



Reliability, Construct Validity, and Measurement Invariance of the PROMIS Physical Function 8b - Adult Short Form v2.0

Du Feng, Fimbel Laurel, Dorothy Castille, Alma Knows His Gun McCormick, Suzanne Held

Feng, D., Laurel, F., Castille, D., McCormick, A. K. H. G., & Held, S. (2020). Reliability, construct validity, and measurement invariance of the PROMIS Physical Function 8b—Adult Short Form v2.0. *Quality of Life Research*, 29(12), 3397-3406.

<https://doi.org/10.1007/s11136-020-02603-5>

This is a post-peer-review, pre-copyedit version of an article published in *Quality of Life Research*. The final authenticated version is available online at: <https://doi.org/10.1007/s11136-020-02603-5>. The following terms of use apply: <https://www.springer.com/gp/open-access/publication-policies/aam-terms-of-use>.

Made available through Montana State University's [ScholarWorks](https://scholarworks.montana.edu)
scholarworks.montana.edu

Published in final edited form as:

Qual Life Res. 2020 December ; 29(12): 3397–3406. doi:10.1007/s11136-020-02603-5.

Reliability, Construct Validity, and Measurement Invariance of the PROMIS Physical Function 8b - Adult Short Form v2.0

Du Feng, Ph.D. [Professor and Associate Dean for Research],

University of Nevada, Las Vegas, School of Nursing, Las Vegas, Nevada, USA

Fimbel Laurel [Undergraduate Student and Researcher],

Montana State University, Department of Health and Human Development, Bozeman, Montana, USA

Dorothy Castille, Ph.D. [Health Scientist Administrator/Project Scientist],

National Institute on Minority Health and Health Disparities/National Institutes of Health, Bethesda, Maryland, USA

Alma Knows His Gun McCormick [Executive Director],

Messengers for Health, Crow Agency, Montana, USA

Suzanne Held, Ph.D. [Professor]

Montana State University, Department of Health and Human Development, Bozeman, Montana, USA

Abstract

Purpose—The National Institutes of Health established the Patient-Reported Outcomes Measurement Information System (PROMIS) to assess health across various chronic illnesses. The standardized PROMIS measures have been used to assess symptoms in studies that included Native American participants, although the psychometric properties of these measures have not been assessed among a solely Native American population. This study aimed to assess the reliability, construct validity, and measurement invariance of a widely used PROMIS Physical Function survey among Native Americans residing on or near the Apsáalooke (Crow) Reservation who were living with chronic illnesses.

Methods—Participants aged 24 to 82 years and living with at least one chronic illness were recruited for a community-based participatory research project. Baseline data were used for the current study ($N = 210$). The 8-item PROMIS Physical Function 8b – Adult Short Form v2.0 was used to assess the function of upper and lower extremities, central core regions, and the ability to complete daily activities on a 5-point Likert scale.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Corresponding author: Du Feng; du.feng@unlv.edu.

Publisher's Disclaimer: This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Results—Results indicated that the above PROMIS survey had high internal consistency (Cronbach's $\alpha = .95$) and split-half ($r = .92$, $p < .001$) reliabilities. Confirmatory factor analyses supported construct validity among females of the above population and when the two sex groups were combined. Results also indicated that corresponding thresholds and factor loadings were invariant across male and female groups.

Conclusions—The above PROMIS measure had good psychometric properties in females and when the two sex groups were combined among Native American adults living on or near the Apsáalooke reservation with chronic illnesses. Thresholds and factor loadings appeared to be invariant by sex. Future studies with a larger sample size among males and more studies on the psychometric properties of other PROMIS measures among Native American populations are needed.

INTRODUCTION

PROMIS® Measures of Physical Function

Patient-reported outcomes (PROs) are “self-report” forms that are used commonly by researchers and clinicians. PROs allow patients who live with chronic illnesses to provide their perspectives to physicians in an efficient way [1] and have been found to be effective for self-reporting on emotional and physical well-being [2]. By self-reporting on their emotional, social, and physical well-being, PRO measures allow the patient experience to be an integral and important part of treatment intervention [3]. There are numerous PRO measures available for clinicians and researchers to use, and PRO measures of physical function (PF) are commonly used in research on patient symptoms, especially among patients living with illnesses that seriously limit their strength or mobility. Although many PRO measures have undergone extensive validation, they have not all been widely tested for reliability and validity among non-white populations.

Inconsistency between different PRO measures made it difficult to compare studies designed to measure the same latent constructs. Thus, the National Institutes of Health (NIH) supported the development of the Patient-Reported Outcomes Measurement Information System (PROMIS) to be used as standardized measures to assess health across various chronic illnesses and across data sets [4, 5]. Developed from a variety of widely validated “legacy” measures, PROMIS provides a large bank of survey questions to measure latent constructs pertaining to various domains of health and well-being, including physical, emotional, and social health [6, 7]. PROMIS measures have been used to assess symptoms in studies that included Native American participants among participants of other ethnicities [8], although the psychometric properties of these measures have not been assessed among a solely Native American population. The reliability and validity of PROMIS measures of physical function have been documented in epidemiological studies [9, 10], and among various patient populations. For example, the PF measures were validated among patients with liver cirrhosis [11], cancer [12, 13], orthopedic injury [14], and arthritis [15, 16]. However, the psychometric properties of PROMIS PF measures have been predominately tested among white populations, with some research in the black and Hispanic populations [9, 12]. Our literature search did not find any existing studies that assessed the reliability and

validity of any version of the PROMIS Physical Function Adult Short Form in Native American populations.

Physical Function of Native Americans

Native American populations are at an elevated risk of developing physically debilitating health conditions due to a variety of historical, environmental, and socioeconomic factors. Studies have shown that Native American populations suffer from an increased risk of diabetes, hypertension, and metabolic illnesses [17–19]. One physical function study within an elderly Native American population determined that lower body physical function abilities decreased drastically with age, comorbidity, and an inactive lifestyle [20]. The study used the Short Physical Performance Battery (SPPB) physical function test to collect data from participants and stated that the SPPB form is an appropriate screening test for underserved communities [20]. Having measurement instruments that are valid and reliable for specific populations is critical to understanding and improving health outcomes.

The Apsáalooke

The Apsáalooke (Crow) people traditionally lived a healthy lifestyle and maintained a rich culture and spiritual connection to the land and environment. The Apsáalooke Reservation, located in a rural setting in southeastern Montana, covers 2.3 million acres and is the fifth largest Native American (NA) reservation in the United States. The vast majority of the reservation lies within the borders of Big Horn County, and it is the largest reservation in the state. Approximately 14,000 individuals are enrolled as members of the Apsáalooke Nation, and around 75% of those individuals live on the reservation [21].

The Apsáalooke, like many Native American communities, have undergone a prolonged history of discrimination, mistreatment, and exploitation. The Fort Laramie Treaty established the Apsáalooke Reservation in 1851. Following this treaty, the Apsáalooke were entitled to supplies and rations, which they were frequently denied [22]. Conditions worsened following the second Fort Laramie Treaty, signed in 1868, which decreased the 38-million acre reservation to 2.25 million acres and restricted the movement of the Apsáalooke [22]. Currently, the Apsáalooke reservation reports a relatively high prevalence of chronic illnesses and unemployment; however, actions to improve community health and wellness through community-engaged research, outreach, and education are making a positive impact. Apsáalooke students and community members are taking the lead in bringing about positive changes [23].

Messengers for Health, a partnership between members of the Apsáalooke Nation and faculty and students at Montana State University, began in 1996. In 2008, the Apsáalooke arm of the partnership transitioned into an Indigenous 501(c)(3) non-profit organization. Messengers for Health is currently implementing the Báa nnilah program, a community-based research intervention focused on improving the quality of life for individuals living with chronic illness on or near the Apsáalooke reservation. In this study, we assessed the measurement properties (i.e., reliability, construct validity, and measurement invariance by sex) of the PROMIS Physical Function 8b - Adult Short Form v2.0 in a sample of Báa nnilah participants.

METHODS

Study Sample

Participants ($N = 211$) were recruited from on or near the Apsáalooke reservation in Southeastern Montana through social events, brochures, and research team member's invitations. Inclusion criteria included being Native American, 24 years or older and living with a chronic illness. The project was approved by the Montana State University Institutional Review Board (IRB) due to the Apsáalooke Nation not having an active IRB. Consented participants completed a self-report health survey using computer tablets at baseline, post-intervention, and six months after the post-test. Baseline data were used in the current study.

One participant did not answer any PROMIS Physical Function items (i.e., had missing data for all eight items) and was excluded from the current study, resulting in an analytical sample size of $N = 210$, with more female participants ($n = 151$, 72%) compared to males ($n = 59$, 28%). The age of participants ranged from 24 to 82 years (median = 54 years; mean = 52.01 years, SD = 13.45 years). More details on the demographic characteristics of the study sample are in Table 1 in the Results section.

Instrument

The 8-item PROMIS Physical Function 8b – Adult Short Form v2.0 is a self-report survey on physical function capabilities. This survey assesses one's function of upper and lower extremities, central core regions, and the ability to complete daily activities. Questions assessing upper extremity movement relate to activities that require hand, arm, and shoulder coordination, such as vacuuming. Questions assessing lower extremity movements relate to mobility, such as the ability to walk for at least 15 minutes. Questions assessing central movement regions relate to the movement of the back and neck and bending, such as scrubbing floors. This survey was created for adults with chronic illnesses and ongoing health conditions. Apsáalooke research partners reviewed each survey item used in the study to determine if the item would be clear, understandable, easy to answer and appropriate for use among the community members. There were no suggestions for changes for any items in the PROMIS 8-item physical function scale, although there were changes in other surveys (e.g., the team worked with the scientists responsible for the SF-12 survey and made changes to items measuring physical function to make activities more relevant to the community).

The above mentioned eight questions were asked using a Likert scale response format inquiring about the participant's individual evaluation of their current physical abilities (see Table 1 for details). The Likert scale response range is 1 = "unable to do/not at all" to 5 = "without any difficulty" for each question. The web-based application HealthMeasures Scoring Service powered by Assessment CenterSM [24] was used to calculate T-scores for the scale score along with the standard error (SE) [25]. The raw item scores were submitted online using a specialized template. Response pattern scoring was used to calculate the scale score, where all valid item scores were used, and the scale score was adjusted for missing data on individual items [26]. Thus, this response pattern scoring allowed for the calculation of a scale score for each participant even if some individual items were not completed.

Data Analysis

Descriptive statistics, reliability analyses, confirmatory factor analyses (CFA), and measurement invariance (ME/I) tests were conducted using *R* version 4.0.0. Sex difference in the group means of the T-score of the Physical Function scale was tested by an independent *t*-test at the .05 alpha level, two-tailed. For item-level correlations, polychoric correlations, which are appropriate for ordinal data collected using a Likert scale, were obtained along with significance tests using SAS/STAT® software. Internal consistency reliability (Cronbach's α) and split-half reliability were computed, where reliability coefficient $\geq .90$ was considered "good" [27].

The *R* package "*lavaan*" Version 0.6–6 [28] was used for the confirmatory factor analyses (CFA) to test construct validity. The chi-square (χ^2) test, comparative fit index (CFI), Tucker Lewis index (TLI), and root mean squared error of approximation (RMSEA) were used to assess the goodness-of-fit of the measurement models. Non-significant χ^2 , CFI and TLI $\geq .95$, and RMSEA $\leq .05$ indicate good fit, although an upper limit of .07 for RMSEA has been suggested for a well-fitting model [29, 30].

Measurement invariance across sex groups was tested using the Wu & Estabrook procedure (ID.cat = "Wu.Estabrook.2016") of the *R* package *semTools* Version 0.5–3 [31]. A typical approach to testing ME/I is through evaluation and statistical comparison of a series of multiple-group CFA (MGCFA) models with increasing restrictions on the parameters. For example, a review and synthesis of the ME/I literature [32] suggested that MGCFA models representing configural invariance (where only the number and pattern of parameters are identical across groups), metric invariance (where factor loadings are constrained across groups), scalar invariance (where factor loadings and item intercepts are constrained), and strict invariance (where equal factor variances are established and then constraint on error variances is added) should be tested successively. The MGCFA approach to establishing ME/I was found to perform better with polytomous data when compared with the item response theory (IRT) approach [33]. However, Wu and Estabrook (2016) argued that the traditional MGCFA approach to testing ME/I is not optimal for ordinal data (e.g., data collected using Likert-type items), and proposed a series of solutions to ME/I model identifications [34]. A recent article provided steps to apply the Wu and Estabrook approach, which focused on 1) establishing a baseline model that is equivalent to configural invariance, 2) testing a model where the thresholds are assumed invariant across all groups, and 3) testing a model where the thresholds and loadings assumed invariant across all groups [35]. Following the updated guidelines to testing multiple-group ME/I with categorical outcomes [35], we tested a series of three single-factor MGCFA models which represented the baseline/configural invariance, threshold invariance (Proposition 4 by Wu and Estabrook, 2016), and threshold and loading invariance (Proposition 7 by Wu and Estabrook, 2016) across two sex groups (males and females), respectively. Each model (except for the baseline model) was compared to the previous model using a χ^2 difference test.

RESULTS

Demographic Characteristics of the Study Sample

Table 1 shows the demographic characteristics of the study sample ($N = 210$). A majority (72%) of the participants were female ($n = 151$), and 49% ($n = 102$) were married or living in a marriage-like relationship. Approximately one-third of the participants (33%, $n = 68$) had an annual household income less than \$10K and another one-third (33%, $n = 69$) had an annual household income between \$10K and \$24,999 (25%, $n = 51$), whereas only 8% ($n = 19$) had an annual household income of \$50K or greater. All participants had at least one ongoing chronic illness, with the majority (70%) having comorbidities. As seen in Table 1, the most common chronic illnesses were diabetes (57%), high blood pressure (55%), arthritis (34%), and chronic pain (35%). On average, a participant had approximately four doctor visits in the past four months ($mean = 4.13$, $SD = 6.36$), and traveled about 19 miles ($mean = 18.90$ miles, $SD = 19.68$ miles) one-way to see their health care provider. There is a large range of distance to the clinic, with 28 participants living within a mile from the clinic, 17 participants traveling 50+ miles each way to the clinic, and the longest distance being 90 miles. Participants were contacted to ensure the accuracy of their self-report data for outlier responses.

Descriptive Statistics of the Items

The actual range of each item was 1 to 5. Missing data on individual items ranged from 0 to 4 (i.e., less than 2%). The distribution of all items was negatively skewed, but none of the distribution was highly skewed, with the highest absolute value of skewness being 1.12. The negative skewness (ranging from $-.20$ to -1.12), along with item means (see Table 2) that ranged from 3.46 (doing 2 hours of physical labor and heavy lifting, respectively) to 4.20 (running errands), indicates that the majority of participants scored closer to the upper (i.e., high physical functioning) end of the distribution, but there are some extreme cases at the lower end of the distribution. An examination of the histograms and box-plots (results not shown) confirmed the above descriptive statistics findings. No distribution was multiple modal (kurtosis ranged from -1.03 to $.31$).

The total scale score in T-scores was 45.22 ± 8.75 (skewness = $.20$, kurtosis = $-.56$) for the full sample, 47.49 ± 9.07 (skewness = $.02$, kurtosis = $-.89$) among males, and 44.34 ± 8.49 (skewness = $.26$, kurtosis = $.32$) among females. An independent t -test revealed that the above male group means was significantly higher than that of the females, $t = 2.37$, $df = 208$, $p = .02$.

As can be seen in Table 3, inter-item polychoric correlations based on the full sample, ranging from $.68$ to $.85$, were high, positive, and significant ($p < .001$). Similarly, inter-item polychoric correlations in each sex group (see Table 4) were high (ranging from $.71$ to $.94$ in males, and $.63$ to $.85$ in females). No systematic sex difference in the pattern of the item-level correlation matrix was detected.

Reliability

Internal consistency reliability was high (Cronbach's $\alpha = .95$). Split-half reliability, $r = .92$, $p < .001$, which was calculated as the correlations (Pearson's r) between the mean of odd-numbered items and the mean of even-numbered items, was above the .9 common cutoff for high split-half reliability.

Confirmatory Factor analysis

Model fit indices of a single-factor CFA model with all eight items (which were identified as ordinal variables) loading on the same latent variable (Physical Function) are reported in Table 5. The χ^2 test of the fit statistic based on the diagonally-weighted least squares (DWLS) method was significant ($p = .025$) for the full sample, but not significant at the .05 level for either sex ($p = .862$ for males; $p = .105$ for females). Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) were .999 or above, RMSEA was below .06, and the χ^2/df ratio was below the cutoff of 2 in the full sample and each of the male and female subsets. Overall, these results indicated adequate/good fit of the single-factor CFA model for the full sample and for the female subsample. This model was used in further analyses of measurement invariance.

However, it should be noted that the variance-covariance matrix of the estimated parameters did not appear to be positive definite for the male subsample, where the smallest eigenvalue = $-7.166567e-17$. Although the above negative but near-zero eigenvalue (with 16 zeros after the decimal place) could be due to a machine error, the tested CFA model may not be identified among males. Additionally, an out-of-range TLI (TLI = 1.001 for males) coupled with a near zero RMSEA indicated that the results based on the male subsample should be interpreted with caution.

Measurement Invariance

To test the degree to which the eight items shown in Table 1 measured the same latent factor, Physical Functioning, in males and females, three single-factor MGCFA models were performed according to the current guidelines for testing multiple-group ME/I with categorical outcomes [35]. Table 6 shows the model fit indices of the configural invariance model, the multiple group threshold invariance model, and the multiple group threshold and loading invariance model (i.e., the baseline model, Proposition 4, and Proposition 7 proposed by Wu and Estabrook, respectively [34]), as well as the significance of the χ^2 difference tests comparing each model to the previous less restrictive model. As seen in Table 6, based on CFI and TLI, all measurement invariance models tested fit the data well (i.e., all CFI and TLI are .999 or 1.000). Additionally, RMSEA was less than .07 for all models, with the baseline model having the highest RMSEA (.063) whereas the threshold and loading invariance model had the lowest RMSEA (.022). Although the χ^2 test was significant for the baseline model ($p = .042$), it was not significant for either of the other two ME/I models, and the χ^2/df ratio was below the cutoff of 2 for all three models. As seen in Table 6, none of the χ^2 difference ($\Delta\chi^2$) tests were significant, indicating that corresponding thresholds and factor loadings were invariant across male and female groups.

CONCLUSION

Results of this study indicated that the 8-item Adult Version 2.0 of the PROMIS Physical Function Short Form had good internal consistency and split-half reliabilities among Native Americans living on or near the Apsáalooke reservation with chronic illnesses. Current findings also provided evidence for construct validity of this PROMIS measure among females of the above population and when the two sex groups were combined, but the subsample of males was not large enough to yield conclusive findings on construct validity of the tested measure. Despite the limitation of the small n for males, findings based on the two-group ME/I tests suggested that the thresholds and factor loadings were invariant across male and female groups. In sum, the tested PROMIS Physical Function Short Form seemed to measure males' and females' Physical Function similarly in this Native American population.

DISCUSSION

The goal of this research is to increase awareness and provide information for future studies focused on the physical function wellness and abilities of community members within rural Native American populations to increase the quality of life. It is important to note that there are 574 federally recognized Native American tribes with different languages and cultures and that it is not appropriate to generalize findings from one tribe to all others [36]. Many Native American populations are at an increased risk of developing chronic illnesses; however, there has been a lack of psychometric testing of widely used PROMIS measures among Native American populations [37, 38]. This study is unique in that it applies the PROMIS survey to analyze data within this community using a community-based and community-driven process. These findings can also be useful in health care settings for community members to share information about their physical well-being [1]. Results of a literature search suggested that the current study was the first to test the reliability and construct validity of the 8-item Adult Version 2.0 of the PROMIS Physical Function Short Form based on a Native American sample. Therefore, this research begins to fill an important gap in the literature regarding the applicability of using these standardized measures among Native Americans.

Limitations and Suggestions for Future Research

This study has several limitations. The total sample size is relatively small ($n = 210$), especially for the male subsample ($n = 59$), and it only included individuals over the age of 24 years who were living with a chronic illness. However, similar studies have been published using small and restrictive sample sizes [39]. The current sample size was determined through a combination of power calculations, feasibility of recruiting and retaining participants, and available resources. Active involvement of community gatekeepers in the research team, a deep understanding of the benefits of the intervention program to the target community, strong rapport between the academic and community partners are the main contributing factors that allowed us to achieve the current sample size, despite the past history of negative research and external imposition conducted on this and other Native American communities. The dearth of studies evaluating psychometric

properties of PROMIS measures among Native Americans limits the use of these otherwise validated measures. Therefore, we argue that studies such as the current one are important for the utilization of PROMIS measures among Native American populations. We hope that this research inspires similar studies and that researchers from diverse disciplines will add to this body of literature.

An additional limitation is that we did not conduct formal cognitive debriefing or formal qualitative surveys with Apsáalooke community members to determine the appropriateness of the test items as has been done with some PROMIS surveys [6]. Apsáalooke research team members reviewed each item and deemed them to be appropriate for the community. This limitation may impact the findings of this study and is an important area for future research for this team.

Futhermore, an examination of the item distributions showed that for 50% of the items (namely, Light Chores, Walk, Errands, and Light Lifting) the median was 5, the highest possible item response score on the 5-point Likert scale among males. The combination of the small n and a ceiling effect among males seemed to limited the interpretability of the results obtained based on male subsample. As mentioned in the Results section, the smallest eigenvalue was negative/near-zero and the variance-covariance matrix of the estimated parameters was not positive definite in the males sample, indicating that the tested CFA model may not be identified among males. Additionally, the TLI of 1.001 was out of range for males (see Table 5). However, existing simulation studies found that out of range TLI values (which was called RHO by the authors) are common with small sample sizes [40]. The near-zero (.000) RMSEA (see Table 5) also indicated that the CFA model may not be identified in the male subsample. In sum, results based on the male subsample does not provide full support of construct validity, and the multiple group ME/I results need to be interpreted with caution. The ceiling effect is at a lesser degree in the female subsample. Although all items are negatively skewed (skewness ranged from $-.23$ to -1.12) with more participants reporting on the higher end of the scale compared to the lower end, only one item (Errands) had a median of 5 in females. Additionally, CFA models based on both the female subsample and the full sample were identified and the findings supported construct validity of the tested PROMIS measure. Future studies with a larger sample size, especially among males, are needed.

Another limitation is that we have only tested the internal consistency reliability, split-half reliability, and construct validity of one physical function short form. There are additional methods for assessing reliability and validity that this study did not perform. For example, test-retest reliability and content validity were not evaluated due to the lack of data to conduct those tests. Additionally, several short forms exist. To what extent the results of this study can be generalized to other short forms is unknown. Some of the measures used in this short form overlap with other forms as well as the longer PROMIS Physical Function questionnaire. Therefore, this study is just a start to testing the psychometric properties of the various forms of PROMIS measures of Physical Function in Native American populations. More research is needed to establish the psychometric properties of hundreds of other widely used PROMIS measures.

ACKNOWLEDGMENTS

We thank Dr. Mark Schure, Dr. Jillian Inouye, Lucille Other Medicine, Trevor Pollom, and the Messengers for Health Research Team who provided comments on an earlier version of this work.

References

1. Fries JF, Bruce B, & Cella D (2005). The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clinical and Experimental Rheumatology*, 23(5 Suppl 39), S53–S57.
2. Black N (2013). Patient reported outcome measures could help transform healthcare. *BMJ : British Medical Journal*, 346, f167 10.1136/bmj.f167 [PubMed: 23358487]
3. Boyce MB, Browne JP, & Greenhalgh J (2014). The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ quality & safety*, 23(6), 508–518. 10.1136/bmjqs-2013-002524
4. Gershon R, Rothrock NE, Hanrahan RT, Jansky LJ, Harniss M, & Riley W (2010). The development of a clinical outcomes survey research application: Assessment Center. *Quality of life research*, 19(5), 677 10.1007/s11136-010-9634-4 [PubMed: 20306332]
5. Cella D, Yount S, Rothrock N, Gershon R, & et al. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. *Medical care*, 45(5), S3 10.1097/01.mlr.0000258615.42478.55
6. DeWalt DA, Rothrock N, Yount S, & Stone AA (2007). Evaluation of Item Candidates: The PROMIS Qualitative Item Review. *Medical care*, 45(5), S12 10.1097/01.mlr.0000254567.79743.e [PubMed: 17443114]
7. Klem M, Saghaei E, Abromitis R, Stover A, Dew M, & Pilkonis P (2009). Building PROMIS Item Banks: Librarians as Co-Investigators. *Quality of Life Research*, 18(7), 881–888. 10.1007/s11136-009-9498-7 [PubMed: 19548118]
8. Irwin DE, Stucky B, Langer MM, Thissen D, DeWitt EM, Lai J-S, et al. (2010). An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Quality of Life Research*, 19(4), 595–607. 10.1007/s11136-010-9619-3 [PubMed: 20213516]
9. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of clinical epidemiology*, 63(11), 1179–1194. 10.1016/j.jclinepi.2010.04.011 [PubMed: 20685078]
10. Rose M, Bjorner JB, Becker J, Fries JF, & Ware JE (2008). Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *Journal of Clinical Epidemiology*, 61(1), 17–33. 10.1016/j.jclinepi.2006.06.025 [PubMed: 18083459]
11. Bajaj J, Thacker L, Wade J, Sanyal A, Heuman D, Sterling R, et al. (2011). PROMIS computerised adaptive tests are dynamic instruments to measure health-related quality of life in patients with cirrhosis. *Alimentary pharmacology & therapeutics*, 34(9), 1123–1132. 10.1111/j.1365-2036.2011.04842.x [PubMed: 21929591]
12. Jensen R, Potosky A, Reeve B, Hahn E, Cella D, Fries J, et al. (2015). Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. *Quality of Life Research*, 24(10), 2333–2344. 10.1007/s11136-015-0992-9 [PubMed: 25935353]
13. Quach C, Langer M, Chen R, Thissen D, Usinger D, Emerson M, et al. (2016). Reliability and validity of PROMIS measures administered by telephone interview in a longitudinal localized prostate cancer study. *Quality of Life Research*, 25(11), 2811–2823. 10.1007/s11136-016-1325-3 [PubMed: 27240448]
14. Hung M, Clegg DO, Greene T, Weir C, & Saltzman CL (2012). A Lower Extremity Physical Function Computerized Adaptive Testing Instrument for Orthopaedic Patients. *Foot & Ankle International*, 33(4), 326–335. 10.3113/FAI.2012.0326 [PubMed: 22735205]

15. Bartlett SJ, Ana-Maria O, Duncan T, DeLeon E, Ruffing V, Clegg-Smith K, et al. (2015). Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PLoS One*, 10(9), e0138543 10.1371/journal.pone.0138543 [PubMed: 26379233]
16. Broderick JE, Schneider S, Junghaenel DU, Schwartz JE, & Stone AA (2013). Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. *Arthritis care & research*, 65(10), 1625–1633. 10.1002/acr.22025 [PubMed: 23592494]
17. Benyshek DC (2001). The political ecology of diabetes among the Havasupai Indians of northern Arizona [Arizona State University].
18. Denny CH, Holtzman D, Goins RT, & Croft JB (2005). Disparities in chronic disease risk factors and health status between American Indian/Alaska Native and white elders: Findings from a telephone survey, 2001 and 2002. *American Journal of Public Health*, 95(5), 825–827. 10.2105/AJPH.2004.043489 [PubMed: 15855458]
19. Walls ML, Sittner KJ, Aronson BD, Forsberg AK, Whitbeck LB, & Mustafa a.A. (2017). Stress Exposure and Physical, Mental, and Behavioral Health among American Indian Adults with Type 2 Diabetes. *International Journal of Environmental Research and Public Health*, 14(9), 1074 10.3390/ijerph14091074
20. Goins RT, Innes K, & Dong L (2012). Lower body functioning prevalence and correlates in older American Indians in a southeastern tribe: the Native Elder Care Study. *Journal of the American Geriatrics Society*, 60(3), 577–582. 10.1111/j.1532-5415.2011.03869.x [PubMed: 22316130]
21. Crow Apsalooke. (2020). Retrieved June 29, 2020, from <https://www.visitmt.com/placesto-go/indian-nations/apsaalooke-crow.html>
22. Medicine Crow J (1992). From the heart of the Crow country : The Crow Indians' Own Stories (1st ed.). Orion Books.
23. Hallett J, Held S, McCormick AKHG, Simonds V, Real Bird S, Martin C, et al. (2017). What Touched Your Heart? Collaborative Story Analysis Emerging From an Apsaalooke Cultural Context. *Qualitative Health Research*, 27(9), 1267–1277. 10.1177/1049732316669340 [PubMed: 27659019]
24. HealthMeasures Scoring Service powered by Assessment CenterSM. (n.d). Retrieved June 29, 2020, from https://www.assessmentcenter.net/ac_scoring-service
25. David C, Gershon R, Bass M, & Rothrock N (2020). Assessment center scoring services user manual. https://www.assessmentcenter.net/ac_scoring-service/templates/UserManual.pdf
26. PROMIS Physical Function Scoring Manual (2019, June 10). Retrieved June 29, 2020, from http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Physical_Function_Scoring_Manual.pdf
27. Charter RA (2003). A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. *The Journal of general psychology*, 130(3), 290–304. 10.1080/00221300309601160 [PubMed: 12926514]
28. Rosseel Y, Oberski D, Byrnes J, Vanbrabant L, Savalei V, Merkle E, Hallquist M, Rhemtulla M, Katsikatsou M, Barendse M, Chow M, & Jorgensen T (2020, May 13). Package 'lavaan'. <https://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
29. Hu L & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. 10.1080/10705519909540118
30. Steiger JH (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893–898. 10.1016/j.paid.2006.09.017
31. Jorgensen TD, Pornprasertmanit S, Schoemann AM, & Rosseel Y (2018). *SemTools: Useful tools for structural equation modeling*. R package version 0.5–1. <https://github.com/simsem/semTools.wiki.git>
32. Vandenberg RJ & Lance CE (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1), 4–70. 10.1177/109442810031002

33. Stark S, Chernyshenko OS, & Drasgow F (2006). Detecting Differential Item Functioning with Confirmatory Factor Analysis and Item Response Theory. *Journal of Applied Psychology*, 91(6), 1292–1306. 10.1037/0021-9010.91.6.1292 [PubMed: 17100485]
34. Wu H & Estabrook R (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. 10.1007/s11336-016-9506-0 [PubMed: 27402166]
35. Svetina D, Rutkowski L, & Rutkowski D (2020). Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using M plus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111–130. 10.1080/10705511.2019.1602776
36. Bureau of Indian Affairs (2020, January 1). Indian entities recognized and eligible to receive services from the United States Bureau of Indian Affairs. <https://www.federalregister.gov/documents/2020/01/30/2020-01707/indian-entities-recognized-by-and-eligible-to-receive-services-from-the-united-states-bureau-of>
37. George S, Duran N, & Norris K (2014). A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *American journal of public health*, 104(2), e16–e31. 10.2105/AJPH.2013.301706
38. Mitchell TL & Baker E (2005). Community-building versus career-building research: The challenges, risks, and responsibilities of conducting research with Aboriginal and Native American communities. *Journal of Cancer Education*, 20(S1), 41–46. 10.1136/bmj.319.7212.774 [PubMed: 15916520]
39. Barun N (2010). Understanding the relevance of sample size calculation. *Indian Journal of Ophthalmology*, 58(6), 469–470. 10.4103/0301-4738.71673 [PubMed: 20952828]
40. Anderson JC & Gerbing DW (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49(2), 155–173. 10.1007/BF02294170

Table 1.

Descriptive Characteristics of the Full Sample and by Gender.

	Full Sample (N = 210) <i>Mean ± SD or % (frequency)</i>	Female (N = 151) <i>Mean ± SD or % (frequency)</i>	Male (N = 59) <i>Mean ± SD or % (frequency)</i>
Demographics			
Age at Baseline	52.01 ± 13.45	52.59 ± 13.32	50.51 ± 13.79
# of miles to the clinic	18.90 ± 19.68	18.93 ± 18.67	18.81 ± 22.30
# of doctor visits in the past 4 months	4.13 ± 6.36	4.26 ± 6.72	3.78 ± 5.36
Numbers of people in the household	4.82 ± 2.65	4.88 ± 2.66	4.67 ± 2.65
Sex- Female	72% (151)	100%	--
Marital Status			
Married/In marriage-like relationship	49% (102)	43% (65)	63% (37)
Separated or Divorced	23% (48)	25% (37)	19% (11)
Widowed	11% (23)	15% (22)	2% (1)
Single, never married	17% (36)	17% (26)	17% (10)
Other	1% (1)	1% (1)	--
Education			
Eighth grade or less	2% (4)	3% (4)	--
Some high school	11% (24)	11% (16)	14% (8)
High school graduate or diploma	22% (45)	19% (29)	27% (16)
At least Some technical/vocational school or Some college	31% (65)	32% (48)	29% (17)
Associate's degree	19% (40)	21% (32)	14% (8)
Bachelor's degree	9% (18)	7% (11)	12% (7)
Post-graduate/professional degree	6% (13)	7% (10)	5% (3)
Annual household income at Baseline			
Under \$10,000	33% (68)	33% (50)	31% (18)
\$10,000 – \$14,999	14% (29)	13% (19)	17% (10)
\$15,000 – \$24,999	19% (40)	19% (29)	19% (11)
\$25,000 – \$34,999	14% (29)	15% (22)	12% (7)
\$35,000 – \$49,999	11% (22)	9% (14)	14% (8)
\$50,000 – \$74,999	4% (9)	4% (6)	5% (3)
\$75,000 – \$99,999	3% (7)	5% (7)	--
\$100,000 and higher	1% (3)	1% (2)	2% (1)
Insurance or health coverage (can check multiple)			
Medicare/Medicaid	59% (123)	60% (90)	56% (33)
Private Insurance	11% (23)	11% (17)	10% (6)
Indian Health Service	45% (95)	47% (70)	42% (25)

	Full Sample (N = 210) <i>Mean ± SD or % (frequency)</i>	Female (N = 151) <i>Mean ± SD or % (frequency)</i>	Male (N = 59) <i>Mean ± SD or % (frequency)</i>
Food assistance programs			
Yes	55% (116)	58% (87)	49% (29)
Ongoing Illness(es) (can check multiple)			
Diabetes	57% (119)	56% (85)	58% (34)
Arthritis	34% (71)	38% (57)	24% (14)
Heart Disease	10% (20)	11% (17)	5% (3)
Blood Pressure	55% (116)	53% (80)	61% (36)
Asthma/Lung disease/COPD	14% (29)	17% (25)	7% (4)
Cancer	6% (12)	5% (8)	7% (4)
Chronic Pain	35% (73)	36% (55)	31% (18)
Other	25% (53)	27% (41)	20% (12)
Type of Diabetes			
Type 1	1% (1)	1% (1)	--
Type 2	51% (106)	52% (78)	48% (28)
Don't know	2% (4)	1% (2)	3% (2)
Did not answer	47% (99)	46% (70)	49% (29)

Table 2.

Descriptive Statistics of the PROMIS Physical Function 8b Adult Short Form v2.0 Items.

	N	Mean (SD)	Median	Skewness
Light Chores – Are you able to do chores such as vacuuming or yard work?	207	4.03±1.15	4	−1.00
Stairs – Are you able to go up and down stairs at a normal pace?	208	3.81±1.18	4	−.61
Walk – Are you able to go for a walk of at least 15 minutes?	207	4.08±1.11	4	−1.04
Errands - Are you able to run errands and shop?	206	4.21±1.00	5	−1.12
Labor - Does your health now limit you in doing two hours of physical labor?	209	3.46±1.24	3	−.20
Moderate Chores - Does your health now limit you in doing moderate work around the house like vacuuming, sweeping floors, or carrying in groceries?	210	3.95±1.09	4	−.73
Light Lifting - Does your health now limit you in lifting or carrying groceries?	210	3.84±1.23	4	−.67
Heavy Lifting - Does your health now limit you in doing heavy work around the house like scrubbing floors, or lifting/moving heavy furniture?	209	3.46±1.32	4	−.37

Table 3.

Polychoric Correlation between Individual Items for the Full Sample.

	Light Chores	Stairs	Walk	Errands	Labor	Moderate Chores	Light Lifting
Stairs	.84 ***						
Walk	.81 ***	.84 ***					
Errands	.77 ***	.75 ***	.79 ***				
Labor	.71 ***	.74 ***	.72 ***	.68 ***			
Moderate Chores	.85 ***	.81 ***	.79 ***	.74 ***	.85 ***		
Light Lifting	.82 ***	.75 ***	.79 ***	.78 ***	.77 ***	.85 ***	
Heavy Lifting	.75 ***	.71 ***	.75 ***	.68 ***	.78 ***	.84 ***	.85 ***

 $p < .001$

$N = 210$

Table 4.

Polychoric Correlation between Individual Items by Gender.

	Light Chores	Stairs	Walk	Errands	Labor	Moderate Chores	Light Lifting	Heavy Lifting
Light Chores		.84 ***	.79 ***	.80 ***	.69 ***	.85 ***	.81 ***	.75 ***
Stairs	.85 ***		.81 ***	.76 ***	.71 ***	.80 ***	.72 ***	.66 ***
Walk	.86 ***	.92 ***		.82 ***	.69 ***	.76 ***	.76 ***	.71 ***
Errands	.71 ***	.74 ***	.73 ***		.63 ***	.73 ***	.78 ***	.59 ***
Labor	.82 ***	.84 ***	.87 ***	.82 ***		.82 ***	.74 ***	.75 ***
Moderate Chores	.90 ***	.88 ***	.89 ***	.77 ***	.91 ***		.83 ***	.83 ***
Light Lifting	.82 ***	.83 ***	.84 ***	.83 ***	.87 ***	.91 ***		.81 ***
Heavy Lifting	.76 ***	.81 ***	.82 ***	.90 ***	.87 ***	.87 ***	.94 ***	

 $p < .001$

The cells above the diagonal are for females ($n = 151$), and those below the diagonal for males ($n = 59$).

Table 5.

Model Fit Indices of the Single-Factor Measurement Model for the Full Sample and by Gender.

Group	<i>Fit Statistic^a</i>	<i>df</i>	<i>p</i> (χ^2)	<i>CFI</i>	<i>TLI</i>	<i>RMSEA</i>
Full Sample, <i>N</i> = 210	34.161	20	.025 *	.999	.999	.058
Male, <i>N</i> = 59	13.338	20	.862	1.000	1.001	.000
Female, <i>N</i> = 151	28.195	20	.105	.999	.999	.052

^aThe model fit test statistic reported are based on the diagonally-weighted least squares (DWLS) method.

*
 $p < .05$

Table 6.

Baseline (Configural). Thresholds, and Thresholds and Loadings Measurement Invariance by Gender.

Model	χ^2	df	p	CFI	TLI	RMSEA	$\Delta\chi^2$	Δdf	p
Baseline/Configural	56.694	40	.042 *	.999	.999	.063	--	--	--
Thresholds	61.109	55	.266	1.000	1.000	.033	13.154	15	.590
Thresholds & Loadings	65.231	62	.365	1.000	1.000	.022	7.191	7	.409

*
 $p < .05$

