

Engaging stakeholders to improve presentation of patient-reported outcomes data in clinical practice

Katherine C. Smith^{1,4} · Michael D. Brundage² · Elliott Tolbert³ · Emily A. Little³ · Elissa T. Bantug⁴ · Claire F. Snyder^{3,4} · PRO Data Presentation Stakeholder Advisory Board

Received: 25 November 2015 / Accepted: 24 April 2016 / Published online: 10 May 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract

Purpose Patient-reported outcomes (PROs) can promote patient-centered care, but previous research has documented interpretation challenges among clinicians and patients. We engaged stakeholders to improve formats for presenting individual-level PRO data (for patient monitoring) and group-level PRO data (for reporting comparative clinical studies).

Methods In an iterative process, investigators partnered with stakeholder workgroups of clinicians and patients to address previously identified interpretation challenges. Candidate approaches were then tested in semi-structured, one-on-one interviews with cancer patients and clinicians. Interpretation issues addressed included conveying score meaning (i.e., what is good/bad) and directional inconsistency (whether higher scores are better/worse). An additional issue for individual-level PROs was highlighting potentially concerning scores and, for group-level PROs, identifying important between-group differences (clinical, statistical).

Results One-on-one interviews in a purposive sample of clinicians ($n=40$) and patients ($n=39$) provided insights regarding approaches to address issues identified. For example, adding descriptive labels to the Y-axis (none, mild, moderate, severe) helps address directional inconsistency and aids interpretation of score meaning. Red circles around concerning data points or a threshold line indicating worse-than-normal scores indicate possibly concerning scores for individual-level PRO data. For group-level PRO data, patients and some clinicians are confused by confidence limits and clinical versus statistical significance, but almost all clinicians want p values displayed.

Conclusions Variations in interpretation accuracy demonstrate the importance of presenting PRO data in ways that promote understanding and use. In an iterative stakeholder-driven process, we developed improved PRO data presentation formats, which will be evaluated in further research across a large population of patients and clinicians.

Keywords Cancer · Oncology · Patient-reported outcomes · Communication · Mixed methods · Stakeholder engagement

✉ Katherine C. Smith
katecsmith@jhu.edu

¹ Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Rm 726, Baltimore, MD, USA

² Queen's Cancer Research Institute, Queen's University, Kingston, ON, Canada

³ Division of General Internal Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA

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

Introduction

The ever-greater focus on patient-centered care demands data on patient-reported outcomes (PROs). PROs represent patients' subjective reports on health and treatment including symptoms, functioning, and health-related quality of life [1, 2]. PROs facilitate clinical decision-making by informing patients and providers about the impact of different treatments on patient functioning and well-being [3, 4]. PROs can also be used for

individual patient monitoring when patients complete PRO questionnaires and data are shared with clinicians to aid identifying and addressing patient concerns [5–8]. Using PROs to improve patient-centered care requires that both patients and clinicians understand PRO data as presented. Prior research has established that variation in the scoring, scaling, and presentation of PRO data can create challenges for both patients and clinicians in terms of interpretation and application in patient care [9–12].

We are conducting a three-part study to determine best practices for presenting PRO data graphically to patients and clinicians. Using a mixed-methods approach, part I assessed the patient and clinician understanding of current PRO data presentation approaches. The study evaluated various formats (e.g., line graphs, bar charts) for communicating *group-level* clinical trial/research study data and *individual-level* patient data for monitoring and management in clinic [12]. Both patients and clinicians preferred line graphs for individual-level data presentation, but preferences for group-level data presentation differed between patients and clinicians [12]. Patients typically preferred simple line graphs, whereas clinicians often wanted greater detail (e.g., confidence intervals, *p* values, normed scores).

A number of interpretive challenges for data presentation formats emerged from part I of our study. For both individual-level and group-level data, patients and clinicians were confused by variation in whether higher scores indicated better or worse outcomes. Directional inconsistency occurs when, for example, higher scores represent better functioning but more severe symptoms which is common in PRO measures [13, 14]. Other issues that emerged across individual- and group-level data included conveying a score's meaning (i.e., what is a good/bad score) and comparison to reference groups (e.g., national averages). Patients also required simplified language. A specific issue for individual-level data was how to highlight possibly concerning scores for clinical attention. For group-level data, a key challenge was displaying between-group differences that were statistically significant and/or clinically meaningful.

To address these issues, in part II, we engaged with patient and clinician stakeholders to develop improved presentation approaches. Given the difference between patient and clinician preferences for presenting group-level data, we stratified study methods into three separate topic areas: presenting individual-level data to both clinicians and patients, presenting group-level data to patients (e.g., educational materials, decision aids), and presenting group-level data to clinicians (e.g., peer-reviewed publications). This paper presents part II's innovative, stakeholder-driven methods and the key findings that informed the development of candidate presentation formats for testing in part III.

Methods

Study design

We conducted an iterative, stakeholder-centered, mixed-methods study to improve PRO presentation formats (see Fig. 1). We invited patient and clinician participants from part I to join workgroups to assist with developing approaches to address the previously identified interpretation challenges. A nine-member stakeholder advisory board (SAB) comprised of patients/caregivers, clinicians, and PRO researchers has provided guidance throughout the three-part project.

In part II, three separate workgroups were formed: patients and clinicians who addressed individual-level data presentation, patients and clinicians who addressed group-level data presented to patients, and clinicians who addressed group-level PRO data presented to clinicians. For each topic, we conducted the following iterative process: (1) investigators reviewed findings from part I and discussed approaches for addressing the interpretation challenges; (2) a workgroup met and provided feedback on the investigator-generated suggestions and generated other ideas; (3) alternative approaches were tested in semi-structured, one-on-one interviews with new patients and clinicians; and (4) results were summarized and presented to the SAB.

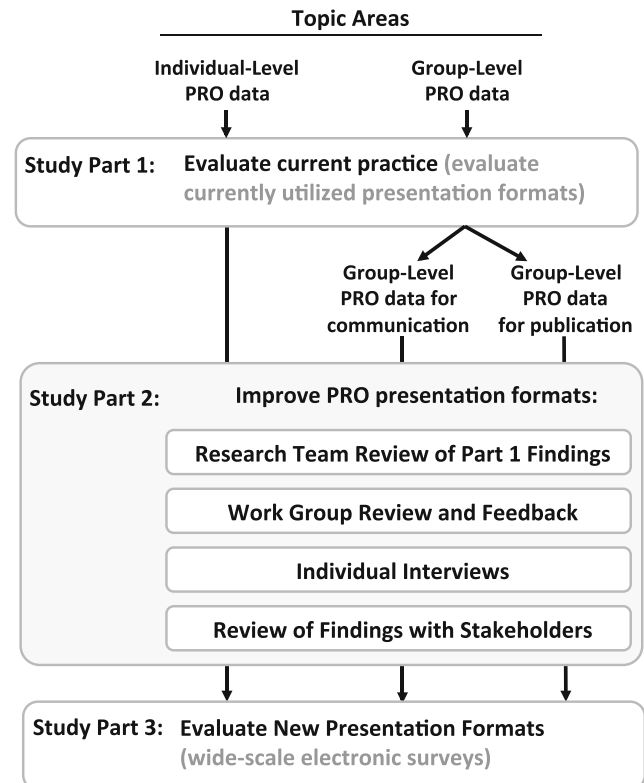


Fig. 1 Data collection for the three-part study of stakeholder understanding of PRO data presentation approaches

The Johns Hopkins School of Medicine Institutional Review Board approved the study. Workgroup *research collaborators* received \$70 gift cards and one-on-one interviewees \$35 gift cards, and all participants provided signed consent.

Participants and setting

Participants were recruited via the Johns Hopkins Clinical Research Network (JHCRN), a consortium of academic and community medical institutions representing diverse populations in the US mid-Atlantic. Workgroup collaborators were participants in part I [12] who agreed to be re-contacted for future research. For the one-on-one interviews, we purposively sampled “naïve” (not involved in part I) participants.

Patients

Eligible patients were ≥ 21 years, able to communicate in English, diagnosed with cancer (except non-melanoma skin) ≥ 6 months previously, and not currently receiving chemotherapy/radiation or within 6 months of surgery (long-term adjuvant treatment was acceptable). Our purposive sampling ensured no more than 30 % of participants had any given cancer type. Moreover, ≥ 30 % of participants were from Johns Hopkins and ≥ 30 % were from JHCRN sites. Finally, we purposively sampled by education to ensure that ≥ 10 % of patient participants did not have a college degree.

Clinicians

Clinicians in active practice treating adult cancer patients at participating sites were eligible. Specialties represented include medical, radiation, and surgical oncologists; gynecologic oncologists/urologists; oncology nurse practitioners/physician assistants; and oncology fellows, with ≥ 30 % of clinicians from Johns Hopkins and ≥ 30 % from JHCRN sites. The sample included ≤ 30 % from any specialty.

Workgroups

We recruited six patients and two clinicians for the individual-level data workgroup, three patients and two clinicians for the group-level data presented to patients workgroup, and five clinicians for the group-level data presented to clinicians workgroup. Two researchers with extensive focus group moderation experience facilitated the workgroups. Each workgroup began with a review and discussion of part I findings followed by feedback on potential modifications and alternative formats to address identified problems. Notes and workgroup recordings/transcripts served as resources in format development and refinement.

One-on-one interviews

Based on the workgroup feedback, the research team selected and refined formats to test in one-on-one, semi-structured interviews with newly recruited targeted end users (clinicians and/or patients). Participants first completed a sample PRO questionnaire, and were then shown between three and six formats in a randomly determined order displaying hypothetical results. For each format, participants answered two questions requiring data interpretation (in order to engage them in the formats) and also rated the format’s ease of understanding. Following the self-directed portion of the exercise, the interviewer conducted a semi-structured debrief, prompting participants to describe their interpretation of the formats and aspects they found helpful and confusing. During the interview, we showed participants a series of slightly altered formats to ascertain preferences for various alternatives (e.g., comparing shading concerning scores in red rather than shading normal scores in green).

Table 1 summarizes the interpretation challenges addressed and approaches tested for each topic. For individual-level data, we addressed line graphs of scores over time only, as that was the preferred format from part I. For group-level data, we addressed both the presentation of average scores/changes and proportions meeting a responder definition (i.e., improved, stable, worsened), as the appropriate format is driven by the conceptualization of the endpoint in the research design.

In part I, directional inconsistency (i.e., higher scores sometimes indicating better outcomes and sometimes worse) emerged as an interpretation challenge for both individual- and group-level data line graphs. Across all line graph formats, therefore, we added a header beneath the domain title to indicate whether higher or lower scores were better or worse. In addition, we evaluated various alternatives in the individual-level data interviews (see Table 1). This topic was not addressed specifically in the group-level data interviews due to other necessary issues for discussion; we expect findings from the individual-level data on this issue to also inform about the group-level data.

The second issue addressed for individual-level data was highlighting potentially concerning scores requiring clinical attention (both poor scores in absolute terms and important worsening from the previous assessment). For the group-level data presented to patients, format interpretation challenges included explaining the meaning of scores (i.e., whether scores are good/bad, what *normal* is) and highlighting important differences between treatments. Because part I suggested that patients found the statistical significance confusing, we evaluated approaches to convey simple *importance*, rather than

Table 1 Summary of interpretation challenges addressed and approaches tested for each data presentation topic

Topic	Formats included	Interpretation challenges addressed	Approaches tested
Individual-level data	Line graphs of scores over time	Directional inconsistency (i.e., sometimes higher scores are better and sometimes worse)	Standard descriptive <i>y</i> -axis labels (e.g., none, mild, moderate, severe) <ul style="list-style-type: none"> • Alternative: numbers without labels on the <i>y</i>-axis • Alternative: descriptive labels using questionnaire's response options Shading the normal range of the graph in green <ul style="list-style-type: none"> • Alternative: shading concerning scores in red • Alternative: spectrum shading the graph in green (normal), yellow (borderline), and red (concerning) Reversing scores for symptoms such that across symptom and function domains, lines going up reflect better outcomes
		Highlighting potentially concerning scores (poor in absolute terms or important worsening)	Red circles around possibly concerning data points based on the absolute score <ul style="list-style-type: none"> • Alternative: using a threshold line across the graph to indicate whether scores are potentially concerning • Alternative: shading to indicate whether scores are in normal or concerning range • Alternative: adding exclamation points to red circles, threshold line, or shading as an indication of possibly concerning change scores
Group-level data for patients	Average changes	Meaning of scores (i.e., good, bad, normal)	Line graphs of average changes over time with descriptive <i>y</i> -axis labels (e.g., severe fatigue, moderate fatigue, mild fatigue, no fatigue) <ul style="list-style-type: none"> • Alternative: numbers without labels on the <i>y</i>-axis
	Proportions meeting a responder definition	Highlighting important differences between treatments	Line graphs with the normal region shaded green <ul style="list-style-type: none"> • Alternative: shade the confidence limit around lines such that non-overlapping shading indicates important differences Bar charts with treatment labels in legend <ul style="list-style-type: none"> • Alternative: putting the treatment label directly on bars Icon array with treatment legend at the top <ul style="list-style-type: none"> • Alternative: treatment legend between the arrays
Group-level data for clinicians	Average changes	Meaning of scores	Pie chart <ul style="list-style-type: none"> • Alternative: numbers without labels on the <i>y</i>-axis
		Representations of statistical significance	Line graphs of average changes over time with descriptive <i>y</i> -axis labels (e.g., very high, moderate, poor, very poor) <ul style="list-style-type: none"> • Alternative: remove <i>p</i> values
	Representation of clinical importance	Line graphs of average changes over time with indication of clinical significance <ul style="list-style-type: none"> • Alternative: use arrows to illustrate clinically meaningful differences • Alternative: shade the confidence limit around lines such that non-overlapping shading indicates clinically meaningful difference • Alternative: define clinically meaningful differences in legend • Alternative: use asterisks to indicate clinically meaningful differences Bar charts of average changes at a single time point with asterisk indicating statistical significance next to domain labels <ul style="list-style-type: none"> • Alternative: asterisk next to bars • Alternative: add clinical significance description to legend 	
	Proportions meeting a responder definition		Bar charts of proportions improved, stabilized, and worsened at a single time point with legend indicating treatment represented by bar color <ul style="list-style-type: none"> • Alternative: putting treatment label directly on bars

focusing on statistical versus clinical significance. In contrast, for the group-level data presented to clinicians, we sought to identify approaches to indicate both

statistical significance and clinical importance, in addition to conveying score meaning (i.e., bad, good, normal).

Analysis of the one-on-one interviews

Using a mixed-methods approach, we concurrently reviewed quantitative results on the preference and ease of understanding of formats with qualitative data from the semi-structured interviews. The quantitative data (ease of understanding and preferences) were summarized descriptively using medians and proportions. For the qualitative data, a coding system based on each of the prompts in the interview was employed. Using ATLAS.ti [15], one team member coded each transcript and a second reviewed. The entire team discussed coding differences, until consensus was reached. A summary report of all coded texts from the interviews for each topic was reviewed by all team members. Quantitative and qualitative summaries were then presented at an in-person meeting of the stakeholder advisory board.

Results

One-on-one interview study sample

We recruited 39 patients and 40 clinicians to participate in the one-on-one interviews (Table 2).

Individual-level data

We conducted 29 one-on-one interviews (19 patients and 10 clinicians) to test individual-level data formats. Ease-of-understanding ratings were high for all formats, with medians ranging from 9 to 10 for patients and 8.5 to 10 for clinicians.

For the directional inconsistency issue, participants found descriptive labels (e.g., “mild,” “moderate,” or “severe”) to be helpful, though some questioned the labels’ meaning. “Someone else’s definition of what moderate is and what mild is...not sure how meaningful that is” [C202]. “I mean you could sit here and read this, very high, moderate, and a lot of people might not understand what it means you know?” [P210]. Overall, 79 % of patients and 90 % of clinicians preferred some kind of descriptive label (either standard or based on the questionnaire response options) versus *y*-axes labeled with numbers alone. Some participants also found the shading helpful. “Let’s you know immediately visually whether good is at the top or the bottom...” [C206], but others did not understand what the shading indicated. “I think that there may be better ways of indicating concerning results than the shading” [P209]. Overall, 74 % of patients and 80 % of clinicians preferred green shading of normal scores “less threatening” [C204] or red shading of concerning scores “a

Table 2 One-on-one interview participant demographics

Group	Characteristic	Individual-level data	Group-level data for patients	Group-level data for clinicians
Patients		(<i>N</i> =19)	(<i>N</i> =20)	N/A
	Age: median (range)	62 (26–78)	66 (36–76)	
	Sex: <i>N</i> (%), female	8 (42)	11 (55)	
	Years from most recent diagnosis: median (range)	1 (0–2)	2 (0–17)	
	Ongoing treatment: <i>N</i> (%)	11 (58)	1 (5)	
	Race: <i>N</i> (%), white	15 (79)	14 (70)	
	Education: <i>N</i> (%)			
	High school graduate or less	4 (19)	5 (25)	
	Some college	2 (11)	6 (30)	
	College graduate	4 (21)	2 (10)	
Post-graduate	9 (47)	7 (35)		
Clinicians		(<i>N</i> =10)	(<i>N</i> =10)	(<i>N</i> =20)
	Age: median (range)	56 (34–58)	41 (35–62)	51 (29–72)
	Sex: <i>N</i> (%), female	3 (30)	4 (40)	11 (55)
	Race: <i>N</i> (%), white	8 (80)	7 (70)	17 (85)
	Specialty: <i>N</i> (%)			
	Medical oncology	1 (10)	2 (20)	2 (10)
	Radiation oncology	3 (30)	2 (20)	1 (5)
	Surgical oncology	3 (30)	2 (20)	5 (25)
	Gyne-oncology urology	2 (20)	0 (0)	4 (20)
	Oncology NP/PA	1 (10)	2 (20)	6 (30)
Oncology fellow	0 (0)	(20)	2 (10)	
Years in practice: median (range)	30 (10–33)	14 (4–24)	15 (2–48)	

call to action” [C202] over spectrum shading “too busy” [C206]. Participants were split regarding whether they wanted the axes reversed to make the direction consistent. Among those who preferred scores reversed, comments included “I...like...where they’re all the same” [C204] and “don’t have to think as much” [C201]. Among those who did not like the scores reversed, either another alternative was preferred (e.g., shading adequately indicated where on the graph was better and worse) or they were confused by higher scores indicating less severe symptoms “going up to equal less doesn’t make sense to me” [P202].

In terms of highlighting important scores, 69 % of patients and 70 % of clinicians preferred the red circles or the threshold line, rather than shading. The red circles were helpful in identifying concerning scores as they “catch your eye right away” [P207], but there was some question of whether patients would be distressed to see them on the score reports. The threshold line “gives me a little bit more information...how far below the threshold that score is” [C202], but there was some concern that “unless it was really explained to them, [patients] wouldn’t know what the line was there for” [P210]. Among patients, 79 % preferred having the exclamation points to indicate changes in scores that were possibly concerning, but only 40 % of clinicians supported including the exclamation points. In some cases, it was not clear whether patients understood the difference between poor scores in absolute terms versus important worsening.

Group-level data presented to patients

Based on part I findings, for the group-level data presented to patients, we simplified the language and presentation throughout. For example, we replaced “randomization” with “start of treatment” and “mean” with “average.” Group labels were placed directly at the end of the corresponding data line, and color was added (although differentiating line symbols were retained). We also presented only four domains in the array, rather than the six-domain array that was used previously. Patients’ median ease-of-understanding ratings ranged from 8 to 9 and clinicians’ from 7.5 to 10.

From the interviews with 20 patients and 10 clinicians, 80 % of patients and 100 % of clinicians preferred descriptive *y*-axis labels for line graphs to numbers only, “this is easier to follow because you can tell what is bad and what is good” [P2202]. Providing the headers of which direction was better/worse was also helpful, “It does say here, and that’s very important, line going up means better emotional functioning” [P214]. Using shading to indicate a normal score was not particularly effective. While some patients liked the green-shaded region, “I really like that green, it really gives an idea of where a patient should be over time” [P2212], others did not notice or understand it, “I didn’t even pay attention to the green shading” [P2204], found it too busy, or questioned what normal indicated, “Normal fatigue scores. Is that for people on treatment or is that for the whole of the population?” [P2216].

As found in part I, clinicians valued indications of importance (statistical or clinical) and missed them in line graph formats without such an indication, “I think these could... hoodwink a patient, because...this looks like it could be meaningful, and it may not be meaningful at all” [C2202]. While some patients found the asterisk a helpful indication of important differences, “the asterisk tells you, for example, that emotional function is real, the pain difference isn’t” [P2219], many did not notice them or understand their meaning. Overall, 90 % of clinicians preferred formats with an asterisk indicating important differences, but only 45 % of patients wanted this. In addition, 80 % of clinicians preferred the asterisk to shading to indicate important differences, but patients were split 50–50 % between asterisks and shading.

For proportional data formats, the majority of both clinicians (70 %) and patients (55 %) preferred pie charts for presenting data of the proportion improved, stable, worsened, “Pie charts are really nice. This is easy to understand” [P2213], though there were some exceptions, “I don’t particularly care for pie charts” [P2206]. The icon arrays were the second choice of patients (25 %) and clinicians (20 %). In general, patients and clinicians were less familiar with icon arrays, some found them too busy, and there were challenges finding color schemes that provided sufficient contrast and did not have negative connotations. Few patients (20 %) and clinicians (10 %) preferred bar charts.

Group-level data presented to clinicians

Clinicians’ median ease-of-understanding ratings for the group-level data presented as for clinicians ranged from 6.5 to 8. For line graphs of average scores over time, there was a weak preference for the addition of labels to numeric values on the *y*-axis (55 vs. 45 %). One clinician commented on those labels, “Give some perspective on what the numbers mean” [C2302], whereas others thought the labels were too subjective or unnecessary, “I know that 70 is higher than 30” [C2307].

For indications of statistical significance, 95 % of clinicians preferred having *p* values in addition to confidence limits. Notably, not all clinicians could interpret confidence limits correctly, “I don’t remember enough from stats to know how that’s helpful or how it changes any of these answers” [C2317]. If there was an indication of clinical significance, 90 % of clinicians wanted this to be in the graph rather than in the legend (45 % preferred asterisks; 45 % preferred shading), but clinicians were mixed in whether they wanted the information. Some found it useful, “The *p*-value is always important...but whether they’re clinically significant or not it’s very difficult to tell just from a *p*-value” [C2310], but others did not differentiate between statistical significance and clinical importance, “I just jumped to the *p*-values” [C2308], or missed the description in the legend, “Oh I didn’t read that...OMG!” [C2314].

For data demonstrating the proportion of patients who improved on a treatment, we found some difficulty with the underlying construct (worsened, stable, improved) as well as some discomfort seeing only a single time point.

In terms of overall preferences, when choosing among the line graph options only, 60 % of clinicians selected one of the versions with an indication of clinical significance and 30 % selected one of the versions with confidence limits. When asked to choose between their preferred line graph and the bar chart of average changes at 9 months, 75 % preferred the line graph. Finally, the bar chart of proportions meeting a responder definition was preferred by only 10 % of clinicians when given as an option.

Discussion

For PRO measures to promote patient-centered care, patients and clinicians must understand the data presented and be able

to apply them in clinical practice. Prior research has pointed to interpretation challenges with presentation of PRO data. We employed an innovative, stakeholder-driven, iterative approach to examine the nature of the challenges for patients and clinicians and develop improved formats for displaying PRO data graphically.

Across the three topics, the part II study results provided critical insights regarding various alternatives for presenting PRO data. Part II was not intended to “pick a winner” for presentation formats, but rather to inform the design of formats that will be tested in a much larger Internet survey to be fielded in populations of cancer patients, cancer providers, and PRO researchers. Importantly, this iterative, stakeholder-driven approach was required to improve and optimize the candidate formats to be tested in a larger study. This systematic approach to format improvement is novel and, to our knowledge, has not been previously reported. As one example of how the part II findings are being applied, Fig. 2a shows the individual-level line graph format tested in part I and Fig. 2b

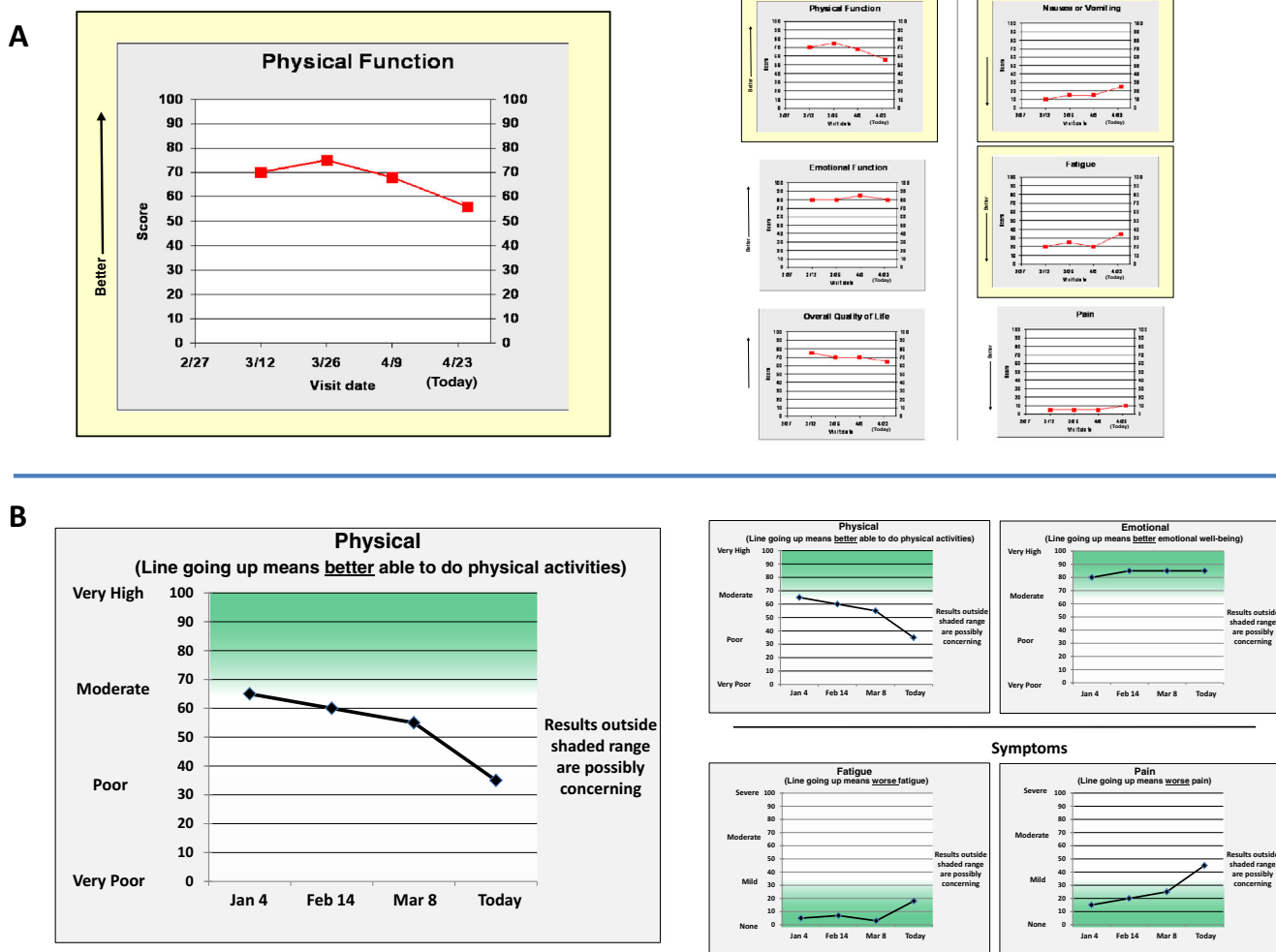


Fig. 2 a, b Individual-level line graph format tested in part I, b individual-level line graph format to be tested in Part III

shows the individual-level line graph format to be tested in part III. Figure 3a, b shows the same for the group-level data presented to patients.

Stakeholder engagement played a key role throughout this process. Because this research focuses on best practices for presenting PRO data in clinical practice, patients, clinicians, and PRO researchers are all important stakeholders. Each is represented among our investigator team and our stakeholder advisory board. In addition, we employed innovative approaches such as recruiting a subset of part I participants to collaborate in workgroups to develop improved data presentation formats. Interviews with additional patients and clinicians provided further insights.

The project’s findings should be interpreted in the context of its strengths and limitations. Two key limitations are the relatively small sample required by the study’s qualitative nature and the challenge of addressing contradictory input from patients and clinicians in some cases. First, because detailed and iterative input was required from workgroups and additional interviews, we could include only a limited number of

stakeholders. However, our purposive sampling of patients focused on cancer type, education, and treatment location also resulted in diversity in terms of age, gender, and time since diagnosis. Similarly, our clinician purposive sampling focused on specialty and practice location also produced a sample with a broad range of practice experience. Second, the findings from this study indicate that stakeholders are likely to be split on certain issues (e.g., red vs. green shading). Our challenge is to identify formats that, even when they are not most preferred, are consistently interpreted accurately and efficiently by patients and clinicians.

The part III internet survey results regarding ease of understanding and accuracy of interpretation will provide critical information to determine best practices for data display in different arenas and for different audiences. The communicative needs and capacities of the primary audience in each case are prioritized (e.g., clinicians and researchers are the primary audience for group-level data in biomedical journals, whereas patients’ needs for interpretation of group data are most appropriately considered in relation to decision aids that they

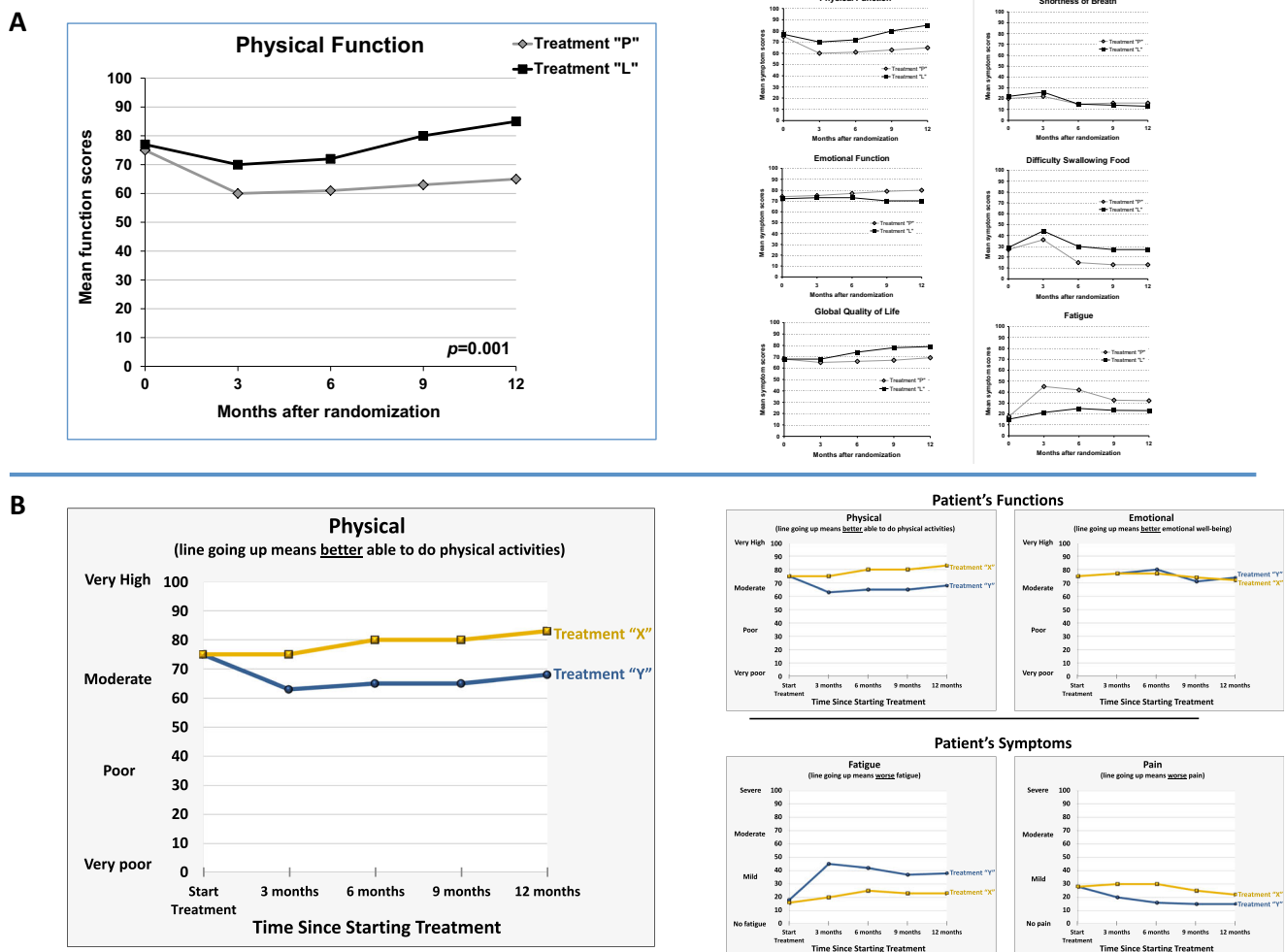


Fig. 3 a, b Individual-level format for patients tested in part I, b individual-level format for patients to be tested in Part III

would use with their clinician). Thus, the part III internet survey will enable a comparison of interpretation accuracy across formats and also determine absolute interpretation accuracy rates. The insights generated from part II of the project, in combination with its stakeholder-driven approach, provide important information on formats for presenting PRO data to be tested in part III. Given the value of PROs for promoting patient-centered care, these insights are critical.

Acknowledgments 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 PCORI and its Board of Governors or Methodology Committee. Dr. Smith and Dr. Snyder are members of the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins (P30 CA 006973). The funding sources had no role in the study design, data collection, analysis, interpretation, writing, or decision to submit the manuscript for publication. 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); and 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); and Carolyn Hendricks, MD, Melissa Hyman, RN, BSN, OCN, and Barbara Squiller, MSN, MPH, CRNP (Suburban Hospital). Finally, we are most grateful to the patients and clinicians who participated in this study. In particular, we extend our gratitude to the patients from part I of the study who agreed to contribute to the workgroups as research partners (rather than participants) for this phase of the research.

Compliance with ethical standards The Johns Hopkins School of Medicine Institutional Review Board approved the study. All participants provided signed consent.

References

1. U.S. Food and Drug Administration (2009) Guidance for industry. Patient reported outcome measures: use in medical product development to support labeling claims. *Fed Regist* 74:65132–65133
2. Acquadro C, Berzon R, Dubois D et al (2001) 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* 6:522–531
3. Brundage MD, Feldman-Stewart D, Bezjak A et al (2011) The value of quality of life information in a cancer treatment decision. ISOQOL 11th annual conference, San Francisco, 2005
4. Brundage M, Bass B, Jolie R et al (2011) A knowledge translation challenge: clinical use of quality of life data from cancer clinical trials. *Qual Life Res* 20:979–985. doi:10.1007/s11136-011-9848-0
5. Snyder CF, Aaronson NK (2009) Use of patient-reported outcomes in clinical practice. *Lancet* 374:369–370. doi:10.1016/S0140-6736(09)61400-8
6. Greenhalgh J (2009) The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res* 18:115–123. doi:10.1007/s11136-008-9430-6
7. Velikova G, Booth L, Smith AB et al (2004) Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 22: 714–724. doi:10.1200/JCO.2004.06.078
8. Detmar SB, Muller MJ, Schornagel JH et al (2002) Health related quality of life assessments and patient physician communication. *J Am Med Assoc* 288:3027–3034. doi:10.1001/jama.288.23.3027
9. Snyder CF, Jensen R, Courtin SO et al (2009) PatientViewpoint: a website for patient-reported outcomes assessment. *Qual Life Res* 18:793–800. doi:10.1007/s11136-009-9497-8
10. Jones JB, Snyder CF, Wu AW (2007) Issues in the design of Internet-based systems for collecting patient-reported outcomes. *Qual Life Res* 16:1407–1417. doi:10.1007/s11136-007-9235-z
11. Snyder CF, Blackford AL, Wolff AC et al (2013) Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. *Psycho-Oncology* 22:895–901. doi:10.1002/pon.3087
12. Brundage M, Smith KC, Little EA, Bantug ET, Snyder CF, PRO Data Presentation Stakeholder Advisory Board (2015) Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. *Qual Life Res Online First*. doi:10.1007/s11136-015-0974-y
13. PROMIS: Dynamic tools to measure health outcomes from the patient perspective. Available at nihpromis.org. Accessed 21 Aug 2015
14. Aaronson NK, Ahmedzai S, Bergman B et al (1993) The European organization for research and treatment of cancer QLQC30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 85:365–376
15. AtlasTi V7. Available at <http://atlasti.com>. Accessed 5 May 2016