles   Orthopedics / Bone and Joint Center

8-7-2022

Calculating ex-ante Utilities From the Neck Disability Index Score: Quantifying the Value of Care For Cervical Spine Pathology

Eric X. Jiang

Joshua P. Castle

Felicity E. Fisk

Kevin Taliaferro

Markian A. Pahuta

Calculating *ex-ante* Utilities From the Neck Disability Index Score: Quantifying the Value of Care For Cervical Spine Pathology

Eric X. Jiang, MD[1] , Joshua P. Castle, MD[1], Felicity E. Fisk, MD[1], Kevin Taliaferro, MD[1], and Markian A. Pahuta, MD, PhD[2]

Abstract

Study Design: General population utility valuation study.

Objective: To develop a technique for calculating utilities from the Neck Disability Index (NDI) score.

Methods: We recruited a sample of 1200 adults from a market research panel. Using an online discrete choice experiment (DCE), participants rated 10 choice sets based on NDI health states. A multi-attribute utility function was estimated using a mixed multinomial-logit regression model (MIXL). The sample was partitioned into a training set used for model fitting and validation set used for model evaluation.

Results: The regression model demonstrated good predictive performance on the validation set with an AUC of .77 (95% CI: .76-.78). The regression model was used to develop a utility scoring rubric for the NDI. Regression results also revealed that participants did not regard all NDI items as equally important. The rank order of importance was (in decreasing order): pain intensity = work; personal care = headache; concentration = sleeping; driving; recreation; lifting; and lastly reading.

Conclusions: This study provides a simple technique for converting the NDI score to utilities and quantify the relative importance of individual NDI items. The ability to evaluate quality-adjusted life-years using these utilities for cervical spine pain and disability could facilitate economic analysis and aid in allocation of healthcare resources.

Keywords

Health economics, utilities, quality-adjusted life years, cervical spine, resource allocation, heath related quality-of-life

Introduction

The number of cervical spine procedures performed for common pathologies such as cervical radiculopathy and cervical spondylosis in the United States have been steadily increasing from the mid 1990s.[1,2] Given the potential risks of surgery, it is critical to demonstrate the value of these procedures to patients and policy makers. The ability to calculate quality-adjusted life-years (QALYs) for patients undergoing cervical spine surgery would help in this regard.

Quality-adjusted life-years analysis could help patients and clinicians jointly assess the trade-offs between prognosis, health-related quality-of-life (HRQoL) benefits, recovery, and potential complications to reach an optimal treatment decision.[3,4] QALYs also aid in economic analysis as economic decisions are based on the incremental cost-effectiveness ratio, which is the cost per QALY gained.[3] QALYs are calculated using utilities, or HRQoL weights. Utilities are a number,

---

[1] Department of Orthopedic Surgery, Henry Ford Hospital, Detroit, MI, USA
[2] Division of Orthopaedic Surgery, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Markian A. Pahuta, MD PhD, Division of Orthopaedic Surgery, McMaster University, 237 Barton Street East, Floor 5N, Hamilton, ON L8L 2X2, Canada.
Email: pahutam@mcmaster.ca

typically between 0 and 1, that quantifies the preference for (ie desirability of) a health state.[3] The utility of perfect health is set at 1 and the utility of a "dead" state is set at 0. If a patient's current health state is measured at a utility of .7, it means that the general population would regard 10 years of life in Patient A's health as equivalent to 7 years of life in perfect health (10 years x .7 = 7 years).

Utility values provide the foundation for determining and comparing health-care interventions. For example, how does the federal health agency of a country determine whether a carotid artery endarterectomy or an anterior cervical discectomy and fusion is more valuable to society and thus which procedure to prioritize for funding? To do this, one needs a metric to compare "apples and oranges" and utilities and its translation to QALYs are currently the most optimal method of doing so. Although utilities can be calculated using generic outcome measures such as the SF-36,[5] there is concern that these generic measures have psychometric limitations.[6,7] Furthermore, disease-specific measures may better capture smaller changes and are more sensitive and responsive for certain conditions.[7-9]

Utilities calculated from an instrument designed for neck pain and disability, such as the Neck Disability Index (NDI) score,[10,11] could increase the sensitivity and specificity of HRQoL assessments for this condition. The NDI, modeled from the Oswestry Low Back Pain Index,[12] is the most widely used self-report instrument to assess neck pain and disability.[13,14] Patients rate symptoms on a 10-item scale (pain intensity, personal care, lifting sleep, driving, sex life, headache, concentration, reading, work) with each item scored out of 5 for a maximum score of 50 (complete disability).[10,11] The NDI score has been psychometrically validated across multiple cultural groups, proved to be highly reliable and valid, and has had minimal clinically important difference values established (3-5 points).[11,13]

In this paper we develop and validate a technique for directly calculating utilities for the NDI score using a discrete choice experiment with a general population sample.

Material and Methods

Subjects

Participants were recruited from an online market research panel (Toluna Influencers, Wilton, CT).[15] Panel members were recruited from across the United States (US) through random-digit-dialing, internet banner advertisements, and partnerships with corporations.[16] We did not provide an incentive for participating in our study; however, the market research company managing the panel does award monthly prizes to panel members based on the number and length of surveys completed. Quota sampling was used to ensure that the study sample was representative of the general US population with respect to region, gender, and age based on 2017 United States Census Bureau Population Estimates Program data.[17]

Health States

The NDI scale (Table 1) was converted to a set of distinct health states consisting of (i) five attributes corresponding to each of the NDI items, and (ii) the duration of survival in the given health state.[18] Duration of survival was set at: 1 year, 2 years, 5 years, and 10 years.[19] Phrasing from the original NDI instrument was modified to the second person and structured as declarative sentences. Following published guidelines, text was modified systematically to achieve a Flesch-Kincaid readability score of United States (US) grade 6 or lower (supplemental material Table S1).[20-23]

Discrete Choice Experiment Valuation Task

Utility valuation was conducted using an online self-administered discrete choice experiment (DCE) questionnaire.[18] DCE methodology is simpler than traditional utility valuation with standard gamble and time-trade-off methods and is therefore better suited for online studies.[24,25] In the DCEs for this study, participants were presented with pairs of health states (choice sets) and asked to select the more desirable health state. Choice sets were presented in a table with differing attributes highlighted (Figure 1).[19]

Choice Set Selection

As there exist over 700-trillion[1] unique choice sets, it was necessary to select a manageable subset for this study. A D-efficient collection of 120 non-dominated choice sets was organized into blocks of 12 using the modified Federov algorithm with Ngene software (supplemental material Table S2).[24,26] The design was developed using parameter values from a general population utility valuation study for the Spine Oncology Study Group Outcomes Questionnaire using DCE methodology.[27] To assess whether participants understood the DCE task, 1 dominated choice set (ie where 1 health state is clearly preferable) was added to each block to test for logic. To assess whether participants engaged in the DCE task and test for internal consistency, 1 choice set was repeated in each block with health state order reversed. Therefore, there were a total of 12 choice sets in each block (10 experimental choice sets, 1 dominated choice set, and 1 repeated choice set). There were 3 levels of randomization in the survey. First, participants were randomized to 1 of the 12 blocks. Second, the order of choice sets in each block was randomized. Third, the health state order was randomized among the participants.

Survey Procedures

The market research company sent panel members an e-mail invitation to participate in our study. Interested panel members were redirected to a secure website hosting the utility valuation exercise.[15,28] Participants

Table 1. Neck Disability Index.

Item (Abbreviation)	Level	Descriptor
Pain intensity (Pain)	0	I have no pain at the moment
	1	The pain is very mild at the moment
	2	The pain is moderate at the moment
	3	The pain is fairly severe at the moment
	4	The pain is very severe at the moment
	5	The pain is the worst imaginable at the moment
Personal care (PerC)	0	I can look after myself normally without causing extra pain
	1	I can look after myself normally but it is very painful
	2	It is painful to look after myself and I am slow and careful
	3	I need some help but manage most of my personal care
	4	I need help every day in most aspects of self care
	5	I do not get dressed, wash with difficulty and stay in bed
Lifting (Lift)	0	I can lift heavy weights without extra pain
	1	I can lift heavy weights but it gives me extra pain
	2	Pain prevents me from lifting heavy weights off the floor but I can manage if they are conveniently positioned, eg on a table
	3	Pain prevents me from lifting heavy weights off the floor but I can manage light to medium weights if they are conveniently positioned
	4	I can lift only very light weights
	5	I cannot lift or carry anything at all
Reading (Read)	0	I can read as much as I want to with no pain in my neck
	1	I can read as much as I want to with slight pain in my neck
	2	I can read as much as I want with moderate pain in my neck
	3	I can't read as much as I want because of moderate pain in my neck
	4	I can hardly read at all because of severe pain in my neck
	5	I cannot read at all
Headaches (Head)	0	I have no headaches at all
	1	I have slight headaches, which come infrequently
	2	I have moderate headaches, which come infrequently
	3	I have moderate headaches, which come frequently
	4	I have severe headaches, which come frequently
	5	I have headaches almost all the time
Concentration (Conc)	0	I can concentrate fully when I want to with no difficulty
	1	I can concentrate fully when I want to with slight difficulty
	2	I have a fair degree of difficulty in concentrating when I want to
	3	I have a lot of difficulty in concentrating when I want to
	4	I have a great deal of difficulty in concentrating when I want to
	5	I cannot concentrate at all
Work (Work)	0	I can do as much work as I want to
	1	I can only do my usual work, but no more
	2	I can do most of my usual work, but no more
	3	I cannot do my usual work
	4	I can hardly do any work at all
	5	I can't do any work at all
Driving (Drive)	0	I can drive my care without any neck pain
	1	I can drive my car as long as I want with slight pain in my neck
	2	I can drive my car as long as I want with moderate pain in my neck
	3	I can't drive my care as long as I want because of moderate pain in my neck
	4	I can hardly drive at all because of severe pain in my neck
	5	I can't drive my car at all
Sleeping (Sleep)	0	I have no trouble sleeping
	1	My sleep is slightly disturbed (less than 1 hour sleepless)
	2	My sleep is mildly disturbed (1-2 hours sleepless)
	3	My sleep is moderately disturbed (2-3 hrs sleepless)
	4	My sleep is greatly disturbed (3-5hrs sleepless)
	5	My sleep is completely disturbed (5-7hrs sleepless)

Table 1. (continued)

| Item (Abbreviation)	Level	Descriptor
Recreation (Rec)	0	I am able to engage in all my recreation activities with no neck pain at all
	1	I am able to engage in all of my recreation activities, with some pain in my neck
	2	I am able to engage in most, but not all of my usual recreation activities because of pain in my neck
	3	I am able to engage in a few of my usual recreation activities because of pain in my neck
	4	I can hardly do any recreation activities because of pain in my neck
	5	I can't do any recreation activities at all
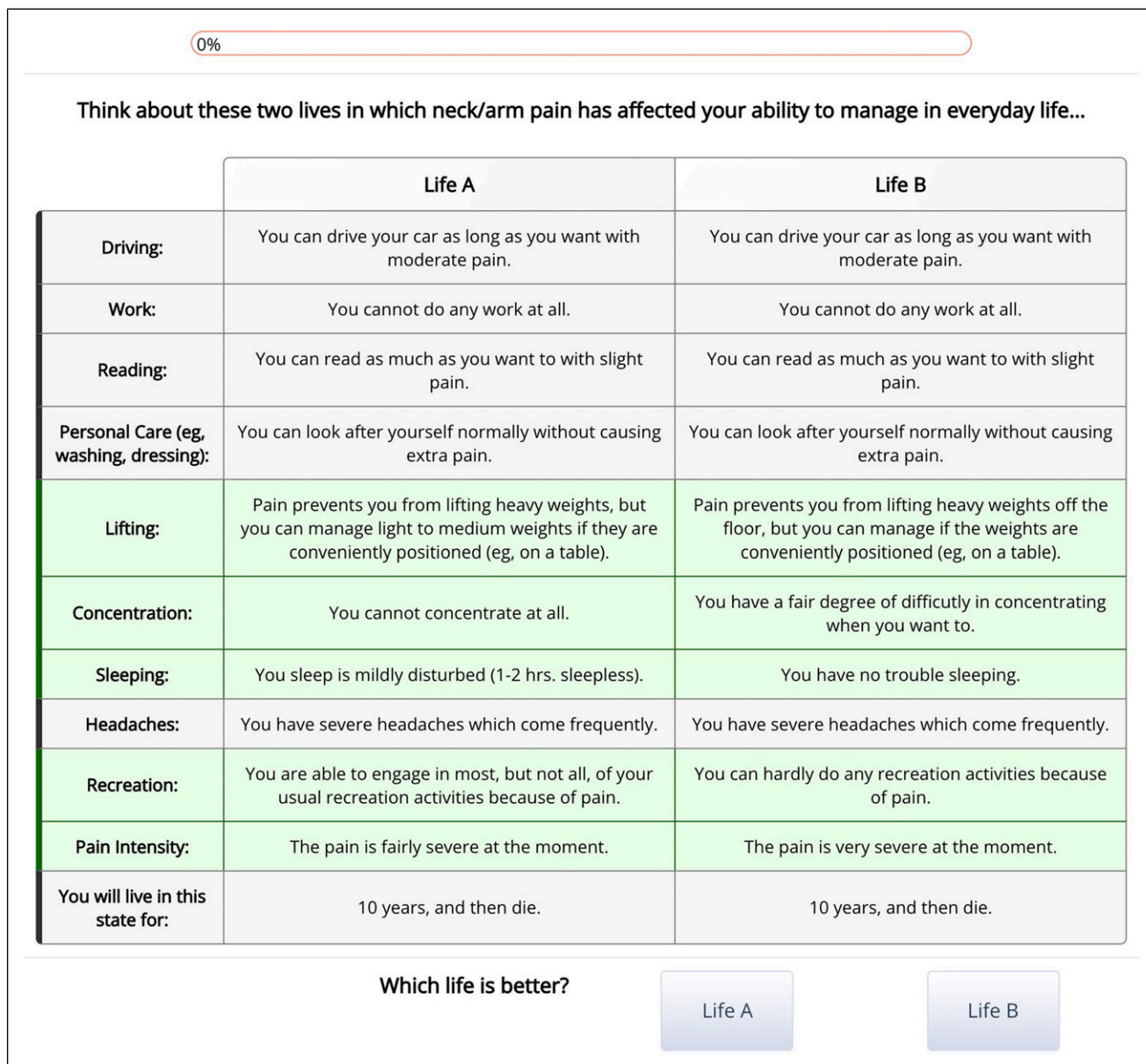
Figure 1. Choice set presentation in online discrete choice experiment. Differing attributes are highlighted in green.

first read brief background information on neck pain and disability that had been scaled to a US grade 6 level. Next, they were provided with an explanation of DCEs and shown a worked example. Participants then completed a practice DCE and provided feedback before completing the study DCEs. At the end of the survey, participants were asked to provide a five-point Likert rating for the statement "this survey was difficult."

Statistical Analysis

Participants who spent an average of at least 8 seconds per choice set (to screen those responses derived from limited engagement), selected the clearly preferable alternative in the dominated choice set, and provided consistent responses for the repeat choice set were deemed to have engaged in and understood the DCE tasks. Only these participants were included in analyses.[15,18,29]

A multi-attribute utility function was estimated from DCE responses using a mixed multinomial-logit regression model (MIXL) using the "mixl" library in the statistical programming language R.[18,30-33] The regression model incorporated the main survival duration effect, and two-way interactions between survival duration and each NDI item.[18] Each parameter was treated as a random effect to account for participant heterogeneity in the repeated DCE tasks. The random effects were modeled with 1000 draws from a normal distribution. In the base regression model, all NDI items were coded as nominal categorical (dummy) predictors to avoid assumptions of linear or extra-linear effects. The base regression model was simplified by removing non-significant predictors and combining adjacent predictors to maintain a monotonic decreasing relationship. For example, for the reading attribute, no levels greater than 0 were significantly different from 0 and so were excluded (supplemental material Table S3). Model performance during the simplification procedure was monitoring using McFadden's ρ^2[34]. Values between .2 to .4 indicate very good model fit and are analogous to an R^2 value between .7 to .9 for linear regression.

In an effort to strengthen the generalizability of the regression analysis, we implemented validation by allocating participants to a training set and validation set in a 1:1 ratio.[35] Regression models were fit using only the training set. The performance of the simplified regression model was assessed via prediction accuracy for choice set selections by participants in the validation set using 1000 draws from the MIXL model. Prediction accuracy was quantified using the area under the curve (AUC) interpreted using the following thresholds: excellent, .9 – 1; good, .8 – .9; fair, .7 – .8; poor, .6 – .7; and failed, .5 – .6.[36]

Regression coefficients quantify the impact of dysfunction in a particular NDI item on utility. Since the lowest level for all NDI items is non-dysfunctional, this level (0) imparts no change in utility. Under this scheme, utilities can be calculated by substituting the sum of the product of predictors and coefficients for each NDI item in the formula:

$$\text{Utility} = 1 - \text{pain} - \text{PerC} - \text{Lift} - \text{Read} - \text{Head} - \text{Conc} - \text{Work} - \text{Drive} - \text{Sleep} - \text{Rec} \quad (1)$$

A worked example is provided in the Results section.

Since, the MIXL model treats each coefficient as a normal ("bell-curve") random variable, regression results consisted of a mean and standard deviation for each coefficient. In this way MIXL techniques model heterogeneity (differences between individuals) of the utility impact of dysfunction in the NDI items. Thus, in order to predict how a single individual values the utility of each NDI item, a random draw is made from the normal distributions estimated by the MIXL model. The mean coefficient values are the expected values for a single individual. In accordance with best practices in health economics, a NDI utility scoring rubric was developed using mean values.[3] The importance of individual NDI items was quantified by calculating the difference in utilities between the best and worst levels of the attribute (importance score).[37]

Sample Size Calculation

Three estimates of sample size were considered. S-efficiency is a measure of the minimum sample size to estimate statistically significant regression parameters at the 95% level.[38] Based on S-efficiency, the minimum sample size for the DCE design shown in Supplemental material Table S3 is 192 participants. Johnson and Orme proposed a simple rule of thumb that considers the number of attribute levels, number of choice sets and alternatives.[39] Based on this rule, the minimum sample size is 600. Furthermore, as we planned to implement a test set and validation set in a 1:1 ratio, we required a total 1200 participants.

Results

We recruited a total of 2875 participants and 1675 were excluded either due to failing the internal consistency test or not demonstrating understanding of DCEs, resulting in a total of 1200 remaining participants. All geographic, gender and age quotas based on the 2017 United States Census Bureau Population Estimates Program were met.[17] There were no statistically or qualitatively significant differences between the training and validation sets in terms of sex, age, or census region (Table 2). Most of the participants in the training and

Table 2. Respondent Demographic Characteristics.

	Training Set N = 600	Validation Set N = 600
Sex – no. (%)		
Female	302 (50.3)	316 (52)
Male	298 (49.7)	284 (48)
Age – no. (%)		
18 – 29 yrs	129 (22)	127 (21)
30 – 39 yrs	112 (18)	94 (16)
40 – 49 yrs	93 (16)	99 (17)
50 – 59 yrs	101 (17)	103 (17)
60 – 69 yrs	79 (13)	97 (16)
≥ 70 yrs	86 (14)	80 (13)
Census Region – no. (%)		
Northeast	134 (22)	116 (19)
Midwest	104 (17)	108 (18)
South	222 (37)	234 (39)
West	140 (24)	142 (24)

validation sets did not agree with the statement that "this survey was difficult" (58% and 53%, respectively).

Multiple adjacent coefficients for several NDI items were collapsed to simplify the base regression: pain 3 and 4; lifting 3/4/5; headache 1/2, headache 3/4; concentration 2/3/4; work 1/2, work 3/4; drive 3/4/5; sleep 4/5; recreation 3/4. Model simplification did not have an adverse effect on performance with the training set as McFadden's ρ^2 remained unchanged at .31, which is indicative of an excellent fit. The simplified regression model had excellent external validity as it predicted DCE choices in the validation set well with an AUC of .77 (95% CI: .76-.78). The final MIXL regression results are shown in Supplemental material Table S3. Statistically significant standard deviation for the majority of coefficients indicated the presence of heterogeneity between participants; therefore, use of a MIXL model was appropriate. Regression results revealed that participants did not regard all NDI items equally important. The rank order of importance for the mean coefficient values for each of the NDI items (in decreasing order of importance) was: pain intensity = work; personal care = headache; concentration = sleeping; driving; recreation; lifting; and lastly reading.

To calculate utilities with equation (1), NDI responses from Table 1 must first be converted to numerical utility levels using Table 3. To illustrate the use of equation (1) and the scoring rubric, we will calculate the utility for Patient A whose NDI scores are: pain intensity – "the pain is very severe at the moment" (the corresponding number in Table 3 is level 4, utility value .05); personal care – "it is painful to look after my and I am slow and careful" (level 2, utility 0); lifting – "I cannot lift or carry anything at all" (level 5, utility .02); reading – "I can read as much as I want to with slight pain in my neck" (level 1, utility 0); headaches – "I have moderate headaches, which come frequently" (level 3, utility .07); concentration – "I can concentrate fully when I want to with no difficulty" (level 0, utility 0); work – "I cannot do my usual work" (level 3, utility .08); driving – "I can hardly drive at all because of severe pain in my neck" (level 4, utility .06); sleeping – "My sleep is mildly disturbed (1-2 hours sleepless)" (level 2, utility 0); recreation – "I am able to engage in a few of

my usual recreation activities because of pain in my neck" (level 3, utility .03). The corresponding values are then substituted into equation (1):

$$
\begin{aligned}
\text{Utility} = {}& 1 - 0.05 - 0 - 0.02 - 0 - 0.07 - 0 \\
& - 0.08 - 0.06 - 0 - 0.03 = 0.69
\end{aligned}
$$

Discussion

In this study, we estimated a multi-attribute utility function for the NDI score for neck pain and disability for the US general population using DCE methods. We validated the regression model by assessing prediction accuracy on an independent set of DCE responses that were not used to develop the regression model. The regression model demonstrated fair prediction accuracy with an AUC of .77 (95% CI: .76-.78). This paper makes 2 clinically useful contributions.

First, we provide a technique for calculating utilities for NDI health states. We have shown a worked example for a hypothetical patient to illustrate how to calculate utilities using equation (1) and Table 3. This utility value quantifies the desirability of patient A's health state relative to perfect health (pain 4, personal care 2, lifting 5, reading 1, headaches 3, concentration 0, work 3, driving 4, sleeping 2, recreation 3) from the perspective of the general population. An overall utility of .69 in our example means that the general population would regard 10 years of life in Patient A's health as equivalent to 6.9 years of life in perfect health (10 years × .69 = 6.9 years). In other words, if given the option between living 10 years in Patient A's health state, or only living 6.9 years plus 1 day in perfect health, members of the general population would, on average, choose to live a shorter duration with better health (the latter option). Since utilities are anchored on perfect health and dead, our data can be used to compare the value of health care interventions across diseases and conditions to aid in prioritization and resource allocation.

The second contribution is the quantification of the importance of each NDI item. Importance scores are listed in Supplemental material Table S3 and quantify the how much

Table 3. NDI Utility Scoring Rubric.

Level					NDI Item					
	Pain	PerC	Lift	Read	Head	Conc	Work	Drive	Sleep	Rec
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	.06	0	.04	0	0	0
2	0	0	0	0	.06	.03	.04	0	0	0
3	.05	0	.02	0	.07	.03	.08	.06	.06	.03
4	.05	.07	.02	0	.07	.03	.08	.06	.07	.03
5	.12	.08	.02	0	.08	.07	.11	.06	.07	.05

To use this table, NDI (Neck Disability Index) responses by subcategory must be converted to numerical levels using Table 1. The appropriate values from this table are then substituted in Equation (1) to calculate utilities. Pain, Pain Intensity; PerC, Personal Care; Lift, Lifting; Read, Reading; Head, Headaches; Conc, Concentration; Work, Work; Drive, Driving; Sleep, Sleeping; Rec, Recreation.

individuals discount life in the worst level of each NDI item relative to the best. For example, an importance score .11 for pain intensity items means that individuals would be willing to trade 11% of their remaining life to reverse pain intensity from its worst state to its best state. In contrast, individuals are only willing to trade 8% of their remaining life to reverse personal care ability (importance score .08) from its worst state to its best state. It is important to note that for the reading attribute, because no levels greater than 0 were significantly different from 0 in terms of their utility (ie Levels 0-5 all had a utility of 0), it was determined to be the least important in the eyes of the general public. Based on our data, the general US population ranks the importance of the NDI items (from most important to least important) as pain intensity = work; personal care = headache; concentration = sleeping; driving; recreation; lifting; and lastly reading. Clinicians should heed these findings and offer treatments that maximize function in the most important attributes.

Utility conversions for the Neck Disability Index (NDI),[40-42] Oswestry Disability Index (ODI),[43] and Scoliosis Research Society 22-item (SRS-22r)[44] that have been previously developed use an indirect "cross-walk" protocol. This protocol involves collecting patient responses using both the condition-specific PROM and a generic PROM and fitting a regression model relating the 2 scores. This allows another regression model to be used to convert the predicted generic PROM score to a utility.[3] This cross-walk protocol has 2 important limitations. First, this technique is complicated and may introduce errors through the use of serial regression models. Second, by only considering the aggregate condition-specific score, this technique cannot differentially weigh the importance of individual items in the condition-specific PROM. It is important to appreciate that *ex ante* utilities are not equivalent to *ex post* utilities obtained from patients who have experienced the health states. Patients tend to provide higher valuations for health states which predominantly affect physical health than the general population for the same health state.[45] Previous work by Richardson et al. presented a regression model for translating NDI scores to *ex post* utility values (the study used a population who had previously undergone surgical treatment of cervical disc disease).[41] These utility values may not be appropriate for global healthcare decision making. Although it may seem that applying lower *ex ante* utilities may infringe on patient autonomy and deny care, healthcare system decision making impacts patients with various conditions. If the objective of healthcare decision making is to maximize the benefit of all patients, utilities across different disease must be comparable to set priorities. Rawls argues that *ex ante* utilities can be used ethically if valued under a "veil of ignorance".[46] If we assume that the general population providing *ex ante* utility valuations may eventually develop the condition of interest, out of self-interest, they should provide fair valuations. Utilities obtained from generic health surveys such as the EuroQol-5D,

Short Form-6D, and Health Utilities Index 3 are actually *ex ante* valuations.[3] Therefore, the *ex ante* utilities derived in this study may not be appropriate for use for individual patient decisions because they do not quantify patient preferences, but they are highly appropriate for facilitating population level healthcare decision making.[47]

One important limitation to our study is that these results are unlikely to be applicable to other countries as median inter-country utility differences for identical health states is over .4.[48] Although differences between value sets within geographic regions are smaller than differences between geographic regions,[49] attempts at explaining these differences through sociodemographic factors, methods of utility valuation, and cultural values have been unsuccessful.[50] Consequently, our results are applicable to the US general population only, and the NDI scale multi-attribute utility functions need to be developed in other regions of the world for use in those areas. Another limitation of our study is that our methodology excludes people who do not have access to the internet. As of 2021, 7% of the United States population do not have access to the internet and therefore could not participate in this study.[51] Lack of internet access is associated with lower socioeconomic status and education.[52] Therefore, due to this "digital divide," these demographic groups may be underrepresented in our sample.

We have quantified and validated a general population multi-attribute utility function for the NDI used for neck pain and disability. equation (1) and Table 3 can be used together to covert NDI responses to utilities. The regression modeling exercise revealed the relative importance of NDI items to the general population. Together, these data can be used to inform population level healthcare decision making, such as the allocation of limited resources for specific treatments.

Declaration of Conflicting Interests

Funding

IRB Approval/Exemption

ORCID iDs

Eric X. Jiang, MD ⓘ https://orcid.org/0000-0002-6271-2172
Markian A. Pahuta, MD, PhD ⓘ https://orcid.org/0000-0002-4808-5459

Supplemental Material

Supplemental material for this article is available online.

References

1. Wang MC, Kreuter W, Wolfla CE, Maiman DJ, Deyo RA. Trends and variations in cervical spine surgery in the United States: Medicare beneficiaries, 1992 to 2005. *Spine (Phila Pa 1976)*. 2009; 34(9):955-961. doi:10.1097/BRS.0b013e31819e2fd5.

2. Yeung CM, Schoenfeld AJ, Lightsey HM, Kang JD, Makhni MC. Trends in spinal surgery performed by American board of orthopaedic surgery part II candidates (2008 to 2017). *J Am Acad Orthop Surg*. 2021;29(11):e563-e575. doi:10.5435/JAAOS-D-20-00437.

3. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. New York, NY: Oxford University Press; 2005.

4. Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. *Value Health*. 2009;12(suppl 1):S27-30. doi:10.1111/j.1524-4733.2009.00519.x.

5. McHorney CA, Ware JEJ, Raczek AE. The MOS 36-Item short-form health survey (SF-36): II. psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31(3):247-263. doi:10.1097/00005650-199303000-00006.

6. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. *Pharmacoeconomics*. 2017;35(Suppl 1):33-41. doi:10.1007/s40273-017-0546-9.

7. Brazier J, Dixon S. The use of condition specific outcome measures in economic appraisal. *Health Econ*. 1995;4(4): 255-264. doi:10.1002/hec.4730040402.

8. Kontodimopoulos N, Stamatopoulou E, Brinia A, Talias MA, Ferreira LN. Are condition-specific utilities more valid than generic preference-based ones in asthma? Evidence from a study comparing EQ-5D-3L and SF-6D with AQL-5D. *Expert Rev Pharmacoecon Outcomes Res*. 2018;18(6):667-675. doi:10.1080/14737167.2018.1505506.

9. Kim H, Cook G, Goodall S, Liew D. Comparison of EQ-5D-3L with QLU-C10D in metastatic melanoma using cost-utility analysis. *PharmacoEconomics - Open* 2021;5(3): 459-467. doi:10.1007/s41669-021-00265-8.

10. Vernon H, Mior S, The neck disability index: A study of reliability and validity. J Manipulative Physiol Ther. *9AD* 1991; 14(7):409-415.

11. Vernon H The Neck Disability Index: state-of-the-art, 1991-2008. *J Manipulative Physiol Ther*. 2008;31(7):491-502. doi:10.1016/J.JMPT.2008.08.006.

12. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980; 66(8):271-273.

13. MacDermid Joy C, Walton DM, Avery S, et al. Measurement properties of the neck disability index: A systematic review. *J Orthop Sports Phys Ther*. 2009;39(5):400-416. doi:10.2519/jospt.2009.2930.

14. Bobos P, Macdermid JC, Walton DM, Gross A, Santaguida PL. Patient-reported outcome measures used for neck disorders: An overview of systematic reviews. *J Orthop Sports Phys Ther*. 2018;48(10):775-788. doi:10.2519/jospt.2018.8131.

15. Pahuta MA, Wai EK, Werier J, van Walraven C, Coyle D. A general population utility valuation study for metastatic epidural spinal cord compression health states. *Spine (Phila Pa 1976)*. 2019;44(13):943-950. doi:10.1097/BRS.0000000000002975.

16. Qualtrics. ESOMAR 28: 28 Questions to Help Research Buyers of Online Samples. Published 2014. https://esomar.org/uploads/attachments/ckqqecpst00gw9dtrl32xetli-questions-to-help-buyers-of-online-samples-2021.pdf

17. United States Census Bureau. State Population by Characteristics: 2010-2018. Population Estimates Program.

18. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *J Health Econ*. 2012;31(1):306-318. doi:10.1016/j.jhealeco.2011.11.004.

19. Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: Feasibility and sensitivity to presentation format. *Qual Life Res*. 2016;25(3):637-649. doi:10.1007/s11136-015-1115-3.

20. U.S. Centers for Medicare & Medicaid. Understanding and using the "Toolkit Guidelines for Writing".

21. El-Daly I, Ibraheim H, Rajakulendran K, Culpan P, Bates P. Are patient-reported outcome measures in orthopaedics easily read by patients? *Clin Orthop Relat Res*. 2016;474(1):246-255. doi:10.1007/s11999-015-4595-0.

22. Perez JL, Mosher ZA, Watson SL, et al. Readability of orthopaedic patient-reported outcome measures: Is there a fundamental failure to communicate? *Clin Orthop Relat Res*. 2017; 475(8):1936-1947. doi:10.1007/s11999-017-5339-0.

23. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: Current role in orthopaedics. *Clin Orthop Relat Res*. 2010;468(10):2572-2580. doi:10.1007/s11999-010-1380-y.

24. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Heal*. 2013;16(1):3-13. doi:10.1016/j.jval.2012.08.2223.

25. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: How can we assess validity and reliability? *Expert Rev Pharmacoeconomics Outcomes Res*. 2017;17(6):531-542. doi:10.1080/14737167.2017.1389648.

26. ChoiceMetrics. Ngene 1.1.1 user manual & reference guide. Published online. 2018.

27. Pahuta MA, Fisk F, Versteeg AL, et al. Calculating utilities from the spine oncology study group outcomes questionnaire: A necessity for economic and decision analysis. *Spine (Phila Pa 1976)*. 2021;46(17):1165-1171. doi:10.1097/BRS.0000000000003981.

28. Pahuta M, Formbach A, Mitera G, Coyle D, Werier J, Wai E. Validation of the self-administered online assessment of preferences (SOAP) utility elicitation tool. *Can J Surg.* 2016;59(suppl 2-3):S40. S63, doi:10.1503/cjs.006916.

29. Harrison M, Marra C, Shojania K, Bansback N. Societal preferences for rheumatoid arthritis treatments: Evidence from a discrete choice experiment. *Rheumatol (United Kingdom).* 2015;54(10):1816-1825. doi:10.1093/rheumatology/kev113.

30. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: Past, present and future. *Pharmacoeconomics.* 2019;37(2):201-226. doi:10.1007/s40273-018-0734-2.

31. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. *Value Heal.* 2016;19(4):300-315. doi:10.1016/j.jval.2016.04.004.

32. R Core Team. R: A Language and Environment for Statistical Computing. Published online 2018.

33. Molloy J, Schmid B, Becker F. Mixl : An open-source R package for estimating complex choice models on large datasets; 2019. doi:10.3929/ethz-b-000334289.

34. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications.* 1st ed. .Cambridge, UK:Cambridge University Press; 2000.

35. Steyerberg EW, Harrell FE, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001;54(8):774-781.

36. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law Hum Behav.* 2005; 29(5):615-620. doi:10.1007/s10979-005-6832-7.

37. Kromer C, Schaarschmidt M-L, Schmieder A, et al. Patient preferences for treatment of psoriasis with biologicals: A discrete choice experiment. *PLoS One.* 2015;10(6):e0129120. doi:10.1371/journal.pone.0129120.

38. Rose JM, Bliemer MCJ. Sample size requirements for stated choice experiments. *Transportation (Amst)* 2013;40(5): 1021-1041. doi:10.1007/s11116-013-9451-z.

39. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: A practical guide. *Patient.* 2015;8(5):373-384. doi:10.1007/s40271-015-0118-z.

40. Carreon LY, Anderson PA, McDonough CM, Djurasovic M, Glassman SD. Predicting SF-6D utility scores from the neck disability index and numeric rating scales for neck and arm pain.

Spine (Phila Pa 1976). 2011;36(6):490-494. doi:10.1097/BRS.0b013e3181d323f3.

41. Richardson SS, Berven S. The development of a model for translation of the Neck Disability Index to utility scores for cost-utility analysis in cervical disorders. *Spine J.* 2012; 12(1):55-62. doi:10.1016/j.spinee.2011.12.002.

42. Zheng Y, Tang K, Ye L, Ai Z, Wu B. Mapping the neck disability index to SF-6D in patients with chronic neck pain. *Health Qual Life Outcomes.* 2016;14, 21(1). doi:10.1186/s12955-016-0422-x.

43. Carreon LY, Glassman SD, McDonough CM, et al. Predicting SF-6D utility scores from the Oswestry disability index and numeric rating scales for back and leg pain. *Spine (Phila Pa 1976).* 2009;34(19):2085-2089. doi:10.1097/BRS.0B013E3181A93EA6.

44. Wong CKH, Cheung PWH, Samartzis D, et al. Mapping the SRS-22r questionnaire onto the EQ-5D-5L utility score in patients with adolescent idiopathic scoliosis. *PLoS One.* 2017;12(4):e0175847. doi:10.1371/journal.pone.0175847.

45. Nord E, Daniels N, Kamlet M. QALYS: Some challenges. *Value Heal.* 2009;12(suppl 1):S10-S15. doi:10.1111/j.1524-4733.2009.00516.x.

46. Rawls J. *A Theory of Justice: Revised Edition.* 2nd ed. Cambridge, MA: Belknap Press; 1999.

47. Drummond M, Brixner D, Gold M, et al. Toward a consensus on the QALY. *Value Health.* 2009;12(suppl 1):S31-S35. doi:10.1111/j.1524-4733.2009.00522.x.

48. Gerlinger C, Bamber L, Leverkus F, et al. Comparing the EQ-5D-5L utility index based on value sets of different countries: impact on the interpretation of clinical study results. *BMC Res Notes.* 2019;12(1):18. doi:10.1186/s13104-019-4067-9.

49. Olsen JA, Lamu AN, Cairns J. In search of a common currency: A comparison of seven EQ-5D-5L value sets. *Health Econ.* 2018;27(1):39-49. doi:10.1002/hec.3606.

50. Roudijk B, Donders ART, Stalmeier PFM, et al. Cultural Values Group. Cultural Values: Can They Explain Differences in Health Utilities between Countries? *Med Decis Making.* 2019; 39(5):605-616. doi:10.1177/0272989X19841587.

51. Demographics of Internet and Home Broadband Usage in the United States | Pew Research Center. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/. Accessed September 20, 2021

52. Haight M, Quan-Haase A, Corbett BA. Revisiting the digital divide in Canada: The impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Inf Commun Soc.* 2014;17(4):503-519. doi:10.1080/1369118X.2014.891633.