Title
Comprehension, utility, and preferences of prostate cancer survivors for visual timelines of patient-reported outcomes co-designed for limited graph literacy: meters and emojis over comics.

Journal
Journal of the American Medical Informatics Association, 29(11)

Authors
Snyder, Lauren
Phan, Daniel
Williams, Kristen
et al.

Publication Date
2022-10-07

DOI
10.1093/jamia/ocac148

Peer reviewed

Journal of the American Medical Informatics Association, 29(11), 2022, 1838–1846
https://doi.org/10.1093/jamia/ocac148
Advance Access Publication Date: 30 August 2022
Research and Applications

Research and Applications

Comprehension, utility, and preferences of prostate cancer survivors for visual timelines of patient-reported outcomes co-designed for limited graph literacy: meters and emojis over comics

Lauren E. Snyder[1], Daniel F. Phan[1,2], Kristen C. Williams[3], Eduardo Piqueiras[3], Sarah E. Connor[3], Sheba George[4,5], Lorna Kwan[3], Jefersson Villatoro Chavez[3], Megha D. Tandel[3], Stanley K. Frencher Jr [3], Mark S. Litwin [3,4,6,7], John L. Gore [8], and Andrea L. Hartzler [1]

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, USA, [2]Department of Biobehavioral Nursing and Health Informatics, School of Nursing, University of Washington, Seattle, Washington, USA, [3]Department of Urology, University of California, Los Angeles, California, USA, [4]Department of Community Health Sciences, The Fielding School of Public Health, University of California, Los Angeles, California, USA, [5]Department of Preventive & Social Medicine and the Center for Biomedical Informatics, Charles R. Drew University of Medicine & Science, Los Angeles, California, USA, [6]School of Nursing, University of California, Los Angeles, California, USA, [7]Department of Health Policy & Management, The Fielding School of Public Health, University of California, Los Angeles, California, USA, and [8]Department of Urology, University of Washington, Seattle, Washington, USA

Corresponding Author: Andrea L. Hartzler, PhD, Department of Biomedical Informatics and Medical Education, University of Washington, 850 Republican St. Building C, Seattle, WA 98109, USA; andreah@uw.edu

Received 8 May 2022; Revised 10 August 2022; Editorial Decision 11 August 2022; Accepted 15 August 2022

ABSTRACT

Objective: Visual timelines of patient-reported outcomes (PRO) can help prostate cancer survivors manage longitudinal data, compare with population averages, and consider future trajectories. PRO visualizations are most effective when designed with deliberate consideration of users. Yet, graph literacy is often overlooked as a design constraint, particularly when users with limited graph literacy are not engaged in their development. We conducted user testing to assess comprehension, utility, and preference of longitudinal PRO visualizations designed for prostate cancer survivors with limited literacy.

Materials and methods: Building upon our prior work co-designing longitudinal PRO visualizations with survivors, we engaged 18 prostate cancer survivors in a user study to assess 4 prototypes: Meter, Words, Comic, and Emoji. During remote sessions, we collected data on prototype comprehension (gist and verbatim), utility, and preference.

Results: Participants were aged 61–77 ($M = 69$), of whom half were African American. The majority of participants had less than a college degree (95%), had inadequate health literacy (78%), and low graph literacy (89%). Among the 4 prototypes, Meter had the best gist comprehension and was preferred. Emoji was also preferred, had the highest verbatim comprehension, and highest rated utility, including helpfulness, confidence, and satisfaction. Meter and Words both rated mid-range for utility, and Words scored lower than Emoji and Meter for comprehension. Comic had the poorest comprehension, lowest utility, and was least preferred.

Discussion: Findings identify design considerations for PRO visualizations, contributing to the knowledge base for visualization best practices. We describe our process to meaningfully engage patients from diverse and hard-to-reach groups for remote user testing, an important endeavor for health equity in biomedical informatics.

Conclusion: Graph literacy is an important design consideration for PRO visualizations. Biomedical informatics researchers should be intentional in understanding user needs by involving diverse and representative individuals during development.

Key words: user-centered design, health education, consumer health information, health literacy, patient-reported outcome measure

INTRODUCTION

With an expected 268 490 new cases in 2022, prostate cancer is the most common cancer in men in the United States.[1] Understanding health data over time is important in cancer care; many survivors will live for many years with prostate cancer, alongside symptoms, and corresponding treatment decisions, which impact their quality of life. Instruments capturing patient-reported outcomes (PRO) among prostate cancer survivors[2] assess the impact of treatment on quality of life.[3] Although PRO are effective measures of the impact of prostate cancer treatments on quality of life,[4] many patients with limited literacy and numeracy find traditional reporting formats (eg, bar, line, pie charts) difficult to understand,[5] which can inhibit informed decision-making[6] and communication with providers and loved ones.

Although prior research suggests that the use of emojis[7] and visual analogies[8] could improve PRO comprehension, these visualizations are not used in practice and many people with limited literacy still have worse PRO outcomes[9] and ask providers fewer questions.[10] As evidenced by a 2019 systematic review of patient-facing health visualizations,[11] more work is needed to inform the design of PRO visualizations for people who are less familiar with traditional charts and graphs in ways they understand and find acceptable. This systematic review found that 76% of studies tested line graphs, bar graphs, and number lines—challenging formats for some users—yet, few studies assessed graph literacy, numeracy, or general literacy. Graph literacy impacts both gist (ie, overall) and verbatim (ie, specific) understanding, yet preferences for and understanding of health data visualizations are not necessarily related.[12] Previous studies have found correlations between graph literacy and numeracy,[13] socioeconomic status,[14] and ethnic and racial background.[15] Yet, there is still a need to better understand the relationship between graph literacy and visualization outcomes, such as comprehension, perceptions of utility, and preference.

Engaging people from vulnerable groups in the design of health interventions is essential to successful intervention development and evaluation but is frequently neglected in research.[16] The exclusion of vulnerable and hard-to-reach populations in biomedical informatics research exacerbates inequities.[17] For example, much prior work to develop PRO visualizations for people with limited literacy has not deeply engaged those users in the design process. Thus, a guiding principle of our project was inclusion of prostate cancer survivors with limited graph literacy. Throughout this project, we closely engaged a patient advisory board (PAB) of survivors recruited from vulnerable groups. Through this engagement, we co-designed web-based, interactive PRO visualizations (ie, "prototypes") with these survivors.[18]

In this article, we report on our user study to assess comprehension, utility, and preference of the co-designed prototypes in a group of racially diverse prostate cancer survivors, of whom the majority had inadequate health literacy and limited graph literacy. We describe our prototype designs, then report on findings from user testing, which required methodological adjustments for remote access among this hard-to-reach population during the COVID-19 pandemic.

OBJECTIVE

The purpose of this user study was to investigate the following research questions among racially diverse survivors with limited education:

- RQ1. What is the comprehension of the PRO visualizations?
- RQ2. What is the utility of the PRO visualizations?
- RQ3. Which PRO visualizations are preferred?

A secondary objective of this study was to perform a preliminary investigation of the usability of web-based, interactive versions of the PRO prototypes in this vulnerable group.

MATERIALS AND METHODS

Study design and procedures

From February 2021 to April 2021, we conducted a remote user study with prostate cancer survivors to assess comprehension, utility, and preferences among 4 prototype PRO visualizations: "Meter," "Words," "Comic," and "Emoji." User testing involved individual sessions lasting 60–90 min to assess the prototypes and literacy for each participant. We pilot-tested the procedures with 2 PAB members and incorporated their feedback into the final session guide (Multimedia Supplementary Material S1). Three members of the research team (LES, DFP, and EP) shared in conducting individual sessions on Zoom, either by phone or web video conferencing, with weekly debriefs and discussion with co-investigators (ALH, SG, SEC, and KCW).

COVID-19 restrictions required shifting from previously planned in-person to remote user testing sessions via telephone or video calls. All participants were mailed a *participant packet* with a session guide and paper copies of the prototypes in advance of their session and asked not to look at the materials until their scheduled session (Multimedia Supplementary Material S1). We included several views of each prototype so that page turns would simulate clicking through an interactive version. The packet grounded the session with a description of a fictional patient to contextualize the prototypes in a concrete scenario. The scenario detailed a fictional patient "John," an African American prostate cancer survivor and truck driver in Los Angeles who has trouble reading graphs. During user testing sessions, the interviewer asked participants to "put themselves in John's shoes" while engaging each prototype to assess comprehension, utility (including helpfulness, confidence, and satisfaction), and preferences. Procedures were approved by Institutional Review Boards at the University of California, Los Angeles, and the University of Washington.

Participant recruitment

We recruited prostate cancer survivors through the UCLA Men's Health Study (MHS), which recruits enrollees from the Improving Access, Counseling, and Treatment for Californians with Prostate Cancer (IMPACT) Program.[19] IMPACT is funded by the California Department of Health Care Services and is administered by UCLA Urology. IMPACT provides free prostate cancer treatment to California residents 18 years or older with little or no insurance and an income under 200% of the Federal Poverty Level. MHS conducts telephone interviews, including quality of life surveys administered at program enrollment and every 6 months thereafter for 5 years. As of December 2021, MHS demographics were 85% nonWhite, 57% non-English-speaking, 46% less than high school education, and 31% without any income. From the 1328 participants with data in the UCLA MHS database from 2002 to 2020, 849 were contactable (ie, alive, in good health, remained in the study, and not lost to follow-up). Of the contactable participants, 267 speak English, reported having no college degree and are racially diverse (ie, Asian, Black, Hispanic/Latino). We limited recruitment to 120 most recently engaged in MHS who we attempted to contact. As we sought to engage individuals with limited graph and health literacy, we screened MHS data as a proxy to identify survivors who did not graduate from college (ie, limited education), and then measured health literacy and graph literacy during the study (see Literacy assessment). Utilizing the UCLA MHS database, we recruited until the list of 120 eligible individuals was exhausted. Verbal consent was obtained during the participant screening phone call for inclusion (ie, ability to read and speak English, no college degree, racially diverse). Once consented, participants were scheduled for a remote user testing session and mailed the participant packet. Of the 33 individuals we contacted who consented (28% response rate), 15 were lost to follow-up (eg, failed to respond to rescheduling requests and follow-up phone calls), resulting in 18 participants who completed the study.

Prototypes

We tested 3 co-designed prototypes: Meter, Words, and Comic (Figure 1). In prior work,[13] we co-designed these PRO timeline visualizations with the PAB through several iterative rounds to display longitudinal Expanded Prostate Cancer Index Composite (EPIC) urinary function scores[20] over 48 months.[18] *Meter* evokes a speedometer and uses minimal text, color, and arrows to encode PRO data. Meter design was sketched out and favored by the PAB; prior research suggests that visual analogies, such as the Meter, improve comprehension of PRO visualizations.[8] In the interactive, web version of Meter, users can select different timepoints on the timeline to display corresponding PRO scores. *Words* uses vertical positioning and color to encode the severity of scores and horizontal positioning to encode time. Words includes textual information about the symptoms that contribute to scores, which PAB members indicated was important. The web version of Words features a clickable (ie, interactive) timeline. *Comic* was suggested by PAB as a pictorial representation of conversations between patient and provider. The characters' expressions reflected scores: when scores are lower, expressions are sad; when scores are higher, expressions are joyous. In the web version of Comic, users press an audio "play" button to hear the characters' conversation aloud. We included a fourth alternative—*Emoji*—because it was shown to be preferred and improved comprehension among people with limited literacy in prior work.[7] By including an example design from the literature with the 3 championed by our PAB, we presented a varied set of visual formats to situate user testing and expand our findings beyond our co-design work. We sought to understand if co-designed formats, as alternatives to traditional graphic formats, could improve performance and understanding among prostate cancer survivors with limited education.

Data collection

Remote user testing sessions consisted of 2 parts: (1) paper prototype assessment and (2) literacy assessment. All participants received the mailed participant packet containing the paper prototypes and materials for literacy assessment. A subset of participants who had access to a smartphone or computer also completed a third part to assess the usability of interactive, web versions of the prototypes. Participants who completed the usability assessment accessed the interactive prototypes through screenshare on Zoom. Verbal responses were coded onto the data collection sheet, which is shown with all tasks and interview prompts in the interviewer Session Guide (Multimedia Supplementary Material S2). All sessions were recorded and transcribed for analysis.

1. *Paper prototype assessment* collected quantitative and qualitative data regarding participant interactions with, and perceptions of, paper versions of the 4 prototypes. We counterbalanced the order of prototypes across participants and balanced the order of positive and negative framing (eg, better or worse scores than "other patients like John"). We did this by alternating the display of PRO data that was better or worse than population benchmarks. For each prototype, we collected data on comprehension, utility, and preferences.

We assessed *comprehension* for gist and verbatim understanding. For *gist* comprehension, we followed the interpretation of Arcia et al[21] of the ISO 9186 standard by asking participants for each prototype, "How would you describe what is happening to a loved one?" During interviews, participants were encouraged to continue explaining their understanding of the prototype until it was clear to the interview if the participant understood the visualization correctly (eg, describe trend in urinary scores getting better or worse over time). Two researchers (LES and DFP) scored gist accuracy as a "clear understanding" or "unclear understanding" of the prototype. Responses in which there was ambiguity in the participant's explanation were discussed with the larger research team for consensus on scoring gist comprehension. For *verbatim* comprehension, we evaluated whether participants responded correctly to the following questions aligned with 3 levels of graph literacy: "reading the data" (ie, to find specific information on the graph), "reading between the data" (ie, to find relationships in data shown on a graph), and "reading beyond the data" (ie, to make inferences or predictions from data on a graph).[22] To assess each level, we asked participants: (1) "What was John's score X months after treatment?"; (2) "At X months after treatment, was John's score better or worse than people like John?"; and (3) "X months after treatment, what score do you expect John to have?". For the "reading the data" and "reading between the data" questions, there was only one response scored as correct, such as 86 and better. For the "reading beyond the data" question, we considered responses ±3 of the correct score as accurate. An open-ended comprehension

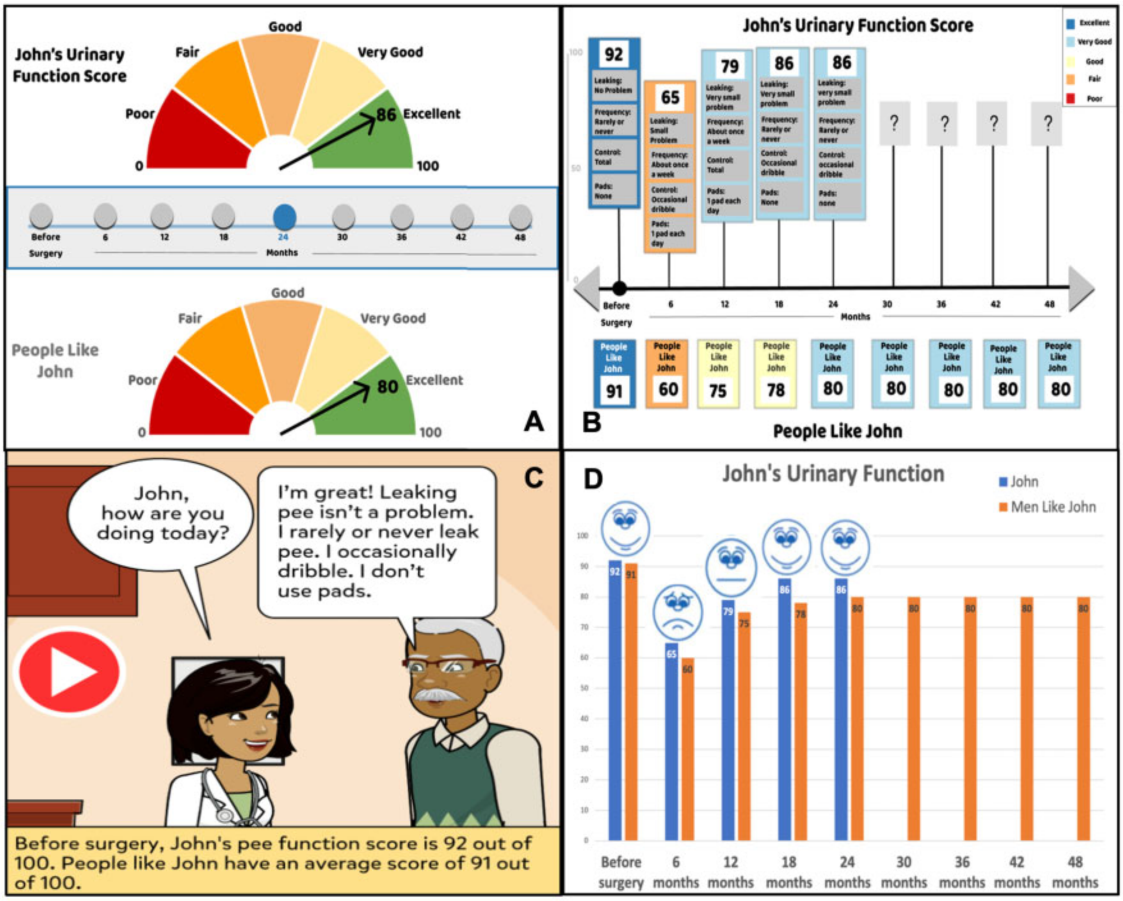

Figure 1. Meter (**A**), Words (**B**), Comic (**C**), and Emoji (**D**) prototypes. Emoji is our own adaptation of a previously published visualization.[7] Only one view of the interactive visualizations is shown. Additional views depicting entire timelines are shown in the participant packet (Multimedia Supplementary Material S1).

question probed participants on how the prototype could be made clearer and easier to understand.

We assessed *utility* by asking participants to rate each prototype for helpfulness, confidence, and satisfaction. Responses were captured using a Likert scale options for helpfulness (1 = "Not helpful at all" to 4 = "Extremely helpful"), confidence (1 = "Not confident at all" to 4 = "Extremely confident") and satisfaction (1 = "Worst" to 5 = "Best"). Open-ended utility questions prompted participants on what could make the prototype easier to use and how they saw themselves using it.

We assessed *preference* by asking participants to rank the 4 prototypes from best to worst, and then asked them to respond to open-ended questions about what they liked and disliked about each prototype.

2. *Literacy assessment* used the Newest Vital Sign (NVS).[23] NVS is a validated measure of health literacy, defined as an understanding of words (prose), numbers (numeracy), and forms (documents). Administration of the 6-item NVS takes 3 min and has been validated against more time-consuming literacy assessment tools.[24] We assessed graph literacy using the Short Graph Literacy scale (SGL).[25] The SGL is a validated, concise graph literacy instrument that consists of 4 items and is adapted from the 13-item Graph Literacy Scale.[22]

3. *Usability assessment*: Participants who could access Zoom using a Smartphone or computer were invited to take part in the usability assessment. We deployed interactive versions of the Meter, Words, and Comic prototypes to a publicly accessible website: http://tinyurl.com/grasp2021. Because the comparator from the literature, Emoji, was not designed to be interactive, it was excluded from usability testing. We asked participants to access the website, share their screen using Zoom, and complete tasks, while we observed their interactions with the prototype. An example task was "Make the chart show John's symptoms 12 months after treatment." After interacting with each prototype, we verbally administered the System Usability Scale (SUS).[26]

Data analysis

We summarized quantitative data with descriptive statistics. We compared the highest and lowest comprehension scores with Fisher's exact tests and compared utility ratings across the 4 prototypes with Friedman tests in R.[27] We summarized qualitative comments through deductive template analysis,[28] focused on comprehension, utility, preference, and usability of the prototypes. Four members of the research team (LES, DFP, EP, and ALH) applied an *a priori* codebook based on the research questions. After reviewing transcripts, we created qualitative memos for each participant following the codebook template to record notes, contextual information, and qualitative impressions of comprehension, utility, preferences, and when applicable usability. The 4 members met regularly to review and discuss qualitative data to provide context to the quantitative results and resolve any discrepancies through group consensus.

Table 1. Participant characteristics

	All participants (*n* = 18)	Paper prototype assessment only (*n* = 14)	Paper prototype and usability assessment (*n* = 4)
Age, mean (SD), range	69 (4.8), 61–77	69 (4.3), 64–77	68 (6.7), 61–75
Not available	1	1	0
Race/Ethnicity, *n* (%)			
African American/Black	9 (50%)	7 (50%)	2 (50%)
Hispanic/Latino	3 (16.7%)	2 (14.3%)	1 (25%)
Non-Hispanic White	5 (27.8%)	4 (28.6%)	1 (25%)
Not available	1 (5.6%)	1 (7.1%)	0 (0%)
Highest education level, *n* (%)			
College graduate[a]	1 (5.6%)	1 (7.1%)	0 (0%)
Some college	7 (38.9%)	6 (42.9%)	1 (25%)
High school or technical school graduate	9 (50%)	6 (42.9%)	3 (75%)
Grade school or less	1 (5.6%)	1 (7.1%)	0 (0%)
Health literacy, *n* (%)			
Adequate literacy	4 (22.2%)	3 (21.4%)	1 (25%)
Possibility of limited literacy	6 (33.3%)	4 (28.6%)	2 (50%)
High likelihood of limited literacy	8 (44.4%)	7 (50%)	1 (25%)
Graph literacy, *n* (%)			
High graph literacy	2 (11.1%)	2 (14.3%)	0 (0%)
Low graph literacy	16 (88.9%)	12 (85.7%)	4 (100%)

[a]One participant earned a college degree after enrolling in IMPACT study, and thus was not screened out at recruitment and was included in our analysis. This participant's assessments showed low graph literacy and high likelihood of limited literacy and their responses were consistent with other participants.

Table 2. Outcomes

	Meter	Words	Comic	Emoji
Comprehension (RQ1), number correct (%) out of 18 participants				
Gist	13 (72.2%)	6 (33.3%)	9 (50.0%)	10 (55.6%)
Verbatim				
Reading the data	15 (83.3%)	13 (72.2%)	9 (50.0%)	14 (77.8%)
Reading between the data	12 (66.7%)	10 (55.6%)	14 (77.8%)	16 (88.9%)
Reading beyond the data	13 (72.2%)	13 (72.2%)	10 (55.6%)	13 (72.2%)
Utility (RQ2), mean rating (SD) across 18 participants				
Helpfulness (out of 4)	3.1 (0.8)	2.9 (0.9)	2.8 (0.9)	3.2 (0.9)
Confidence (out of 4)	3.0 (0.9)	3.1 (0.7)	3.0 (0.8)	3.2 (0.9)
Satisfaction (out of 5)	4.3 (0.9)	4.1 (0.9)	3.8 (1.4)	4.5 (0.7)
Preference (RQ3), *n* (%) out of 18 participants				
First choice	1 (5.6%)	6 (33.3%)	4 (22.2%)	7 (38.9%)
Second choice	13 (72.2%)	1 (5.6%)	1 (5.6%)	3 (16.7%)
Third choice	2 (11.1%)	7 (38.9%)	4 (22.2%)	5 (27.8%)
Fourth choice	2 (11.1%)	4 (22.2%)	9 (50%)	3 (16.6%)

RESULTS

Participants

Eighteen participants completed the study (P1–P18). Of those, 4 (22%) completed usability testing. Table 1 summarizes participant characteristics, health literacy, and graph literacy. Participants were survivors in their 60s and 70s, of whom half were African American, half were high school graduates or less, and the majority had inadequate health literacy and low graph literacy. Participants who completed the usability assessment had a lower proportion of limited health literacy but were otherwise similar to other participants.

Prototype comprehension, utility, and preference

Table 2 shows comprehension, utility, and preference outcomes across the 4 prototypes for all participants. Gist comprehension differed across the prototypes, varying from a low of 33% for Words to a high of 72% for Meter. In contrast, verbatim comprehension was generally higher but also varied. Utility scores, including helpfulness, confidence, and satisfaction, were highest for Emoji. Prototype preferences rankings varied across participants. Multimedia Supplementary Material S3 summarizes participant outcomes stratified by health literacy level and graph literacy level.

Comprehension

Comprehension varied by prototype, with gist comprehension highest for Meter and lowest for Words ($P = .044$, Fisher's exact). Meter had the highest "reading the data" verbatim comprehension, whereas Comic had the lowest ($P = .075$, Fisher's exact). In contrast, Emoji had the highest "reading between the data," whereas Words had the lowest ($P = .060$, Fisher's exact). For "reading-beyond-the-data," Meter, Words, and Emoji formats were similar, but

Comic was lower. Using the ISO 9186 threshold of at least 66% accuracy,[29] the only prototype with adequate gist comprehension across all measures was Meter.

Adequate comprehension for Meter could be related to participants' familiarity with this type of information display, as P18 explained:

> I think the graphs that we have here on [this page] will be better because ...—even though if you couldn't hardly read them, you would know the difference—, looking at a gas gauge or speedometer, you can read this graph much better because it's going to let you know the category where you are.

Although Emoji was considered the simplest visualization, some participants misinterpreted it. For example, P2 conveyed poor gist comprehension: "It's just giving the different urinary functions of two different types of males." Whereas Meter performed well for "reading the data," Emoji performed well for "reading between the data" to compare one's own score ("John's score") to population averages ("Men like John").

Several participants expressed confusion understanding the Comic. For example, P4 noted: "It's consistent but it's confused. I mean it's kind of confusing for someone that's, you know. It's dealing with and trying to get [someone viewing Comic] to read a story to—it doesn't make sense. Just look at the number and the number is there."

Similarly, the crowded placement of the text in Words may have contributed to lower comprehension. P9 noted: "[Words] is ... a little more difficult again just because it's got all the wording around the chart". P6 added, "...[Words] is little congested, but other than that, the rest of them are pretty much, you know, you recognize what's going on fairly quickly." Some participants described the amount of text, as well as how that text was laid out in Words, added to their confusion.

Utility

Utility ratings varied by the prototype. Although prototype ratings did not differ significantly for helpfulness ($X^2 = 3.52$, $P = .32$) or confidence ($X^2 = 2.05$, $P = .56$), ratings for satisfaction varied significantly across prototypes ($X^2 = 10.71$, $P = .01$). Emoji scored highest in perceived helpfulness, confidence, and satisfaction, whereas Comic had the lowest scores. The utility of Words was rated in the middle. Although some participants felt that Words contained too much or poorly placed text, others appreciated the additional information. P9 shared the value of contextual information along with the numeric scores that made Words easier to use:

> Yes, extra information is really important for people like me cause the more information they get, the better they feel about it. When it's real vague and stuff then they're like well, but they can read something and see it and then read it and then they can say oh, that happens to me. That's what I'm doing too. That's my problem. I'm having those symptoms too and the more information the better.

Preference

Concordant with the utility ratings, Emoji ranked highest with 38.9% of participants ranking it first as most preferred and 16.7% ranking it second. Meter was also preferred by participants with 72.2% of participants ranking it second. Comic was ranked last by half of participants.

Some participants did not like that Comic seemed juvenile. P3 explained his low ranking of Comic was due to confusion and lack of seriousness of the topic:

> No, I love cartoons. I watch cartoons all the time. It's just as far as getting data across; I think I personally needed it to be more straightforward. This looks like, if John had 6 months to pee how many buckets would he pee in 6 months in math. That's kind of what it reminds me of. If a train going 45.... that kind of thing. It's...I don't want to say juvenile but it's overly simplified.

Others described their use of similar visuals in everyday life. P14 described his preference for Emoji due to its association with his smartphone:

> The other, still, on your phone, it would probably be good to go with the Emoji because that's what people are, that's what the world is revolving around. People talk with Emojis now.

Prototype usability

Our preliminary usability results from testing the web-based, interactive prototypes with 4 participants align with findings from assess-

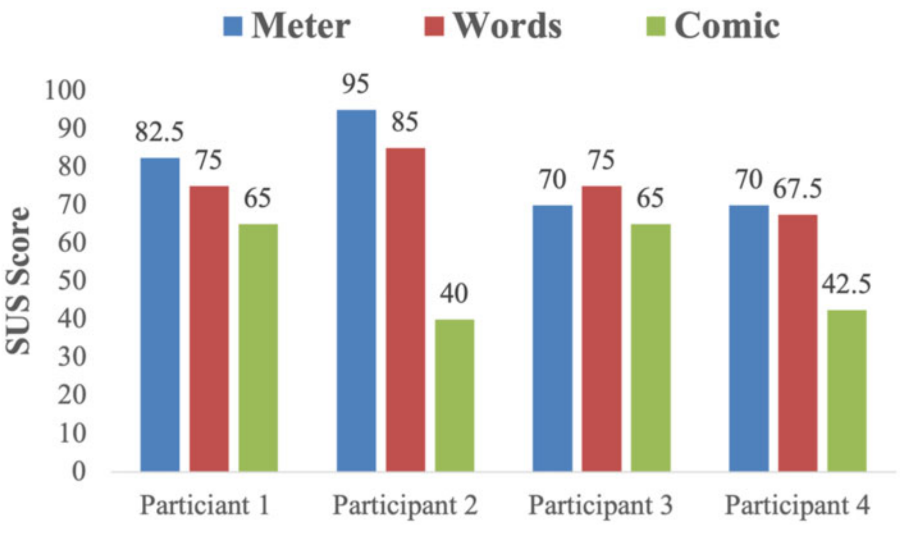

Figure 2. SUS scores by usability participant.

ing paper prototypes (Figure 2). Meter and Words had mean SUS scores of 79.4 and 75.6, respectively (considered "grade B"). Comic performed notably worse, with a mean score of 53.1 (considered "grade D"). The learnability and usability SUS subscales follow the same pattern as the overall SUS scores with Meter and Words scoring higher and Comic scoring lower.

DISCUSSION

Our findings provide insight into co-designed longitudinal PRO visualizations that were best comprehended, found useful, and preferred by prostate cancer survivors with limited education. Meter had the best gist comprehension and was highly preferred. Emoji was highly preferred, had the highest verbatim comprehension, and highest rated utility, including helpfulness, confidence, and satisfaction. Comic had the poorest comprehension, lowest utility, and was least preferred.

Participants provided qualitative insights to explain these findings. They valued simplicity and ease of understanding, but also wanted complete and informative depictions. Some participants felt that using a comic format made light of an incredibly challenging cancer journey. This balance of familiar association and simplicity was perhaps better achieved by other prototypes. Participants noted the familiarity of Meter and Emoji as justifications for high utility and preference rankings. Meter and Emoji could have reduced cognitive barriers or resistance to challenging information, by using everyday objects (ie, speedometer) or symbols (eg, used during texting).

A strength of this study was the intentional involvement and inclusion of racially diverse participants, including the involvement of the PAB.[18] It is critical that user testing involves representative users, but diverse populations are hard to reach, and, unfortunately, biomedical research often lacks inclusion.[30] Especially in prostate cancer research, where disparities exist in incidence and health outcomes, it is critical to improve equitable representation. Another strength was our study design included multiple prototypes to provide a balanced picture of the user perspective. Compared with single prototype testing, including multiple prototypes can foster more substantive participant feedback during user testing.[31]

Despite our intention to engage a diverse group, we only reached a small number of participants capable of remote videoconference sessions for usability assessment, reflecting the digital divide that affects this research. Biomedical informatics researchers during the COVID-19 pandemic have adapted user testing to remote-based methods. A previous study reports challenges with remote user and usability testing, including trust-building, extended length of setup, and technical challenges.[32] We experienced similar challenges, such as longer than anticipated session times and barriers with participant's technology use and access. To increase rapport and shorten individual sessions, we met with participants multiple times for scheduling, setup, before the user testing session. We also maintained communication and focused on building trust with the PAB who served as advisors for prototype development and pilot testing. We made every effort to create the conditions for study participants to share their own experiences and provide their opinions openly. COVID-19 has increased social isolation and feelings of loneliness, especially among older adults.[33] Although isolation entails potential harms, we found that it can also promote participants' willingness to communicate in-depth about their prostate cancer experiences and design perspectives with researchers. Despite this potential benefit, our challenges recruiting participants for web-based interactive testing could signal disparities in remote methods that require greater technology access and digital literacy. It is important that researchers consider connectivity and accessibility barriers of vulnerable groups, as well as impacts of COVID-19 on research participation and study design.

One surprising finding was the poor outcomes of Comic, which was enthusiastically favored by the PAB during co-design.[34] This finding could be related to differences between the PAB and study participants, the shift from web-based to paper prototypes, or changing cultural preferences. We did not assess the graph literacy or health literacy of the PAB who had high school education and greater. Beyond comprehension and utility, differences in preference for Comic could be related to differences in literacy levels. The text component of Comic required more reading than the other prototypes, so the poor outcomes could reflect literacy challenges. Another reason for the poor Comic outcomes may have been that the intended interactive design plays audio of the conversation between patient and provider to describe PRO trends over time. However, this functionality was only feasible in the interactive, web-based, prototype experienced by 4 of 18 participants. Another constraint of the paper-based version of the Comic was that participants may have struggled to remember one panel of the Comic to the next when flipping through the pages. Although there is a growing body of literature that suggests visualizations using a story-telling approach effectively contextualizes information for users,[35] the use of comics or cartoons for users with limited graph literacy is underexplored. For example, a study among cancer patients with low socioeconomic status found a disconnect between outcomes; comprehension of comics was high, but acceptability was low.[36] The preference for Emoji over Comic could reflect familiarity with emojis due to their high proliferation in smartphones and social media. In fact, 93% of the overall MHS population from which we recruited used a cell phone and 61% used text messaging. Over the past decade, we have witnessed a stark decrease in the circulation of newspapers—a historically common medium from which to access comics—but steady increase in cell phone use across all races, ethnicity, and age groups,[37] Understanding underlying cultural shifts in graphic preferences, including the potential shift from favoring comics to emojis, is an important direction for future work on the acceptability of health data visualizations. Additionally, specific elements of the Comic design might be particularly important for personalizing visualizations, such as skin color and gender of the physician, which was highlighted by some study participants and the PAB. Furthermore, it is important to note that although we found poor outcomes associated with Comic in this study, comics or storytelling formats may be valuable tools with other data and use cases.[38]

A primary limitation of this study was the small number of participants in our interactive usability session ($n = 4$). Although participants provided rich qualitative data regarding their prostate cancer experience, the qualitative data we collected regarding the prototypes was sparse and limited our qualitative findings. This was compounded by our sole focus on English-speaking prostate cancer survivors. Future studies should explore visual PRO timelines in other languages, such Spanish-speaking populations and other language groups, to improve diverse representation. Additionally, our findings indicate that visual formats have strengths and weaknesses that could reflect well-known perceptual differences impacting performance.[38] For example, the unfamiliar Words may have required more time to learn than the more familiar Meter, Comic, and Emoji bar chart formats, reflected by poorer gist comprehension. Meter makes it easy to determine the value of a data point (ie, "reading the data"), but harder than Emoji to compare 2 data points (ie, "reading between the data") since differences in angles are more difficult to discern than differences in lengths of bars. Such perceptual differences in these formats likely impacted comprehension scores. Informaticists should carefully assess the mapping between visual format and the user's task, especially when designing tools for limited graph literacy users.

Although our sample was small, our results suggest that remote, interactive usability testing with vulnerable populations is possible. In particular, mailing paper packets allowed us to reach a population that may be technically or financially challenged and missed by other research studies. However, there were a number of disadvantages: (1) mailing packets added time between screening and data collection, (2) participants found it challenging to envision interactivity based on paper prototypes, and (3) some participants lost their paper packets and needed to be re-mailed and the interview rescheduled. We attribute much loss to follow due to the complexities of the onboarding process for remote sessions (eg, responding to scheduling requests and follow-up phone calls, testing Zoom in a pre-session before usability testing). Given pandemic-related disparities known to impact racially diverse and older people, we imagine that many individuals experienced competing demands during COVID-19 pandemic, making this a particularly challenging time to recruit. Thus, engaging vulnerable populations remotely to evaluate biomedical informatics interventions necessitates additional time, training, and connectivity considerations as well as significant recruitment efforts.

Overall, our research provides an initial exploration into usability, but future work should compare perceptions and performance in a larger sample of individuals with varied backgrounds and literacy levels. Additionally, future research is needed to understand the detailed use cases, where, and how prostate cancer survivors use the PRO visualizations to elucidate functional requirements. Design researchers should carefully consider all tasks associated with a visualizations' use cases to ensure that those tasks map to the appropriate visual format.[38] Finally, a comparison of in-person and remote methods would be advantageous to understand strengths and weaknesses of both approaches.

CONCLUSION

Longitudinal visualizations are an important tool in communicating trends in PRO. However, graph literacy should be considered a design constraint, necessitating thorough user testing of alternative prototypes to ensure users understand the information displayed *and* find these tools acceptable. Among diverse prostate cancer survivors engaged in user testing of co-designed prototypes, we found that Meter and Emoji performed well for comprehension, utility, and preference, whereas the Comic performed worse for all 3 outcomes. Our study demonstrates the potential for intentional engagement of a hard-to-reach population through remote methods. This work highlights the critical need to consider all facets of the user population—including broad definitions of literacy (eg, health literacy and graph literacy), socio-economic status, race and ethnicity—in design and evaluation of patient-facing health technologies. Considering the complex patient journey during prostate cancer, it is critical that patient-facing visualizations are designed and tested in collaboration with patients from vulnerable groups.

FUNDING

The research reported in this publication was supported by the US Department of Defense Office of the Congressionally Directed Medical Research Programs (CDMRP) grant, W81XWH-18-1-0465 (PI: MSL).

AUTHOR CONTRIBUTIONS

LES developed interactive prototypes, drafted the manuscript, and incorporated revisions from all authors. ALH supervised the project and provided substantive edits to the manuscript. LES, ALH, and DFP designed the data collection instrument with input from SEC, KCW, and SG. LES and DFP piloted the instrument. LES, DFP, and EP conducted interviews. LES, DFP, EP, and ALH analyzed the data with input from KCW, SEC, SG, and KCW. JVC, MDT, and KCW recruited participants and provided demographic data. LES, DFP, and ALH performed the literature review. JLG, MSL, SKF, KCW, SEC, SG, ALH, EP, KCW, and DFP contributed input into design of the study, substantive review, and final approval.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

The authors would like to thank all study participants, who generously gave their time and efforts for this research, especially the GRASP Study Patient Advisory Board (PAB). The research was facilitated by University of California, Los Angeles, and the University of Washington. The content is solely the responsibility of the authors.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

1. National Cancer Institute: Surveillance, Epidemiology, and End Results Program. Prostate Cancer—Cancer Stat Facts. https://seer.cancer.gov/statfacts/html/prost.html Accessed April 30, 2022.
2. Schmidt S, Garin O, Pardo Y, *et al.*; EMPRO Group. Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments. *Qual Life Res* 2014; 23 (8): 2169–81.
3. Ávila M, Patel L, López S, *et al.* Patient-reported outcomes after treatment for clinically localized prostate cancer: a systematic review and meta-analysis. *Cancer Treat Rev* 2018; 66: 23–44.
4. Donovan JL, Hamdy FC, Lane JA, *et al.*; ProtecT Study Group. Patient-reported outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med* 2016; 375 (15): 1425–37.
5. Snyder CF, Smith KC, Bantug ET, Tolbert EE, Blackford AL, Brundage MD; PRO Data Presentation Stakeholder Advisory Board. What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. *Cancer* 2017; 123 (10): 1848–59.
6. Kilbridge KL, Fraser G, Krahn M, *et al.* Lack of comprehension of common prostate cancer terms in an underserved population. *JCO* 2009; 27 (12): 2015–21.
7. Stonbraker S, Porras T, Schnall R. Patient preferences for visualization of longitudinal patient-reported outcomes data. *J Am Med Inform Assoc* 2020; 27 (2): 212–24.
8. Reading Turchioe M, Grossman LV, Myers AC, Baik D, Goyal P, Masterson Creber RM. Visual analogies, not graphs, increase patients' compre-

hension of changes in their health status. *J Am Med Inform Assoc* 2020; 27 (5): 677–89.

9. Katz P, Dall'Era M, Trupin L, *et al.* Impact of limited health literacy on patient-reported outcomes in systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* 2021; 73 (1): 110–9.
10. Aboumatar HJ, Carson KA, Beach MC, Roter DL, Cooper LA. The impact of health literacy on desire for participation in healthcare, medical visit communication, and patient reported outcomes among patients with hypertension. *J Gen Intern Med* 2013; 28 (11): 1469–76.
11. Turchioe MR, Myers A, Isaac S, *et al.* A systematic review of patient-facing visualizations of personal health data. *Appl Clin Inform* 2019; 10 (4): 751–70.
12. van Weert JCM, Alblas MC, van Dijk L, Jansen J. Preference for and understanding of graphs presenting health risk information. The role of age, health literacy, numeracy and graph literacy. *Patient Educ Couns* 2021; 104 (1): 109–17.
13. Nayak JG, Hartzler AL, Macleod LC, Izard JP, Dalkin BM, Gore JL. Relevance of graph literacy in the development of patient-centered communication tools. *Patient Educ Couns* 2016; 99 (3): 448–54.
14. Durand M-A, Yen RW, O'Malley J, Elwyn G, Mancini J. Graph literacy matters: Examining the association between graph literacy, health literacy, and numeracy in a Medicaid eligible population. *PLoS One* 2020; 15 (11): e0241844.
15. Okan Y, Galesic M, Garcia-Retamero R. How people with low and high graph literacy process health graphs: evidence from eye-tracking. *J Behav Dec Making* 2016; 29 (2–3): 271–94.
16. Willis A, Isaacs T, Khunti K. Improving diversity in research and trial participation: the challenges of language. *Lancet Public Heal* 2021; 6 (7): e445–6.
17. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc* 2018; 25 (8): 1080–8.
18. Snyder L, Saraf AA, Casanova-Perez R, *et al.*Visualization Co-Design with Prostate Cancer Survivors who have Limited Graph Literacy. In: *2020 Workshop on Visual Analytics in Healthcare (VAHC)*. 2020: 17–23; Chicago, IL.
19. IMPACT-IMProving Access, Counseling, and Treatment for Californians with Prostate Cancer | Index Page. https://www.california-impact.org/pages/ Accessed April 30, 2022.
20. Wei JT, Dunn RL, Litwin MS, Sandler HM, Sanda MG. Development and validation of the Expanded Prostate Cancer Index Composite (EPIC) for comprehensive assessment of health-related quality of life in men with prostate cancer. *Urology* 2000; 56 (6): 899–905.
21. Arcia A, Grossman LV, George M, Turchioe MR, Mangal S, Creber RMM. Modifications to the ISO 9186 method for testing comprehension of visualizations: successes and lessons learned. In: *2019 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*. IEEE; 2019: 41–7; Vancouver, BC, Canada.
22. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. *Med Decis Making* 2011; 31 (3): 444–57.
23. Pfizer. The Newest Vital Sign: A Health Literacy Assessment Tool. 2011. https://cdn.pfizer.com/pfizercom/health/nvs_flipbook_english_final.pdf Accessed January 13, 2022.
24. Weiss BD, Mays MZ, Martz W, *et al.* Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med* 2005; 3 (6): 514–22.
25. Okan Y, Janssen E, Galesic M, Waters EA. Using the short graph literacy scale to predict precursors of health behavior change. *Med Decis Making* 2019; 39 (3): 183–95.
26. Brooke J. SUS—a quick and dirty usability scale. *Usability Eval Ind* 1996; 189: 4–7.
27. R Studio Team. RStudio: Integrated Development for R. 2020. https://www.rstudio.com/ Accessed February 10, 2022.
28. Crabtree B, Miller W. *Doing Qualitative Research*. 2nd ed. Thousand Oaks, CA: Sage Publications; 2000.
29. ISO/TC 145 SC/1. Testing According to ISO 9186-1—Criteria of Acceptability. 2014. https://isotc.iso.org/livelink/livelink/fetch/-8903172/8903808/8903819/9945810/Testing_according_to_ISO_9186-1_-_criteria_of_acceptability.pdf?nodeid=10136076&vernum=-2 Accessed August 8, 2022.
30. Oh SS, Galanter J, Thakur N, *et al.* Diversity in clinical and biomedical research: a promise yet to be fulfilled. *PLoS Med* 2015; 12 (12): e1001918.
31. Tohidi M, Buxton W, Baecker R, Sellen A. Getting the right design and the design right: testing many is better than one. In: *CHI 2006 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2006: 1243–51; Montreal.
32. Andreasen MS, Nielsen HV, Schrøder SO, Stage J. What happened to remote usability testing? An empirical study of three methods. In: *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. 2007: 1405–14; San Jose, CA.
33. Hwang T-J, Rabheru K, Peisah C, Reichman W, Ikeda M. Loneliness and social isolation during the COVID-19 pandemic. *Int Psychogeriatr* 2020; 32 (10): 1217–20.
34. Daniel H, Snyder Sulmasy L. Policy Recommendations to Guide the Use of Telemedicine in Primary Care Settings: An American College of Physicians Position Paper. *Ann Intern Med*. 2015; 163 (10): 787–9.
35. Yen PY, Bakken S. Review of health information technology usability study methodologies. *J Am Med Inform Assoc.* 2012; 19 (3): 413–22.
36. Alam S, Elwyn G, Percac-Lima S, Grande S, Durand M-A. Assessing the acceptability and feasibility of encounter decision aids for early stage breast cancer targeted at underserved patients. *BMC Med Inform Decis Mak* 2016; 16: 147.
37. Pew Research Center. Demographics of Mobile Device Ownership and Adoption in the United States | Pew Research Center. 2021. https://www.pewresearch.org/internet/fact-sheet/mobile/#who-owns-cellphones-and-smartphones Accessed January 12, 2022.
38. Franconeri SL, Padilla LM, Shah P, Zacks JM, *et al.* The science of visual data communication: what works. *Psychol Sci Public Interest* 2021; 22 (3): 110–61.