A Comparison of PROMIS Physical Function and Pain Interference Scores in Patients With Carpal Tunnel Syndrome: Research Collection Versus Routine Clinical Collection

David N. Bernstein¹ , Bilal Mahmood¹, Constantinos Ketonis¹, and Warren C. Hammert¹

Abstract

Background: There is a concern that patients may answer patient-reported outcome (PRO) questionnaires differently depending on the purpose—clinical care or research (eg, “Hawthorne effect”). We sought to determine whether Patient-Reported Outcomes Management Information System (PROMIS) scores differ at the same clinic visit based on whether a patient was completing the PRO tool for study or clinical care purposes. **Methods:** Patients presenting to one surgeon at an academic medical center hand clinic were asked to complete PROMIS Physical Function (PF) and Pain Interference (PI) questionnaires as part of routine care. Those diagnosed with carpal tunnel syndrome from February 2015 to April 2017 were then asked to participate in a clinical research project, which had them complete PROMIS PF and PROMIS PI again. Data from those who completed both routine and research PROs at each visit were compared. Between the 2 settings, test-retest reliability was determined using Pearson correlation coefficients (r), and internal consistency was evaluated using Cronbach α . **Results:** A total of 128 unique office visits representing 67 patients fit our inclusion criteria. There was a strong correlation between PROMIS PF and PI in the research and patient care setting (PF: $r = 0.82$, $P < .01$; PI: $r = 0.83$, $P < .01$). Both domains had a Cronbach α of 0.90. The PROMIS PF scores were not different between the 2 groups ($P = .19$), but the PROMIS PI scores were slightly different ($P < .01$). **Conclusions:** Patients appear to be consistent when completing PROMIS for both clinical care and research, supporting the idea that data obtained in either setting are generalizable and appropriate for research purposes.

Keywords: hand surgery, patient-reported outcomes, PROMIS, test-retest reliability, bias

Introduction

With the increased shift toward a health care system focused on delivering higher value—defined as health outcome achieved per dollar spent¹—patient-reported outcome (PRO) instruments have become more common in clinical care settings. Patient-Reported Outcomes Management Information System (PROMIS),² a generic PRO tool developed with the support of the National Institutes of Health (NIH), is one option. Prior research has demonstrated that preoperative PROMIS scores can predict notable score improvement postoperatively under traumatic and atraumatic conditions.³ In addition, within the upper extremity literature, a number of studies have compared PROMIS domains with upper extremity and hand-specific PRO tools such as the Quick Disabilities of the Arm, Shoulder and

Hand (QuickDASH)^{2,4} and Disabilities of the Arm, Shoulder and Hand.⁵ The conclusions of this previous research suggest PROMIS domains capture hand and upper extremity conditions as well as these more traditional PRO instruments, offer insight into clinical improvement, and may be more favorable given their shorter time to completion.

Within the orthopedic surgery literature, prior research has evaluated whether certain situations alter PRO instrument scores provided by patients. One major concern is

¹University of Rochester Medical Center, NY, USA

Corresponding Author:

Warren C. Hammert, Department of Orthopaedics & Rehabilitation, University of Rochester Medical Center, 601 Elmwood Avenue, Box 665, Rochester, NY 14642, USA.
Email: warren_hammert@urmc.rochester.edu

the impact of the “Hawthorne effect,” which suggests that patients may answer questions differently if they are aware they are part of a study. Similar to the current study, Buckley et al⁶ evaluated a cohort of 74 patients undergoing carpal tunnel release to determine whether there were differences in QuickDASH scores between patients who were versus patients who were not aware of being a part of an observational study; no significant differences were noted. In a study of nearly 80 000 patient visits to an hand surgeon, Guattery et al⁷ found that the PROMIS Depression domain had a notable floor effect, but the PROMIS Physical Function (PF) domain was normally distributed. These results suggest that the situation matters, as the authors concluded that patients may be reluctant to report depressive symptoms to an orthopedic surgeon but are more comfortable with completing accurate PROMIS PF information.⁷ Given the increasing research involving PROMIS, as well as the fact that the clinical application of PROMIS is gaining traction, it becomes important to ensure PROMIS scores are consistent in the 2 most common situations—clinical research and clinical care settings. Our work would build upon the prior research by focusing on a general PRO instrument (ie, PROMIS) that is growing in popularity.

By understanding whether there are differences between study scenarios and patient care settings, future research design can be even more informed. This study aimed to determine the test-retest reliability of PROMIS PF and Pain Interference (PI) scores completed on the same data based on whether a patient was completing the PRO questionnaire for study purposes or for clinical care purposes. In addition, this study looked to evaluate internal consistency of the PROMIS PF and PI domains between the 2 questionnaire administrations.

Materials and Methods

The institutional review board (IRB) approved the prospective study in which the research data were collected; a separate IRB was obtained to review the data collected as part of clinical care.

Between February 2015 and April 2017, all patients presenting to a single, urban academic medical center hand clinic to seek care from one fellowship-trained hand surgeon were asked to complete PROMIS PF and PI domains using Apple iPads as part of routine clinical care at clinic check-in.⁸ Within hand clinics, the use of Apple iPads to collect PROs has been shown to be more efficient and preferable than traditional pen-and-paper methods.⁹ The PROMIS uses item response theory as part of a computerized adaptive test, and in-depth analysis of PROMIS has shown that it is reliable and precise as a PRO tool.¹⁰ The PROMIS domains have a mean *t* score of 50 with a standard deviation (SD) of 10. Of note, higher PROMIS PF scores indicate

Table 1. Patient and PROMIS Score Characteristics (N = 128).

Characteristic	Description of the Characteristic	P value
Age, y, mean (SD)	58 (13)	
Sex, No. (%)		
Men	49 (38)	
Women	79 (62)	
PROMIS PF score, mean (SD)		
PF, study	47.95 (7.33)	P = .19
PF, clinic	47.45 (7.27)	
PROMIS PI scores, mean (SD)		
PI, study	54.56 (8.77)	P < .01
PI, clinic	56.65 (7.35)	
PROMIS score changes		
Delta PF, mean (SD)	0.50 (4.34)	
Delta PI, mean (SD)	−2.09 (4.85)	

Note. PROMIS = Patient-Reported Outcomes Management Information System; PF = Physical Function; PI = Pain Interference.

increased physical function, whereas lower PROMIS PI scores represent less pain.

While in office, patients diagnosed with carpal tunnel syndrome were asked to participate in a clinical research project in addition to the PROMIS data gathered as part of our institution’s standard clinic care. This typically occurred within 30 minutes of completing the PROMIS domain questionnaires for part of routine clinical care at check-in. This timing was used for one key reason: to remove bias introduced by a change in symptoms over the days following the routine clinical data collection. If a patient agreed to participate, additional PRO instruments, including PROMIS PF, PROMIS PI, Michigan Hand Questionnaire, and Boston Carpal Tunnel Questionnaire were completed. The PRO results were then stored using REDCap. Only patients who completed both routine and research PROs at each visit were compared. The data were collected for a study to compare change in PROMIS scores with traditional legacy measures for carpal tunnel syndrome.

Descriptive statistics were calculated and reported. Paired *t* tests were used to compare the means between clinic and study scenario PROMIS domain scores. Similar to previously published work, test-retest reliability and internal consistency were determined using Pearson correlation coefficients (*r*) and Cronbach α .¹¹ Consistent with prior research, a reliability of >0.70 was considered acceptable¹¹ and a Cronbach α of at least 0.80 was considered evidence of good internal consistency.¹² Significance was set a priori at *P* < .05.

Results

A total of 128 unique office visits representing 67 patients fit our inclusion criteria. Our sample was a majority women

Table 2. Pearson Correlation Coefficients (*r*) and Cronbach α for PROMIS Scores in Each Setting.

Comparison	<i>r</i>	Reliable?	Cronbach α
Patient Care PI/Study PI	0.83 ($P < .01$)	Yes	0.90
Patient Care PF/Study PF	0.82 ($P < .01$)	Yes	0.90

Note. PROMIS = Patient-Reported Outcomes Management Information System; PI = Pain Interference; PF = Physical Function.

(62%), and the average age was 58 years (SD = 13) (Table 1). As part of the study protocol, the average PROMIS PF and PI scores were 47.95 (SD = 7.33) and 54.56 (SD = 8.77), respectively. As part of the standard clinic care, the average PROMIS PF and PI scores were 47.45 (SD = 7.27) and 56.65 (SD = 7.35), respectively. The PROMIS PF and PI average changes in scores between the clinic and study scenarios were 0.50 (SD = 4.34) and -2.09 (SD = 4.85), respectively. Overall, PROMIS PF scores were not significantly different between scenarios (47.95 [95% confidence interval (CI): 47.78-48.13] versus 47.45 [95% CI: 47.27-47.62], $P = 0.19$) However, PROMIS PI scores were significantly lower (ie, less pain interference) in the study scenario (54.56 [95% CI: 54.39-54.73] vs 56.65 [95% CI: 56.48-56.83], $P < .01$).

Pearson correlation coefficients (*r*) between study and clinic settings for PROMIS PF and PI were 0.82 ($P < .01$) and 0.83 ($P < .01$), respectively (Table 2). When comparing each PROMIS domain in the clinic versus study scenarios, Cronbach α was 0.90 for both PROMIS PF and PI.

Discussion

Our work builds upon previous research that suggests that orthopedic patients do not answer physical function and symptom PRO questionnaires differently based on whether or not they are aware they are part of a research study.⁶ However, this is the first study we are aware of that conducts such an analysis using PROMIS. There is a concern that some patients may respond differently when completing information for research purposes due to the desire to please their physician, but this has not been documented. In the current study, the strong correlation between PROMIS PF and PI scores acquired on the same day in a study scenario and patient care setting suggests test-retest reliability. Furthermore, there was good internal consistency between the 2 settings. The PROMIS PF did not demonstrate a difference between the 2 scenarios, whereas PROMIS PI did. The difference found in PROMIS PI scores may be due to a myriad of factors biasing the results in each setting. In this case, the small difference, although statistically significance, likely does not represent a clinical difference. This is especially true given the larger importance of change in PROMIS scores before and after treatment instead of scores at one

point in time. To put into perspective, according to the distribution method, a change of half one SD following treatment has been accepted by many in the literature as the minimal clinically important difference or the smallest PRO score increase associated with true clinical improvement.¹³ For PROMIS, which is designed to have a mean *t* score of 50 and an SD of 10, this would be about 5; this understanding helps support the idea that the 2-point difference in PROMIS PI scores based on setting is not of notable clinical importance.

Our study results must be evaluated with the limitations of our work in mind. First, data were gathered from patients who sought care from a single surgeon at a hand clinic located at one urban tertiary care center. Second, only those with diagnosed carpal tunnel syndrome were included in our analysis. These 2 limitations may constrain the generalizability of our findings, as the results for patients with more traumatic conditions may differ. Future research can verify that our findings are consistent across a greater range of hand diagnoses. Third, while our study design allowed the clinic check-in process, including PROMIS completion, not to be altered from baseline clinical care, it could have biased our results as the order of questionnaire completion (research vs clinical care) was not randomized, and this could affect our findings. Future work can better determine whether patients act differently if they fill out PROMIS questionnaires in reverse of this study's order. Finally, we were limited by the availability of the PROMIS domains at our institution at the time of the study. We used PROMIS PF because scores were routinely collected in our study protocol, as well as in the patient care setting. In contrast, while PROMIS Upper Extremity (UE) was collected as part of our study protocol, the domain was not actively collected in patient care settings at that time. Thus, we were unable to ask the same study questions of the PROMIS UE results; future work can better evaluate PROMIS UE test-retest reliability and internal consistency.

A valuable feature of a PRO tool is the reproducibility of the same results if no health change has occurred. However, the question that arises is what timeframe between the first test administration and second test administration is appropriate. In our study, we asked patients to complete the PROMIS PF and PI domains for research and clinical care during the same office visit on the same day. Previous test-retest reliability studies with PRO instruments, such as the Quick Foot and Ankle Ability Measure, used a 4-week time interval,¹⁴ while another evaluating PROMIS social health domains in caregivers of civilians and veterans with traumatic brain injury used a 3-week interval.¹⁵ As our research data were collected separately and after clinical care data and on the same day, we felt that evaluating PROMIS scores on the same day would help clarify any differences without confounding factors, such as change in overall health or their change in symptoms following carpal tunnel

release surgery. This further supported our belief that evaluating PROMIS scores from the same day may be best, as this protected against bias from other conditions affecting our results.

However, there is concern that conducting test-retest reliability using results from the same day may introduce bias, as patients may remember the questions asked and alter answers. Specifically, this may be evident in the fact that patients demonstrated significant (albeit small) improved pain scores when PROMIS PI was completed for research purposes and not clinical care. While the timeframe of the analysis may be one reason for this finding, another may be that patients—consciously or subconsciously—wanted to present worse off to the surgeon to get what was perceived to be more and/or better care. Nonetheless, it is important to note that the difference, while statistically significant, is quite small and likely not clinically relevant. Furthermore, the fact that only one of the PROMIS domains demonstrated a slight difference based on scenario further supports this belief. In addition, our results suggest that PROMIS PF and PI have test-retest reliability and internal consistency, but the calculated values are not perfect (eg, $r \neq 1$; $\alpha \neq 1$). This may also account for the minor changes in PROMIS PI scores noted between the study scenario and patient care settings. Prior research in orthopedics, including in hand and foot and ankle care, asking similar questions of PRO tools also found slight differences as well despite test-retest reliability and internal consistency findings.^{16,17}

The PROMIS PF and PI have been validated across a wide range of clinical settings.^{18,19} We sought to ensure and confirm that these 2 domains demonstrated adequate test-retest reliability (ie, stability) and internal consistency within a population seeking care for carpal tunnel syndrome. With an inconsistent definition of health care quality in orthopedic care²⁰ and the current shift toward value-based medicine, there is an increased call to incorporate PROs into patient care as a means to help improve outcomes and offer patients a greater say in their care.²¹ However, for these changes to occur, the research conducted must use reliable data. Our work helps support the idea that data obtained as part of routine care or in a study scenario can be used for clinical research.

Ethical Approval

The Institutional Review Board of the University of Rochester Medical Center approved this study.

Statement of Human and Animal Rights

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5).

Statement of Informed Consent

This is a database review. We have institutional review board approval for review of the data in the hand repository, but individual patient consent is not required.

Statement on Research Location

The work was performed at the Department of Orthopedics and Rehabilitation, University of Rochester Medical Center, Rochester, New York

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: DNB received the AOA Carolyn L. Kuckein Student Research Fellowship. WCH received personal fees from ExsoMed and is Editor at Hand-e (American Society for Surgery of the Hand [ASSH]). All other authors (BM, CK) certify that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements) that might pose a conflict of interest in connection with the submitted article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

David N. Bernstein <https://orcid.org/0000-0002-1784-3288>

References

- Porter ME. A strategy for health care reform—toward a value-based system. *N Engl J Med*. 2009;361(2):109-112.
- Jones RS, Stukenborg GJ. Patient-Reported Outcomes Measurement Information System (PROMIS) use in surgical care: a scoping study. *J Am Coll Surg*. 2017;224:245-254.e1.
- Bernstein DN, Houck JR, Gonzalez RM, et al. Preoperative PROMIS scores predict postoperative PROMIS score improvement for patients undergoing hand surgery [published online ahead of print August 3, 2018]. *Hand (N Y)*. doi:10.1177/1558944718791188.
- Doring AC, Nota SP, Hageman MG, et al. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *J Hand Surg Am*. 2014;39(6):1160-1165.
- Tyser AR, Beckmann J, Franklin JD, et al. Evaluation of the PROMIS physical function computer adaptive test in the upper extremity. *J Hand Surg Am*. 2014;39(10):2047-2051.e4.
- Buckley T, Mitten D, Elfar J. The effect of informed consent on results of a standard upper extremity intake questionnaire. *J Hand Surg Am*. 2013;38(2):366-371.
- Guattery JM, Dardas AZ, Kelly M, et al. Floor effect of PROMIS depression CAT associated with hasty completion in orthopaedic surgery patients. *Clin Orthop Relat Res*. 2018;476(4):696-703.
- Papuga MO, Dasilva C, McIntyre A, et al. Large-scale clinical implementation of PROMIS computer adaptive testing with

- direct incorporation into the electronic medical record. *Health Syst.* 2017;6(15):1-13.
9. Yaffe M, Goyal N, Kokmeyer D, et al. The use of an iPad to collect patient-reported functional outcome measures in hand surgery. *Hand (N Y)*. 2015;10(3):522-528.
 10. 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(11):1179-1194.
 11. Bartlett SJ, Orbai AM, Duncan T, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. *PLoS ONE*. 2015;10(9):e0138543.
 12. Prey JE, Qian M, Restaino S, et al. Reliability and validity of the patient activation measure in hospitalized patients. *Patient Educ Couns.* 2016;99(12):2026-2033.
 13. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003;41(5):582-592.
 14. Hoch JM, Powden CJ, Hoch MC. Reliability, minimal detectable change, and responsiveness of the Quick-FAAM. *Phys Ther Sport.* 2018;32:269-272.
 15. Carlozzi NE, Ianni PA, Lange RT, et al. Understanding health-related quality of life of caregivers of civilians and service members/veterans with Traumatic Brain Injury: establishing the reliability and validity of PROMIS social health measures [published online ahead of print July 31, 2018]. *Arch Phys Med Rehabil*. pii: S0003-9993(18)30884-0. doi: 10.1016/j.apmr.2018.06.026.
 16. Briet JP, Bot AG, Hageman MG, et al. The pain self-efficacy questionnaire: validation of an abbreviated two-item questionnaire. *Psychosomatics.* 2014;55(6):578-585.
 17. Garratt AM, Naumann MG, Sigurdson U, et al. Evaluation of three patient reported outcome measures following operative fixation of closed ankle fractures. *BMC Musculoskelet Disord.* 2018;19(1):134.
 18. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. *Pain.* 2010;150(1):173-182.
 19. Schalet BD, Hays RD, Jensen SE, et al. Validity of PROMIS physical function measured in diverse clinical samples. *J Clin Epidemiol.* 2016;73:112-118.
 20. Bernstein DN, Mesfin A, Bozic KJ. Total joint arthroplasty quality ratings: how are they similar and how are they different? *Am J Orthop (Belle Mead NJ)*. 2018;47(7). doi: 10.12788/ajo.2018.0060.
 21. Baumhauer JF. Patient-reported outcomes: are they living up to their potential? *N Engl J Med.* 2017;377(1):6-9.