



In proportion: approaches for displaying patient-reported outcome research study results as percentages responding to treatment

Elliott Tolbert^{1,2}  · Michael Brundage³ · Elissa Bantug⁴ · Amanda L. Blackford^{4,5} · Katherine Smith^{2,4,6} · Claire Snyder^{1,2,4,7} · PRO Data Presentation Stakeholder Advisory Board

Accepted: 22 November 2018 / Published online: 29 November 2018
© Springer Nature Switzerland AG 2018

Abstract

Purpose Patient-reported outcome (PRO) data from clinical trials can promote valuable patient–clinician communication and aid the decision-making process regarding treatment options. Despite these benefits, both patients and doctors face challenges in interpreting PRO scores. The purpose of this study was to identify best practices for presenting PRO results expressed as proportions of patients with changes from baseline (improved/stable/worsened) for use in patient educational materials and decision aids.

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 saw clinical trial data comparing two treatments as proportions changed using three different formats: pie charts, bar graphs, icon arrays. Interpretation accuracy, clarity, and format preference were analyzed quantitatively and online survey comments and interviews, qualitatively.

Results The internet sample included 629 patients, 139 clinicians, and 249 researchers; 10 patients and 5 clinicians completed interviews. Bar graphs were less accurately interpreted than pie charts (OR 0.39; $p < .0001$) and icon arrays (OR 0.47; $p < .0001$). Bar graphs and icon arrays were less likely to be rated clear than pie charts (OR 0.37 and OR 0.18; both $p < .0001$). Qualitative data informed interpretation of these findings.

Conclusions For communicating PROs as proportions changed in patient educational materials and decision aids, these results support the use of pie charts.

Keywords Cancer · Patient-reported outcomes (PROs) · Decision-making · Communication · Educational materials · Decision aids

The Members of PRO Data Presentation Stakeholder Advisory Board are listed in Acknowledgements.

✉ Elliott Tolbert
etolber2@jhmi.edu

Michael Brundage
michael.brundage@krcc.on.ca

Elissa Bantug
ebantug1@jhmi.edu

Amanda L. Blackford
ablackf1@jhmi.edu

Katherine Smith
ksmit103@jhu.edu

Claire Snyder
csnyder@jhu.edu

² Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Room 725, Baltimore, MD 21205, USA

³ Cancer Clinic of Southeastern Ontario, Queens Cancer Research Institute, 25 King Street West, Kingston, ON K7L 5P9, Canada

⁴ Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, 1650 Orleans Street, Baltimore, MD 21287, USA

⁵ Johns Hopkins School of Medicine, 550 N. Broadway, Room 1111, Baltimore, MD 21205, USA

⁶ Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Room 726, Baltimore, MD 21205, USA

⁷ Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, Room 649, Baltimore, MD 21205, USA

¹ Johns Hopkins School of Medicine, Baltimore, MD, USA

Introduction

Because patient-reported outcomes (PROs) are increasingly assessed in clinical trials and other types of comparative research studies [1–3], data regarding patients' perspectives about symptoms, functioning, and health-related quality of life (HRQOL) are more readily available to help patients understand the impact of health conditions and their treatments [3–11]. These data are valued by patients and clinicians and inform decision-making [8, 9, 11]. There are different approaches to analyzing PRO research study results, including comparing mean scores over time for the different treatment groups and comparing the proportion of patients in each treatment group meeting a responder definition (i.e., the proportion who improved, stayed about the same, or worsened). Regardless of the analytic approach, for the PRO data to aid education and decision-making most effectively, the results need to be interpretable by both patients and their clinicians.

We have conducted a multi-phase project to evaluate different approaches for displaying PRO data to improve understanding and use. A nine-member Stakeholder Advisory Board (SAB) guided the development and implementation of the entire project. This SAB was comprised of cancer patients and caregivers, cancer clinicians, and PRO researchers (see listing and affiliations in Acknowledgement).

During the first phase of the project, we obtained patient and clinician perspectives on currently existing approaches for PRO data display, including both individual patient data for use in monitoring and management, as well as comparative research study results. Participants evaluated formats for presenting average scores over time (line graphs of mean scores/line graphs of normed mean scores/line graphs with confidence intervals), proportions responding (improved/stable/worsened), bar charts of average changes, and cumulative distribution functions for interpretation accuracy, ease-of-understanding, and usefulness. Both clinicians and patients preferred line graphs of mean scores over the alternative formats [12].

In the second phase, we partnered with patients and clinicians to develop PRO display formats that capitalized on the attributes that were found helpful and minimized the attributes that were considered confusing [13]. Although patients and clinicians preferred the presentation of mean scores over time provided by line graphs, it is commonly recognized that a summary of the proportion of patients meeting a responder definition provides an alternative representation of the study findings. For this reason, we developed presentation display candidates for both of these analytic approaches based on findings from the first phase. In addition, results from phase one suggested that patients

and clinicians differ in their perspectives regarding displays of research study results. For example, clinicians value statistical details (e.g., *p* values), whereas patients find this information confusing [12]. Consequently, we addressed display formats for research study findings for patient educational materials/decision aids separately from display formats aimed at clinicians (e.g., in peer-reviewed publications) [14].

The third part of the study evaluated interpretation accuracy, clarity, and preferences among the candidate data display formats developed during the second phase. As noted above, we evaluated both line graph formats representing average scores over time and formats reporting the proportion meeting a responder definition. Regarding average scores over time, our results supported using line graphs with higher scores consistently indicating better outcomes [15]. Here we report on the findings regarding presentation of proportions changed (from baseline) to patients, for use in educational materials or decision aids. Our approach was guided by existing literature focused on quantitative risk communication. A recent systematic review of that literature concluded that visual aids (e.g., icon arrays or bar charts) can improve patients' understanding of probabilistic information over numeric or narrative approaches, but that the superiority of any single visual method could not be established [16]. Based on this literature, we hypothesized that participants' preferences for visual display would vary, but that one particular strategy might prove to be more accurately interpreted than others across respondents.

Methods

Study design

This cross-sectional, mixed-methods study employed a survey to evaluate the accuracy of interpretation and clarity of an array of PRO graphs. The survey was developed to present the results of a hypothetical clinical trial and displayed graphs comparing the symptom and function results of two different treatments. The survey was presented online via an internet access link and assessed interpretation accuracy and clarity ratings for each data display format. Online survey participants also could provide free-text comments as they completed the survey. The qualitative portion incorporated: (1) the free-text comments online survey participants submitted and (2) the same online survey administered face-to-face, after which the participant completed a cognitive debriefing interview. Analysis of free-text comments supplemented interpretation of quantitative results [17].

In the online portion of the study, an introductory screen stated that completion of the survey would indicate consent to study participation. Written informed consent was

obtained from all interview participants included in the study. The Johns Hopkins School of Medicine Institutional Review Board (IRB-X) reviewed both the online and interview portions of the study and assigned exempt and approved statuses, respectively.

Population

Online survey participants

Online survey eligibility included cancer patients and survivors, cancer clinicians, and PRO researchers who were at least 21 years of age. To recruit these participants, we reached out to organizations that would be able to share the survey's link electronically, often through social media. The SAB was influential in making these connections, as many members have affiliations with organizations that serve or represent our target populations. Snowball sampling was also utilized, as recipients had the opportunity to forward the survey link to individuals who fit the eligibility criteria. Before the survey began, eligibility was determined using a series of screening questions. The online survey and cognitive interview were only available in English.

Cognitive interview participants

Cognitive interview participants were cancer patients and clinicians recruited via flyer from within the Johns Hopkins Clinical Research Network (JHCRN). The JHCRN is a consortium of academic and community health systems in the US mid-Atlantic. The initial recruitment targets were ten patients and five clinicians to complete interviews. These targets could increase if thematic saturation (across both interviews and online free-text comments) was not attained. Patients were eligible if they were at least 21 years of age, diagnosed with any cancer (excluding non-melanoma skin cancer), ≥ 6 months post-diagnosis, not currently undergoing acute treatment, and able to communicate in English. Clinician eligibility included active oncologists (i.e., medical, radiation, surgical, gynecologic/urologic, nurse practitioners/physician assistants, and fellows). Purposive sampling was used for both groups of participants, such that the sample included patients with different cancer types, education levels, and who were treated at different clinical sites, as well as a sample of clinicians who represented varying specialties and clinical sites.

Study procedures

Online survey data collection

The online survey showed each participant all three different formats (bar graphs, pie charts, icon arrays), one at a time, in

a randomly assigned order (Figs. 1, 2, 3). Each format presented four charts that displayed the proportion of patients who had improved, stayed about the same, and worsened at 9 months compared to baseline for two hypothetical clinical trial treatments. The charts represented two function domains (ability to do physical activities and emotional well-being) and two symptom domains (pain and fatigue). Regardless of which format was seen first, the first format was evaluated using two accuracy questions; the second and third formats were evaluated using one accuracy question each (the question wording is summarized in Fig. 4). To account for possible order effects, both the data each format displayed and the accuracy questions asked remained constant across all surveys, so that differences found in accuracy and clarity could be ascribed to the format rather than the data presented or questions asked (Fig. 5). To assess clarity, participants were provided an opportunity to rate each format as either “very confusing,” “somewhat confusing,” “somewhat clear,” or “very clear.” In addition, a text box allowed an opportunity to provide additional information. As explained earlier, this phase of the large project examined display formats both for average scores over time and for the proportion meeting a responder definition. For this reason, half of all survey participants evaluated a set of randomly assigned line graphs before they began evaluating each of the proportions, while the other half evaluated proportions first. Once participants completed the survey, they could enter for a chance to receive a \$100 Amazon giftcard, with winners randomly chosen.

Cognitive interview data collection

Participants who completed cognitive interviews completed one of the same randomly assigned surveys as online participants. To obtain qualitative feedback, participants were recorded while they completed the survey and asked to think aloud as they answered questions, made decisions, and rated clarity for each format. Upon completion, a research team member asked the participant to share any overall feedback and additional thoughts that came up during survey completion. Participants were given a \$35 giftcard at the end of the interview. Interview recordings were transcribed in order to be analyzed using qualitative software.

Analysis

Quantitative data

Descriptive statistics included sample characteristics for online participants, as well as their responses to accuracy and clarity questions. Multivariable generalized estimating equation (GEE) logistic regression models with the individual as the cluster unit tested differences in accuracy and

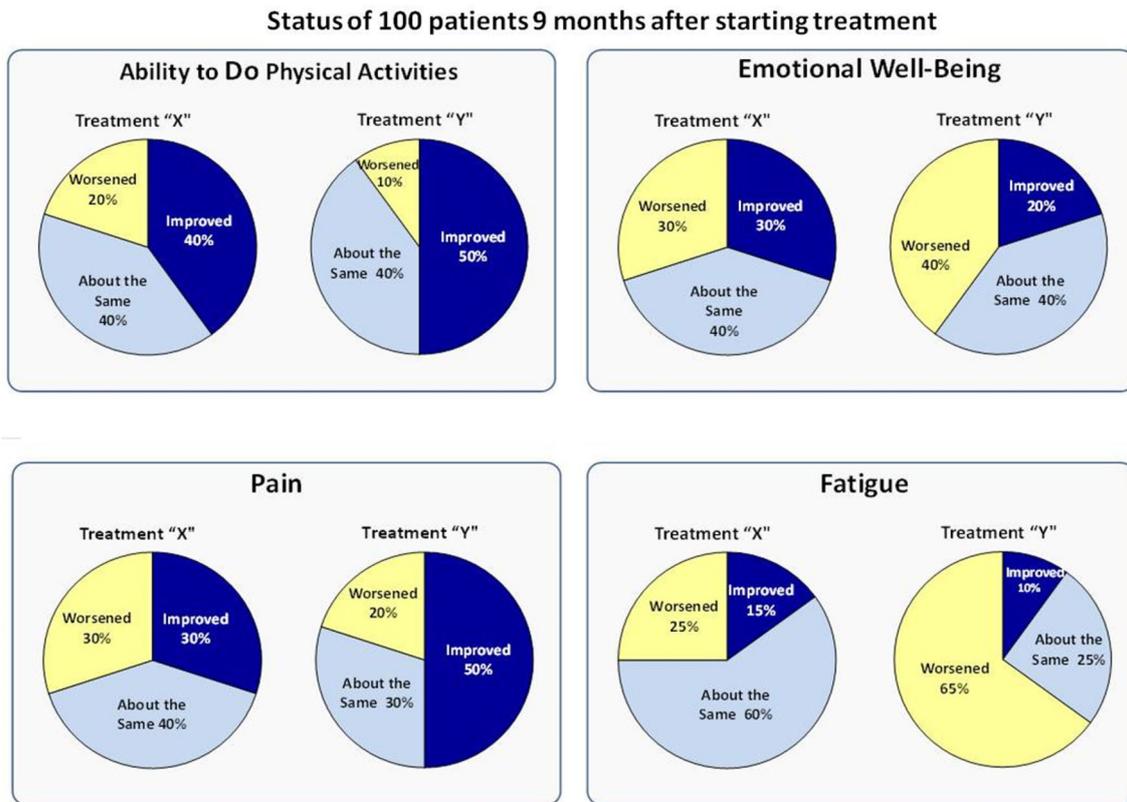


Fig. 1 Example of pie chart format tested

clarity by format. The following two outcomes evaluated interpretation accuracy: (1) accuracy regarding the two questions asked for the first format seen and (2) accuracy regarding all questions on each format seen, across all orders (and, therefore, all participants). Fixed effects for the specific questions were included in the model that included all questions, and all models adjusted for participant type (patient, clinician, researcher) and order the format was seen. Clarity was evaluated using the following two outcomes: (1) those rating the format “very” or “somewhat” clear and (2) those rating the format “very” clear only. These categorizations were based on the distribution of the responses.

Qualitative data

Qualitative data obtained from the cognitive interview transcripts were analyzed by a deductive coding scheme based on the study objectives, interview structure, and interview content. The codebook, which was piloted and subsequently revised, was developed by a member of the research team and reviewed by another team member who specializes in qualitative research. The codebook was designed to capture the broad coding categories of positive or negative comments, misinterpretations, and preferences for each presentation format. One member of the research team coded

each transcript using ATLAS.ti [18] and a second member reviewed each coded transcript for consensus. A report was generated from the coded transcripts by format to identify themes. Quotes reflecting these themes are included in the results, with “P” indicating comments made by patients and “C” designating clinicians.

Text box comments provided by online survey participants were also analyzed qualitatively. These comments were organized, by participant and format type, into the preexisting categories of “positive” and “negative.” Illustrative text box comments are also included in the results, with “PA” indicating patient comments, “CL” for clinicians, and “RE” for researchers.

Results

Study sample

Online survey sample

The online participant sample included 629 patients, 139 clinicians, and 249 PRO researchers, for an overall total of 1017 (Table 1). Patients were 58 years of age on average, and were predominately female (87%), White (94%),

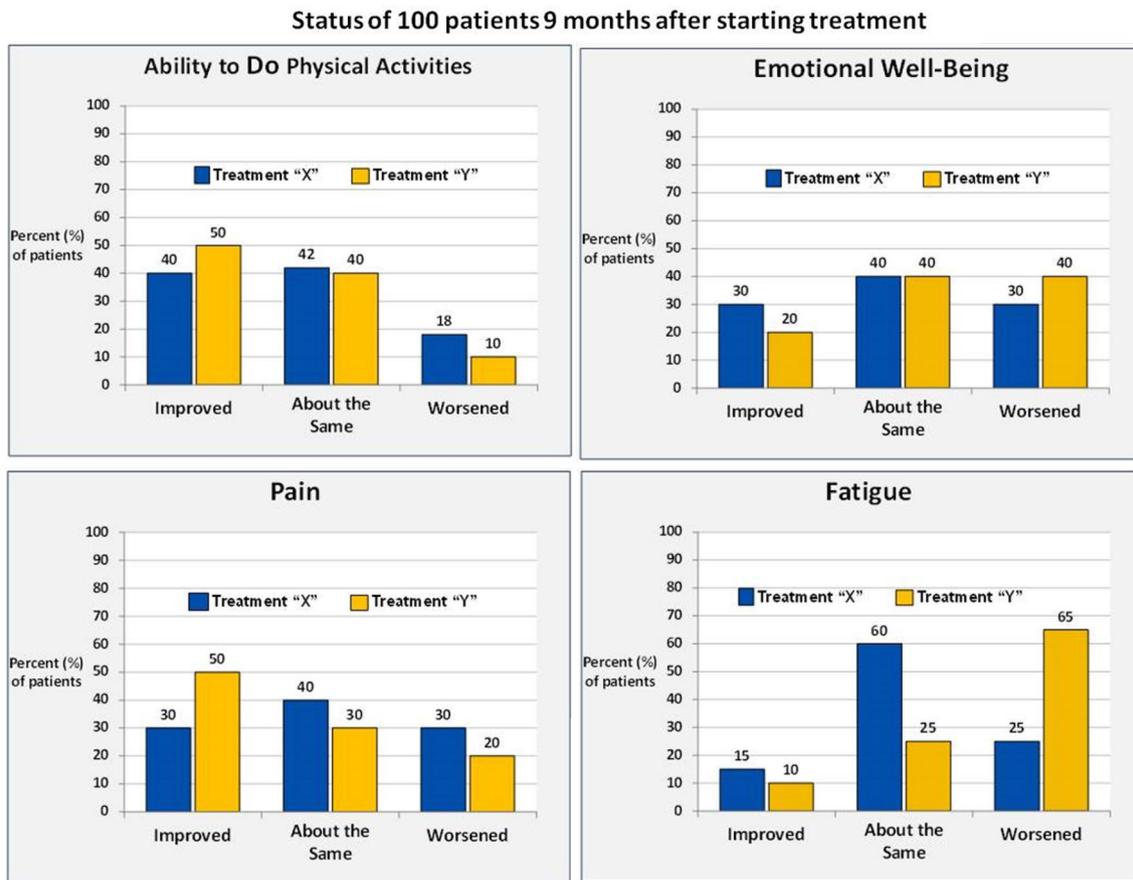


Fig. 2 Example of bar graph format tested

breast cancer patients and survivors (56%). Twenty-three percent of patients had less than a college degree. Clinicians had a mean age of 44 years and had been in practice for an average of 16 years. The plurality practiced medical oncology (44%). Researchers had an average age of 45, and 46% had over 10 years of experience as a PRO researcher.

Cognitive interview sample

Purposive sampling led to ten patient participants of whom 70% were survivors of cancer types other than breast, 70% were outside Johns Hopkins, and 30% had less than a college degree. Among the five clinician participants, 60% were from outside Johns Hopkins and specialties included an oncology fellow; nurse practitioner; and surgical, radiation, and medical oncologists. As no new issues related to question or concept comprehension emerged, saturation was considered achieved. For this reason, no additional interview participants were recruited.

Findings

Accuracy and clarity: quantitative findings

Figures 6 and 7 summarize patients’, clinicians’, and researchers’ accuracy of interpretation and clarity ratings for proportion formats. Among patients, accuracy was highest for pie charts and icon arrays. Clinicians and researchers were more likely to accurately answer questions about pie charts. Pie charts received the best clarity ratings among all three participant groups. Table 2 summarizes the multi-variable model accuracy and clarity outcome results. Across all four accuracy questions, bar graphs were less accurately interpreted than pie charts (OR 0.39, $p < .0001$) and icon arrays (OR 0.47, $p < .0001$). The analyses focusing on the two accuracy questions for the first format seen demonstrated an even larger difference (OR 0.22, $p < .0001$ for bar graphs vs. pie charts; OR 0.30, $p < .0001$ for bar graphs vs. icon arrays). In addition, bar graphs and icon arrays were less likely to be rated “somewhat” or “very” clear than pie charts (OR 0.37 and OR 0.18, respectively; both $p < .0001$).

Status of 100 patients 9 months after starting treatment

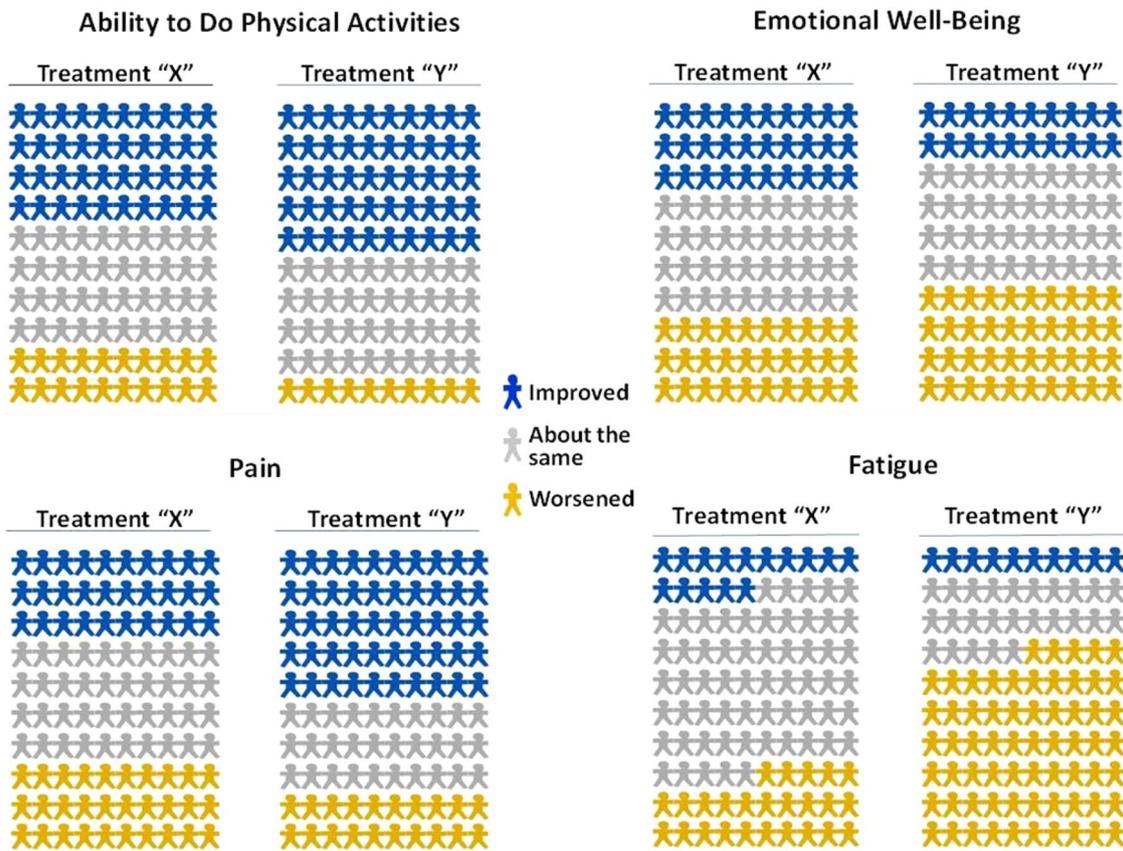


Fig. 3 Example of icon array format tested

Fig. 4 Accuracy questions and answer choices as seen by randomized format order

First Format	Second Format	Third Format
1. At 9 months, on which treatment did more patients <u>improve</u> with regard to doing PHYSICAL activities?	1. At 9 months, on which treatment did more patients <u>improve</u> with regard to EMOTIONAL well-being?	1. At 9 months, on which treatment did more patients <u>stay about the same</u> with regard to FATIGUE?
2. At 9 months, on which treatment did more patients <u>worsen</u> with regard to PAIN?	Not applicable	Not applicable

Answer Choices:	Treatment "X"	Treatment "Y"	Treatments are about the same
-----------------	---------------	---------------	-------------------------------

Fig. 5 Format display order by survey version

	Format 1	Format 2	Format 3	Format 4
Version 1	Pies	Bars	Icons	Line Graphs
Version 2	Bars	Icons	Pies	Line Graphs
Version 3	Icons	Pies	Bars	Line Graphs
Version 4	Line Graphs	Pies	Bars	Icons
Version 5	Line Graphs	Bars	Icons	Pies
Version 6	Line Graphs	Icons	Pies	Bars

We also analyzed accuracy by clarity rating. Participants who rated a format as “very clear” were 2–3 times more likely to get both accuracy answers correct than those who did not rate the format as “very clear” (icon arrays OR 1.9, $p = .13$; bar charts OR 2.0, $p = .01$; and pie charts OR 2.8, $p = .02$).

Accuracy and clarity: qualitative findings

In general, qualitative comments were categorized as being positive or negative for each format. Participants were much more likely to discuss and describe negative aspects of each format, which is evident throughout all qualitative findings. Strong qualitative support for pie charts emerged from participants’ comments. Patients found the pie charts easy to read and were able to obtain information quickly: “Very easy to see at a quick glance, and I wished I had been privy to more graphs like these; it may well have helped me to make treatment decisions regarding my care” [PA]. Clinicians and researchers had similar comments: “It’s easy to see with the side by side with the different colors and especially with the percentages laid out to correlate with the colors to see kind of, you know, how many people improved, worsened or about the same is quick to sort out between the 2 and the colors along with the percentages work well to quickly figure out what’s going on” [C05]. There were, however, participants within each group that found limitations in pie charts: “Pie graphs ok if only three or less choices, but confusing if more than three (ten to fifteen)” [PA]; “Too hard to quickly review and note any substantive differences—they look too similar to one another” [CL]. When the general categories of positive and negative were considered, approximately 61% of patient comments and 40% of clinician and researcher comments in regard to pie charts described positive aspects.

Participants noted helpful aspects of bar charts: “Side by side comparisons are much easier to read and comprehend” [PA]; “Now I can quantify the difference and decide for myself whether I find this is small or a big difference. I also think this approach is more standardized (you cannot ‘cheat’ as a researcher in presenting your data) and therefore less prone to bias” [RE]. There were also comments that offered insight as to why bar charts were less accurately interpreted and less likely to be rated clear: “It takes a second to read the graph and then connect the color to the corresponding graph” [PA]; “Sometimes a little difficult to say when you compare overall to each one, treatment X and treatment Y, somewhat confusing is why I said that” [P02]. Approximately 21% of patient comments and 38% of clinician and researcher comments were coded as positive.

Icon arrays, similar to bar graphs, received mixed comments, providing further support for quantitative results. All participant types noted that icon arrays would be easy for patients to understand: “Cute and pleasant, and manage to convey the information in a clear and concise way” [PA]; “Pictorial representations are better understood by patients than graphs” [CL]. However, many of the comments, across participant types, focused on the difficulties of icon arrays: “These seem most effective for large changes—I wouldn’t want to have to sit and count the little people figures to find out if there was one more or one less or the same number” [PA]; “Definitely takes more work to look up the different colors and what they represent. Also a bit more visually complicated” [CL]. “I’m not used to dealing in gingerbread people, it should be numbers here, um, yeah I, worse, you know, I would like to see numbers” [P06]. Among comments describing icon arrays, approximately 35% of patients’ and 43% of clinicians’ and researchers’ feedback was positive.

Table 1 Sample characteristics

Characteristic	Survivors (<i>n</i> = 629)	Clinicians (<i>n</i> = 139)	Researchers (<i>n</i> = 249)
Age			
Mean (SD)	58.1 (11.3)	43.8 (12.56)	45.0 (11.92)
Male <i>n</i> (%)	70 (13.3)	58 (46.0)	74 (33.3)
Race <i>n</i> (%)			
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 <i>n</i> (%)	16 (3.1)	9 (7.3)	9 (4.1)
Country <i>n</i> (%)			
United States	450 (85.6)	62 (49.2)	107 (48.6)
Education <i>n</i> (%)			
< 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 <i>n</i> (%) <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 <i>n</i> (%)			
< 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 <i>n</i> (%)		12 (9.5)	18 (8.2)
Provider specialty <i>n</i> (%)			
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)	
Provider years in practice			
Mean (SD)		15.7 (11.24)	
PRO researcher expertise <i>n</i> (%) <i>all that apply</i>			
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)

Table 1 (continued)

Characteristic	Survivors (<i>n</i> = 629)	Clinicians (<i>n</i> = 139)	Researchers (<i>n</i> = 249)
Regulator/health administrator			3 (1.2)
Other			16 (6.4)
PRO research experience <i>n</i> (%)			
Student			22 (10.1)
Post-doc			16 (7.3)
< 5 years of experience			34 (15.6)
5–10 years of experience			45 (20.6)
> 10 years of experience			101 (46.3)

Fig. 6 Accuracy of interpretation for the first proportion format seen

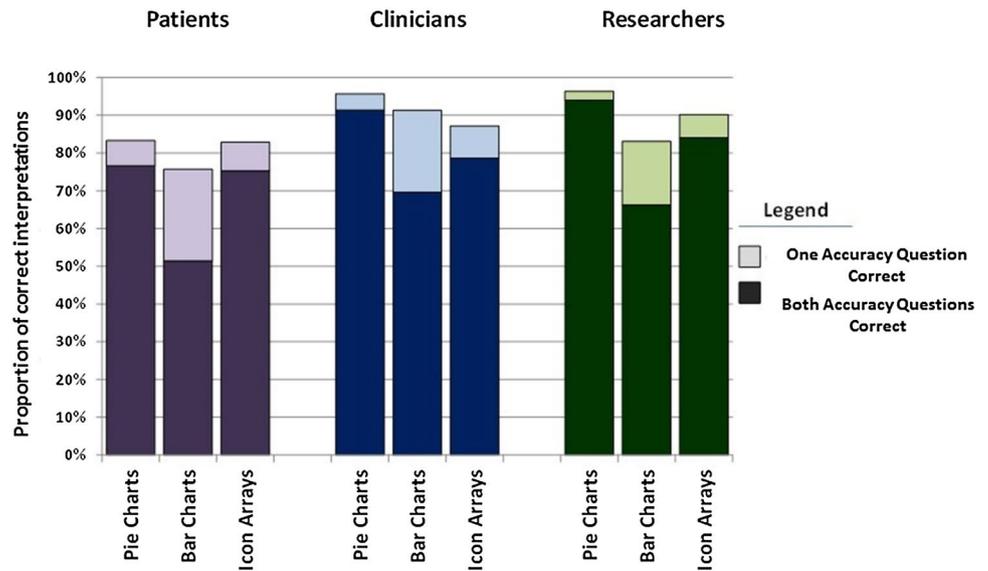


Fig. 7 Clarity ratings for proportion formats

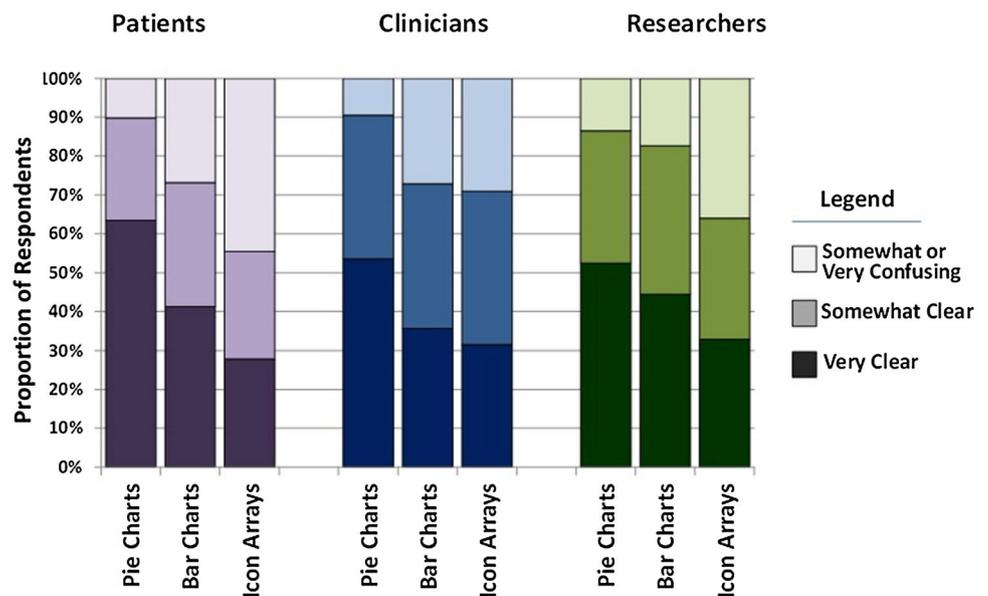


Table 2 Adjusted odds ratios for the association between format and outcome (giving the correct answer and clarity ratings)

Comparison	Correct answer, all four questions	<i>p</i>	Correct answer, first two questions	<i>p</i>	Rated very or somewhat clear	<i>p</i>	Rated very clear	<i>p</i>
	OR [95% CI]		OR [95% CI]		OR [95% CI]		OR [95% CI]	
Bars vs. pies	0.39 [0.30, 0.52]	<0.0001	0.22 [0.14, 0.35]	<0.0001	0.37 [0.29, 0.49]	<0.0001	0.48 [0.40, 0.58]	<0.0001
Icons vs. pies	0.83 [0.63, 1.10]	0.1913	0.74 [0.42, 1.28]	0.2794	0.18 [0.14, 0.23]	<0.0001	0.29 [0.24, 0.35]	<0.0001
Bars vs. icons	0.47 [0.36, 0.62]	<0.0001	0.30 [0.20, 0.46]	<0.0001	2.08 [1.70, 2.55]	<0.0001	1.67 [1.37, 2.04]	<0.0001

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 participant. Terms for the fixed effects of the specific questions were included in the larger model for accuracy. Adjusted for participant type (survivor, clinician, researcher) and whether proportion formats were seen before or after line graphs

Discussion

It is essential for PRO data collected during clinical trials and other comparative research studies to be presented in a way that people can understand in order to effectively contribute to the decision-making process. This study aimed to contribute to the knowledge base regarding the effectiveness of visual PRO data presentations, as little previous research exists [19]. Specifically, this work tested candidate formats that had been developed over the course of a multi-phase project for displaying PRO data as proportions changed. Our findings show that pie charts were more accurately interpreted than bar charts, equally well interpreted as icon arrays, and rated the clearest for communicating proportions changed from baseline.

Our findings are similar to those of other studies on communicating probabilistic outcomes to patients, in that different formats (bar graphs, pie charts, and icon arrays) have been found to be useful in different contexts [20, 21]. For example, between-treatment differences in PRO findings in this study were not intended to be shown with precision, and are best characterized as ‘gist’ rather than ‘verbatim’ estimates [22]; the PRO setting may have made the precision of icon arrays seem less useful to some viewers.

Although there exist a great number of established International Patient Decision Aids Standards (IPDAS) for decision aids, the optimal choice of visual format for presentation of probabilities is not specifically addressed in the IPDAS checklist [23, 24]. Our study was designed to inform the development of best practices for consistent communication of PROs, through future Delphi-based consensus-building exercises similar to those used to evolve the IPDAS checklist [23].

An important strength of this study is that we randomized the order in which the different formats were seen, held the data and interpretation accuracy questions constant across the orders, and varied only the format for data display. This design helps ensure that the differences found can be attributed to the format. In addition, the iterative approach and embedded mixed-methods design strengthen this study by adding validity to the results. The cognitive

interviews, for example, allowed purposive sampling and allowed us to determine that there were no systematic differences within feedback obtained from online and cognitive interview participants. In addition, the online platform resulted in a large sample that included participants from a wide variety of locations.

In addition to its strengths, this study has limitations that should be recognized and considered along with its findings. The quantitative and a major portion of the qualitative data were obtained using an online platform, which may have excluded possible participants who did not have internet access. Furthermore, the online platform combined with subsequent snowball sampling likely skewed the sample to be more highly educated and predominately female, which limited analysis by different subgroups. Future research of this topic should be structured to ensure good representation of participants with low educational attainment as well as ethnic and racial diversity. Online completion also meant that there was no way to ensure that self-screening was accurate and that multiple participation did not take place.

On review of our study findings, the SAB suggested that the evidence base developed through this research could inform recommendations for displaying PRO data in educational materials and decision aids. They advised engaging a broader group of stakeholders to develop recommendations that are both evidence-based and stakeholder-driven. Thus, 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, and to identify areas of uncertainty and opportunities for further research.

The findings of this study will be useful to promote more effective communication of comparative study PRO data with patients, thereby informing discussions with clinicians that result in better informed decisions and in turn, more effective patient-centered care and enhanced quality of life. The next step of our research is to use these results as part of an evidence base as we partner 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 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 (R-1410-24904). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. U.S. Food and Drug Administration: Guidance for Industry. (2009). Patient reported outcome measures: Use in medical product development to support labeling claims. *Federal Register*, 74, 65132–65133.
2. Acquadro, C., Berzon, R., Dubois, D., Leidy, N. K., Marquis, P., Revicki, D., & Rothman, M. (2003). 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(5), 522–531.
3. Lipscomb, J., Gotay, C. C., & Snyder, C. (Eds.). (2005). *Outcomes assessment in cancer: Measures, methods and applications*. Cambridge, MA: Cambridge University Press.
4. Greenhalgh, J. (2009). The applications of PROs in clinical practice: What are they, do they work, and why? *Quality of Life Research*, 18(1), 115–123.
5. Bruner, D. W., Bryan, C. J., Aaronson, N., Blackmore, C. C., Brundage, M., Cella, D., et al. (2007). Issues and challenges with integrating patient-reported outcomes in clinical trials supported by the National Cancer Institute-sponsored clinical trials networks. *Journal of Clinical Oncology*, 25(32), 5051–5057.
6. Till, J. E., Osoba, D., Pater, J. L., & Young, J. R. (1994). Research on health-related quality of life: Dissemination into practical applications. *Quality of Life Research*, 3(4), 279–283.
7. Au, H. J., Ringash, J., Brundage, M., Palmer, M., Richardson, H., & Meyer, R. M. (2010). Added value of health-related quality of life measurement in cancer clinical trials: The experience of the NCIC CTG. *Expert Review of Pharmacoeconomics & Outcomes Research*, 10(2), 119–128. <https://doi.org/10.1586/erp.10.15>.
8. Brundage, M. D., Feldman-Stewart, D., Bezjak, A., Leis, A., Degner, L., Fleming, S. et al. (2005). The value of quality of life information in a cancer treatment decision. In *ISOQOL 11th annual conference*, San Francisco
9. Brundage, M., Bass, B., Jolie, R., & Foley, K. (2011). A knowledge translation challenge: Clinical use of quality of life data from cancer clinical trials. *Quality of Life Research*, 20(7), 979–985.
10. Snyder, C. F., & Aaronson, N. K. (2009). Use of patient-reported outcomes in clinical practice. *Lancet*, 374(9687), 369–370.
11. Bezjak, A., Ng, P., Skeel, R., Depettrillo, A. D., Comis, R., & Taylor, K. M. (2001). Oncologists' use of quality of life information: Results of a survey of eastern cooperative oncology group physicians. *Quality of Life Research*, 10(1), 1–13.
12. Brundage, M. D., Smith, K. C., Little, E. A., Bantug, E. T., Snyder, C. F., & PRO Data Presentation Stakeholder Advisory Board. (2015). Communicating patient-reported outcome scores using graphic formats: Results from a mixed methods evaluation. *Quality of Life Research*, 24(10), 2457–2472.
13. Smith, K. C., Brundage, M. D., Tolbert, E., Little, E. A., Bantug, E. T., Snyder, C. F., & PRO Data Presentation Stakeholder Advisory Board. (2016). Engaging stakeholders to improve presentation of patient-reported outcomes data in clinical practice. *Support Care in Cancer*, 24(10), 1–9. <https://doi.org/10.1007/s00520-016-3240-0>.
14. Brundage, M., Blackford, A., Tolbert, E., Smith, K., Bantug, E., Snyder, C., & PRO Data Presentation Stakeholder Advisory Board. (2018). Presenting comparative study PRO results to clinicians and researchers: Beyond the eye of the beholder. *Quality of Life Research*, 27(1), 75–90.
15. Tolbert, E., Brundage, M., Bantug, E., Blackford, A. L., Smith, K., Snyder, C., & PRO Data Presentation Stakeholder Advisory Board (2018). Picture this: Presenting longitudinal patient-reported outcome research study results to patients. *Medical Decision Making*. <https://doi.org/10.1177/0272989X18791177>.
16. Zipkin, D. A., Umscheid, C. A., Keating, N. L., et al. (2014). Evidence-based risk communication: A systematic review. *Annals of Internal Medicine*, 161, 270–280.
17. Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). *Advanced mixed methods research designs. Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks: Sage
18. AtlasTi, in, ATLAS.ti Scientific Software Development GmbH. 2014.
19. Bantug, E. T., Coles, T., Smith, K. C., Snyder, C. F., Rouette, J., & Brundage, M. D. (2016). Graphical displays of patient-reported outcomes (PRO) for use in clinical practice: What makes a picture worth a thousand words? *Patient Education and Counseling*, 99(4), 483–490. <https://doi.org/10.1016/j.pec.2015.10.027>.
20. Feldman-Stewart, D., Brundage, M. D., & Zotov, V. (2007). Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. *Medical Decision Making*, 27(1), 34–43. <https://doi.org/10.1177/0272989X06297101>.

21. Le, T., Aragon, C., Thompson, H. J., & Demiris, G. (2014). Elementary graphical perception for older adults: A comparison with the general population. *Perception*, *43*(11), 1249–1260.
22. Corbin, J. C., Reyna, V. F., Weldon, R. B., & Brainerd, C. J. (2015). How reasoning, judgment, and decision making are colored by gist-based intuition: A fuzzy-trace theory approach. *Journal of Applied Research in Memory and Cognition*, *4*(4), 344–355.
23. Joseph-Williams, N., Newcombe, R., Politi, M., Durand, M. A., Sivell, S., Stacey, D., et al. (2014). Toward minimum standards for certifying patient decision aids a modified Delphi consensus process. *Medical Decision Making*, *34*(6), 699–710.
24. McDonald, H., Charles, C., & Gafni, A. (2014). Assessing the conceptual clarity and evidence base of quality criteria/standards developed for evaluating decision aids. *Health Expectations*, *17*(2), 232–243.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.