10-16-2021

Criterion validation and interpretability of the Single Assessment Numerical Evaluation (SANE) of self-reported recovery in patients with neck pain

Elizabeth Oakley

Chad E. Cook

Bryan O'Halloran

Musculoskeletal Science and Practice 56 (2021) 102467

Contents lists available at ScienceDirect

Musculoskeletal Science and Practice

journal homepage: www.elsevier.com/locate/msksp

Original article

Criterion validation and interpretability of the Single Assessment Numerical Evaluation (SANE) of self-reported recovery in patients with neck pain

Bryan O'Halloran [a,*], Chad E. Cook [b], Elizabeth Oakley [c]

[a] School of Physical Therapy, Samson College of Health Sciences, University of the Sciences, 600 S 43rd St Philadelphia, PA, 19104, USA
[b] Doctor of Physical Therapy Division, DUMC 104002311 Trent Drive, Duke University, Durham, NC, 27710, USA
[c] Physical Therapy Building 135, Andrews University, 8975 Old 31, Berrien Springs, MI, 49104, USA

ARTICLE INFO

Keywords:
Single assessment numerical evaluation
Patient reported outcome measures
Neck disability index
Neck pain

ABSTRACT

Background: The SANE is a PROM of recovery, which may assist clinicians in clinical decision-making and discharge planning. The psychometric measurement properties of the SANE have yet to be determined for neck pain.
Objectives: Threefold objectives included: 1)determine the numerical threshold for the SANE at which patients with neck pain determine their symptoms are acceptable; 2)determine the association between scores for the NDI and VAS, with the SANE; 3)determine the average number of visits, costs and value associated with the management of neck pain.
Design: Longitudinal repeated measures cohort design.
Methods: Threshold measures for self-reported recovery with the SANE anchored to the PASS were examined using ROC. PCC determined the relationship between the VAS/pain and NDI raw/percentage change scores and the SANE at discharge. Descriptive statistics were used for number of visits and cost. Value was calculated as the proportion of change on the NDI and VAS/$100 US dollars spent.
Results: 57 subjects completed full observation. ROC analysis indicates a threshold value of 82.5%(Sn = 56.0, Sp = 85.7,+LR = 1.68,-LR = 0.29) on the SANE with an AUC of 0.820(95%CI = 0.638, 1.00). A weak correlation was found between raw NDI($r = 0.39$ $p < 0.05$)/Pain($r = 0.45$ $p < 0.05$) scores and the SANE with a moderate correlation between percent change scores of NDI($r = 0.52$ $p < 0.05$)/PAIN($r = 0.54$ $p < 0.05$) and the SANE. The value proposition indicated cost of care amounted to a 10.5% and 12.9%; improvement in the NDI and pain scores/$100 spent.
Conclusions: Patients reporting greater than 82.5% on the SANE are likely to find their present status acceptable and potentially stop seeking care.

1. Introduction

Neck pain is the fourth leading cause of disability in the United States with an annual prevalence rate in adults exceeding 30% (Cohen, 2015). In the United States, the costs of treating low back and neck pain are higher than any other musculoskeletal condition (Dieleman et al., 2020). Whereas the cost effectiveness of treating neck pain conservatively is still unknown (Driessen et al., 2012), knowing when to stop management or when the patient has reached their perceived acceptable rate of improvement has the potential to reduce unnecessary care. Clinicians, such as physical therapists, who manage non-communicable diseases, are responsible for gathering meaningful information involving change in function/pain when providing care to each patient. Nonetheless, challenges exist within clinical care pathways when measuring patient reported outcome measures (PROM). These challenges include the administrative burden associated with instrument completion and scoring, as well as the relevance of items within the selected PROM measure to each unique individual presentation (Warren and Smith, 2018; Black, 2013).

In order to justify ongoing care, inform decision-making, and justify discharge planning, third party payers require demonstration of meaningful improvement (Williams et al., 2007). As such, many facilities

* Corresponding author. 57 W. Eagle Rd, Havertown, Pa, 19083, USA
E-mail addresses: bohalloran@usciences.edu (B. O'Halloran), chad.cook@duke.edu (C.E. Cook), oakleye@andrews.edu (E. Oakley).

https://doi.org/10.1016/j.msksp.2021.102467
Received 4 June 2021; Received in revised form 11 October 2021; Accepted 13 October 2021
Available online 16 October 2021

have implemented standard functional outcome instruments (i.e., legacy measures) that are representative of either a specific disease state or the body part under management (Bolt and Wendland, 2020). These tools are designed to measure the complex and multidimensional health status of each patient during their care process, including at discharge (Maitland, 2013, 2014).

The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) (Mokkink et al., 2010) suggests that in research and clinical practice, the selection of PRO measures be based on the measurement properties of the instrument (Portney and Watkins, 2004). An example includes the Patient-Reported Outcomes Measurement Information System® (PROMIS®), which is a relatively new multidimensional outcome instrument, and is endorsed by the National Institutes of Health. At present, PROMIS has been subjected to limited analysis of threshold values such as Minimal Clinically Important Difference (MCID) scores, with a wide range reported, depending on the method of calculation (Hung et al., 2018a). This influences its interpretability in clinical practice (Vaishnav et al., 2020). The NDI also demonstrates a wide range of reported MCID from 6 to 22 points presenting a challenge for clinician interpretability of change scores associated with this instrument (Hung et al., 2018a).

The majority of legacy instruments (including PROMIS measures) quantify current health status, but it is unclear whether they measure the patient's interpretation of their own *recovery*, or whether they are likely to pursue continued care (Williams et al., 2007; Razaeian et al., 2020). For patients with musculoskeletal pain, recovery is an individualist concept that varies across patients. Some will experience resolution of symptoms, some will modify their activities or readjust, while others will simply adapt to living with the complaint or redefine it. The limited research on what constitutes *recovery* from neck pain has largely examined patient perception of the healthcare provided rather than recovery itself (Beaton et al., 2001). The concept of recovery likely drives future healthcare utilization use (Hush et al., 2012) and is likely a driver of the total number of visits and cost per visit within an episode of care. The Single Assessment Numerical Evaluation (SANE) is a multidimensional single item tool that has been used by clinicians to measure patients' self-assessment of recovery (Nazari et al., 2020, 2021). It has undergone criterion validation for a number of conditions including the shoulder (Williams et al., 1999), back (Kvamme et al., 2010), knee and ankle (Garcia et al., 2019) but has not been subjected to validation for neck pain. The SANE has value, since shorter outcome measures can effectively capture meaningful information and are preferred in a traditional clinical setting (Nayak et al., 2015). The Patient Acceptable Symptom State (PASS) is also a single item PROM measure. The PASS is indicative of a status which the patient considers *acceptable* (Kvamme et al., 2010) and beyond which they are unlikely to seek further care.

Understanding the patient's recovery level with a simple, single question may assist clinicians in making definitive care-related decisions. In addition, knowing the number of visits required and cost of care to reach an acceptable symptom state of self-reported recovery can help reduce unnecessary care. The objectives of this study were to: 1) determine the numerical threshold for the SANE at which patients with mechanical neck pain determine their symptoms are acceptable; 2) determine the association between scores reported for the Neck Disability Index (NDI) and Visual Analog Scale (VAS) for pain, with the SANE, and 3) determine the average number of visits, costs and value associated with the conservative management of neck pain. Based on prior work performed on low back pain using the ODI and VAS and also subacromial impingement syndrome using the QuickDash and VAS, we anticipated a weak positive correlation between the SANE and these measures a priori (Wright and Cook, 2013; O'Halloran et al., 2013). Lastly, to collect cost of care data to determine the efficiency of physical therapy in the management of neck pain.

2. Method

This study is a longitudinal repeated measures cohort design which collected data on patients with mechanical neck pain and used anchor-based methodology to relate the score of the SANE to the external criterion or independent measure of the PASS (Devji et al., 2020). The PASS is a current, interpretable and relevant measure of the status of a patient's perceived recovery (Wright and Cook, 2013; O'Halloran et al., 2013). The study was approved by Andrews University Institutional Review Board under IRB approval number IRB Protocol#:20–076 with data collection performed between August of 2020 through January 2021.

3. Participants and centers

Consecutive subjects presenting for physical therapy management of neck pain, were screened for inclusion by the treating physical therapist during their first visit for therapy at one of ten physical therapy clinics across seven states in the United States and was a purposive sample of clinicians interested in patient reported outcome measures. These States covered the Mid Atlantic, New England, West Coast and Southern regions of the country. Patients were included and informed consent obtained if they met the following criteria: 1) Currently presenting with neck pain, 2) Over 18 years of age, 3) scores of greater than or equal to 2/10 on the VAS for pain and 4) English speaking. Exclusion criteria included a past history of cervical surgery, absent upper extremity reflexes, weakness of neurological origin in the upper extremities (e.g., cervical radiculopathy), upper motor neuron signs including positive clonus, positive Babinski or Hoffmann's sign as well as the presence of any red flags (i.e., fracture, osteoporosis, cancer, cervical instability etc.).

All therapists involved in data collection were either Fellows of the American Academy of Orthopedic Manual Physical Therapy (FAAOMPT), or Fellows in Training for AAOMPT Fellowship programs. These therapists performed the screening and participated in individual training via video link with the Primary Investigator for screening and data collection procedures and the forms used. It has been shown that Fellowship trained therapists are more efficient in their management of patients (Rodeghero et al., 2015), have a requirement for clinical research in their fellowship training (Furze et al., 2016), and therefore may be more practiced in the screening and data collection process.

4. Procedures

The study used a pragmatic interventional approach allowing therapist discretion and clinical reasoning to guide care. Consequently, each patient received the care that was identified as both preferential by the patient and necessary by the treating clinician. Baseline demographic data were collected, in conjunction with completion of patient reported outcome measures (PROMs) on paper including the NDI and VAS for pain at the first visit. Every 5th visit, the VAS for pain, SANE and PASS were collected until the patient indicated a "yes" response on the PASS, at which time the NDI would then be repeated or this would be collected at discharge from care. The total number of visits and total costs of care were also collected upon discharge.

4.1. Variables

Assessment Variables: The SANE was designed to identify the patient's impression of their own recovery percentage. The measure ranges from 0 to 100, with 100 being fully recovered. The single item included "How would you rate your neck today as a percentage of normal using a 0%–100% scale with 100% being normal?". Criterion validity has been reported for the SANE in the shoulder (O'Halloran et al., 2013) and for mechanical low back pain (Wright and Cook, 2013) with both studies using the PASS as an anchor for the SANE. Minimal clinically important

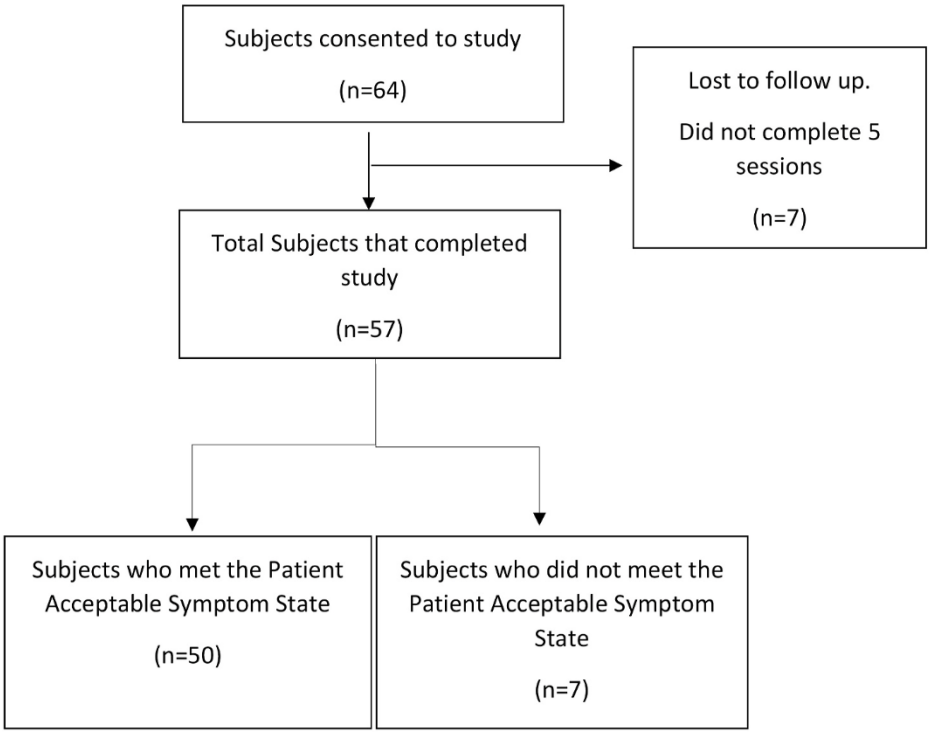

Fig. 1. STROBE flow diagram for subject recruitment and retention.

threshold recovery scores of 87% (Garcia et al., 2019) and 82.5%[43] were identified in those two studies.

The NDI is a self-report measure with 10 sections, each with a 5-point Likert scale by which participants rate their perceived disability related to neck pain. It has been demonstrated as valid and reliable for patients with neck pain (Bobos et al., 2018). The VAS for pain is a 10-cm line with anchors ranging from "no pain" to "worst imaginable pain". This has been demonstrated as a valid and reliable tool for scoring pain (Lundeberg et al., 2001). Participants were asked to rate their average neck pain intensity over the past week. Percentage change scores for the VAS and the NDI were calculated by taking the final value and subtracting it from the initial value then dividing by the initial value; that value was then multiplied by 100 to create a whole number.

Cost of Care and Value Variables: Upon discharge, the number of visits and cost of care was also captured and recorded. Cost of care was calculated based upon the summation of patient payment plus third party payment and was collected by each therapist in consultation with their practice reimbursement specialist upon closure of the billing/administrative file. This was deemed an accurate reflection of total cost of care rather than billable charges. This allowed the opportunity to determine the value proposition of the physical therapy management of neck pain by examining the metric of value defined as the change in outcome per dollar spent (Horn et al., 2016). The calculation of metric values was performed on the NDI and also the VAS pain scale. The metric of value for disability and for pain was calculated as described by Horn et al.(2016) (Horn et al., 2016) and allows for an interpretation of change in disability, and also pain score per 100 dollars spent on care. Smaller numbers indicate lower value and larger numbers higher value.

Anchor Variable: The PASS is a dichotomous, self reported patient rate of recovery score. The question includes "Taking into account all the activities you have during your daily life, your level of pain, and also your functional impairment, do you consider your current state is satisfactory?- yes/no" (Deyo et al., 1998). The PASS is a patient self-reported status beyond which the patient considers their state "acceptable" (Williams et al., 2007). This allows for a patient determined status report at discharge defined by a single question (Wright and Cook, 2013; O'Halloran et al., 2013). Distribution based methods consider the measurement precision of the instrument but because they are entirely based on standard deviation, they are sample specific whereas anchor based approaches consider the clinical relevance of the measure but not the precision of the instrument (Jayadevappa et al., 2012). As a result of these characteristics, it has been suggested that anchor based, rather than distribution based methods of analysis are the most useful methodology for practical applications (Hung et al., 2018a).

4.1.1. Sample size estimate

A priori analysis for sample size was performed using MEDCALC20 (MedCalc Statistical Softw, 2020). The sample size estimate was based on the following parameters: Area under the ROC curve 0.725, Null hypothesis value 0.5, Type I Error (Alpha Significance) = 0.05, Type II Error (Beta, 1-Power) = 0.20, which provided a sample size of 57. When assuming a dropout rate of 10%, the total sample size required was estimated to be 63 subjects enrolled.

4.2. Data analyses

All statistical analyses, apart from the a priori analysis, were performed using *Statistical Package for the Social Sciences* (SPSS) version 27. (IBM SPSS Statistics for Windows) Descriptive statistics for age, sex, duration of symptoms, raw score, change score for VAS and NDI, number of visits and cost of care were expressed as means and standard deviations or proportions and frequencies. A Pearson correlation matrix was used to assess the level of correlation between raw scores and percentage change scores of the VAS and NDI with the SANE. Strength of association was determined according to the following criteria, 0.00–0.19 was considered very weak, 0.20–0.39 as weak, 0.40–0.59 as moderate, 0.60–0.79 as strong and 0.80–1.0 as very strong (Portney and Watkins, 2004). A $p < 0.05$ was considered significant for all associations. An area under the curve (AUC) receiver operating characteristic (ROC) (Hajian-Tilaki, 2013) curve was used to identify the optimal cut-point for the SANE when anchored to the PASS. The method used to identify the optimal cut-point was the point closest to-(0,1) corner in the

Table 1
Characteristics of the sample including those who met and did not meet the PASS.

s	TOTAL			MET PASS			DID NOT MEET PASS		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Age	64	50.6	17.9	50	51.8	17.9	7	38.4	15.0
Duration (weeks)	64	27.3	74.6	50	29.6	82.8	7	21.3	26.8
Pain Initial	64	6.1	2.2	50	6.2	2.1	7	5.3	1.7
NDI Initial	64	32.3	17.8	50	30.4	16.9	7	40.0	20.1
Raw Pain Change	57	4.3	2.6	50	4.7	2.4	7	1.6	1.8
Raw NDI Change	56	18.0	17.6	50	18.3	17.0	6	16.0	23.6
% change Pain	57	70.5	31.7	50	75.8	27.4	7	30.2	31.2
% change NDI	56	55.7	32.6	50	60.2	30.0	6	40.0	41.7
Total Visits	64	7.3	6.7	50	7.7	7.2	7	16.4	16.0
Total cost US$	64	813	1102	50	846	1182	7	1182	867

PASS; Patient Acceptable Symptom State, NDI; Neck Disability Index, N; Sample size.
SD; standard deviation.

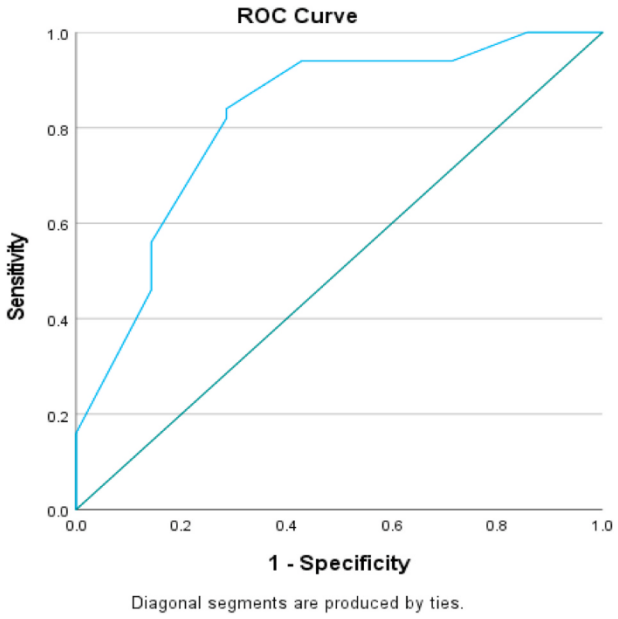

Fig. 2. Area under the curve = 0.820 (95%CI = 0.638, 1.00).

ROC plane which defines the optimal cut-point as the point minimizing the Euclidean distance between the ROC curve and the (0,1) point (Unal, 2017). For AUC-ROC measures, values of 0.5–1.0 suggest the ability to distinguish a meaningful threshold, with higher values suggesting better performance of the model. A value of 0.5 suggests no discrimination, 0.7–0.8 is considered acceptable, 0.8–0.9 is considered excellent and greater than 0.9 is considered outstanding (Hosmer and Lemeshow, 2004). Value was then calculated as the percent change and raw score change on the NDI and the VAS per $100 collected for care.

5. Results

Sixty-four subjects were consented and seven were lost to follow up prior to the fifth visit for follow up data collection leaving fifty-seven subjects total (Fig. 1). The mean age of those enrolled was 50.6 years (SD = 17.9) with 72% being women. The average duration of symptoms since time of onset to presentation for care was 27.3 weeks (SD = 74.6). The mean baseline score for the VAS was 6.1 (SD = 2.2) and for the NDI 32.3 (SD = 17.8), indicative of moderate to severe levels of pain and impairment. There was marked improvement from baseline to discharge with an average raw pain score improvement of 4.3 (SD = 2.6) for a percent change of 70.5% (SD = 31.7). NDI raw scores also demonstrated improvement of 18.0 points (SD = 17.6) for a percent change of 55.7% (SD = 32.6). At discharge, 50 (88%) subjects met the "acceptable" status on the PASS (Table 1). Seven subjects dropped out of care for unknown reasons prior to the 5th visit and one subject did not complete the NDI at the final visit due to data collection oversight.

The Area Under the Curve (AUC) was 0.820 (95%CI = 0.638, 1.00) and Receiver Operating Characteristic curve modelling was significant ($p < 0.01$) indicating excellent strength in the ability of the model to distinguish the threshold of the SANE to the PASS (Fig. 2). The optimal cut point on the ROC curve for self-reported recovery as demonstrated by the SANE, was >82.5/100 which provided a Sensitivity of 56.0 and a Specificity of 85.7. The Positive Likelihood Ratio was 1.95 and the Negative Likelihood Ratio was 0.52, with 51% of subjects who completed data collection meeting the cut point of 82.5.

Analysis of the Pearson correlation matrix demonstrated significant relationships ($p < 0.05$) between all variables including SANE raw scores and percent change scores for VAS for pain and the NDI. There was a moderate positive correlation between the SANE and percent change for both measures and a weak positive correlation between the SANE and raw scores when comparing the initial visit to the final visit (Table 2). The mean for total visits was 7.28 (SD = 6.69) with a median of 5 visits (IQR = 4–10) and the mean cost of care was $813 (SD = 1102) with a median of $511.50 (IQR = 281.5–917). Using the median value of cost

Table 2
Pearson correlation matrix.

Variable		SANE	Raw Change Pain	% Change Pain	Raw Change NDI	% Change NDI
SANE (Raw score)	Pearson Correlation	1				
	N	57				
Raw Change Pain	Pearson Correlation	r = 0.45*	1			
	N	57	57			
% Change Pain	Pearson Correlation	r = 0.54*	r = 0.84*	1		
	N	57	57	57		
Raw Change NDI	Pearson Correlation	r = 0.39*	r = 0.62*	r = 0.47*	1	
	N	56	56	56	56	
% Change NDI	Pearson Correlation	r = 0.52*	r = 0.49*	r = 0.54*	r = 0.64*	1
	N	56	56	56	56	56

SANE; Single Assessment Numerical Evaluation, NDI; Neck Disability Index, N; Sample size, N/A; Not applicable, *; correlation is significant at the 0.01 level (2-tailed).

of care, this produces a 10.5% improvement on the NDI per $100 spent and 12.9% improvement in the VAS Pain score per $100 spent (3.53 raw NDI score improvement and 0.84 raw pain score improvement per $100 spent).

6. Discussion

In this study involving patients with neck pain, we identified 82.5% as the optimal threshold for the SANE, which is consistent with previous findings for the same instrument in low back pain (82.5%) (Wright and Cook, 2013) and subacromial impingement syndrome (87%) (O'Halloran et al., 2013). As a continuous measure that ranges from 0 to 100, the SANE may provide interpretability advantages over the binary PASS. Interpretability is an important characteristic of a measurement instrument and can be described as the degree to which one can assign qualitative meaning (that is, clinical or commonly understood connotations) to an instrument's quantitative scores or change in scores (Mokkink et al., 2010). Interpretability may be considered a domain, equal to reliability, validity and responsiveness and if we do so, then the SANE can be considered as an interpretable patient reported measure indicative of clinically meaningful change (Unal, 2017) as defined by recovery, independent of change in status.

We anchored to the PASS as a measure that reflects on an individual level (O'Halloran et al., 2013). Our threshold value is designed to reflect the average patient's acceptable state (Crosby et al., 2003), and is a group based value. This has similar disadvantages to values that are pre-determined from a group mean, selected by the therapist, identified by an insurance company, or calculated from a legacy outcome measure. For patients in our study, the threshold of 82.5% for the SANE can be interpreted as a point where the patient has recovered enough that they feel their current state is acceptable. The clinician, can use this for discharge planning or advocacy for additional care in the presence of mechanical neck pain patients, as this is indicative of patient perceived interpretation of *recovery* rather than *improvement,* as measured by the NDI. Using a SANE allows the clinician to advocate for more care, or termination of care, even when values are markedly variable with traditional legacy measures (Maltenfort and Díaz-Ledezma, 2017; Hung et al., 2018b).

Our secondary objective was to look at the association between the scores of the Neck Disability Index (NDI) and Visual Analog Scale (VAS) for pain with the SANE. We found a moderate association between the SANE and the percent change score of the NDI and VAS for pain and a weak association between the SANE and raw scores. This confirmed our a priori hypothesis of a weak positive correlation for raw scores. The moderate positive association demonstrated for percent change scores suggests that percent change scores from baseline may be a more important measure of recovery than raw scores. Raw score calculations are influenced by the level of the initial score. This conversion from raw scores (interval data) into percent values, allows for the SANE (ratio data), to be compared with percent values for the NDI and VAS (ratio data) and may explain some of the difference in the level of association. NDI and VAS pain raw scores may vary based upon the severity of the initial presentation and often do not have the same discriminative properties as percent change scores or the SANE.

This is also the first study to have examined the number of visits and the *cost* of care rather than *charges* for care, determined by patient self-reported recovery rather than improvement, over a wide geographic catchment comprising ten separate data collection sites in seven states with substantial demographic differences. The identification of the cost of conservative care associated with the management of mechanical neck pain as defined by the change in function/pain scores per 100 dollars spent also has significant potential for clinical utility. We found a notable change in outcome for payments received.

With a median cost (we defined costs by the dollar amount actually collected) of $120.20/visit and a median of 5 visits to meet the PASS, patients demonstrated a 10.5% (3.53 points) and 12.9% (0.84 points) improvement on the NDI and the VAS for pain respectively. A similar analysis by Horn et al., demonstrated only a 2.27% improvement in the NDI per 100 dollars in total charges, and reported a 0.38 point improvement in raw pain scores per 100 dollars in total charges. These values are demonstrably lower than ours (Horn et al., 2016). It is likely related to the use of charges, which are typically much higher than payments received, and it may be related to the unreported experience level of the clinicians. In our study, all therapists involved in data collection were either Fellows or Fellows in Training for AAOMPT. It has been suggested that Fellowship trained physical therapists are more likely to achieve greater treatment effect size, greater functional status change and greater efficiency than those without fellowship training (Furze et al., 2016).

7. Limitations

We did not capture the level of education or socioeconomic status in our patient demographics. This data may provide meaningful insight into the contribution of social gradient on interpretability of the SANE and may be useful to include in future studies of the psychometric properties of this instrument. Clinician recruitment involved a purposive selection of Fellowship trained therapists which may limit generalizability of these findings to generalist outpatient orthopedic physical therapy settings. It would also have been useful to capture the self-reported perception of recovery at the beginning of care in order to clarify improvement that has already occurred as a result of previous management and as a result of the natural history of the disorder compared with that improvement which occurs during the time frame of conservative care. This would have allowed the characterization of cost associated with point change on the SANE, an additional measure of value associated with the conservative management of neck pain. Capture of the EuroQol- 5 Dimension (EQ-5D) would also have allowed for the opportunity not just to calculate value based on cost of care but also provide a conversion into Quality Adjusted Life Years (QALY's) and consequent provision of Incremental Cost-Effectiveness Ratios (ICERs), a cost comparison for which further research is warranted (Alvin et al., 2019). There also appear to be two to three prognostic phenotypes within the mechanical neck pain population with different trajectories of recovery. Accounting for these confounders may improve the likelihood of predicting those who may or may not meet the threshold of recovery (Lee et al., 2020).

8. Conclusions

A majority of legacy outcomes measures do not reflect recovery and may provide administrative burden in application to clinical practice. The SANE is a single item tool, easily incorporated in clinical practice and research, in which a threshold value of 82.5% reflects a groups' unlikelihood of seeking further care. The SANE reflects recovery, and is only marginally associated with other legacy measures of disability and pain. Cost of care data provided a value determination indicating an expectation of change on the order of 10.5% improvement on the NDI per $100 spent and 12.9% improvement in the VAS Pain score per $100 spent. As we have no published comparator cost data for the conservative management of neck pain, this study provides a unique measure of value for care. Future research expanding this work into other body regions and conditions is warranted.

Declaration of competing interest

None.

Acknowledgements

The authors would like to thank and acknowledge the contribution of the physical therapists involved in data collection: Bob Fleming, Damon

Daura, Dhaval Talati, Bernie Povanda, Robert Price, Karena Wu, Michael Gray, Ethan Brown, Addison Lerner and Travis Stone.

References

Alvin, M.D., Mehta, V., Halabi, H.A., Lubelski, D., Benzel, E.C., Mroz, T.E., 2019. Cost-effectiveness of cervical epidural steroid injections: a 3-month pilot study. Global Spine J. 9 (2), 143–149. https://doi.org/10.1177/2192568218764913.

Beaton, D.E., Tarasuk, V., Katz, J.N., Wright, J.G., Bombardier, C., 2001 Jun. Are you better?" A qualitative study of the meaning of recovery. Arthritis Rheum. 45 (3), 270–279. https://doi.org/10.1002/1529-0131(200106)45:3<270::AID-ART260>3.0.CO;2-T.

Black, N., 2013. Patient reported outcome measures could help transform healthcare. BMJ 346, f167. https://doi.org/10.1136/bmj.f167. Published 2013 Jan 28.

Bobos, P., MacDermid, J.C., Walton, D.M., Gross, A., Santaguida, P.L., 2018 Oct. Patient-reported outcome measures used for neck disorders: an overview of systematic reviews. J. Orthop. Sports Phys. Ther. 48 (10), 775–788. https://doi.org/10.2519/jospt.2018.8131.

Bolt, W., Wendland, D., 2020. Implementation of the comprehensive care for joint replacement model: a post-acute physical therapy case report. Physiother. Theory Pract. 36 (3), 450–458. https://doi.org/10.1080/09593985.2018.1488190.

Cohen, S.P., 2015. Epidemiology, diagnosis, and treatment of neck pain. Mayo Clin. Proc. 90 (2), 284–299. https://doi.org/10.1016/j.mayocp.2014.09.008.

Crosby, R.D., Kolotkin, R.L., Williams, G.R., 2003. Defining clinically meaningful change in health-related quality of life. J. Clin. Epidemiol. 56 (5), 395–407. https://doi.org/10.1016/s0895-4356(03)00044-1.

Devji, T., Carrasco-Labra, A., Qasim, A., et al., 2020. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ 369, m1714. https://doi.org/10.1136/bmj.m171425. Published 2020 Jun 4.

Deyo, R.A., Battie, M., Beurskens, A.J., et al., 1998. Outcome measures for low back pain research. A proposal for standardized use [published correction appears in Spine 1999 Feb 15;24(4):418]. Spine 23 (18), 2003–2013. https://doi.org/10.1097/00007632-199809150-00018.

Dieleman, J.L., Cao, J., Chapin, A., et al., 2020. US health care spending by payer and health condition, 1996-2016. J. Am. Med. Assoc. 323 (9), 863–884. https://doi.org/10.1001/jama.2020.0734.

Driessen, M.T., Lin, C.W., van Tulder, M.W., 2012. Cost-effectiveness of conservative treatments for neck pain: a systematic review on economic evaluations. Eur. Spine J. 21 (8), 1441–1450. https://doi.org/10.1007/s00586-012-2272-5.

Furze, Jennifer A., Jo Tichenor, Carol, Fisher, Beth E., Jensen, Gail M., 1 July 2016. Mary jane rapport, physical therapy residency and fellowship education: reflections on the past, present, and future. Phys. Ther. 96 (7), 949–960. https://doi.org/10.2522/ptj.20150473.

Garcia, A.N., Cook, C., Lutz, A., Thigpen, C.A., 2019. Concurrent validity of the single assessment numerical evaluation and patient-reported functional measures in patients with musculoskeletal disorders: an observational study. Musculoskelet Sci Pract 44, 102057. https://doi.org/10.1016/j.msksp.2019.102057.

Hajian-Tilaki, K., 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 4 (2), 627–635. PMID: 24009950.

Horn, M.E., Brennan, G.P., George, S.Z., Harman, J.S., Bishop, M.D., 2016. A value proposition for early physical therapist management of neck pain: a retrospective cohort analysis. BMC Health Serv. Res. 16, 253. https://doi.org/10.1186/s12913-016-1504-5. Published 2016 Jul 12.

Hosmer, D.W., Lemeshow, S., 2004. In: Applied Logistic Regression, second ed. John Wiley & Sons, New York.

Hung, M., Saltzman, C.L., Kendall, R., et al., 2018a. What are the MCIDs for PROMIS, NDI, and ODI instruments among patients with spinal conditions? Clin. Orthop. Relat. Res. 476 (10), 2027–2036. https://doi.org/10.1097/CORR.0000000000000419.

Hung, M., Saltzman, C.L., Kendall, R., et al., 2018b. What are the MCIDs for PROMIS, NDI, and ODI instruments among patients with spinal conditions? Clin. Orthop. Relat. Res. 476 (10), 2027–2036. https://doi.org/10.1097/CORR.0000000000000419.

Hush, J.M., Kamper, S.J., Stanton, T.R., Ostelo, R., Refshauge, K.M., 2012. Standardized measurement of recovery from nonspecific back pain. Arch. Phys. Med. Rehabil. 93 (5), 849–855. https://doi.org/10.1016/j.apmr.2011.11.035.

IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp.

Jayadevappa, R., Malkowicz, S.B., Wittink, M., Wein, A.J., Chhatre, S., 2012 Oct. Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. Health Serv. Res. 47 (5), 1902–1925. https://doi.org/10.1111/j.1475-6773.2012.01395.x. Epub 2012 Mar 14.

Kvamme, M.K., Kristiansen, I.S., Lie, E., Kvien, T.K., 2010. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. J. Rheumatol. 37 (1), 26–31. https://doi.org/10.3899/jrheum.090449.

Lee, J.Y., Walton, D.M., Tremblay, P., May, C., Millard, W., Elliott, J.M., MacDermid, J. C., 2020 Sep 17. Defining pain and interference recovery trajectories after acute non-catastrophic musculoskeletal trauma through growth mixture modeling. BMC Muscoskel. Disord. 21 (1), 615. https://doi.org/10.1186/s12891-020-03621-7.

Lundeberg, T., Lund, I., Dahlin, L., et al., 2001. Reliability and responsiveness of three different pain assessments. J. Rehabil. Med. 33 (6), 279–283. https://doi.org/10.1080/165019701753236473.

Maitland, G.D., 2013. Vertebral Manipulation. Churchill-Livingstone, London, UK.

Maitland, G.D., 2014. Peripheral Manipulation. Churchill-Livingstone, London, UK.

Maltenfort, M., Díaz-Ledezma, C., 2017. Statistics in brief: minimum clinically important difference-availability of reliable estimates. Clin. Orthop. Relat. Res. 475 (4), 933–946. https://doi.org/10.1007/s11999-016-5204-6.

MedCalc statistical software version 19.2.6 (MedCalc software bv, Ostend, Belgium. https://www.medcalc.org, 2020.

Mokkink, L.B., Terwee, C.B., Gibbons, E., et al., 2010. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Med. Res. Methodol. 10, 82. https://doi.org/10.1186/1471-2288-10-82. Published 2010 Sep. 22.

Nayak, N.R., Coats, J.M., Abdullah, K.G., Stein, S.C., Malhotra, N.R., 2015. Tracking patient-reported outcomes in spinal disorders. Surg. Neurol. Int. 6 (Suppl 19), S490–S499. https://doi.org/10.4103/2152-7806.166892. Published 2015 Oct 8.

Nazari, G., MacDermid, J.C., Bobos, P., Furtado, R., 2020 Dec. Psychometric properties of the Single Assessment Numeric Evaluation (SANE) in patients with shoulder conditions. A systematic review. Physiotherapy 109, 33–42. https://doi.org/10.1016/j.physio.2020.02.008. Epub 2020 Mar 4.

Nazari, G., Bobos, P., Lu, S., MacDermid, J.C., 2021 Jul. Psychometric properties of the single assessment numeric evaluation in patients with lower extremity pathologies. A systematic review. Disabil. Rehabil. 43 (15), 2092–2099. https://doi.org/10.1080/09638288.2019.1693641.

O'Halloran, B., Wright, A., Cook, C.E., 2013. Criterion validation of the rate of recovery, a single alphanumeric measure, in patients with shoulder pain. Int J Sports Phys Ther 8 (6), 784–792. PMID: 24377064; PMCID: PMC3867071.

Portney, L., Watkins, M., 2004. Foundations of clinical research applications to practice. third ed. In: 19b. Hosmer DW, Lemeshow S. *Applied logistic Regression*, second ed. Oxford University Press, New York, pp. 101–109. John Wiley & Sons.

Razaeian, S., Wiese, B., Zhang, D., Krettek, C., Meller, R., Hawi, N., 2020. Correlation between Oxford elbow score and single assessment numeric evaluation: is one simple question enough? J. Shoulder Elbow Surg. 29 (6), 1223–1229. https://doi.org/10.1016/j.jse.2020.01.067.

Rodeghero, J., Wang, Y.C., Flynn, T., Cleland, J.A., Wainner, R.S., Whitman, J.M., 2015. The impact of physical therapy residency or fellowship education on clinical outcomes for patients with musculoskeletal conditions. J. Orthop. Sports Phys. Ther. 45 (2), 86–96. https://doi.org/10.2519/jospt.2015.5255.

Unal, I., 2017. Defining an optimal cut-point value in ROC analysis: an alternative approach. Comput Math Methods Med 3762651. https://doi.org/10.1155/2017/3762651, 2017.

Vaishnav, A.S., Gang, C.H., Iyer, S., McAnany, S., Albert, T., Qureshi, S.A., 2020. Correlation between NDI, PROMIS and SF-12 in cervical spine surgery. Spine J. 20 (3), 409–416. https://doi.org/10.1016/j.spinee.2019.10.017.

Warren, M., Smith, H.L., 2018 Dec 1. Exploration of functional limitation codes for outpatient physical therapy in the medicare population: a retrospective cohort study. Phys. Ther. 98 (12), 980–989. https://doi.org/10.1093/ptj/pzy104. PMID: 30184120.

Williams, G.N., Gangel, T.J., Arciero, R.A., Uhorchak, J.M., Taylor, D.C., 1999. Comparison of the Single Assessment Numeric Evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery. Am. J. Sports Med. 27 (2), 214–221. https://doi.org/10.1177/03635465990270021701.

Williams, N.H., Hendry, M., Lewis, R., Russell, I., Westmoreland, A., Wilkinson, C., 2007. Psychological response in spinal manipulation (PRISM): a systematic review of psychological outcomes in randomised controlled trials. Compl. Ther. Med. 15 (4), 271–283. https://doi.org/10.1016/j.ctim.2007.01.008.

Wright, A.A., Cook, C.E., 2013. Criterion validation of the rate of recovery, single alphanumeric measure, in patients with low back pain. Physiother. Res. Int. 18 (2), 124–129. https://doi.org/10.1002/pri.1538.