Published in final edited form as:

Med Decis Making. 2018 November ; 38(8): 994–1005. doi:10.1177/0272989X18791177.

Picture This: Presenting Longitudinal Patient-reported Outcome Research Study Results to Patients

Elliott Tolbert, PhD^{1,2},

624 N. Broadway, Room 725, Baltimore, MD 21205, 410-955-9867, etolber2@jhmi.edu

Michael Brundage, MD, MSc³,

Cancer Clinic of Southeastern Ontario, 25 King Street West, Kingston, ON K7L 5P9, Canada, 613-533-6000, michael.brundage@krcc.on.ca

Elissa Bantug, MHS⁴,

Sidney Kimmel Comprehensive Cancer Center, 1650 Orleans Street, Baltimore, MD 21287, 410-502-3472, ebantug1@jhmi.edu

Amanda L. Blackford, ScM⁴,

Johns Hopkins School of Medicine, 550 N. Broadway, Room 1111, Baltimore, MD 21205, 410-614-0361, ablackf1@jhmi.edu

Katherine Smith, PhD^{2,4},

624 N. Broadway, Room 726, Baltimore, MD 21205, 410-502-0025, ksmit103@jhu.edu

Claire Snyder, PhD^{1,2,4}, and

624 N. Broadway, Room 649, Baltimore, MD 21205, 443-287-5469, csnyder@jhu.edu

PRO Data Presentation Stakeholder Advisory Board

¹Johns Hopkins School of Medicine, Baltimore, MD

²Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

³Queens Cancer Research Institute, Kingston, ON, Canada

⁴Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD

Abstract

Background: Patient-reported outcome (PRO) results from clinical trials and research studies can inform patient-clinician decision-making. However, data presentation issues specific to PROs, such as scaling directionality (higher scores may represent better or worse outcomes) and scoring strategies (normed vs. non-normed scores) can make the interpretation of PRO scores uniquely challenging.

Correspondence to: Elliott Tolbert.

Study Locations: Johns Hopkins Hospital, Johns Hopkins Sibley Memorial Hospital, Johns Hopkins Suburban Hospital, Anne Arundel Medical Center, Greater Baltimore Medical Center, Inova Health System, and Peninsula Regional Medical Center

Presented in part at the 2016 Health Literacy Research Conference and 2016 International Society for Quality of Life Research Annual Meeting

Conflicts of Interest: none

Objective: To identify the association of PRO score directionality, score norming, and other factors on a) how accurately PRO scores are interpreted and b) on how clearly they are rated to be by patients, clinicians, and PRO researchers.

Methods: We electronically surveyed adult cancer patients/survivors, oncology clinicians, and PRO researchers, and conducted one-on-one cognitive interviews with patients/survivors and clinicians. Participants were randomized to one of three line graph formats showing longitudinal change: higher scores indicating “better”, “more” (better for function, worse for symptoms), or “normed” to a population average. Quantitative data evaluated interpretation accuracy and clarity. Online survey comments and cognitive interviews were analyzed qualitatively.

Results: The internet sample included 629 patients, 139 clinicians, and 249 researchers; 10 patients and 5 clinicians completed cognitive interviews. “Normed” line graphs were less accurately interpreted than “more” (OR=0.76; p=.04). “Better” line graphs were more accurately interpreted than both “more” (OR=1.43; p=.01) and “normed” (OR=1.88; p=.04). “Better” line graphs were more likely to be rated clear than “more” (OR=1.51; p=0.05). Qualitative data informed interpretation of these findings.

Limitations: The survey relied on the online platforms used for distribution and consequent snowball sampling. We do not have information regarding participants’ numeracy/graph literacy.

Conclusions: For communicating PROs as line graphs in patient educational materials and decision aids, these results support using graphs with higher scores consistently indicating better outcomes.

INTRODUCTION

The recent emphasis on patient-centered care has increased demand for patient-reported outcomes (PROs), that is, data collected directly from patients about health conditions and treatments (e.g., symptoms, functioning, health-related quality of life).^{1,2} PROs are commonly collected in clinical trials and comparative research studies, and the results from these studies can contribute to patient-centered care by informing discussions and decision-making regarding the PRO impacts of different treatments.^{3–11} Oncologists have endorsed the use of PROs for this purpose, and there is evidence that PRO results can influence treatment choices.^{8,9,11}

For educational materials and decision aids that present PRO data to patients to be most effective, patients (and their clinicians) have to be able to interpret them accurately. As with any quantitative data, the education and numeracy of the interpreter will influence the accuracy with which the data are interpreted. PRO data, however, have unique characteristics that create additional challenges to their accurate interpretation. There exists a myriad of PRO questionnaires creating a variety of data structures and summaries.¹² Variability exists in PRO scoring (e.g., higher scores indicating better or worse outcomes), scaling (e.g., scores ranging from 0–100 indicating the worst-to-best possible scores or scores “normed” to a defined population), and whether data are presented as mean scores over time or proportions meeting a responder definition (i.e., the proportions improved, stable, worsened). Evidence has documented that this variability in PRO data scoring and analyses make PRO data difficult to interpret for both patients and clinicians.^{9, 10, 13, 14} Efforts to

improve understanding should incorporate and address these attributes, as they are inherent to PRO questionnaires.

Despite a large body of literature addressing graphical presentation of clinical data, only nine empirical studies have addressed graphic presentation of PROs specifically¹⁵; general findings of these studies include that no single format is accurately interpreted by all patients and clinicians but that some formats are more accurately interpreted than others. Notably, accuracy of interpretation, perceived level of understanding, and preferences for the different formats were often discordant. Across studies, patient age and education sometimes predicted the accuracy of PRO graph interpretation, and patients tended to prefer simpler graphs compared to more complex annotations.

We undertook a three-part research program designed to improve graphical formats for presenting PRO data, and to then systematically determine the key aspects of PRO data that impact on their interpretation and perceived clarity. Part 1 was a mixed-method evaluation of existing approaches for presenting PRO data research study results,¹⁴ showing that line graphs were seen as useful and easy to understand, a finding consistent with the common use of line graphs to convey changes over time, and with previous research findings.^{14,16} Patients and clinicians rated line graphs of mean scores highest for ease-of-understanding and perceived clarity when compared to proportions responding (improved/stable/worsened), bar charts of average changes, and cumulative distribution functions, but patients typically differed from clinicians in that clinicians valued information such as p-values and confidence intervals, whereas patients tended to find such information confusing¹⁴. Thus, in Part 2 of our study, we evaluated the presentation of PRO data to patients separately from that presented to clinicians. We partnered with stakeholder workgroups of clinicians and patients to develop candidate PRO data display formats that improved presentation of directionality (whether higher scores are better or worse) and interpretability, such as clear labeling and descriptive axis labels (along with numeric scores).¹³

Here, we present findings from Part 3 of our study, where three different line graph versions (developed in study Part 2) were tested across a broad population of patients, clinicians, and PRO researchers. Informed by our previous work, we developed three main research questions. First, given the improvements in graphic formats made in Part 2, we wished to describe how accurately patients interpret PRO data, and which line-graph versions they find the easiest to understand. Second, for PROs that report symptom scores (e.g., shortness of breath or pain scores), we observed that some participants intuitively interpret upward trending scores as reporting patients' symptoms worsening (more symptomatic), while others interpret the trend as patients improving (higher quality of life).¹⁴ We therefore designed the study to explicitly address the impact of PRO score symptom directionality on interpretation accuracy and perceived clarity. Third, we observed that group PRO data normed to a mean score of 50 with a standard deviation of 10 (as is reported by commonly used PRO questionnaires¹⁷) was easily understood by some participants, whereas others found this approach confusing,¹⁴ preferring scaled results from 0–100 (an approach also commonly used¹⁸). Our study design, therefore, also addressed the impact of score norming on interpretation accuracy and perceived clarity. In sum, we aimed to identify the line graph

format that conveyed the treatment option with more favorable patient reported outcomes most accurately and with the highest perceived clarity.

METHODS

Study Design

In this cross-sectional, mixed-methods study, we conducted an online survey of adult cancer patients/survivors, cancer clinicians, and PRO researchers evaluating graphs presenting PRO results from a hypothetical clinical trial comparing Treatment X and Treatment Y. The online survey included both quantitative data (interpretation accuracy, clarity), as well as free-text comments that provided insights regarding close-ended question responses. In addition, we obtained qualitative feedback on the different formats tested by administering the same survey in-person to oncology patients and clinicians using a one-on-one cognitive interview structure. Thus, this study incorporated an embedded mixed-methods design, with both the free-text comments from the online survey and the interviews supplementing the quantitative survey data.¹⁹ We initially targeted 10 patient and 5 clinician interview participants, with the option to add interviews if thematic saturation (taking both the interviews and online free-text comments into account) was not achieved. This mixed methods approach allowed qualitative data to provide useful insight concerning the quantitative results by illuminating users' perspectives and how they differ.

A nine-member Stakeholder Advisory Board (SAB) guided the three-part project (see listing and affiliations in Acknowledgement). The SAB included members from three key stakeholder groups: patients/caregivers (who use PRO data for their own care), clinicians (who use PRO data to improve their own understanding and to educate patients), and PRO researchers (who produce PRO data).

The Johns Hopkins School of Medicine Institutional Review Board deemed the online survey exempt and approved the interview study. The instructions to the online survey indicated that survey completion represented consent; interview participants provided written consent. The funding source had no role in study design, data collection, analysis, interpretation, writing, or decision to submit the manuscript for publication.

Population and Setting

To recruit adult cancer patients and survivors, cancer clinicians, and PRO researchers, we partnered with the study's SAB to distribute the survey link to our target populations. Recipients were also encouraged to forward the survey link to individuals who fit the eligibility criteria. Eligibility, defined as 21 years of age or older and self-identified as a cancer patient/survivor, health care clinician to cancer patients, or PRO researcher (PhD, MD, Masters-level or otherwise experienced researcher who develops and investigates methods associated with PROs [not necessarily in cancer]), was screened through initial survey questions.

Participants who completed the cognitive interview version were recruited from the Johns Hopkins Clinical Research Network (JHCRN), a consortium of academic and community health systems in the US mid-Atlantic. Eligible patients were 21 years, diagnosed with any

cancer (excluding non-melanoma skin cancer), 6 months post-diagnosis, not currently undergoing acute treatment, and able to communicate in English. We purposively sampled patients to include a variety of education levels, cancer types, and JHCRN sites. Eligible clinicians were active oncologists, including medical, radiation, surgical, gynecologic/urologic, nurse practitioners/physician assistants, and fellows. Purposive sampling among clinicians was structured so as to include various JHCRN sites and specialties.

Study Procedures

Participants were randomized to view one of three versions of line graphs presenting mean scores over time for two clinical trial treatments: lines going up indicating “better” outcomes, lines going up indicating “more” (better for function domains, worse for symptoms), or lines that had been “normed” to a general population average of 50 (Figs. 1, 2, and 3). As illustrated in the figures, each format showed two function and two symptom PRO domains.

Participants were asked three accuracy questions about the line graph type to which they were randomized: “At 12 months, on which treatment are patients better able to do PHYSICAL activities?”; “At 12 months, on which treatment do patients report better EMOTIONAL well-being?”; and “At 12 months, on which treatment do patients report worse FATIGUE?” Answer choices for each question included “Treatment X”, “Treatment Y”, and “Treatments are about the same”. The data were displayed on the same screen as the questions, so no recall was required. The questions asked and the data displayed remained the same for all assigned formats (“better,” “more,” or “normed”). After completing the accuracy questions, participants rated the line graph format’s clarity by answering: “How clear are these graphs to you?” Response options were “Very confusing”, “Somewhat confusing”, “Somewhat clear”, and “Very clear”. The last question provided an opportunity for open-ended feedback about the graphs in a text box. Upon conclusion, participants had the opportunity to enter for a chance to receive a \$100 Amazon giftcard.

Cognitive interview participants completed the online survey and were randomly assigned to one of the three line-graph versions. Before beginning, participants were told that the interview would be recorded and that they would be asked to think aloud as they completed the survey. Instead of the text boxes, the survey prompted interview participants to state their open-ended feedback about the graphs to the interviewer with them. At the end of the interview, participants were given an opportunity to share any overall feedback that was not captured during survey completion. Participants received a \$35 giftcard as compensation.

Analysis

Quantitative data included demographic characteristics and accuracy and clarity responses from online survey participants, and were summarized descriptively by respondent type (patient, clinician, researcher). In our primary analysis, we compared the accuracy of interpretation and clarity ratings of the three different line graph versions in multivariable generalized estimating equation (GEE) logistic regression models with the individual as the cluster unit and adjusting for respondent type. The interpretation accuracy was based on the responses to the 3 accuracy questions, and fixed effects for the specific questions were

included in the model. There were two clarity rating outcomes: (1) those rating the format “very” or “somewhat” clear and (2) those rating the format “very” clear only. In secondary analyses, we conducted GEE models stratified by respondent subgroup (patient, clinician, researcher).

Qualitative analysis included engagement with information obtained from both cognitive interviews and open-ended text box comments obtained from online survey participants. Data obtained from the interviews were analyzed using a coding scheme developed by one member of the research team and reviewed by the wider team, and similar to the schemes we had applied in previous parts of the study.^{13,14} Using a deductive approach, codes were structured to facilitate comparisons of areas of understanding versus confusion, and to capture positive or negative comments about the various presentations, as well as specific misinterpretations, and directionality concerns (i.e., interpretation of what “higher” means). The codebook was piloted twice by the research team before being finalized. A single coder initially coded each transcript using ATLAS.ti²⁰ and a second member reviewed each coded transcript, such that all data were fully reviewed twice. In instances of disagreement, another member of the team with expertise in qualitative research provided a third review. After coding was complete, a thematic report was generated that served as the foundation for the analysis presented. Open-ended text box comments obtained from online survey participants were sorted into broad, preexisting categories of “positive” and “negative”, and were integrated into the analysis of the interview data.

RESULTS

Study Sample

The internet sample (n= 1017) included 629 patients, 139 clinicians, and 249 PRO researchers (Table 1). Patients had a mean age of 58, were 87% female, 94% White, 23% <college graduate and 56% breast cancer survivors. Among clinicians, the average age was 44 years, 44% practiced medical oncology, and the average time in practice was 16 years. Researchers had a mean age of 45, and 46% had more than 10 years of experience in the field. For the interviews, patient participants were sampled by cancer type, clinical site, and education level; clinician sampling characteristics included specialty and hospital site. Based on the previous parts of this research agenda,^{13, 14} we expected to achieve thematic saturation by 15 interviews. This was indeed the case (though interviews would have continued had saturation not been met). With the initial target of 10 patients and 5 clinicians, no themes were identified that were not also identified from the substantial qualitative information provided by the survey comments. The 10 patients who participated in cognitive interviews reflect the sampling characteristics that were designated for interview participants throughout this research program and include: 30% breast cancer survivors, 30% from Johns Hopkins, and 30% <college graduate. The 5 clinicians who participated in interviews were 40% from Johns Hopkins and included an oncology fellow, nurse practitioner, and surgical, radiation, and medical oncologists.

Findings

Patients were more likely to accurately interpret “better” line graphs than “more” or “normed” line graphs. The proportion of patients answering all questions correctly was 56% for “better” formatted lines, compared to 41% for “more” and 39% for “normed” graphs. Similar patterns were seen for the proportion answering two questions correctly (81%, 73% and 60% respectively). Clinician results showed a similar pattern; the proportions getting all answers correct were 70%, 65%, and 65% for “better”, “more” and “normed” respectively. In contrast, researchers were more likely to answer all questions correctly when presented with “more” formatted line graphs (75%) than for “better” (65%) or “normed” (40%) line graphs.

Respondents’ ratings for the clarity of each format were similar across respondent types and across format types. The “better” formatted line graphs were consistently rated most often as “very clear” or “somewhat clear” by all three groups (84% by patients, 81% by clinicians, and 85% by researchers). However, the range in the proportion rating each format “very clear” or “somewhat clear” was narrow: 77% (clinicians’ ratings of “more” format) to 85% (researchers’ rating of “better” formats) indicating the similarity across groups.

Multivariable considerations of these results are shown in Table 2. “Normed” line graphs were less accurately interpreted than “more” line graphs ($OR=0.76$; $p=.04$) and “better” line graphs were more accurately interpreted than both “more” ($OR=1.43$; $p=.01$) and “normed” line graphs ($OR=1.88$; $p=.04$). In addition, “better” line graphs were more likely to be rated somewhat or very clear than “more” line graphs ($OR=1.51$; $p=0.05$). The results of the analysis stratified by subgroup were largely similar (Table 3). Patients more accurately interpreted line graphs with higher always “better” versus “more” ($OR=1.69$; $p=.003$) or “normed” ($OR=1.81$; $p<.001$). Researchers more accurately interpreted “better” line graphs than “normed” ($OR=2.42$; $p=.005$) and “normed” line graphs less accurately than “more” line graphs ($OR=0.33$; $p<.001$). Multivariable modelling of the association between responder type and accuracy rates showed a significant difference between the patients vs. researchers ($OR=0.62$, $p<.001$) whereas the odds ratios for clinicians vs. researchers ($OR=1.21$) was not statistically significant.

Qualitative comments from the online survey and cognitive interviews are provided in Table 4. Participants randomized to “normed” line graphs did not provide very much feedback. The “normed” value was seen as helpful to clinicians, but confused patients. Participants randomized to “more” and “better” line graphs provided the most feedback, the majority of which focused on directionality. “More” line graph comments noted that up could mean different things, which could lead to errors from having to switch how directionality should be interpreted. “Better” line graph comments pointed out how changing the scale could result in interpretation errors as one must orient to the direction of the scale each time in order to correctly interpret the graph. Overall, “more” line graphs received the most negative feedback and the majority of comments focused on the issue of directionality.

DISCUSSION

In order for PRO data collected from clinical trials and research studies to be most useful in educating patients and helping them make decisions about their care, efforts must be made to present the PRO results in an understandable way. This study contributed to this process by evaluating the interpretation accuracy and clarity of different ways to present longitudinal PRO data to patients. Our findings clearly show that overall, line graphs with higher always indicating better average outcomes over time were more accurately interpreted than were the competing formats. However, significant differences were seen in the multivariable models among patient respondents (who followed the general trend of interpreting “higher always better” more accurately) and researchers (who tended to interpret the “more” formatted line graphs more accurately). The reasons for these observed differences cannot be obtained from the data collected in our study, but presumably relate to the *a priori* experience of PRO researchers in the interpretation of PRO findings on scales using the “more” presentation format.

Little previous research has explicitly examined visual formats for the presentation of PRO data, and further synthesis is required.¹⁵ Previous research has found that line graphs displaying PROs were more useful and easy to understand than other methods,^{14, 16} but the development of best practice recommendations for the presentation of PRO data are more complex than simply recommending that line graphs be used, following the general principles of numeric presentation for decision support.^{21,22} The optimal communication of PRO data presents unique challenges, in that the numeric outcomes themselves are inherently unfamiliar to patients and physicians, the wide variety of PRO measures used vary in how their numeric outcomes are scaled, and the outcomes themselves are multidimensional (and may indicate both patient improvement and worsening across different domains, cross-sectionally or over time). Despite a great number of established International Patient Decision Aids Standards (IPDAS) for decision aids, many of which focus on effective communication of outcome probabilities, the optimal choice of visual format for presentation of these probabilities remains unclear and is not specifically addressed in the IPDAS checklist.^{21,22}

Our findings suggest that consideration of varying directionality within PRO measures, and that the use of normed measures, reduce the accuracy of interpretation among patients, clinicians, and researchers. These findings have important implications for the effective communication of PRO information between clinicians and patients at the time of a clinical encounter, and for the effective creation of patient educational materials and decision aids. Moreover, the findings have implications for the effective communication of PRO data in other contexts such as clinician’s interpretation of PRO-determined adverse events, performance reports, policy development, and other applications of PROs generated by comparative studies.

At the conclusion of this project, the study’s SAB suggested that the evidence-base developed through this research was central for informing recommendations for displaying PRO data in educational materials and decision aids – but could not serve as the sole determinant owing to the complexity of implementing our study findings in the context of

widely used legacy measures and the development of new ways (such as computerized adaptive technology) to capture and report PROs. They advised engaging a broader group of stakeholders to develop recommendations that are both evidence-based and stakeholder-driven. In the next step of this research, we are partnering with stakeholders to develop recommendations for PRO data presentation in patient educational materials and decision aids using a modified-Delphi approach and informed by these results.²³ The follow-on project will recommend PRO data display approaches and identify areas requiring further research.

There are limitations that should be noted when considering the results of this study. Although previous stages of this research agenda^{13,14} and the interviews in this study enabled purposive sampling by education level, the online survey did not, which resulted in a more highly educated sample. We also do not have information regarding participants' numeracy/graph literacy; however, past research has shown that patients generally understand the types of displays that were included.^{16,24} The survey relied on the online platforms used for distribution and consequent snowball sampling, resulting in a highly educated and predominantly female convenience sample. Perceived clarity as measured by this study is likely to be related to an individual's familiarity with the formats.²⁵ Also, the data presented to participants were relatively straightforward and did not explore extremes. Online survey participation depended on internet access and familiarity, potentially missing populations that lack these attributes. The online sample also relied on self-assessment for eligibility and had no way to deter repeated participation or participation by ineligible persons.

The study also had several strengths. First, the iterative approach and mixed-methods study design enhances the validity of the results. The large sample that resulted from online distribution included participants and locations that otherwise would not have been accessible. The interviews permitted purposive sampling and no systematic differences were found between the feedback from online and interview participants. Finally, the data that were presented were held constant regardless of format, which helped to isolate the effects of each format.

The findings of this study help address the need to provide patients and clinicians with the most effective PRO data presentations, thereby increasing opportunities to promote patient-centered care. In the next step of this research, we will use these results as part of an evidence base in partnering with stakeholders to develop recommendations for PRO data presentation in patient educational materials and decision aids.

ACKNOWLEDGEMENTS

This analysis was supported by a Patient-Centered Outcomes Research Institute (PCORI) Award (R-1410-24904). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee. Drs. Smith and Snyder are members of the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins (P30 CA 006973). The PRO Data Presentation Stakeholder Advisory Board includes Neil K. Aaronson, PhD (Netherlands Cancer Institute); Patricia A. Ganz, MD (University of California Los Angeles and Jonsson Comprehensive Cancer Center); Ravin Garg, MD (Anne Arundel Medical Center); Michael Fisch, MD (M.D. Anderson Cancer Center); Vanessa Hoffman, MPH (Bladder Cancer Advocacy Network); Bryce B. Reeve, PhD (University of North Carolina at Chapel Hill and Lineberger Comprehensive Cancer Center); Eden

Stotsky-Himelfarb (Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins); Ellen Stovall (National Coalition for Cancer Survivorship); Matthew Zachary (Stupid Cancer). The Johns Hopkins Clinical Research Network (JHCRN) site investigators and staff include: Ravin Garg, MD, and Steven P. DeMartino, CCRC, CRT, RPFT (Anne Arundel Medical Center), Melissa Gerstenhaber, MAS, MSN, RN, CCRN (JHCRN/Anne Arundel Medical Center); Gary Cohen, MD, and Cynthia MacInnis, BS, CCRP (Greater Baltimore Medical Center); James Zabora, ScD, MSW (Inova Health System), and Sandra Schaefer, BSN, RN, OCN (JHCRN/Inova Health System); Paul Zorsky, MD, Lynne Armiger, MSN, CRNP, ANP-C, Sandra L. Heineken, BS, RN, OCN, and Nancy J. Mayonado, MS (Peninsula Regional Medical Center); Michael Carducci, MD (Johns Hopkins Sibley Memorial Hospital); Carolyn Hendricks, MD, Melissa Hyman, RN, BSN, OCN, and Barbara Squiller, MSN, MPH, CRNP (Suburban Hospital). Lastly, we are most grateful to the patients and clinicians who contributed and participated in this study.

Funding: Patient-Centered Outcomes Research Institute (R-1410-24904); Drs. Snyder and Smith are members of the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins (P30CA006973).

Financial support for this study was provided by the Patient-Centered Outcomes Research Institute. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

REFERENCES

1. U.S. Food and Drug Administration: Guidance for Industry. Patient Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Federal Register. 2009;74:65132–65133.
2. Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D, Rothman M. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health*. 2003;6(5): 522–531. [PubMed:] [PubMed: 14627058]
3. Greenhalgh J The applications of PROs in clinical practice: What are they, do they work, and why? *Qual Life Research*. 2009;18(1):115–123. [PubMed:] [PubMed: 19105048]
4. Lipscomb J, Gotay CC, Snyder C, editors. 2005 Outcomes assessment in cancer: measures, methods and applications. Cambridge: Cambridge University Press.
5. Bruner DW, Bryan CJ, Aaronson N, Blackmore CC, Brundage M, Cella D, Ganz PA, Gotay C, Hinds PS, Kornblith AB, Movsas B. 2007. Issues and challenges with integrating patient-reported outcomes in clinical trials supported by the National Cancer Institute-sponsored clinical trials networks. *J Clin Oncol* 2007;10;25(32):5051–7. [PubMed:] [PubMed: 17991920]
6. Till JE, Osoba D, Pater JL, Young JR. Research on health-related quality of life: dissemination into practical applications. *Qual Life Research*. 1994;3(4):279–83. [PubMed:] [PubMed: 7812281]
7. Au HJ, Ringash J, Brundage M, Palmer M, Richardson H, Meyer RM. Added value of health related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG. *Expert Rev Pharmacoecon Outcomes Res* 2010;10(2):119–28. [PubMed:] [PubMed: 20384559]
8. Brundage MD, Feldman-Stewart D, Bezjak A, et al. The value of quality of life information in a cancer treatment decision. ISOQOL 11th annual conference, San Francisco, 2005.
9. Brundage M, Bass B, Jolie R, Foley K. A knowledge translation challenge: clinical use of quality of life data from cancer clinical trials. *Qual Life Research*. 2011;20(7): 979–985. [PubMed:] [PubMed: 21279446]
10. Snyder CF, Aaronson NK. Use of patient-reported outcomes in clinical practice. *Lancet* 2009;374(9687):369–370. [PubMed:] [PubMed: 19647598]
11. Bezjak A, Ng P, Skeel R, Depetrillo AD, Comis R, Taylor KM. Oncologists' use of quality of life information: results of a survey of eastern cooperative oncology group physicians. *Qual Life Research*. 2001;10(1):1–13. [PubMed:] [PubMed: 11508471]
12. PROQOLID. The patient-reported outcome and quality of life instruments database. 2012 <http://www.proquolid.org>
13. Smith KC, Brundage MD, Tolbert E, Little EA, Bantug ET, Snyder CF, PRO Data Presentation Stakeholder Advisory Board. Engaging stakeholders to improve presentation of patient-reported

outcomes data in clinical practice. *Supp Care Cancer*. 2016;24(10):1–9. [PubMed:] [PubMed: 27165054]

14. 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 Research*. 2015;24(10):2457–2472. [PubMed:] [PubMed: 26012839]
15. Bantug ET, Coles T, Smith KC, Snyder CF, Rouette J, Brundage MD. 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. [PubMed:] [PubMed: 26603445]
16. Brundage M, Feldman-Stewart D, Leis A, Bezjak A, Degner L, Velji K, Zetes-Zanatta L, Tu D, Ritvo P, Pater J. Communicating quality of life information to cancer patients: a study of six presentation formats. *J. Clin. Oncol* 2005;23(28):6949–6956. [PubMed:] [PubMed: 16192583]
17. National Institutes of Health. Intro to PROMIS. Available at: <http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis>.
18. Aaronson NK, Ahmedzai S, Bergman B et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Canc Inst* 1993;85:365–376. [PubMed:] [PubMed: 8433390]
19. Creswell JW, Plano Clark VL, Gutmann ML, Hanson WE. Advanced mixed methods research designs. *Handbook of Mixed Methods in Social and Behavioral Research*. 2003:209: 240.
20. AtlasTi, in, ATLAS.ti Scientific Software Development GmbH. 2014.
21. Joseph-Williams N, Newcombe R, Politi M, Durand MA, Sivell S, Stacey D, O'Connor A, Volk RJ, Edwards A, Bennett C, Pignone M. Toward Minimum Standards for Certifying Patient Decision Aids A Modified Delphi Consensus Process. *Med Decis Making*. 2014;34(6):699–710. [PubMed:] [PubMed: 23963501]
22. McDonald H, Charles C, Gafni A. Assessing the conceptual clarity and evidence base of quality criteria/standards developed for evaluating decision aids. *Health Expect*. 2014;17(2):232–243. [PubMed:] [PubMed: 22050440]
23. Brundage M, Bantug E, Smith K, Holzner B, Rivera Y, Snyder C, PRO Data Presentation Delphi Panel. Recommendations for Presenting Patient-Reported Outcomes Data to Patients and Clinicians: Results from a Modified-Delphi Consensus Process. 2017 Annual Meeting of the International Society for Quality of Life Research October 17–21, 2017 Philadelphia, Pennsylvania.
24. McNair AG, Brookes ST, Davis CR, Argyropoulos M, Blazeby JM. Communicating the results of randomized clinical trials: do patients understand multidimensional patient-reported outcomes?. *J. Clin. Oncol* 2010;28(5):738–743. [PubMed:] [PubMed: 20065187]
25. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Educ. Couns* 2008;73(3):448–455. [PubMed:] [PubMed: 18755566]

Patient's Functioning

Patient's Symptoms

Figure 1:
Example of “More” Line Graph Format Tested

Patient's Functioning

Patient's Symptoms

Figure 2:
Example of “Better” Line Graph Format Tested

Patient's Functioning

Patient's Symptoms

Figure 3:
Example of “Normed” Line Graph Format Tested

Table 1:

Sample Characteristics

Characteristic	Patients (n=629)	Clinicians (n=139)	Researchers (n=249)
Age			
Mean(SD)	58.1 (11.3)	43.8 (12.56)	45.0 (11.92)
Male n(%)	70 (13.3)	58 (46.0)	74 (33.3)
Race n(%)			
White	494 (94.1)	87 (70.2)	175 (79.2)
Black/African-American	16 (3.0)	3 (2.4)	4 (1.8)
Asian	6 (1.1)	23 (18.5)	32 (14.5)
Other	9 (1.7)	11 (8.9)	10 (4.5)
Hispanic n(%)	16 (3.1)	9 (7.3)	9 (4.1)
Country n(%)			
United States	450 (85.6)	62 (49.2)	107 (48.6)
Education n(%)			
<High School Graduate	1 (0.2)		
High School Graduate	34 (6.5)		
Some College	88 (16.7)		
College graduate	199 (37.8)		
Any post-secondary work	205 (38.9)		
Cancer Type n(%) <i>all that apply</i>			
Breast	351 (55.8)		
Bladder	44 (7.0)		
Colorectal	44 (7.0)		
Prostate	26 (4.1)		
Lymphoma	21 (3.3)		
Gynecological	20 (3.2)		
Other	103 (16.4)		
Time Since Diagnosis n(%)			
<1 year	28 (5.3)		
1–5 years	215 (41.0)		
6–10 years	130 (24.8)		
11+ years	151 (28.8)		
History of Cancer n(%)		12 (9.5)	18 (8.2)
Provider Specialty n(%)			
Medical Oncology		55 (43.7)	
Radiation Oncology		16 (12.7)	
Surgical Oncology		15 (11.9)	
Gynecologic Oncology/Urology		2 (1.6)	
Oncology Nurse Practitioner/ Physician Assistant		5 (4.0)	
Other		33 (26.2)	

Characteristic	Patients (n=629)	Clinicians (n=139)	Researchers (n=249)
Provider Years in Practice			
Mean(SD)		15.7 (11.24)	
PRO Researcher Expertise n(%) all that apply			
Patient Perspective			35 (14.1)
Clinician			28 (11.2)
Clinician-Scientist			42 (16.9)
PRO Assessment/ Psychology/Sociology			134 (53.8)
Clinical Trial Methods/Analysis			70 (28.1)
Psychometrics			59 (23.7)
Policy/Public Health			59 (23.7)
Journal Editor			17 (6.8)
Frequent Journal Reviewer			78 (31.3)
Regulator/Health Administrator			3 (1.2)
Other			16 (6.4)
PRO Research Experience n(%)			
Student			22 (10.1)
Post-doc			16 (7.3)
<5 years experience			34 (15.6)
5–10 years experience			45 (20.6)
>10 years experience			101 (46.3)

Table 2:

Adjusted odds ratios for the association between format and outcome (giving the correct answer and clarity ratings). All odds ratios in a given column are from a single multivariable logistic regression model estimated using generalized estimating equations (GEE). The cluster unit was the individual respondent. Terms for the fixed effects of the specific questions were included in the model for accuracy. Adjusted for respondent type (patient, clinician, researcher).

	Correct Answer, all 3		Rated Very or Somewhat Clear		Rated Very Clear	
	OR [95% CI]	P	OR [95% CI]	P	OR [95% CI]	P
"Normed" vs. "More"	0.76 [0.59, 0.98]	0.04	1.07 [0.72, 1.58]	0.75	0.86 [0.62, 1.19]	0.37
"Better" vs. "More"	1.43 [1.07, 1.91]	0.01	1.51 [1, 2.29]	0.05	1.12 [0.81, 1.55]	0.50
"Better" vs. "Normed"	1.88 [1.42, 2.5]	0.04	1.42 [0.93, 2.16]	0.75	1.3 [0.93, 1.81]	0.37

Adjusted odds ratios for the association between format and outcome (giving the correct answer and clarity ratings). All odds ratios in a given column are from a single multivariable logistic regression model estimated using generalized estimating equations (GEE) and stratified by respondent type. The cluster unit was the individual respondent. Terms for the fixed effects of the specific questions were included in the model for accuracy.

Table 3:

	Correct Answer, all 3		Rated Very or Somewhat Clear		Rated Very Clear	
	OR [95% CI]	P	OR [95% CI]	P	OR [95% CI]	P
PATIENTS						
"Normed" vs. "More"	0.93 [0.68, 1.28]	0.67	0.98 [0.59, 1.62]	0.93	0.7 [0.45, 1.08]	0.11
"Better" vs. "More"	1.69 [1.2, 2.36]	0.003	1.53 [0.9, 2.6]	0.12	1.03 [0.68, 1.58]	0.88
"Better" vs. "Normed"	1.81 [1.28, 2.55]	<0.001	1.56 [0.91, 2.67]	0.10	1.49 [0.96, 2.31]	0.08
CLINICIANS						
"Normed" vs. "More"	1.06 [0.47, 2.4]	0.88	1.09 [0.4, 2.95]	0.86	1.65 [0.71, 3.85]	0.25
"Better" vs. "More"	1.56 [0.62, 3.93]	0.35	1.35 [0.46, 3.92]	0.58	1.63 [0.68, 3.92]	0.27
"Better" vs. "Normed"	1.46 [0.6, 3.58]	0.41	1.24 [0.44, 3.51]	0.69	0.99 [0.43, 2.29]	0.98
RESEARCHERS						
"Normed" vs. "More"	0.33 [0.18, 0.58]	<0.001	1.3 [0.58, 2.95]	0.53	0.92 [0.48, 1.76]	0.80
"Better" vs. "More"	0.79 [0.39, 1.58]	0.50	1.57 [0.69, 3.56]	0.28	1.07 [0.57, 2]	0.84
"Better" vs. "Normed"	2.42 [1.3, 4.51]	0.005	1.2 [0.49, 2.95]	0.68	1.16 [0.6, 2.25]	0.65

Table 4:

Selected Positive/ Negative Comments on the Three Line Graph Versions by Patients (P), Clinicians (C), and Researchers (R)

“More” Line Graphs	
Positive	<p>“For me it’s a personal preference to understand the, if you dip down that you can come back up and be in the same place rather than looking at what the end result is.” (P)¹</p> <p>“This is what I expect to see when I look at a PRO summary.” (R)</p>
Negative	<p>“Have to read the title explanation carefully. Also tricky because directionality is different in top two graphs and bottom two graphs. Probably greater chance of a wrong interpretation.” (P)</p> <p>“The fact that the positive/negative scale changes between functioning and symptoms (so that “up” means different things) makes error much, much more likely in interpreting these graphs.” (P)</p> <p>“On upper charts, higher up Y axis equals “better” but on lower charts, higher on the Y axis equals “worse” - a bit confusing if you don’t read carefully.” (P)</p> <p>“The positives go up, but the negative responses for fatigue and pain, also increase going up. This is a little confusing.” (P)</p> <p>“If you don’t read the chart header, um, it can be confusing to switch from better being the line going up to worse being the line going up, so consistency in that might be helpful, which is understandably hard considering your, what’s being measured.” (P)¹</p> <p>“The scales were flipped in some.” (C)</p> <p>“For the first two graphs, the top line represented better functioning while for the bottom two graphs, the top line represented more symptom burden; at first glance, one could misinterpret the results if not paying attention to the legend/axis labels.” (C)</p> <p>“Need to make sure the patient is oriented to whether higher is better or worse. I see you’ve tried to do this with the y-axis and the note under the title, but I still think it’s easy to glance quickly and miss those things.” (R)</p> <p>“For me it would be easier if all scales would go from low to high, with low meaning problems and high meaning no problems.” (R)</p>
“Better” Line Graphs	
Positive	<p>“This was easy to understand, since the ‘preferred’ result is on the top.” (P)</p> <p>“The inversion of the axis works well for showing the expected directionality for improvements. It might be further improved by a color gradation (red/yellow/green) from bottom to top.” (C)</p>
Negative	<p>“The lines rising seem to indicate more but it actually means less the Y axis should start at 0 and go to more to seem more familiar.” (P)</p> <p>“Expected the axis on Fatigue and perhaps Pain to be reversed, with No Fatigue or No Pain at the lower/starting point of the scale/axis - so I had to read that to understand what chart was telling me. Physical and Emotional were very clear.” (P)</p> <p>“I don’t like how the fatigue and pain have declining scales on Y axis as opposed to the physical and emotional scales Y axis. If I were using all 4 graphs, that would be confusing to me.” (C)</p> <p>“Don’t like the fact that the symptoms are reversed of the functioning - have to put extra effort into figuring out which way is right. The explanation under the title is very helpful, but still would like them to go the same way.” (R)</p>
“Normed” Line Graphs	
Positive	<p>“I would want to know, oh at 3 months you may not be that much better, but by 12 you will be.” (P)¹</p> <p>“I like the way that the graphs incorporate “average for U.S. adults”, as this is an important reference point when assessing symptom improvement or quality of life. I also like the way that they add clear language to indicate “line going up means more fatigue”. These are unusually clearly displayed graphs for patients (and providers), making the data easy to understand and interpret.” (C)</p> <p>“Average for US Adults VERY helpful.” (C)</p>
Negative	<p>“THE AVERAGE OF US ADULTS IS CONSTANT OVER 12 MONTHS???” (P)²</p>

¹. Cognitive interview participant

². Presented as captured