Independent predictors of reliability between full time employee-dependent acquisition of functional outcomes compared to non-full time employee-dependent methodologies: a prospective single institutional study

Owoicho Adogwa[1], Aladine A. Elsamadicy[2], Joseph Cheng[3], Carlos Bagley[4]

[1]Department of Neurosurgery, Rush University Medical Center, Chicago, IL, USA; [2]Duke University Medical Center, Durham, NC, USA; [3]Department of Neurosurgery, Yale University Medical Center, New Haven, CT, USA; [4]Department of Neurosurgery, University of Texas Southwestern Medical Center, Dallas, TX, USA

Correspondence to: Owoicho Adogwa, MD, MPH. Department of Neurosurgery, Rush University Medical Center, 1725 W, Harrison, Suite 855, Chicago IL, 60612, USA. Email: owoicho.adogwa@gmail.com.

Background: The prospective acquisition of reliable patient-reported outcomes (PROs) measures demonstrating the effectiveness of spine surgery, or lack thereof, remains a challenge. The aims of this study are to compare the reliability of functional outcomes metrics obtained using full time employee (FTE) *vs.* non-FTE-dependent methodologies and to determine the independent predictors of response reliability using non FTE-dependent methodologies

Methods: One hundred and nineteen adult patients (male: 65, female: 54) undergoing one- and two-level lumbar fusions at Duke University Medical Center were enrolled in this prospective study. Enrollment criteria included available demographic, clinical and baseline functional outcomes data. All patients were administered two similar sets of baseline questionnaires—(I) phone interviews (FTE-dependent) and (II) hardcopy in clinic (patient self-survey, non-FTE-dependent). All patients had at least a two-week washout period between phone interviews and in-clinic self-surveys to minimize effect of recall. Questionnaires included Oswestry Disability Index (ODI) and Visual Analog Back and Leg Pain Scale (VAS-BP/LP). Reliability was assessed by the degree to which patient responses to baseline questionnaires differed between both time points.

Results: About 26.89% had a history an anxiety disorder and 28.57% reported a history of depression. At least 97.47% of patients had a High School Diploma or GED, with 49.57% attaining a 4-year college degree or post-graduate degree. 29.94% reported full-time employment and 14.28% were on disability. There was a very high correlation between baseline PRO's data captured between FTE-dependent compared to non-FTE-dependent methodologies (r=0.89). In a multivariate logistic regression model, the absence of anxiety and depression, higher levels of education (college or greater) and full-time employment, were independently associated with high response reliability using non-FTE-dependent methodologies.

Conclusions: Our study suggests that capturing health-related quality of life data using non-FTE-dependent methodologies is highly reliable and maybe a more cost-effective alternative. Well-educated patients who are employed full-time appear to be the most reliable.

Keywords: Patient-reported outcomes (PROs); full time employee-dependent acquisition; phone interviews, prospective outcomes

Introduction

Patient-reported outcomes (PROs) have become increasingly popular in assessing post-surgical interventions, especially post-spine surgeries, and are becoming the new standard for measuring many clinical outcomes (1,2). Multiple factors are incorporated in the PROs collected and assessed for post-surgical interventions. In a systematic review of 13,806 post-surgical patients, Waljee and colleagues found of the PROs available 64% of the patients were assessed for disability and 25% for pain, with only 5% assessed for mood disorders (2).

There are multiple methodologies in post-surgical spine interventions that are used to collect PROs including a heavy reliance on full-time employee (FTE)-dependent (i.e., phone interviews), with uncertain non-FTE-dependent (i.e., emails and clinic questionnaires) methodologies. However, in a recent study, Adogwa *et al.* demonstrated a high correlation between PROs' data captured between FTE-dependent (via phone interviews) and non-FTE-dependent (via hard-copy clinic surveys) methodologies (3). Furthermore, PROs we assessed were Oswestry disability index (ODI) and visual analog back and leg pain scale (VAS-BP/LP).

Even though we have shown there is high reliability between FTE-dependent and non-FTE-dependent methodologies, reliability of the individual responses collected still remains relatively unknown in spine research. The aim of this study is to determine the independent predictors of response reliability using non-FTE-dependent methodologies.

Methods

Patient selection

We enrolled patients who were undergoing surgery and follow-up at Duke University Medical Center for this prospective study. Institutional Review Board approval was obtained prior to study initiation. Enrollment criteria included patients aged 18 years and older with (I) available baseline demographics; (II) who was undergoing one- and two-level lumbar fusions; (III) and who made available baseline and clinical PROs measures. Patients were excluded if patient had confounding medical conditions that either limited their ability to (I) listen and respond to phone interviews or (II) visualize and complete hardcopy questionnaires.

Patient reported outcomes (PROs)

Baseline questionnaires were administered to all patients via (I) phone interview (FTE-dependent) and (II) hardcopy in clinic (patient self-survey, non-FTE dependent) with at least two weeks in between questionnaires respectively. Back pain was assessed using the back pain-visual analog scale (BP-VAS) questionnaire, while leg pain was assessed using the leg pain-visual analog scale (LP-VAS) questionnaire. Functional status was assessed using the ODI questionnaire. These questionnaires have been validated, widely used and accepted in spine research.

Statistical analysis

We compared patient-reported pain measures and functional status between FTE-dependent and non-FTE-dependent patient-reported methodologies. Demographic variables included patient age, gender, education level, employment, medical co-morbidities [diabetes, coronary artery disease (CAD), osteoporosis], and psychological factors (anxiety disorder and depression).

Parametric data was expressed as means ± standard deviation (SD) and compared via Student *t*-test. Factors associated with high reliability using non-FTE-dependent methodologies were assessed via ANOVA measures. All tests were two-sided and were statistically significant if the P value was less than 0.05.

Results

One hundred and nineteen patients were enrolled in this study. We included patients 18 years and older with available demographics who were undergoing one- and two-level lumbar fusions and made available baseline and clinical PROs measures. We excluded patients who had confounding medical conditions that either limited their ability to complete phone interviews or hardcopy questionnaires.

Baseline patient profile

Baseline patient characteristics and education level are illustrated in *Table 1*. The average age of the cohort was 60.22±14.41, with 57.52% male and 42.48% female. 39.49% of patients were smokers, 21% had diabetes and 11.76% had CAD. About 26.89% had a history of an anxiety disorder and 28.57% reported a history of depression. At least 97.47% of

patients had a High School Diploma or GED, with 49.57% attaining a 4-year college degree or post-graduate degree, *Table 1*. 29.94% reported full-time employment and 14.28% were on disability, *Table 1*.

Baseline and clinical PRO measures

There was no statistically significant difference in patients' reporting of baseline pain and functional disability between FTE-dependent and non-FTE-dependent methodologies, *Table 2*. The mean ± SD BP-VAS score of FTE-dependent and non-FTE-dependent was 6.33±2.90 and 6.53±2.48,

P=0.57, respectively, *Table 2*. The mean ± SD LP-VAS score of FTE-dependent and non-FTE-dependent was 6.13±2.78 and 6.46±2.79, P=0.36, respectively, *Table 2*. The mean ± SD baseline ODI score of FTE-dependent and non-FTE-dependent was 47.73±16.77 and 45.81±12.11, P=0.39, respectively, *Table 2*. There was a very high correlation of baseline PROs captured via FTE-dependent versus non-FTE-dependent methodologies, r=0.89.

Independent predictors of response reliability

There were statistically significant differences in the reliability of patients' reporting of baseline pain and functional disability via non-FTE-dependent methodology with independent factors accounted for, *Table 3*. Independent factors associated with statistically significant difference include (I) depression, P=0.045; (II) anxiety disorder, P=0.025; (III) employment, P=0.032; and (IV) education level (college or greater), P=0.02, *Table 3*. Other independent factors that did not yield statistically significant differences include age, BMI, diabetes, coronary artery disease, and race, *Table 3*.

Discussion

In this prospective cohort study assessing the independent predictors of response reliability using non-FTE-dependent methodologies, we demonstrate there are statistically significant differences in high response reliability with the absence of anxiety and depression, and who have higher levels of education (college or greater) and full-time employment.

The reliability of the PROs collected by non-FTE-dependent methodologies is conditional on patients' ability to understand their symptoms, functionality, and pain and consistently report it. Cleeland *et al*. describes that the practicality of PRO measures depends on how patients

Table 1 Baseline demographic characteristics	
Characteristics	Patients (n=119)
Mean age (years)	60.22±14.41
Male (%)	57.52
Smoker (%)	39.49
Diabetes (%)	21.00
CAD (%)	11.76
Osteoporosis (%)	10.90
Anxiety disorder (%)	26.89
Depression (%)	28.57
Level of education	
Less than high school (%)	2.52
High school (%)	35.29
2-year college (%)	13.44
4-year college (%)	26.89
Post-college (%)	22.68
Employment	
Full-time employment (%)	29.94
Disability (%)	14.28

Data expressed as mean ± SD or number (%). CAD, coronary artery disease.

Table 2 Baseline Patient Reported Outcomes in back-pain (VAS-BP), leg-pain (VAS-LP) and functional disability (ODI) showed no statistically significant difference between both cohorts.

Baseline patient reported outcomes measures (mean ± SD)	FTE dependent (n=119)	Non FTE-dependent (n=119)	P value
BP-VAS	6.33±2.90	6.53±2.48	0.57
LP-VAS	6.13±2.78	6.46±2.79	0.36
ODI	47.73±16.77	45.81±12.11	0.39

BP-VAS, back pain visual analog pain scale; LP-VAS, back pain visual analog pain scale; ODI, Oswestry disability index; FTE, full-time employee.

Table 3 Independent predictors of response reliability

Independent variable	Coefficient	P value
Age	−0.054	0.140
BMI	0.63	0.252
Diabetes	0.603	0.294
CAD	0.667	0.673
Depression	−5.232	0.045
Anxiety disorder	−3.246	0.025
Race	2.35	0.98
Education (college or higher)	2.89	0.02
Employment	3.457	0.032
Anxiety disorder	−3.246	0.025
CAD, coronary artery disease.		

perceive their current health status post-treatment in relation to their previous health status pre-treatment (4). Frost *et al.* attributes the efficacy of a PRO by having significant reliability and validity measures, which he describes as two different psychometric properties (5). Furthermore, Frost *et al.* defines reliability as the degree in which a certain measure (i.e., PRO questionnaire) repeatedly produces the same outcome (i.e., number or score) while the construct (i.e., patients) has not changed (5). Therefore, even though a questionnaire may be validated, if a patient is unreliable then the efficacy of the PRO to accurately assess a patient's status is low.

There have been studies demonstrating that differences in race, ethnicity, socioeconomic status, and perceived power differential between health professionals and patients have substantial influences on the patient-reported outcomes (PROs) and affecting the reliability of the PROs being collected (6-8). In a recent prospective study of 128 patients with chronic inflammatory disease, Jamilloux *et al.* demonstrated that there were statistically significant differences in patients who adhered to 6-month period of responding to e-questionnaires once a month dependent on the patient's characteristics (9). The authors identified that the positive predictive factors included patients who were (I) married or living as a couple; (II) greater number of children living at home; and (III) participated in online surveys in the past; while the main negative predictive factor was "too busy to participate" (9). Furthermore, 86% of the patients who adhered to full 6-month period were employed (9). Analogously, we demonstrated that patients who were employed full-time were more reliable in their

responses than the patients who were part-time, retired, or disabled. A patient's education level has been a concern for reliably collecting PROs and accurately responding to questionnaires. In a review on collecting PROs in clinical practice, Rose *et al.* indicates that literacy (ability to read questions) and reading level (ability to understand the wording) may affect the reliability of PROs administered to certain populations (10). Similarly, we demonstrated that patients with a higher education level of college or higher were more significantly reliable than patients who did not have a college degree. Along with employment status, we demonstrated that in the absence of anxiety and depression, patients were more reliable with their PROs responses. In a study of identifying depression, anxiety, and anger of 15,000 respondents in the Patient-Reported Outcomes Measurement Informational System (PROMIS), Pickonis *et al.* demonstrates a high correlation (R=0.81) between respondents with depression and anxiety in how the patients responded to the questionnaires (11). Furthermore, the authors distinguished PROs in two distributions of patients who report changes in health status, (I) people with no distress and (II) people with distress (i.e., depression and anxiety), and described a skewed distribution of patients with distress leading to "artificially large discrimination parameters" (11).

Conversely, there are studies demonstrating that there are minimal differences with reliability in PROs based on patients' characteristics. In a recent study of 4,840 cancer patients evaluating the validity of PROMIS physical function form (PF), Jensen *et al.* found that the reliability of PROMIS was consistent across a diverse socio-economic and demographic population cohort (12). The authors characterized the cohort by (I) Race (White, Black, Hispanic, Asian); (II) age; (III) sex; (IV) number of comorbidities; (V) education level (< high school, high school-, college-, graduate-degree); (VI) income level (VII) current employment status; and (VIII) survey language (English, Spanish, Chinese, other); and demonstrated a reliability Cronbach's α coefficient of 0.96 within all race and age groups with the PF-16 questionnaire (12). It is important to note, that most studies combine PROs for overall assessment of post-surgical spine interventions and that our findings does not suggest eliminating PROs of certain patients when forming the cohort. Our study suggests that spine surgeons should be aware of certain patient characteristics that may yield unreliable PROs on an individual basis of practice. Understanding the variables that may skew or underscore PROs is necessary to ensure adequate post-surgical

treatment and satisfaction with the level of care the patient receives.

This study has limitations, which has implications for its interpretation. Firstly, the sample size is small, and a secondary study with a larger sample size may be warranted. Second, we cannot generalize these results for other questionnaires used to assess PROs in different settings. Third, whether these results are durable through 2 years remains unknown. Despite these limitations, this study demonstrates there are statistically significant differences in high response reliability of ODI, VAS-BP, and VAS-LP in the absence of anxiety and depression, and who have higher levels of education (college or greater) and full-time employment.

Conclusions

Our study suggests that capturing health-related quality of life data using non FTE-dependent methodologies (i.e., in-clinic surveys or email surveys) is highly reliable and patients who are well-educated, employed full-time, and do not have anxiety and depression appear to be the most reliable.

Acknowledgements

Footnote

References

1. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010;63:1179-94.
2. Waljee J, McGlinn EP, Sears ED, et al. Patient expectations and patient-reported outcomes in surgery: a systematic review. Surgery 2014;155:799-808.
3. Adogwa O, Elsamadicy AA, Cheng J, et al. Assessing Patient Reported Outcomes Measures Via Phone Interviews Versus Patient Self-Survey in Clinic: Are we Measuring the Same Thing? World Neurosurg 2015.
4. Cleeland CS. Symptom burden: multiple symptoms and their impact as patient-reported outcomes. J Natl Cancer Inst Monogr 2007;(37):16-21.
5. Frost MH, Reeve BB, Liepa AM, et al. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health 2007;10 Suppl 2:S94-S105.
6. Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: challenges and opportunities. Qual Life Res 2009;18:99-107.
7. Hill ME. Race of the interviewer and perception of skin color: Evidence from the Multi-City Study of Urban Inequality. American Sociological Review 2002;67:99-108.
8. PICKERY J, LOOSVELDT G, CARTON A. The Effects of Interviewer and Respondent Characteristics on Response Behavior in Panel Surveys A Multilevel Approach. Sociological Methods Research 2001;29:509-23.
9. Jamilloux Y, Sarabi M, Kérever S, et al. Adherence to online monitoring of patient-reported outcomes by patients with chronic inflammatory diseases: a feasibility study. Lupus 2015;24:1429-36.
10. Rose M, Bezjak A. Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples. Qual Life Res 2009;18:125-36.
11. Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment 2011;18:263-83.
12. Jensen RE, Potosky AL, Reeve BB, et al. Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res 2015;24:2333-44.