Development and evaluation of patient-reported outcome score visualization to improve their utilization (PROVIZ)

Principal Investigator and Team Members
Heather T. Gold, PhD, MPI (contact)
Enrico Bertini, PhD, MPI (now at Northeastern University)
Nataliya Byrne, Research Coordinator
Racquel Fygenson, Research Assistant

Organizations:
NYU Grossman School of Medicine
NYU Tandon School of Engineering

Structured Abstract (Max 250 words)

Purpose: When building patient-reported outcome measures (PROs) into the electronic health record (EHR), health systems, researchers, and clinicians must investigate the optimal methods and impact of collecting and sharing PRO and patient contextual data with clinicians and patients in order to have positive effects on clinical care, quality of life, and outcomes.

Scope: We iteratively designed multiple prototypes of PRO visualizations for NIH PROMIS scores and aimed to evaluate patient and clinician preferences, understanding, usability, and response time for various PRO score presentations.

Methods: Using mixed-methods and user-centered design approach, we included three components: 1) qualitative interviews with clinicians treating patients with hip/knee pain and patients to understand current perceptions of PROs and preferences for data presentation; 2) subsequent development of data visualization prototypes that attempt to meet clinician and patient preferences; and 3) formal evaluation and iteration of the novel visualizations with stakeholders and the general public.

Results: Novel "smiley-timeline" PRO score visualizations reduced response times for certain interpretations, specifically significantly quicker response times than all other line charts for finding a negative extreme and explaining how to use the graph. The new visualization resulted in longer response times or higher error rates when participants were asked for the steepest rate of change, to estimate numerical score values, and to identify a data point beyond a threshold. More traditional line charts with y-axis directions corresponding to semantic language resulted in quicker response times and lower error rates than line charts with y-axes that used position to encode beneficial change.

Purpose. When building patient-reported outcome measures (PROs) into the electronic health record (EHR), health systems, researchers, and clinicians must investigate the optimal methods and impact of collecting and sharing PRO and patient contextual data with clinicians and patients in order to have positive effects on clinical care, quality of life, and outcomes. The purpose of this study was to iteratively design multiple prototypes of PRO visualizations for NIH PROMIS scores and evaluate *patient* and *clinician* preferences, understanding, usability, and response time for various PRO score presentations.

Scope/background.
Electronic collection and use of patient reported outcomes measures (PROs) hold promise to improve healthcare delivery and response to the individual patient experience. Technical advances have led to the adoption and integration of PROs into the electronic medical record to improve responsiveness to patients' changes in health status.[1–4] PRO measurement provides objective patient data that can lead to personalized, targeted care by identifying patients' experience of illness and wellness, refining patient selection for intervention, and responding to their specific needs to improve quality of life.[5] Examples of the effectiveness of capturing and acting on PROs are found in cancer care,[6–10] orthopedics,[4,11–25] and other clinical areas.[5] Valid, reliable capture of PROs provides important data for person-centered, high-quality, high-value care.

Greater emphasis on value in healthcare creates an imperative to be more responsive to both patient quality of life and efficient healthcare utilization. The Centers for Medicare and Medicaid Services included the collection of PROMIS Global Health scale in its bundled payment program for total joint replacement, which will improve our ability to evaluate the quality, value, and outcomes using data that are patient centered.[26] Alternative payment initiatives, such as bundled payment programs, create incentives for hospitals and physicians to manage the whole patient to optimize length of stay, reduce readmission rates, and track PROs.

Graphical display of patient information affects data interpretability, which is key for using such information to enhance patient-physician communication and shared decision making.[27,28] Research shows that patients and clinicians may struggle to interpret PROs,[27,29,30] which could inhibit their use to improve clinical management or shared decision making. For PRO scores to be valuable, users must understand what PRO scores and score changes represent. Our survey of NYU physicians (n=46) with current access to PRO scores indicates that while 59% agree that PROs can improve quality of care, 39% state that PRO score presentation is unclear. Further, about half state that they cannot easily integrate PROs into patient interaction, but agree that research supports the use of PROs clinically. If clinicians incorporate PROs into conversation during the patient visit, along with clinical factors and the physical exam, patients may become more activated and engage in shared decision making for treatment choices.[5] Higher preoperative patient engagement may be associated with better outcomes,[31] as has been shown for total joint arthroplasty.[32]

There is limited empirical research guiding how to represent PRO scores graphically, particularly for NIH PROMIS assessments. Work has been done to refine presentation of patient-facing[33] and physician-facing[28] test results and other clinical data,[27] as well as PROs for physical function, emotional function, fatigue, and pain[34] and PROs for cancer quality of life.[35] Line graphs, directionality of "higher = better" scores, and threshold lines indicating concern about score values appear to be most effective, according to a large study focused on an oncology population.[34] However, there is little research on presentation of multiple scores or adding clinical context. For example, when multiple scores are presented at once, conflicting directionality may cause confusion (e.g., when pain score is higher, pain is worse, whereas when physical function is higher, function is better). Graphical indicators for different clinical events, and which ones, or score thresholds such as meaningful improvement or decline also are understudied within PRO score visualizations.

PROs allow for patient-centered care by integrating the patient experience into treatment decision-making.[5,36] Further, PROs provide for meaningful symptom monitoring, managing patient expectations for surgical recovery, and minimizing clinical variations in care. However, without optimized usability and interpretability, PROs will not be integrated effectively into care management. There is strong interest in use of PROs at medical centers across the United States, including NYU Langone Health, and this study will improve how we systematically inform clinical care with PROs by improving usability and interpretability.

By developing representations of PRO data in an iterative process, we aimed to identify the most preferred data visualization approach(es) for patients and clinicians.

Our key <u>hypothesis</u> was that by developing representations of PRO data in an iterative process, we would identify the most preferred data visualization approach(es) for patients and clinicians.

We included three populations:

- Patient subjects with hip or knee pain or hip or knee osteoarthritis, ages 18+, women and men using the electronic patient portal at NYU Langone Health in New York City region.
- Clinician subjects who were practicing orthopedic surgeons, physical and occupational therapists, or rheumatologists at NYU Langone Health.
- General public via Prolific survey platform who may be patients of other health systems

Justification for study population: Most available work in the field has focused on PRO data for oncology, but not for general health PRO assessments, such as pain interference, pain intensity, and physical function. We have integrated NIH PROMIS assessments into care in the Departments of Orthopedic Surgery and Rheumatology, and our familiarity with that population helped inform this study. There is no control group, because there was no specific intervention requiring a control.

Methods (Study Design, Data Sources/Collection, Interventions, Measures, Limitations).
This study was conducted in three phases: qualitative interviews regarding current PRO visualizations at NYU Langone Health and via Epic EHR, integration of qualitative findings to develop data visualization prototypes and interview key stakeholders about prototypes, and create revised prototypes and test alongside current visualizations. Our review of related work highlighted key areas that may contribute to confusion and misinterpretation of current PROMIS score visualizations.

Phase 1 Approach. To examine these potential gaps further, we conducted interviews with clinicians and patients at NYU Langone Health system in which we showed them a series of PROMIS score visualizations, one for function (physical function) and one for symptoms (pain severity), and asked them to think aloud while they interpreted the fictional patients' data. In total, we interviewed 7 clinicians and 5 patients and recorded their verbal thought processes as they interpreted 6 different PROMIS score graphs, all visualized in the same way. First, clinicians and patients were given a brief overview of PRO scores and their potential uses. Then they were shown 5 sets of graphs with scores for PROMIS Pain Severity and Physical Function in which the y-axes both started at 0 and increased to 100, and 1 set of graphs in which the Pain Severity y-axis was 100 at the origin and increased to 0 at the top of the chart (see Figure 1b for comparison). For all interviews, clinicians were shown current Epic-designed line charts that were not implemented in the NYU Langone Health system (see Figure 1, right), while the patients were asked to interpret simpler line charts as represented currently in the patient portal (see Figure 1, left), and then shown the clinician-facing EHR-designed line charts and asked for their preference. Clinicians and patients were asked whether they found anything confusing in the charts shown, if there was any other information they would like to see included in the charts, if they would prefer to view PRO scores in two side-by-side graphs or together in one graph, and how important they thought it was to show the date of medical interventions or procedures on the charts.

Figure 1.
Left: An example of the charts shown to interviewed patients representing the current portal visualization.
Right: An example of the charts shown to interviewed clinicians and to patients following the current portal representation.
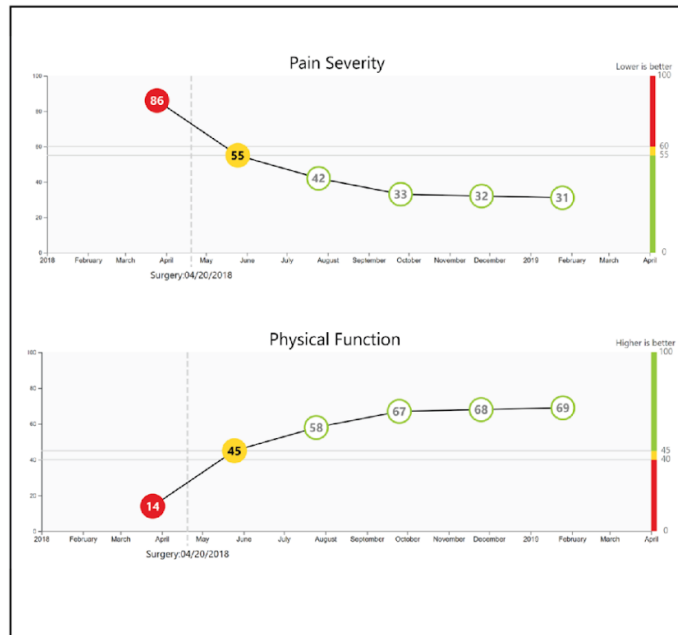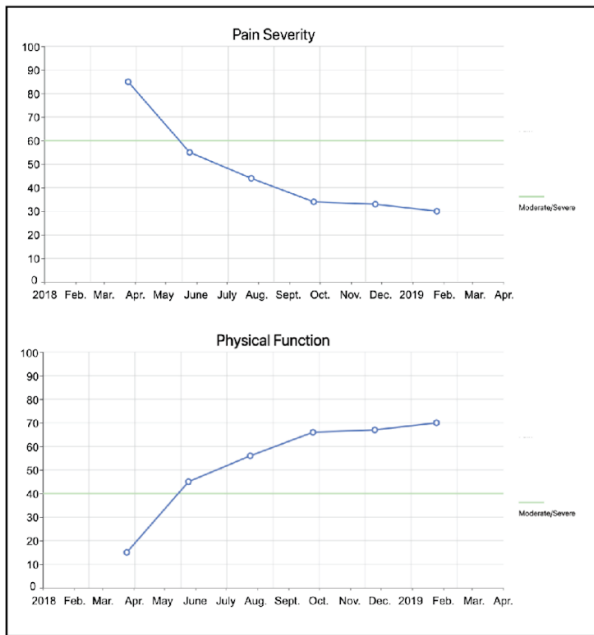
These interviews ranged from 20 to 39 minutes and were conducted by a minimum of two researchers over a video call with the interviewee. The various stimuli were shown to interviewees via screen share of a slide deck that the interviewers controlled. All interviews were recorded, transcribed, and then qualitatively coded by a third researcher.

Figure 1b. An example of the comparative charts shown to interviewed patients, representing the inverse Y-axis on the left and the current visualization of the patient portal on the right. These formats were displayed separately to explore if the inverse Y-axis overcame any confusion in the current visualization.
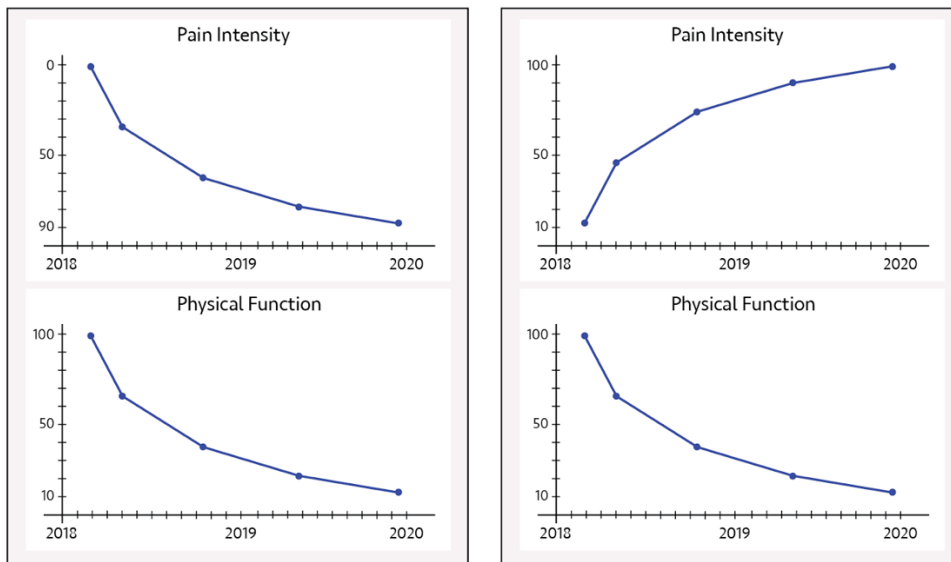
Phase 1 Results.
Across both patient and clinician interviews, the most commonly expressed confusion revolved around the direction of the y-axes for the paired graphs. When the y-axes of both charts started at 0 and increased to 100, interviewees were confused by "up" indicating both a positive change in Physical Function (function) and a negative change in Pain Severity (symptom). When the Pain Severity y-axis was flipped whereby it started at 100 and increased to 0, thus making "up" indicate a beneficial change simultaneously in Pain Severity and Physical Function, interviewees were confused by the semantic mismatch of an "upward" increase in pain indicating pain improvement. Additionally, patients reported confusion with the numerical scoring of the PROMIS measures, asking instead for a y-axis that was ordinal (e.g., Bad, Fair, Good). Across all interviews,

and in logical accordance with the expressed confusion, interviewees expressed a strong preference for doubly encoding Pain Severity and Physical Function scores using color.

Phase 2 Approach.
Motivated by findings from Phase 1 interviews, we developed a novel method for visualizing PROMIS scores and shared them with clinician stakeholders from NYU Langone Health from orthopedic surgery and rheumatology. Our novel visualization has no y-axis, instead encoding PRO scores along a timeline via color and iconography.(Figure 2) Our hypothesis is that removing the y-axis entirely assists readers in avoiding visual and semantic confusion documented in our qualitative study. Further, we simulated visualizations for color-blindness and ensured all prototypes could be well-visualized regardless of this condition. Clinician stakeholders provided feedback and generally appreciated the novel prototypes, although commented that they were so novel compared with traditional x-y graphs of data that some people might take more time to interpret them.

Prior to a larger test (i.e., Phase 3) of the usability and interpretability of the novel prototypes, we pilot tested them on a smaller NYU Langone Health sample. In total, 37 patients and 7 clinicians completed our pilot study. We discarded 1 patient survey response because it was completed on an interim draft of the survey, leaving us with a total sample of 36 patients and 7 clinicians. Subjects were asked about fictional patients' Pain Severity and Physical Function PRO data. First, as a control, subjects were shown numerical PRO data in a table and asked 10 comprehension questions. Then they were either shown a black-and-white version of PRO data plotted on two line charts with y-axes going in the same direction, or black-and-white version of PRO data plotted on two line charts with y-axes going in opposite directions, and asked the same comprehension questions. Next, subjects were shown similar line charts with additional stop-light (red, yellow, green) color encoding and asked the same questions. Finally, subjects were shown a version of our novel "smiley"-timeline visualization – either on a binned or a continuous scale, and with either a stoplight-scheme color encoding or grey-scale encoding – and asked the same questions. (See Figure 2) The four variations of smiley-timelines in our pilot study allowed us to test the importance of color encoding and of representing PROMIS scores as a continuous or discrete visual variable.

Figure 2. Initial prototypes developed for Phase 2.
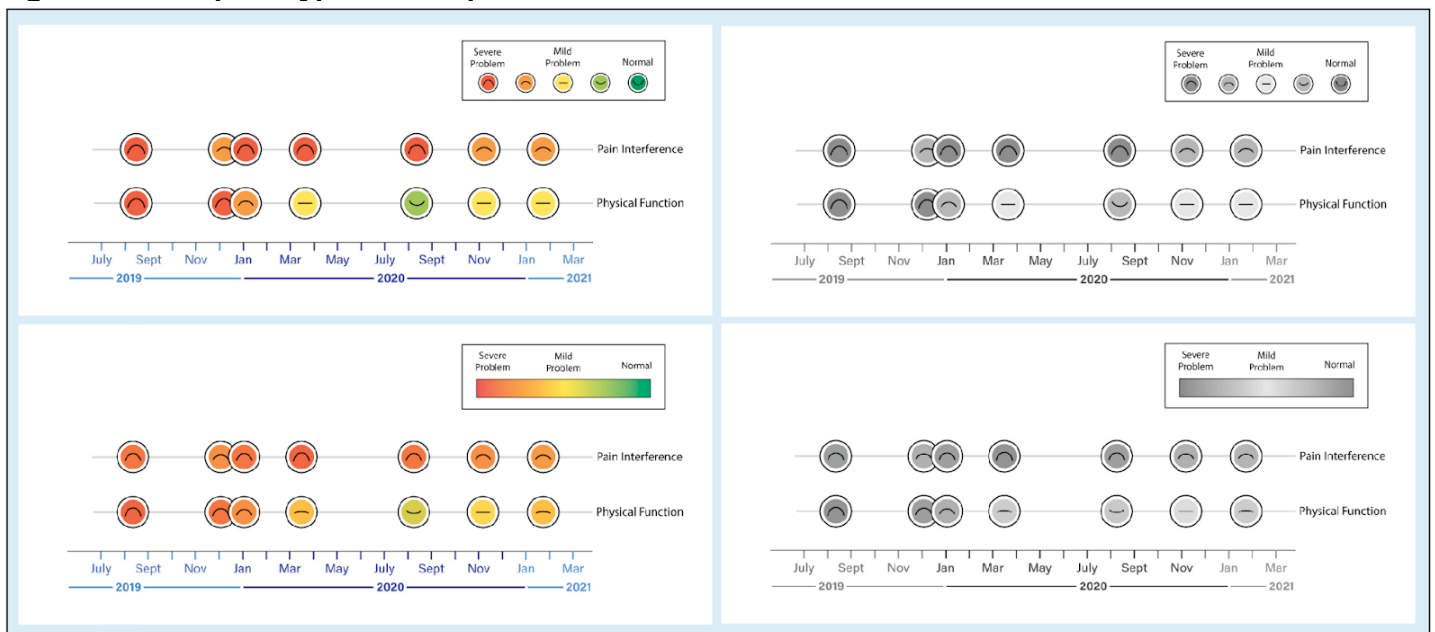
Phase 2 Results.
The pilot study allowed us to modify prototypes and survey questions. We made slight changes to correct question wording and alter the novel visualizations' iconography to be clearer. The most significant change was adding eyes to the smiley face icons to make them more recognizable (see Figure 3).

We evaluated the performance of each stimulus in terms of response error rate and length of time taken to answer questions using the stimulus. When respondents answered using the line charts with additional color encoding or smiley-timelines with color encoding, they answered more questions correctly at a faster rate than when using any other type of chart.

Figure 3. Adaptation to prototype to include eyes to icons, making them more recognizable.

Phase 3 Methods.
We collected the response times and error rates of the 3 charts with the highest performance from our pilot study: the line chart with axes going in the same direction and color encoding, the line chart with axes going in different directions and color encoding, and the smiley-timeline with color encoding. We also compared versions of both line charts without color encoding to better study the impact of color's role in effectively communicating PRO scores. We hypothesized that a stoplight color scheme encoding on any chart will make PRO visualizations more effective because of its prior association with "good" and "bad", which translates directly into the semantics of both symptoms and functions (i.e. green is good and red is bad for both symptoms and functions). The traditionally used line chart position encoding, on the other hand, translates inversely to symptoms and functions (i.e. down is good for symptoms, but bad for functions). Additionally, we hypothesized that removing position encodings for PRO scores altogether would further reduce the chance of readers confusing beneficial and detrimental changes for symptoms and functions. Thus, our overall hypothesis was that smiley-timelines would result in the lowest error rate and the quickest answer time when study participants were asked to interpret general trends without any instruction and to identify the correct method for reading the chart.

We recruited 150 participants from the general public via the website Prolific.com. All participants were fluent in English, residents of the United States at the time, between 44 and 100 years old, and had a >= 97% approval rate on Prolific. Participants were compensated $1.59 based on time required to complete the survey (averaging out to a rate of $12.70/hr). No identifying data was collected by researchers. Via a Qualtrics survey that they were directed to from Prolific, participants answered 9 chart comprehension questions, 4 subjective ranking questions, 5 questions about their history with PROs and clinical care, and 1 question about preference to use PROs in the future. Participants were also asked to self-report if they were colorblind. To minimize respondent burden to enhance sample size, each participant was shown only 1 chart type, populated with the 4 different data discussed above.

Question order was identical across all surveys, while answer order was randomized. We asked participants to interpret overall trends and explain how to read the chart.
    Task 1: Interpreting overall trends
        • "In general, how would you say the patient's pain has been doing?"
        • "In general, how would you say the patient's physical function has been doing?"
    Task 2: Explaining how to read the chart
        • "When the patient's pain gets better, the corresponding line/smiley gets _____"
        • "When the patient's physical function gets better, the corresponding line/smiley gets _____"

Secondly, we asked participants to find specified extrema (e.g., worst pain), compare rate of change (e.g., largest improvement), estimate PRO scores, compare trends across pain and physical function, and interpret thresholds. Finally, we tested for subjective difficulty, intuitiveness, quickness, and likelihood of recommending the charts by asking participants to fill out Likert scales from 1 to 10 at the end of the survey.

We evaluated the efficacy of the various chart stimuli based on the mean amount of time it took participants to answer questions and the number of errors participants made while answering questions. We evaluated the tasks piecewise, finding differences in chart performance for some tasks. We analyzed the data corresponding to different tasks independently. We analyzed both the response time and error rate of each of the 5 graphs by plotting the difference of means and its confidence intervals for each task and by calculating their respective Cohen's d and effect size statistics. All calculations and visualizations were conducted using the Javascript packages d3 and the statistical library JStat, and are hosted via the web-based notebook Observable. To analyze response times of participants, first we calculated the skewness and kurtosis of the distribution of response times for each task. All tasks' distributions had a skewness > 2 and were leptokurtic with a kurtosis > 3, and thus we conducted the rest of our response-time analysis using nonparametric statistical tests. For each task, we conducted a Kruskal-Wallis test using the python package SciPy (Virtanen) on the response times for the various charts to identify any significant differences between their distributions. The following tasks' Kruskal-Wallis tests returned a p-value < 0.05, indicating that at least one of the chart's time responses were significantly different from the other charts':

- o finding specified extrema
- o comparing rate of change
- o estimating scores
- o explaining how to read the chart

To analyze the error rate data, we again separated our data into independent sets based on task and evaluated the likelihood that the error rates associated with using different charts to answer the same question are significantly different. Because our null hypotheses are that each question has the same probability of being answered correctly regardless of the chart used, we can use parametric analyses to evaluate our data. Thus, for each task, we conducted a pairwise series of t-tests between the error rates of questions answered using different charts. Due to the large number of t-tests we conducted on our data, our p-values were again adjusted using a two-stage Benjamini-Hochberg false discovery rate correction with an of alpha = 0.5, implemented via the statsmodel python package (Seabold). We then compared the means of the charts in each pair to determine which chart had a significantly lower error rate.

Phase 3 Results.
Figure 4 shows the task and charts with significantly different performance, response times in the top half and error rate in the bottom half. Five of the 7 tasks are listed. The 2 remaining tasks showed no significant difference in performance. The chart with quicker response times or lower error rates is shown on the left side of every pair.

Figure 4. Pairs of chart comparisons with significant differences in performance.
Top half: Charts with significant difference in response times
Bottom half: Charts with significant differences in error rates.

We also collected 4 user preference metrics via Likert scales at the end of our survey. We compared these preference data using a nonparametric Kruskal Wallis test (Virtanen). All p-values were > 0.05, so the null hypothesis, that all charts' reported preferences are from the same distribution, could not be rejected. There are three key results:

- Line charts with both Symptom and Function axes increasing in the same direction from "bad" to "good" performed less effectively than line charts with the Symptom and Function axes adhering to semantic directions
- Smiley-timeline charts offer significant decreases in the time responses of certain tasks (i.e., finding direction-specific extrema, and explaining how to read the chart), without affecting the error response rate for those tasks.
- For tasks that require more exact numbers, like comparing rate of change and interpreting thresholds, smiley-timeline charts resulted in a significant increase in time response over line charts with color encoding, and a significant increase in error over line charts with improvement encoded via different directions and no color encoding.

The Dunn tests returned significant p-values for 8 pairs of charts across 3 tasks:
- comparing rate of change
- estimating scores
- interpreting thresholds.

Discussion.
Novel "smiley-timeline" PRO score visualizations reduced response times for certain interpretations, specifically significantly quicker response times than all other line charts for finding a negative extreme and explaining how to use the graph. The new visualization resulted in longer response times or higher error rates when participants were asked for the steepest rate of change, to estimate numerical score values, and to identify a data point beyond a threshold. More traditional line charts with y-axis directions corresponding to semantic language resulted in quicker response times and lower error rates than line charts with y-axes that used position to encode beneficial change. Directional and semantic mismatches confuse patients who read PROMIS score data that are visualized traditionally via line charts. Encoding PRO scores via iconography instead of direction is a viable method of addressing this problem, but does not outperform traditional line charts that doubly encode scores with color.

Limitations. Some of the caveats of smiley-timelines could be mitigated in the future, with additional functionality such as scores appearing upon mouseover to eliminate readers' problem estimating scores precisely. Other issues, such as interpreting thresholds, are an inherent limitation of the design. However, supplementary indicators of thresholds could be added to future visualizations, such as a text legend stating that a score in the red zone is highly concerning or that a score change of a certain number of points should be addressed.

We did not attempt to combat familiarity bias in our experiments, and thus comparing response times of line charts to the novel smiley-timelines should acknowledge that line charts are a prominent part of graph literacy education. At the same time, there is a strong familiarity bias associated with smile icons and the stoplight coloring that were used on the smiley-timeline, so it is unclear whether said familiarity bias is enough to offset any error or response time differences.

Implications. This study developed novel visualizations for NIH PROMIS scores that overcomes some problems of traditional line charts. We have discussed the visualizations with members of the Epic EHR data visualization team and will be meeting again to talk about our findings from the testing in phase 3.

References

1. Gold HT, Karia RJ, Link A, et al. Implementation and early adaptation of patient-reported outcome measures into an electronic health record: A technical report. *Health Informatics J*. 2020;26(1):129-140. doi:10.1177/1460458218813710

2. Jensen RE, Rothrock NE, DeWitt EM, et al. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Med Care*. 2015;53(2):153-159. doi:10.1097/MLR.0000000000000289

3. Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice--adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol*. 2013;66(8 Suppl):S12-20. doi:10.1016/j.jclinepi.2013.04.005

4. Slover JD, Karia RJ, Hauer C, Gelber Z, Band PA, Graham J. Feasibility of integrating standardized patient-reported outcomes in orthopedic care. *Am J Manag Care*. 2015;21(8):e494-500.

5. Lavallee DC, Chenok KE, Love RM, et al. Incorporating Patient-Reported Outcomes Into Health Care To Engage Patients And Enhance Care. *Health Aff*. 2016;35(4):575-582. doi:10.1377/hlthaff.2015.1362

6. Basch E, Deal AM, Kris MG, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. *J Clin Oncol*. 2016;34(6):557-565. doi:10.1200/JCO.2015.63.0830

7. Jensen RE, Snyder CF. PRO-cision Medicine: Personalizing Patient Care Using Patient-Reported Outcomes. *J Clin Oncol*. 2016;34(6):527-529. doi:10.1200/JCO.2015.64.9491

8. Wagner LI, Schink J, Bass M, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. *Cancer*. 2015;121(6):927-934. doi:10.1002/cncr.29104

9. Basch E, Artz D, Iasonos A, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *J Am Med Inform Assoc*. 2007;14(3):264-268. doi:10.1197/jamia.M2177

10. Berry DL, Blumenstein BA, Halpenny B, et al. Enhancing Patient-Provider Communication With the Electronic Self-Report Assessment for Cancer: A Randomized Trial. *J Clin Oncol*. 2011;29(8):1029-1035. doi:10.1200/JCO.2010.30.3909

11. Ayers DC, Bozic KJ. The importance of outcome measurement in orthopaedics. *Clin Orthop Relat Res*. 2013;471(11):3409-3411. doi:10.1007/s11999-013-3224-z

12. Hung M, Hon SD, Cheng C, et al. Psychometric Evaluation of the Lower Extremity Computerized Adaptive Test, the Modified Harris Hip Score, and the Hip Outcome Score. *Orthop J Sport Med*. 2014;2(12):2325967114562191. doi:10.1177/2325967114562191

13. Hung M, Franklin JD, Hon SD, Cheng C, Conrad J, Saltzman CL. Time for a paradigm shift with computerized adaptive testing of general physical function outcomes measurements. *Foot Ankle Int*. 2014;35(1):1-7. doi:10.1177/1071100713507905

14. Hung M, Hon SD, Franklin JD, et al. Psychometric properties of the PROMIS physical function item bank in patients with spinal disorders. *Spine (Phila Pa 1976)*. 2014;39(2):158-163. doi:10.1097/BRS.0000000000000097

15. Hung M, Stuart AR, Higgins TF, Saltzman CL, Kubiak EN. Computerized Adaptive Testing Using the PROMIS Physical Function Item Bank Reduces Test Burden With Less Ceiling Effects Compared With the Short Musculoskeletal Function Assessment in Orthopaedic Trauma Patients. *J Orthop Trauma*. 2014;28(8):439-443. doi:10.1097/BOT.0000000000000059

16. Hung M, Baumhauer JF, Latt LD, et al. Validation of PROMIS ® Physical Function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clin Orthop Relat Res*. 2013;471(11):3466-3474. doi:10.1007/s11999-013-3097-1

17. Beckmann JT, Hung M, Voss MW, Crum AB, Bounsanga J, Tyser AR. Evaluation of the Patient-Reported Outcomes Measurement Information System Upper Extremity Computer Adaptive Test. *J Hand Surg Am*. 2016;41(7):739-744.e4. doi:10.1016/j.jhsa.2016.04.025

18. Lyman S, Lee Y-Y, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: A Short-form Hip Replacement Survey. *Clin Orthop Relat Res*. 2016;474(6):1472-1482. doi:10.1007/s11999-016-4718-2

19. Ayers DC, Zheng H, Franklin PD. Integrating Patient-reported Outcomes Into Orthopaedic Clinical Practice: Proof of Concept From FORCE-TJR. *Clin Orthop Relat Res*. 2013;471(11):3419.

20. Ayers DC, Li W, Harrold L, Allison J, Franklin PD. Preoperative pain and function profiles reflect consistent TKA patient selection among US surgeons. *Clin Orthop Relat Res*. 2015;473(1):76-81. doi:10.1007/s11999-014-3716-5

21. Forsberg HH, Nelson EC, Reid R, et al. Using patient-reported outcomes in routine practice: three novel

use cases and implications. *J Ambul Care Manage*. 38(2):188-195. doi:10.1097/JAC.0000000000000052

22. Yu SW, Szulc AL, Walton SL, Davidovitch RI, Bosco JA, Iorio R. Liposomal Bupivacaine as an Adjunct to Postoperative Pain Control in Total Hip Arthroplasty. *J Arthroplasty*. 2016;31(7):1510-1515. doi:10.1016/j.arth.2016.01.004

23. Franklin PD, Lewallen D, Bozic K, Hallstrom B, Jiranek W, Ayers DC. Implementation of patient-reported outcome measures in U.S. Total joint replacement registries: rationale, status, and plans. *J Bone Joint Surg Am*. 2014;96 Suppl 1:104-109. doi:10.2106/JBJS.N.00328

24. 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. doi:10.1016/j.jhsa.2014.06.130

25. Beckmann JT, Hung M, Bounsanga J, Wylie JD, Granger EK, Tashjian RZ. Psychometric evaluation of the PROMIS Physical Function Computerized Adaptive Test in comparison to the American Shoulder and Elbow Surgeons score and Simple Shoulder Test in patients with rotator cuff disease. *J Shoulder Elbow Surg*. 2015;24(12):1961-1967. doi:10.1016/j.jse.2015.06.025

26. Northwestern University. Applications of HealthMeasures: Applications in Evaluating Quality of Care. HealthMeasures. Published 2016. Accessed July 27, 2016. www.healthmeasures.net/applications-of-healthmeasures

27. Bantug ET, Coles T, Smith KC, et al. Graphical displays of patient-reported outcomes (PRO) for use in clinical practice: What makes a pro picture worth a thousand words? *Patient Educ Couns*. 2016;99(4):483-490. doi:10.1016/j.pec.2015.10.027

28. Sittig DF, Murphy DR, Smith MW, Russo E, Wright A, Singh H. Graphical display of diagnostic test results in electronic health Records: a comparison of 8 systems. *J Am Med Informatics Assoc*. 2015;22(4):900-904. doi:10.1093/jamia/ocv013

29. Grossman L, Feiner S, Mitchell E, Masterson Creber R. Leveraging Patient-Reported Outcomes Using Data Visualization. *Appl Clin Inform*. 2018;09(03):565-575. doi:10.1055/s-0038-1667041

30. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract*. 2006;12(5):559-568. doi:10.1111/j.1365-2753.2006.00650.x

31. Baumhauer JF, Bozic KJ. Value-based Healthcare: Patient-reported Outcomes in Clinical Decision Making. *Clin Orthop Relat Res*. 2016;474(6):1375-1378. doi:10.1007/s11999-016-4813-4

32. Andrawis J, Akhavan S, Chan V, Lehil M, Pong D, Bozic KJ. Higher Preoperative Patient Activation Associated With Better Patient-reported Outcomes After Total Joint Arthroplasty. *Clin Orthop Relat Res*. 2015;473(8):2688-2697. doi:10.1007/s11999-015-4247-4

33. Nystrom DT, Singh H, Baldwin J, Sittig DF, Giardina TD. Methods for Patient-Centered Interface Design of Test Result Display in Online Portals. *eGEMs (Generating Evid Methods to Improv patient outcomes)*. 2018;6(1):15. doi:10.5334/egems.255

34. Snyder CF, Smith KC, Bantug ET, Tolbert EE, Blackford AL, Brundage MD. What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. *Cancer*. 2017;123(10):1848-1859. doi:10.1002/cncr.30530

35. Brundage MD, Smith KC, Little EA, Bantug ET, Snyder CF, PRO Data Presentation Stakeholder Advisory Board. Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. *Qual Life Res*. 2015;24(10):2457-2472. doi:10.1007/s11136-015-0974-y

36. Baumhauer JF. Patient-Reported Outcomes — Are They Living Up to Their Potential? *N Engl J Med*. 2017;377(1):6-9. doi:10.1056/NEJMp1702978