PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Use of Patient-Reported Outcome Measurement Information System ® (PROMIS®) measures to characterize health status for patients seeking care from an orthopedic provider: A retrospective cohort study
AUTHORS	Horn, Maggie; Reinke, Emily; Yan, Xiaofang; Luo, Sheng; Bolognesi, Michael; Reeve, Bryce B.; George, Steven

VERSION 1 – REVIEW

REVIEWER	Metcalfe, Andrew University of Warwick Warwick Medical School, Warwick Clinical Trials Unit
REVIEW RETURNED	01-Feb-2021

GENERAL COMMENTS	Overall a well written paper analysing a large dataset using PROMIS scores, which are becoming increasingly important in clinical effectiveness research. Whilst there is limited information that can be gleaned from such a broad study population, covering a wide variety of conditions, I do think there is value in reporting PROMIS scores as a form of baseline for future studies to understand the burden of disease presenting to orthopaedic units. However, I think the authors could do a lot more to examine the influence of baseline characteristics on their results and provide more meaningful data about the population they studied and the PROMIS Scores. The results do not clarify the numbers of exclusions or why. For example, how many people attended the hospital during this time and therefore how large was the eligible population? how many people refused to complete PROMIS and how many undertook the PROMIS scores but had incomplete data that could not be analysed? (Did these proportions change after PROMIS CAT was instituted?). A patient flowchart showing the full population who presented to the unit and clarifying exclusions is needed. There is little explanation of the change between PROMIS and PROMIS CAT? Did the results change when the new system was used? I think this paper would benefit from a multi-variate analysis including baseline factors (even simple things like age and gender) and adjusted for score type (paper or CAT). This would help interpret some of the differences between the groups and provide some adjustment for bias, such as the relationship between age and PROMIS score (do the sports surgery group have higher scores just

	because they are younger?). There is a wealth of data including baseline data here which could provide a much richer picture then is presented in the study at present. I would like to see a better and more nuanced discussion around whether PROMIS fully captures the domains that are relevant in all these conditions - hand surgery has suffered from this challenge for many years and this is only partially touched on. Joint instability is another presentation that is not always well collected in scores (although I don't know PROMIS domains well enough to know if this is true), and young people may be more limited (due to being unable to work, for example) then the elderly with the same problem. So such direct comparisons between quite divergent conditions/specialities should always be considered in light of the wider bio-psycho-social impacts these conditions have on individuals. This is not really touched on in the discussion, and I don't think the pitfalls of PROMIS are really considered. I would like to see the above points addressed, but I do think the paper has value and, with modification, would be worth considering for publication in BMJOpen.

REVIEWER	Repo, Jussi Central Finland Central Hospital, Department of Surgery
REVIEW RETURNED	04-Mar-2021

GENERAL COMMENTS	The authors present an exiting a retrospective cohort research of using Patient-Reported Outcome Measurement Information System ® (PROMIS®) measures to characterize health status for patients seeking care from an orthopedic provider. The sample size of 14,910 patients is extensive. The study setting is clear. Abstract is punctual and informative. The aim and the conclusions meet. Introduction and Methods are explained in sufficient detail. The results are presented with sufficient extent. Please provide a section for strengths and weaknesses in the Discussion section. Also, add a paragraph for conclusions in the discussion section. Congratulations for your great work.

REVIEWER	Gulledge, Caleb Wayne State University School of Medicine
REVIEW RETURNED	09-Mar-2021

GENERAL COMMENTS	Thank you for the opportunity to review your manuscript. It is well written and it is obvious a lot of time and thought went into the preparation of this manuscript. Please see specific comments below. Line 51: it would be helpful to state that the mild, moderate, and severe scores are set at 0.5, 1, and 2 STD's from the mean, so that the reader does not have to dive into those sources to determine this. For the ANOVA tests, I think a post-hoc analysis would aid in the interpretation of the p-value, so it is clear which values are statistically different.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

1. **Overall a well-written paper analyzing a large dataset using PROMIS scores, which are becoming increasingly important in clinical effectiveness research.**

Thank you!

2. **A. Whilst there is limited information that can be gleaned from such a broad study population, covering a wide variety of conditions, I do think there is value in reporting PROMIS scores as a form of baseline for future studies to understand the burden of disease presenting to orthopedic units.**

Thank you, we also agree that this is a very good point and a strength of the paper. This was added briefly to the introduction on lines 62-63 and expanded upon this concept further in the discussion lines 300-301

3. **However, I think the authors could do a lot more to examine the influence of baseline characteristics on the results and provide more meaningful data about the population studied and the PROMIS scores**.

Thank you for the suggestion. Based on your subsequent comment and this comment regarding modeling (comment 6), we have chosen to report table 2 values as adjusted values. More detail and rationale is provided within that comment

After some discussion with the author team our descriptive approach was thought most appropriate because awe didn't plan to present hypotheses for characteristics that could influence variability in PROMIS scores (This could be a paper alone). Our main concern is that without hypotheses we would identify many variables that statistically contribute to variability in PROMIS scores simply because this was a large dataset. Lastly, this is a cross-sectional cohort and we feel like a larger impact will be had by examining these influences in future longitudinal analyses when data are available.

4. **The results do not clarify the number of exclusions or why. For example, how many people attended the hospital during this time, and therefore how large was the eligible population? how many people refused to complete PROMIS and how many undertook the PROMIS scores but had incomplete data that could not be analyzed? (Did these proportions change after PROMIS CAT was instituted?). A patient flowchart showing the full population who presented to the unit and clarifying exclusions is needed.**

Thank you for this comment. We did not initially include a flowchart because this was a targeted EHR data extraction for this study looking at PROMIS scores primarily (that being the major inclusion criteria). However, we agree that a flowchart adds value to the methodology and clarifies inclusion/exclusion criteria. I improved the clarity of the language within the methods and added the flow diagram on lines73-78.

To address your comment re the SF to CAT transition- both measures were collected via the patient portal (Epic MyChart). There was not a change in how the PROMIS measures were administered. Although not the focus of this paper, we found that during the pilot phase of administering the 8 PROMIS domains our evaluation showed that most people who elected to complete the PROMIS measures completed the full battery of assessments. The associated

numbers for each domain and specialty for multivariate analyses are detailed in table 2, last column to help clarify how this affected sample size

5. **There is little explanation of the change between PROMIS and PROMIS CAT? Did the results change when the new system was used?**

I have added in detail the change in the collection of short forms to CAT with a reduced set of domains. This was done as part of the upgrades available in our EHR during this initiative. This change did not fundamentally change the collection process (both instruments were collected via EHR). To account for the difference in instrument type as suggested in a comment regarding the methods I have added this as a covariate in models for table 2. Lines 90-100.

6. **I think this paper would benefit from a multi-variate analysis including baseline factors (even simple things like age and gender) and adjusted for score type (paper or CAT). This would help interpret some of the differences between the groups and provide some adjustment for bias, such as the relationship between age and PROMIS score (do the sports surgery group have higher scores just because they are younger?). There is a wealth of data including baseline data here which could provide a much richer picture than is presented in the study at present.**

Thank you for this recommendation. I have adjusted the models to include the covariates: age, sex, instrument type (SF, CAT), race, and ethnicity. I compared 3 models in this process: unadjusted, partial adjustment(age, sex, instrument type), and robust adjustment(age, sex, instrument type(SF, CAT), race, and ethnicity). All models yielded very similar scores, with less than a one-point difference between models. This is encouraging because a one-point difference in PROMIS scores is well below any clinically relevant thresholds for changes/differences. Below is a sample of the output for PROMIS Physical Function Scores for reference. I have added the changed modeling to the methods on lines 149-154.

	Unadjusted	Partial	Robust	Diff Unadj to Partial	Diff unadjusted to Robust	Diff partial to Robust
Foot and Ankle	39.49418	39.83741	39.83378	-0.34323	-0.3396	0.00363
Hand	41.84946	41.77194	41.67531	0.07752	0.17415	0.09663
Neurosurgery	35.04133	35.14042	35.18148	-0.09909	-0.14015	-0.04106
Ortho Oncology	38.35242	38.38086	38.40825	-0.02844	-0.05583	-0.02739
Spine	35.4485	35.71876	35.76809	-0.27026	-0.31959	-0.04933
Sports Medicine	39.51216	39.13857	39.12597	0.37359	0.38619	0.0126
Total Joint	36.57589	36.59244	36.65705	-0.01655	-0.08116	-0.06461
Trauma	31.59897	31.5918	31.79687	0.00717	-0.1979	-0.20507

7. **I would like to see a better and more nuanced discussion around whether PROMIS fully captures the domains that are relevant in all these conditions - hand surgery has suffered from this challenge for many years and this is only partially touched on. Joint instability is another presentation that is not always**

well collected in scores (although I don't know PROMIS domains well enough to know if this is true), and young people may be more limited (due to being unable to work, for example) then the elderly with the same problem. So such direct comparisons between quite divergent conditions/specialties should always be considered in light of the wider bio-psycho-social impacts these conditions have on individuals. This is not really touched on in the discussion, and I don't think the pitfalls of PROMIS are really considered.

Thank you for this comment. This really brings up a poignant discussion point for this paper. There has been a recent exponential increase in the reporting of PROMIS measures across Orthopaedics (Horn et. al, 2020; https://doi.org/10.1186/s13018-020-02068-9) (Zdziarski-Horodyski L et al. 10.1186/s13063-017-2430-5) (O-Hara et al., https://doi.org/10.1016/j.injury.2019.10.076)

Many studies, including systematic reviews, some cited above, have demonstrated the validity, strengths/limitation and widespread adoption of using PROMIS Measures in Orthopaedics. Yet, I agree on face value that there is some skepticism that PROMIS measures may not be capturing all constructs that are relevant in these patient populations. I now discuss this and I have added additional citations to support the discussion on lines 274-290.

In regards to challenges in upper extremity patients, In this study, we used the PROMIS Physical Function measure which includes mobility and upper body functioning and impacts on daily living. While not included in the study, there is a PROMIS Upper Extremity measure that focuses on the functioning of the upper body (e.g., carrying objects, wash your back, change a light bulb overhead, put on/take off jacket). Currently, there is not a PROMIS measure of fine motor functioning. We agree that the specificity of the PROMIS measures for specific bodyegions is a current limitation of the PROMIS measures. We now identify this in our limitations section and in other parts of the discussion on lines 264-266, 315-318

8. **I would like to see the above points addressed, but I do think the paper has value and, with modification, would be worth considering for publication in BMJOpen.**

Thank you, I appreciate your detailed comments and suggestions which I have incorporated into the revised manuscript which has overall greatly improved the impact of the paper.

Reviewer 2: Dr. Jussi Repo, Central Finland Central Hospital

1. **The authors present an exciting retrospective cohort research of using Patient-Reported Outcome Measurement Information System® (PROMIS®) measures to characterize health status for patients seeking care from an orthopedic provider. The sample size of 14,910 patients is extensive. The study setting is clear. The abstract is punctual and informative. The aim and the conclusions meet. Introduction and Methods are explained in sufficient detail. The results are presented to a sufficient extent**.

Thank you!

2. **Please provide a section for strengths and weaknesses in the Discussion section. Also, add a paragraph for conclusions in the discussion section. Congratulations for your great work.**

Thank you. I have added in a section heading for strengthens and weaknesses on line 292. I have also added in the Conclusions section heading and text on lines 332-342

Reviewer 3: Dr. Caleb Gulledge, Wayne State University School of Medicine

1. **Thank you for the opportunity to review your manuscript. It is well written and it is obvious a lot of time and thought went into the preparation of this manuscript. Please see specific comments below.**

Thank you!

2. **Line 51: it would be helpful to state that the mild, moderate, and severe scores are set at 0.5, 1, and 2 STD's from the mean so that the reader does not have to dive into those sources to determine this.**

Thank you, this was an oversight in the methods section. In the methods on lines 124-129I have added the explanation of the interpretation categories.

3. **For the ANOVA tests, I think a post-hoc analysis would aid in the interpretation of the p-value, so it is clear which values are statistically different.**

Thank you for this methodological suggestion. We agree this would add value in determining the groups that are driving the statistical difference across the demographics. Due to the sample size (>14,000), the observed differences between groups are all statistically significant; and in almost all cases, this does not translate to a clinically significant finding. However, we agree that some of these statistical differences between groups need to be considered in the context of there contribution to PROMIS scores. Therefore, in the spirit of your suggestion and reviewer 1's suggestion, we have added multivariate regression analyses for examining PROMIS scores across the clinical specialties (updated table 2). Our intention here is that we acknowledge the implications of differences in key baseline factors on the difference of PROMIS scores across groups.

VERSION 2 – REVIEW

REVIEWER	Metcalfe, Andrew University of Warwick Warwick Medical School, Warwick Clinical Trials Unit
REVIEW RETURNED	03-May-2021

GENERAL COMMENTS	Thank you for your detailed response and corrections to the paper. The paper has been substantially improved and I think is worthy of publication. I have enjoyed reading this interesting piece of work.