Journal of the American Medical Informatics Association, 0(0), 2019, 1–13
doi: 10.1093/jamia/ocz217
Research and Applications

Research and Applications

Visual analogies, not graphs, increase patients' comprehension of changes in their health status

Meghan Reading Turchioe, [1†] **Lisa V. Grossman** , [2†] **Annie C. Myers,**[1] **Dawon Baik,**[3] **Parag Goyal,**[4] **and Ruth M. Masterson Creber** [1]

[1]Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, New York, USA, [2]Department of Biomedical Informatics, College of Physicians and Surgeons, Columbia University, New York, New York, USA, [3]College of Nursing, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA, and [4]Department of Medicine, Weill Cornell Medicine, New York, New York, USA

[†]These authors contributed equally to this work.

Corresponding Author: Lisa V. Grossman, Department of Biomedical Informatics, Columbia University, 622 W 168th St, PH-20, New York, NY 10032, USA; lvg2104@cumc.columbia.edu

Received 1 July 2019; Revised 25 November 2019; Editorial Decision 9 December 2019; Accepted 12 December 2019

ABSTRACT

Objectives: Patients increasingly use patient-reported outcomes (PROs) to self-monitor their health status. Visualizing PROs longitudinally (over time) could help patients interpret and contextualize their PROs. The study sought to assess hospitalized patients' objective comprehension (primary outcome) of text-only, non-graph, and graph visualizations that display longitudinal PROs.

Materials and Methods: We conducted a clinical research study in 40 hospitalized patients comparing 4 visualization conditions: (1) text-only, (2) text plus visual analogy, (3) text plus number line, and (4) text plus line graph. Each participant viewed every condition, and we used counterbalancing (systematic randomization) to control for potential order effects. We assessed objective comprehension using the International Organization for Standardization protocol. Secondary outcomes included response times, preferences, risk perceptions, and behavioral intentions.

Results: Overall, 63% correctly comprehended the text-only condition and 60% comprehended the line graph condition, compared with 83% for the visual analogy and 70% for the number line ($P = .05$) conditions. Participants comprehended the visual analogy significantly better than the text-only ($P = .02$) and line graph ($P = .02$) conditions. Of participants who comprehended at least 1 condition, 14% preferred a condition that they did not comprehend. Low comprehension was associated with worse cognition ($P < .001$), lower education level ($P = .02$), and fewer financial resources ($P = .03$).

Conclusions: The results support using visual analogies rather than text to display longitudinal PROs but caution against relying on graphs, which is consistent with the known high prevalence of inadequate graph literacy. The discrepancies between comprehension and preferences suggest factors other than comprehension influence preferences, and that future researchers should assess comprehension rather than preferences to guide presentation decisions.

Key words: data visualization, consumer health informatics, human-computer interaction, patient-reported outcome measures, health communication

INTRODUCTION

The potential impact of patient-reported outcomes (PROs),[1–5] coupled with policies encouraging their use[6–8] and patients' satisfaction with them,[1,2,5] explains their recent increase in popularity and use across multiple medical domains. Researchers and administrators use PROs to assess population-level trends,[9–11] whereas clinicians use PROs to inform individual-level care[12] and improve patient-clinician communication,[13–16] which can improve health outcomes.[1–5] As such, the federal financial incentive policy Meaningful Use stage 3 mandates that healthcare organizations collect PROs.[6–8] Government-sponsored initiatives have promoted electronic PRO (e-PRO) integration into major health record systems, including Epic and Cerner,[17] and independent e-PRO systems have developed concurrently.[17–21] Increasingly, patients use PROs, independent of clinician involvement, for self-monitoring and symptom management.[10,12,21–24] In the future, many more patients will likely receive their PROs, as transparency increases[25,26] and e-PRO systems gain functionality.[27]

Although newer and less well researched, e-PRO systems can learn lessons from older and more well-researched patient-facing technologies, specifically patient portals. One major lesson is that displaying raw results, without interpretation or contextualization, is ineffective.[28–30] Without assistance, only the most well-educated and health-literate patients can effectively interpret raw results. This causes intervention-generated inequity, a phenomena where well-intentioned interventions worsen existing health disparities, rather than reduce them.[31–33] To prevent intervention-generated inequity, e-PRO systems must employ strategies to ensure all patients can interpret their results.

Visualization, defined as representation of information using graphs or images, is a uniquely effective strategy to improve interpretation and contextualization.[34–37] In patient portals and other applications, visualization has improved patients' medical decision making by representing risk,[35,38–40] helped patients distinguish between urgent and nonurgent laboratory test results,[36,37,41,42] and enhanced understanding of treatment options.[35,43–45] PROs, like risk modeling or laboratory test results, contain quantitative scores and interpretable ranges. Therefore, the benefits of visualization will likely apply to PROs. Because current e-PRO systems rarely incorporate visualization, any advances have potential for far-reaching impact, such as improved self-monitoring and better communication.

Current research on PRO visualization is limited,[46–50] and several unique challenges remain unsolved. The first challenge is displaying longitudinal changes in PROs, or changes over time. Previously studied PRO visualizations predominantly used line graphs to convey longitudinal changes.[46–50] Unfortunately, inadequate graph literacy is prevalent, comprising around 40% of the U.S. population.[51] Alternative longitudinal visualization strategies to line graphs remain unexplored. The second challenge is representing PRO scores. PRO measures vary in scale, and the same score can have widely different clinical meaning depending on the measure, hindering interpretation.[10] The third challenge is assessing objective comprehension (ie, "What does this mean?"), rather than proxies such as subjective comprehension (ie, "Did you understand this? Yes/no") or subjective preferences (ie, "Did you like this? Yes/no"). The relationship between subjective proxies and objective comprehension is unknown, and subjective comprehension may be inaccurate. The fourth challenge is evaluation in the real-world–like setting. Previous studies were conducted in laboratory-based settings,[46–50] which have fewer distractions and may prompt participants to spend more time scrutinizing visualizations than they would normally.[52]

In this study, we aimed to assess hospitalized patients' objective comprehension (primary outcome) of text-only, non-graph, and graph visualizations that display longitudinal changes in PROs. Given the known benefits of visualization and the known low levels of graph literacy, our primary hypothesis was that non-graph visualizations would help patients comprehend their longitudinal changes in PROs significantly better than would text-only and graph visualizations. We evaluated 2 non-graph visualizations that exploited different cognitive mechanisms: (1) visual analogies and (2) number lines. As a secondary outcome, we assessed subjective preference to observe its correlation with objective comprehension.

MATERIALS AND METHODS

Study design

Between January and May 2019, we recruited a purposive sample of 40 hospitalized heart failure patients diverse on age, gender, and race from an inpatient cardiac unit at a large urban hospital. We recruited heart failure patients because heart failure is prevalent[53,54] and requires rigorous symptom self-management.[55] We assessed baseline characteristics, then each participant viewed 4 visualization conditions: (1) text-only, (2) text plus visual analogy, (3) text plus number line, and (4) text plus line graph (Figure 1). We counterbalanced, or systematically randomized the order and information contained in each condition, to control for potential order effects (Supplementary Appendix 1).[56] The primary outcome was comprehension, assessed using the International Organization for Standardization Method for Testing Comprehension (ISO 9186).[57] Secondary outcomes included response times, preferences, risk perceptions, and behavioral intentions. Additionally, we captured qualitative information related to the visualization components, such as colors, graphics, and text. The Weill Cornell Medicine Institutional Review Board approved the study.

Visualizations

We selected conditions based on effective visualization concepts reported in the literature.[37,42,46–49,58] A visual analogy systematically compares 2 concepts using a graphic, one familiar ("the analog," a graphic) and the other unfamiliar.[59–62] For example, the gas gauge (Figure 1B) serves as an analog for physical function, and compares decreasing function with "running out of gas." A line graph displays the relationship between 2 variables, measured on 2 respective axes, and connects individual points using lines. Here, we display the relationship between PRO score and time (Figure 1D). A number line, unlike a graph, displays one variable on one axis. Here, the number line displays the PRO score (Figure 1C). The text-only condition (Figure 1A) does not contain any graphics, and serves as a control.

A professional visual designer adapted each concept to PROs and created the designs. Each condition displayed the score (eg, "8 out of 10"), the clinical interpretation (eg, "high"), and the date (eg, "today"). Side-by-side comparison of the current and previous score was provided. A header was used to textually interpret the change ("better," "worse," "same," "about the same"). We included the header based on previous research that brief text interpretations combined with visualizations perform better than either alone.[49,58,63] A red-yellow-green color scheme represented meaning,

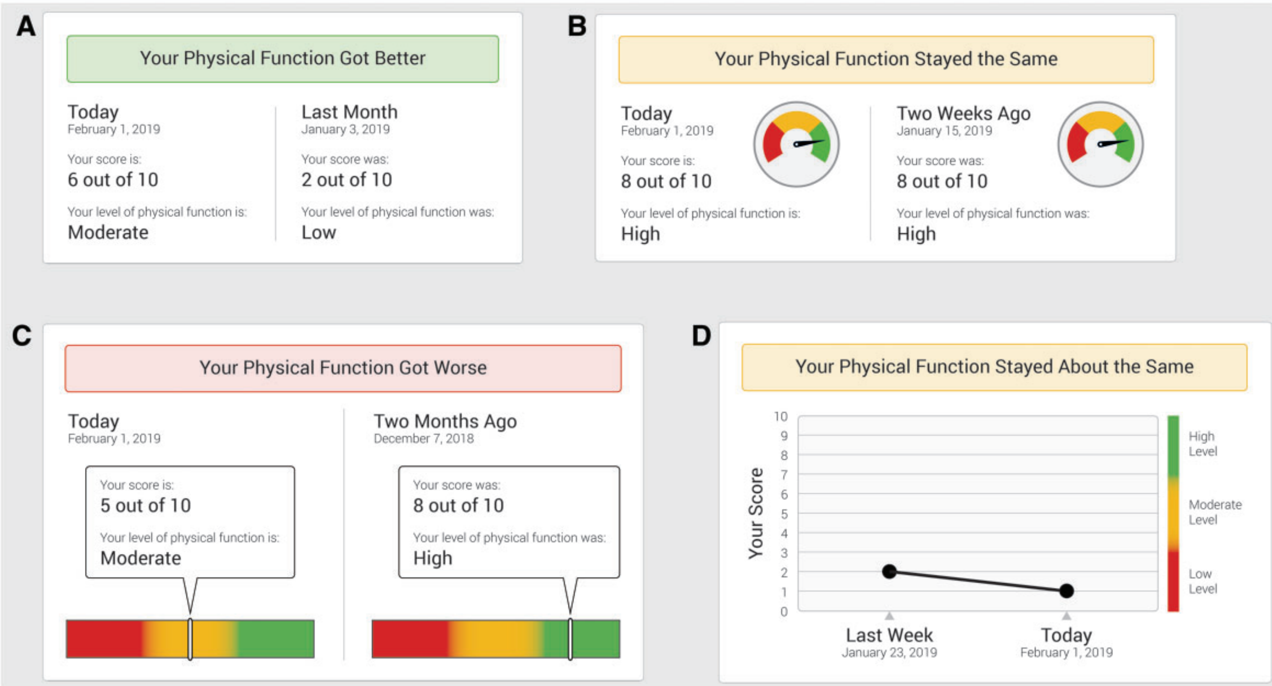

Figure 1. Visualization conditions: (A) text-only; (B) text plus visual analogy; (C) text plus number line; (D) text plus line graph.

where green indicates good, red indicates bad, and yellow indicates in-between. We visualized Patient-Reported Outcome Measurement Information System (PROMIS) constructs, specifically fatigue, depression, and physical function.[64,65] PROMIS is a standardized and broadly applicable set of PRO measures. We choose each visualization condition partly because of its potential generalizability to numerous PROMIS and PRO constructs.

Scores

We displayed scores as a number out of 10. This is in direct contrast with previous work, which typically displayed scores as raw or T-scores.[46–49] To translate T-scores into base-10 scores, we first determined the minimally important difference (MID), defined as the smallest clinically meaningful change in T-score. As a conservative heuristic, PROMIS recommends an MID of 5 for all patients.[66] This is a general recommendation and potentially less relevant for specific disease populations. Therefore, researchers have developed disease-specific MIDs.[66,67] Based on our own clinical data[68,69] and previously reported MIDs,[67] we selected an MID of 2 for heart failure patients, which we then used for conversion to base-10. As an example, Figure 2 demonstrates how we converted T-scores to base-10 scores for PROMIS physical function in heart failure patients.

Our rationale for base-10 conversion is as follows. First, few patients possess familiarity with statistical concepts like T-scores.[70] Second, research has indicated that patients comprehend base-10 scores.[71] Third, base-10 scores may represent change more effectively than base-100 scores,[72] especially when mapped nonlinearly to clinical severity as described above. For example, movement from 7 of 10 to 5 of 10 is a much higher percentage change than movement from 40 of 100 to 36 of 100. However, both movements represent the same severe and clinically important change from New York Heart Association (NYHA) functional class II to NYHA functional class III heart failure.[69] Fourth, each unit change in base-10 score represents clinically meaningful change, while each unit change in T-score does not.[66] For example, movement from 7 of 10 to 6 of 10 indicates worsening physical function, while movement from 40 of 100 to 39 of 100 does not.

Participants

We included participants 21 years of age or older, who spoke English, were diagnosed with heart failure according to the electronic health record, and were willing and able to provide informed consent. We excluded patients with severe cognitive impairment such as clinically diagnosed dementia and patients with major psychiatric illness such as active psychosis. The exclusion criteria were limited to improve generalizability. We purposively balanced the participant pool based on age, gender, and race. We aimed for an appropriate distribution of 4 generation groups (Millennials, Generation X, Boomers, and Silent Generation), as recommended by the Pew Research Center.[73] Generation group is more informative of social context and personal history than an arbitrary biological age category. We also aimed for adequate sampling of racial minorities (20% Latino, 20% Black) and gender (50% women). Participants were compensated $25 for their time.

Procedures

The research coordinators identified new cardiac admissions from the electronic health record, and clinicians confirmed eligibility and an active heart failure diagnosis through chart review. After written informed consent, the coordinator guided each participant through the study protocol, typically 30 to 60 minutes in duration. First, the coordinator collected demographic information and important covariates related to comprehension. Second, the coordinator assessed comprehension for each condition as per ISO 9186. Finally, the coordinator assessed secondary outcomes (preferences, risk perceptions, and behavioral intentions). A second coordinator audio-recorded each session using a Memorex digital voice recorder, dis-

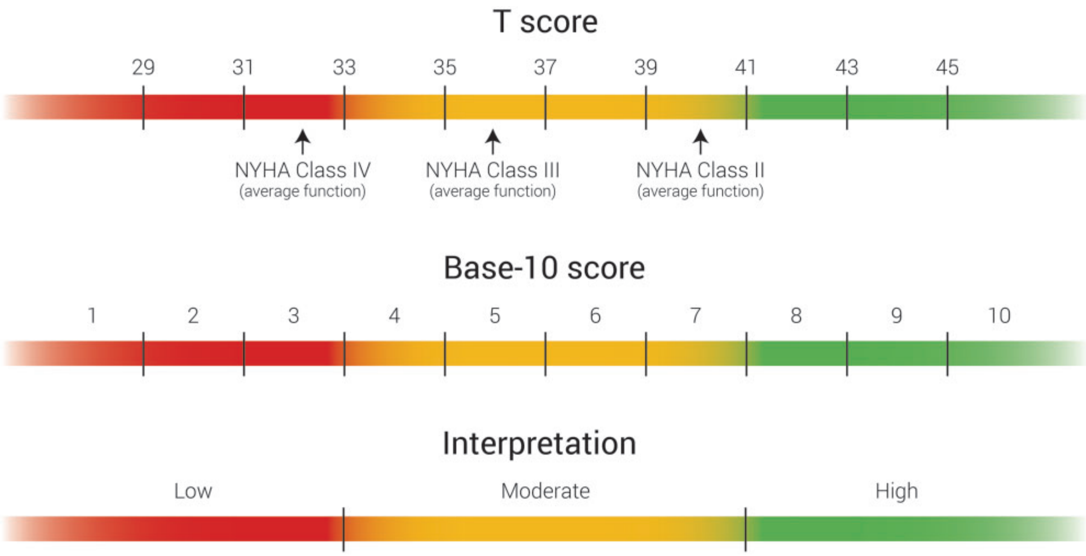

Figure 2. Base-10 conversion chart for Patient-Reported Outcome Measurement Information System physical function in heart failure. NYHA: New York Heart Association.

cretely assessed response times, and concurrently captured quantitative data using REDCap (Research Electronic Data Capture) on an Apple iPad. We conducted individual interviews in 2 rounds, each with 20 people, and modified visualizations to eliminate potential confounding between rounds. A technical document that contains additional details on the design, modifications, randomization, and procedures is available as Supplementary Appendix 1. The recruitment protocol is available as Supplementary Appendix 2.

Measurements

Participants completed a baseline questionnaire to assess demographics, socioeconomic status, technology experience, cognitive status, health literacy, subjective numeracy, graph literacy, and heart failure severity. Demographics included age, generation, gender, race, and ethnicity. Socioeconomic status included education, financial resources, insurance status, and disability status. Technology experience included possession of an email address, computer ownership, smartphone ownership, Internet access, and Internet use based on our previously published questionnaire.[29,74] We assessed cognitive status using the Montreal Cognitive Assessment (MoCA).[75] We screened for inadequate health literacy using the 3-item health literacy screen.[76] We assessed numeracy using the 3-item subjective numeracy scale.[77–80] We assessed graph literacy using the 4-item graph literacy scale.[51,81] Finally, we assessed heart failure severity using the NYHA functional classification questionnaire.[82]

Comprehension, response times, and preferences

We adapted ISO 9186 for verbal administration as previously described.[83] In brief, the coordinator verbally primed participants to the desired granularity of responses, then asked participants to explain each visualization's meaning. If the response was incomplete or incorrect, the coordinator could only say "tell me more," which prevented them from biasing the response. We considered the response correct if the participant correctly stated the direction of change ("better," "worse," or "the same"). A second coordinator recorded response time, defined as the time (in seconds) between the participant hearing the prompt and giving their response. After the participant responded, the coordinator probed for additional information by asking participants to indicate components of each visualization used for comprehension, and by asking for preferences.

Risk perceptions and behavioral intentions

The coordinator showed a single visualization, dependent on what conditions the participant had adequately comprehended (Figure 3). In each case, the single visualization displayed the same severely worsening symptom burden. Participants completed the 3-item subjective risk perceptions questionnaire,[43,84–86] which assessed the perceived likelihood, seriousness, and concern for the worsening symptom burden. Then, the coordinator assessed behavioral intentions by asking "how likely would you be to do something in response?" as derived from ISO 9186.[57,86] The coordinator asked participants to describe any intended actions, and probed for the rationale behind each action or nonaction. For both subjective risk perceptions and behavioral intentions, responses were recorded on a Likert-type scale (1: Not at all, 2: Slightly, 3: Moderately, 4: Very, 5: Extremely). After removing the visualization from view, the coordinator asked participants the 2-item objective risk perceptions questionnaire,[71,78,87] which contained 1 gist and 1 verbatim recall question. Gist recall required remembering the direction of change (ie, worsening), whereas verbatim recall required remembering the amount of change (ie, from 8 to 3).

Statistical analysis

Quantitative

We conducted descriptive and inferential statistical analyses in SAS version 9.4 (SAS Institute, Cary, NC) and R version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria). A professional service transcribed the audio recordings verbatim. Two independent raters scored the transcribed comprehension responses as per ISO 9186 ("correct" or "incorrect"). We conducted 1 round of interrater reliability, and found high initial agreement between raters (Cohen's kappa = 0.74). The study team met to review, discuss, and arrive at consensus for the comprehension scoring. To assess whether comprehension differed between conditions, we used the Cochran Q test and McNemar exact tests. We used the Benjamini-

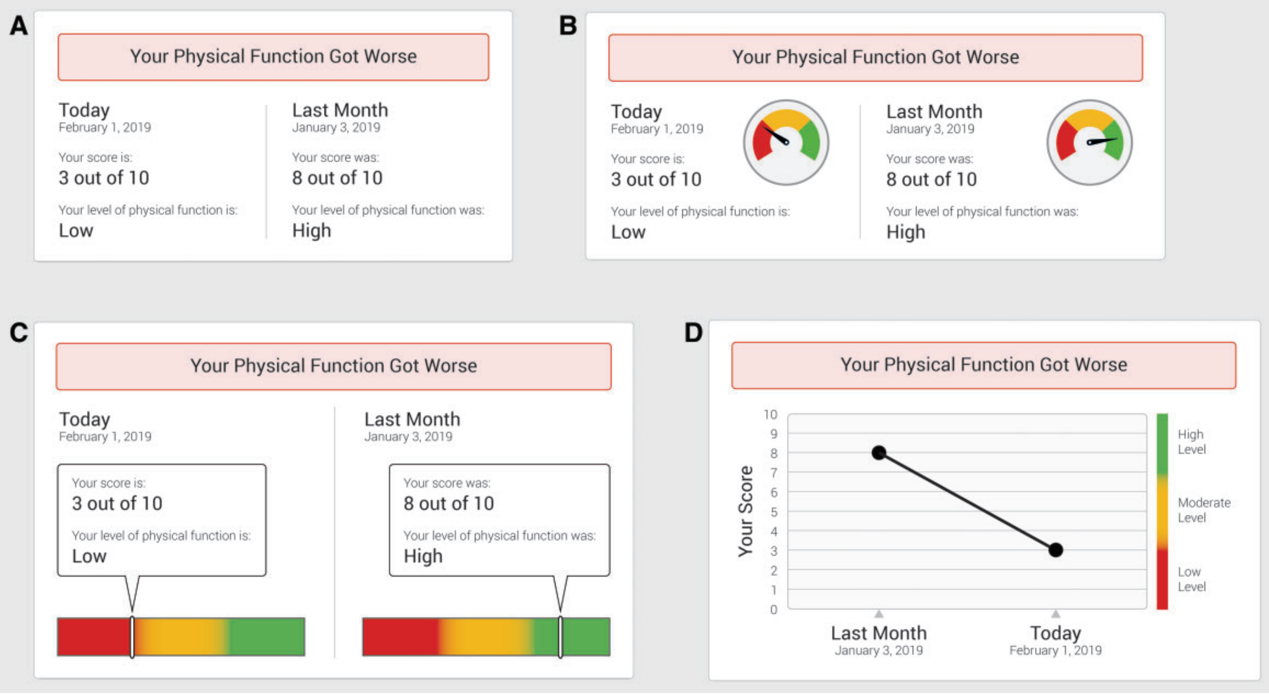

Figure 3. Visualizations for assessing risk perceptions and behavioral intentions: (A) text-only; (B) text plus visual analogy; (C) text plus number line; (D) text plus line graph.

Hochberg procedure for multiple comparisons correction as appropriate for non-independent tests.

To assess whether response times and preferences differed between conditions, we used repeated measures analysis of variance and Fisher's exact tests, respectively. We conducted bivariate analyses to assess whether each baseline characteristic differed between participants with high overall comprehension (>2 conditions comprehended) and low overall comprehension (≤2 conditions comprehended). Nominal variables were compared using chi-squared tests or Fisher's exact tests, while ordinal and numerical variables were compared using Wilcoxon rank-sum tests.

Qualitative

Transcripts were imported into Dedoose 8.0.35 (SocioCultural Research Consultants, Manhattan Beach, CA) for thematic analysis. First, 2 researchers with training in qualitative methods independently used inductive open coding to identify all emerging themes and subthemes. To ensure quality, the researchers met frequently to discuss. Disagreements were resolved by discussion with the study team, and the study team reviewed the final coding dictionary. Then, 2 researchers independently coded all transcripts using codes corresponding to the dictionary. We conducted 1 round of inter-coder comparison queries in Dedoose (Cohen's kappa = 0.64). To enhance confirmability, we shared summaries with the research coordinators for their confirmation or revisions to interpretation, and we triangulated results with the quantitative data.

RESULTS

Table 1 presents the baseline characteristics. Overall, participants had an average age of 61.3 ± 12.5 years (range, 34-85) and were 22% women, 23% Black, 52% White, and 38% Latino. 45% reported an education level of high school or less, and 55% screened positive for inadequate health literacy, comparable with the U.S. mean.[76] On average, participants scored 12.2 ± 4.7 out of 18 (range, 3-18) on subjective numeracy, comparable with the U.S. mean.[78] Most (88%) had cognitive impairment (MoCA score <26). On average, participants scored poorly on graph literacy (mean 1.1 ± 1.1 out of 4), and 37% did not answer any graph literacy items correctly.

Comprehension, response times, and preferences (Table 2)

Overall, 63% correctly comprehended the text-only condition and 60% comprehended the line graph condition, compared with 83% for the visual analogy and 70% for the number line ($P = .05$) conditions, consistent with our primary hypothesis. Participants comprehended the visual analogy significantly better than the text-only ($P = .02$) and line graph ($P = .02$) conditions. Participants did not comprehend the number line significantly better than the text-only ($P = .61$) or line graph ($P = .42$) conditions. An additional table that contains every pairwise comparison is available in Supplementary Appendix 3.

Response times were lowest for the least comprehended condition (line graph: mean 7.8 ± 4.8 seconds), and highest for the most comprehended condition (visual analogy: mean 9.1 ± 5.8 seconds). Response times did not significantly differ between conditions ($P = .69$).

Overall, only 1 (2%) participant preferred the text-only condition, whereas 31% preferred the visual analogy, 41% preferred the number line, and 26% preferred the line graph ($P = .006$). Three (8%) participants did not comprehend any conditions (Table 3). Of participants who comprehended at least 1 condition, 14% preferred

Table 1. Baseline characteristics

Variable	Overall (n = 40)	Comprehension Low (n = 15)	Comprehension High (n = 25)	*P* Value
Demographics				
Age (mean ± SD)	61.3 ± 12.5	63.5 ± 9.5	59.9 ± 14.0	.38
Generation (n, %)				.27
Silent generation	8 (20)	3 (20)	5 (20)	
Baby boomer	21 (52)	10 (67)	11 (44)	
Generation X	10 (25)	2 (13)	8 (32)	
Millennial	1 (3)	0 (0)	1 (4)	
Female gender (n, %)	9 (22)	3 (20)	6 (24)	.77
Race (n, %)				.46
Asian, multirace, or other	9 (23)	2 (13)	7 (28)	
Black or African American	10 (25)	5 (33)	5 (20)	
White or Caucasian	21 (52)	8 (53)	13 (52)	
Hispanic or Latino origin (n, %)	10 (25)	5 (33)	5 (20)	.42
Socioeconomic status				
Education (n, %)				.02[a]
High school graduate or less	18 (45)	10 (67)	8 (32)	
Some college or bachelor's degree	17 (43)	5 (33)	12 (48)	
Graduate degree	5 (12)	0 (0)	5 (20)	
Financial resources (n, %)				.03[a]
Not enough	17 (43)	10 (67)	7 (28)	
Enough	18 (45)	4 (27)	14 (56)	
More than enough	5 (12)	1 (7)	4 (16)	
Insurance status (n, %)				.41
Medicare	24 (60)	10 (67)	14 (56)	
Medicaid	17 (43)	5 (33)	12 (48)	
Employer	11 (28)	3 (20)	8 (32)	
Disability status (n, %)				.57
Physical disability	13 (33)	4 (27)	9 (36)	
Hearing-related disability	1 (3)	0 (0)	1 (4)	
Eyesight-related disability	6 (15)	2 (13)	4 (16)	
Other disability	8 (20)	3 (20)	5 (20)	
No disability	19 (48)	8 (53)	11 (44)	
Technology experience				
No email address (n, %)	8 (20)	5 (33)	3 (12)	.10
No Internet access (n, %)	3 (7)	2 (13)	1 (4)	.39
Computer ownership (n, %)				.82
Desktop	11 (28)	4 (27)	7 (28)	
Laptop	22 (55)	7 (47)	15 (60)	
Tablet	15 (38)	10 (67)	5 (20)	
Smartphone	34 (85)	13 (87)	21 (84)	
Does not own computer	6 (15)	2 (13)	4 (16)	
Internet use per day (n, %)				.71
<30 minutes	18 (45)	8 (53)	10 (40)	
30 minutes to 5 hours	13 (33)	4 (27)	9 (36)	
>5 hours	9 (23)	3 (20)	6 (24)	
Cognitive status				
MoCA score (mean ± SD)[b]	20.0 ± 5.5	15.7 ± 5.4	22.3 ± 4.2	<.001[a]
Cognitive impairment (n, %)				.01[a]
Mild (MoCA <26)	18 (45)	4 (27)	14 (56)	
Moderate (MoCA <18)	8 (20)	6 (40)	2 (8)	
Severe (MoCA <10)	1 (3)	1 (7)	0 (0)	
Did not complete	8 (20)	4 (27)	4 (16)	
Health Literacy				
Inadequate health literacy score (n, %)	22 (55)	12 (80)	10 (40)	.32
Numeracy				
Subjective numeracy score (mean ± SD)[c]	12.2 ± 4.7	10.6 ± 5.7	13.1 ± 3.9	.11
Graph literacy				
Short graph literacy score (mean ± SD)	1.1 ± 1.1	0.83 ± 0.9	1.23 ± 1.1	.32
Short graph literacy score (n, %)				.39
Excellent (4/4 correct answers)	1 (3)	0 (0)	1 (4)	

(continued)

Table 1. continued

Variable	Comprehension			P Value
	Overall (n = 40)	Low (n = 15)	High (n = 25)	
Good (3/4 correct answers)	2 (6)	1 (8)	1 (4)	
Fair (2/4 correct answers)	9 (25)	1 (8)	8 (35)	
Poor (1/4 correct answers)	10 (29)	5 (42)	5 (22)	
Very poor (0/4 correct answers)	13 (37)	5 (42)	8 (35)	
Heart failure severity				
NYHA functional class (n, %)				.74
II	10 (25)	4 (27)	6 (24)	
III	14 (35)	6 (40)	8 (32)	
IV	16 (40)	5 (33)	11 (44)	

Categorical variables reported as n (%). Continuous variables reported as mean ± SD. Percentages exclude missing data. For non-exclusive variables (insurance status, disability status, computer ownership), we report P values for exclusive comparisons only (employer vs nonemployer insurance, disability vs none, computer ownership vs none).

MoCA: Montreal Cognitive Assessment; NYHA: New York Heart Association; SD, standard deviation; n, number.

[a]Significant at an exploratory P value of .05.

[b]MoCA score is out of 30.

[c]Subjective numeracy score is out of 18.

Table 2. Comprehension, response times, and preferences

Variable	Text-Only	Text + Visual Analogy	Text + Number Line	Text + Line Graph	P Value
Primary outcome					
Comprehension (n, %)	25 (63)	33 (83)	28 (70)	24 (60)	.05[a]
Secondary outcomes					
Response times (mean ± SD)	8.1 ± 4.6	9.1 ± 5.8	8.1 ± 5.3	7.8 ± 4.8	.69
Preferences (n, %)	1 (2)	12 (31)	16 (41)	10 (26)	.006[a]

Categorical variables reported as n (%). Continuous variables reported as mean ± SD. Percentages exclude missing data. SD, standard deviation; n, number.

[a]Significant at $P = .05$. All pairwise comparisons reported in Supplementary Appendix 3.

Table 3. Comprehension, risk perceptions, and behavioral intentions

Variable	Overall (n = 40)
Comprehension	
Overall comprehension (n, %)	
4/4 conditions comprehended	15 (38)
3/4 conditions comprehended	10 (25)
2/4 conditions comprehended	8 (20)
1/4 conditions comprehended	4 (10)
0/4 conditions comprehended	3 (8)
Risk perceptions	
Gist recall (n, %)	26 (67)
Verbatim recall (n, %)	4 (10)
Subjective risk perceptions (n, %)	
Likelihood of illness getting worse (very or extremely)	27 (70)
Seriousness of illness getting worse (very or extremely)	18 (46)
Concern about illness getting worse (very or extremely)	22 (54)
Behavioral Intentions	
Intention to act (very or extremely) (n, %)	32 (82)
Intended action (n, %)[a]	
Show to healthcare provider	25 (78)
Attempt to self-manage	3 (10)
Search the Internet	2 (6)
Go to the hospital	2 (6)

Categorical variables reported as n (%). Percentages exclude missing data. HF: heart failure; n, number.

[a]Denominator is those who intended to act (n = 32).

a condition that they did not comprehend. Specifically, 11% who preferred the line graph did not comprehend it, 27% who preferred the number line did not comprehend it, and none who preferred the visual analogy did not comprehend it.

More than half (63%) of participants satisfied criteria for high overall comprehension (>2 conditions comprehended) (Table 3), and the remaining 38% had low comprehension (≤2 conditions comprehended). A comparison of baseline characteristics between high and low comprehension groups is presented in Table 1. Low comprehension was associated with worse cognition (MoCA score 15.7 low vs 22.3 high; $P < .001$), lower education level ($P = .02$), and fewer financial resources ($P = .03$). MoCA scores significantly differed among participants who did and did not understand the line graph ($P = .02$), number line ($P = .03$), and text-only ($P = .04$) conditions, but not the visual analogy ($P = .11$).

The major components of each visualization, and which components participants used for comprehension, are described in Supplementary Appendix 3. Overall, participants most often relied on graphics or scores, and much less often on textual interpretations. About half (51%) used 2 or more components to aid their comprehension.

Risk perceptions and behavioral intentions

The majority (67%) had correct gist recall (general impression), but only 10% had correct verbatim recall (specific numerical). The majority (70%) perceived that the illness got worse, but only 46% perceived it as serious and only 54% reported being worried. However, 82% intended to act in response to the visualization. When asked to describe intended actions, 78% who intended to act reported intent to show their healthcare provider, 10% reported intent to self-manage, 6% reported intent to search the Internet, and 6% reported intent to go to the hospital.

Qualitative analysis

Table 4 describes the main themes from patient interviews. 24 participants (60%) mentioned that color indicated or enhanced the visualization's meaning. Participants associated the following words with *red*: high, severe, worse, danger, inappropriate, not good at all,

Table 4. Qualitative analysis

Category	Theme	Evidence[a]	Example quotes[b]
Colors	Colors indicate or enhance meaning	E	"Red is danger. Yellow is ok. Green is great." (4) "The colors are universal and provide information, like from good to bad." (5)
	Colors used to describe symptom status	S	"He's improved on his fatigue level; it's green." (28) "He's in the red here, and he's in the green here." (36)
	Colors reminiscent of everyday objects	S	"The colors help a lot. It's like the [traffic] light on the street here." (21) "It's very similar to a traffic light, red, yellow, green, obviously green means positive." (6)
	Colors imply need for action	S	"You need red for when something is high and should be paid attention to." (37) "Red is used to warn people; it means seek physical help or medical attention." (5)
Visual analogies	Visual analogies easy to understand or use	S	"It's more understandable. It's very easy on the eyes, and very easy on the brain." (10) "It makes more sense. It's easy to move [the needle] around in your head." (38)
	Visual analogies reminiscent of everyday objects	S	"You're always seeing this chart, since you were a kid, when you cook or drive." (11) "I grew up with dials; I'm used to them, so I think they are better." (38)
	Visual analogies support cognitive analogies	L	"You don't have as much gas as you used to; [your body] is not working like it used to." (12) "You look at the gauge, and you can see you're running out of heart function here." (12)
Components	Single component used for comprehension	E	"I just used the top box, I didn't really need the rest of the picture." (19) "I was just looking at the scores, not the pictures or colors." (33)
	Multiple components used for comprehension	E	"Numbers, colors, lines . . . everything is helpful in its own cohesive way." (32) "Everything helped me figure it out . . . if you can read, everything you need is right here." (14)
	Different components reinforce one another	E	"You can see [the answer] right away; you can go by the numbers or by the color." (34) "I could go back up here and say, does this correlate with that. And it does." (38)
	Different people will use different components	L	"It's a wise choice to leave everything in, because some people don't read completely." (26) "Many people don't deal with numbers so when you see this [picture] it's better." (11)
Context	Need explicit instructions for action	S	"Instead of just leaving him with some basic random score . . . do something to guide him. Tell him when he needs to see the doctor because his score is dropping." (5)
	Need more information on "why" and "how"	S	"What events happened in between that caused the change? I want to know." (28) "It doesn't tell me where I improved, what part of my body improved." (2)
	Visualization needs more description	L	"It would help if there were labels such as severe and moderate on the color scheme." (13) "Add a qualifier to explain that although [the symptom] stayed the same it's still low." (37)
Text	Bigger and more contrast is better	S	"Nice big print here, so I could see it. I could see this without my glasses on." (12) "The text has to be dark; more contrast. So I can see it better." (23)

E: extensive evidence (≥15 mentions); S: substantial evidence (5-14 mentions);
[a]L: limited evidence (2-4 mentions). Cutoffs calculated based on sample size.
[b]The parenthesized number is the participant's study identifier.

warning, hot, blood, and negative. They associated with *yellow*: moderate, middle, okay, appropriate, and so-so. Finally, they associated with *green*: low, improvement, good, great, outstanding, really good, and positive. 9 participants (23%) used color to describe symptom status directly, saying "[the symptom] is in the red" or even "[the symptom] is red."

Participants found visual analogies especially easy to understand, and reminiscent of familiar everyday objects like dials or gauges. Participants used multiple and varied components to comprehend each visualization, consistent with our quantitative results. Some participants wanted additional information, such as explicit instructions for further action based on their results. Although we employed appropriate text sizes as per National Institute on Aging recommendations, 4 participants (10%) reported it could be bigger.

DISCUSSION

Patients increasingly use PROs for self-monitoring and symptom management. Unfortunately, few e-PRO systems interpret or contextualize PROs for patients. Visualization is a known and effective strategy for better interpretation. Our results suggest that non-graph visualizations, specifically visual analogies, enhance patients' comprehension of longitudinal PROs, compared with text-only or graphs. The results support using visualizations rather than text to display longitudinal PROs, but caution against relying on graphs. Additionally, we found discrepancies between objective comprehension and subjective preferences, which suggests factors other than comprehension influence preferences, possibly aesthetics. Finally, our results emphasize the importance of *first impressions* and *decision support* for comprehension and subsequent health decision making.

Visual analogies substantially outperformed the text-only and graph conditions on objective comprehension. By definition, visual analogies systematically compare 2 concepts, one familiar ("the analog") and the other unfamiliar.[59–62] Visual analogies employ different cognitive mechanisms than data graphics such as number lines or graphs,[59] and research indicates that analogies centrally support the human brain's conceptualization of abstractions.[88] Analogies help humans construct *mental models* or cognitive representations of abstract concepts, by basing them upon concrete physical experiences or objects.[59,61,62,88] We observe this phenomenon in our participants' language, such as "he's running out of heart function" (12). Analogies can especially help novices, or persons without preexisting mental models.[59,61,89] This may explain their effectiveness here, as many of our participants had cognitive impairment or limited education. Future research should explore familiar,[62,90] appropriate,[59,62] and effective[61] visual analogies for conveying symptom information.

We found discrepancies between objective comprehension and subjective preferences. One in 7 participants (14%) preferred a visualization that they did not comprehend over one they did. This is consistent with previous research that physicians and patients preferred design features not associated with accurate quantitative judgements.[91,92] Given our results and the previous evidence for discrepancies, future studies should measure objective comprehension directly. This includes user-centered and participatory design research, which may inappropriately conflate subjective preferences with efficacy. Our results suggest that, alongside comprehension, visual aesthetics may influence preferences. Almost all participants preferred conditions with graphics or colors, important components of aesthetics,[93] and more participants mentioned color than any other feature. Future work should explore how aesthetics influence preferences and potentially comprehension as well.

While non-significant, the least comprehended condition had the shortest response times, and the most comprehended condition had the longest response times. One possible explanation is that people may judge their ability to interpret visuals before actually interpreting them.[94] If the visualization looks too complex, people do not attempt to interpret it, regardless of their actual cognitive capacity. This results in a shorter response times. Recent research revealed that people form *first impressions* of visual complexity within the first 17 milliseconds,[94] which influences their intention to interact. A related concept is *positive reinforcement*, which might encourage participants to continue interacting with their health information, even when complexity increases. Future research should explore how first impressions and positive reinforcement may impact patients' interactions with health technologies.

Our behavioral intentions results echo previous research on displaying health information through patient portals.[28–30] Most participants (82%) intended to act, but many expressed uncertainty about how to act. The majority (78%) chose to "show their healthcare provider" to aid their decision, and responses varied dramatically from "search the Internet" to "go to the hospital." This underscores patients' need for health decision support. Comprehension is a necessary first step toward better decision making, but it is insufficient. This is consistent with previous research that visual features which improve quantitative reasoning differ from features that alter behavior or intentions,[92] and that simply displaying information is insufficient to modify behavior.[28–30] Future studies should explore strategies to connect comprehension with healthy actions, and bridge the gap between interpreting and using health information.

Given the known prevalence of low graph literacy,[51] as well as the known cognitive impairment found in hospitalized patients,[95–99] our participants' low graph literacy did not surprise us. Additionally, we found that the text plus line graph had nearly equivalent comprehension to the text-only condition (60% vs 63%; $P = 1.000$). Given these results, communicators should not assume line graphs will be better comprehended than text. This is consistent with previous research that graphs may require more effortful cognitive skills, and patients' interpretations of graphs depend on their expertise and familiarity.[92] The cognitive burden of graphs is especially concerning in hospitalized patients, who arguably have the greatest need to comprehend their health information, yet frequently have cognitive limitations due to stress or illness. The cognitive impairment in our sample (84%) is consistent with previous research describing cognitive impairment (62%-80%) in hospitalized populations,[95–100] and most patients with mild-to-moderate cognitive impairment still make their own healthcare decisions.[101] Future work on displaying health information should therefore include hospitalized patients and consider cognitive status.

The importance of conducting this study in the real-world–like setting cannot be understated. In the controlled laboratory-based setting, participants can review their health information at their leisure, without distractions. In the real-world setting, participants may experience competing priorities or distractions. The reality in the clinical and home environment is that most patients will not sit for an hour and read hospital materials. Instead, they will briefly review materials, and either understand them or not. This "immediate" comprehension is likely different from comprehension after careful consideration in a controlled setting, with greater emphasis on rapid communication and easy learnability.

In our study, we found participants paid greatest attention to visuals, and much less attention to additional information cues, like text. This somewhat contradicts the recent informatics literature suggesting more information cues always improves interpretation.[58,102] The contradiction may result from evaluating real-world "immediate" comprehension versus laboratory-based "extended" comprehension. More research is needed to explore "immediate" comprehension.

Limitations

Our visualizations compared PRO results at 2 time points for simplicity. Results may not generalize to 3 or more time points, particularly unevenly spaced ones.[103] The number line did not perform as well as expected, potentially due to formatting. In the future, the following formatting strategies may improve number line performance: (1) vertical rather than horizontal comparison of 2 number lines, (2) showing 2 time points on 1 number line rather than 2 separate number lines. We assessed comprehension for 3 symptom constructs: physical function, depression, and fatigue. Results may not generalize to other symptom constructs, and certain constructs may require unique visualizations. Future work should offer patients a broader variety of visualizations and strategies for interpreting results.

Our sample demonstrated high levels of cognitive impairment (84%), consistent with being hospitalized.[95–99] Given the urgent information needs of hospitalized patients, and their exclusion from previous studies, our inclusion of hospitalized, cognitively impaired participants is an important strength. However, results may not generalize to non-hospitalized, non-cognitively-impaired populations. Expert review, and not a previously validated phenotyping algorithm, was used to identify heart failure patients. Our participants represented a wide range of demographics; however, a broader sample, including non-English speakers, could further increase our generalizability. Many potential participants were unavailable during recruitment or declined to participate, and findings may differ in non-participants, although our purposive sampling strategy prevented differential attrition based on age, gender, and race. Future work should evaluate objective comprehension in a larger sample, to determine whether each non-significant trend is an artifact or reality, particularly for the number line.

CONCLUSION

Often, electronic systems display longitudinal PROs to patients without contextualizing or interpreting them. We found that visual analogies can enhance patients' comprehension of longitudinal PROs, compared with text-only or graph interpretations. The results support using visualizations rather than text to display longitudinal PROs, but caution against relying on graphs. Additionally, we found discrepancies between comprehension and preferences, which suggests factors other than comprehension influence preferences. Future researchers should assess comprehension rather than preferences to determine efficacy.

FUNDING

This work was supported by National Institute of Nursing Research R00NR016275 (PI: RMMC), National Library of Medicine F31LM054013 (PI: LVG), and Heart Failure Society of America (PI: RMMC).

AUTHOR CONTRIBUTIONS

MRT and LVG are co-first authors. RMMC conceptualized and obtained federal funding for the study. LVG conceptualized and designed the visualizations. MRT, LVG, and RMMC collaboratively refined the study design. MRT designed and refined the recruitment protocol. MRT and ACM conducted recruitment. MRT, LVG, ACM, and DB performed the analyses. LVG drafted the manuscript, and all authors contributed to refining all sections and critically editing the article.

SUPPLEMENTARY MATERIAL

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

ACKNOWLEDGMENTS

Thank you to Jessica Ancker, Suzanne Bakken, George Hripcsak, David Vawdrey, Natalie Benda, and Marianne Sharko for your support regarding the design, analysis, and manuscript. We would also like to acknowledge the support of the dozens of healthcare providers and patients who helped carry out this project.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Kotronoulas G, Kearney N, Maguire R, *et al*. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol* 2014; 32 (14): 1480–501.
2. Basch E, Deal AM, Kris MG, *et al*. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol* 2016; 34 (6): 557–65.
3. Valderas JM, Kotzeva A, Espallargues M, *et al*. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res* 2008; 17 (2): 179–93.
4. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. *J Eval Clin Pract* 2006; 12 (5): 559–68.
5. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract* 1999; 5 (4): 401–16.
6. Wagle N. Implementing patient-reported outcome measures. *NEJM Catal* 2017 Oct 12
7. Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice–adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol* 2013; 66 (8): S12–20.
8. Wu A, Jensen R, Salzberg C, Snyder C. Advances in the use of patient reported outcome measures in electronic health records. Presented at PCORI National Workshop to Advance the Use of PRO Measures in Electronic Health Records; November 19–20, 2013; Atlanta, GA.
9. Devlin NJ, Appleby J. *Getting the Most Out of PROMS: Putting Health Outcomes at the Heart of NHS Decision-Making*. London, United Kingdom: The King's Fund; 2010.
10. Snyder CF, Aaronson NK, Choucair AK, *et al*. Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations. *Qual Life Res* 2012; 21 (8): 1305–14.

11. Au H-J, Ringash J, Brundage M, *et al*. Added value of health-related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG. *Expert Rev Pharmacoecon Outcomes Res* 2010; 10 (2): 119–28.
12. Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res* 2009; 18 (1): 115–23.
13. Velikova G, Booth L, Smith AB, *et al*. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 2004; 22 (4): 714–24.
14. Santana M-J, Feeny D, Johnson JA, *et al*. Assessing the use of health-related quality of life measures in the routine clinical care of lung-transplant patients. *Qual Life Res* 2010; 19 (3): 371–9.
15. Detmar SB, Muller MJ, Schornagel JH, Wever LDV, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA* 2002; 288 (23): 3027–34.
16. Berry DL, Blumenstein BA, Halpenny B, *et al*. Enhancing patient-provider communication with the electronic self-report assessment for cancer: a randomized trial. *J Clin Oncol* 2011; 29 (8): 1029–35.
17. Paul M. Northwestern leads consortium to integrate patient-reported outcomes into electronic health records. October 25, 2016. https://news.northwestern.edu/stories/2016/10/northwestern-leads-consortium-to-integrate-patient-reported-outcomes-into-electronic-health-records/ Accessed April 7, 2019.
18. Baumhauer JF, Dasilva C, Mitten D, Rubery P, Rotondo M. The cost of patient-reported outcomes in medicine. *NEJM Catal* 2018 Jan 25
19. Jensen RE, Snyder CF, Abernethy AP, *et al*. Review of electronic patient-reported outcomes systems used in cancer clinical care. *J Oncol Pract* 2014; 10 (4): e215–22.
20. van Cranenburgh OD, ter Stege JA, de Korte J, *et al*. Patient-reported outcome measurement in clinical dermatological practice: relevance and feasibility of a web-based portal. *Dermatology* 2016; 232 (1): 64–70.
21. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. *CA Cancer J Clin* 2015; 62: 337–47.
22. Grossman LV, Feiner SK, Mitchell EG, Masterson Creber RM. Leveraging patient-reported outcomes using data visualization. *Appl Clin Inform* 2018; 9: 565–75.
23. Hamari J, Koivisto J, Sarsa H. Does gamification work?—a literature review of empirical studies on gamification. In: proceedings of the 2014 47th Hawaii International Conference on System Sciences; January 6–9, 2014; Waikoloa, HI.
24. Cheong C, Cheong F, Filippou J. Quick quiz: a gamified approach for enhancing learning. In: proceedings of the 17th Annual Pacific Asia Conference Information Systems; June 18, 2013; Jeju Island, Korea.
25. Berwick D. Era 3 for medicine and health care. *JAMA* 2016; 315 (13): 1329–30.
26. American College of Physicians. *Improving Health Care Efficacy and Efficiency Through Increased Transparency*. Philadelphia, PA: American College of Physicians; 2017.
27. Baumhauer JF. Patient-reported outcomes-are they living up to their potential? *N Engl J Med* 2017; 377 (1): 6–9.
28. Rudin RS, Bates DW, MacRae C. Accelerating innovation in health IT. *N Engl J Med* 2016; 375 (9): 815–7.
29. Masterson Creber RM, Grossman LV, Ryan B, *et al*. Engaging hospitalized patients with personalized health information: a randomized trial of an inpatient portal. *J Am Med Inform Assoc* 2019; 26 (2): 115–23.
30. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. *J Med Internet Res* 2015; 17 (6): e148.
31. 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.
32. Lorenc T, Petticrew M, Welch V, Tugwell P. What types of interventions generate inequalities? Evidence from systematic reviews. *J Epidemiol Community Health* 2013; 67 (2): 190–3.
33. Hart JT. The inverse care law. *Lancet* 1971; 1 (7696): 405–12.
34. Fekete J-D, van Wijk JJ, Stasko JT, North C. The value of information visualization. In: Kerren A, Stasko JT, Fekete J-D, North C, eds. *Information Visualization: Human-Centered Issues and Perspectives*. Berlin, Germany: Springer-Verlag; 2008: 1–18.
35. Hawley ST, Zikmund-Fisher B, Ubel P, *et al*. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Educ Couns* 2008; 73 (3): 448–55.
36. Solomon J, Scherer AM, Exe NL, *et al*. Is this good or bad? Redesigning visual displays of medical test results in patient portals to provide context and meaning. In: proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. New York, NY: ACM Press; 2016: 2314–20.
37. Zikmund-Fisher BJ, Scherer AM, Witteman HO, *et al*. Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *J Am Med Inform Assoc* 2017; 24 (3): 520–8.
38. Wegier P, Shaffer VA. Aiding risk information learning through simulated experience (ARISE): using simulated outcomes to improve understanding of conditional probabilities in prenatal Down syndrome. *Patient Educ Couns* 2017; 100 (10): 1882–9.
39. Stone ER, Bruine de Bruin W, Wilkins AM, Boker EM, MacDonald Gibson J. Designing graphs to communicate risks: understanding how the choice of graphical format influences decision making. *Risk Anal* 2017; 37 (4): 612–28.
40. Siegrist M, Orlow P, Keller C. The effect of graphical and numerical presentation of hypothetical prenatal diagnosis results on risk perception. *Med Decis Making* 2008; 28 (4): 567–74.
41. Zikmund-Fisher BJ, Exe NL, Witteman HO. Numeracy and literacy independently predict patients' ability to identify out-of-range test results. *J Med Internet Res* 2014; 16 (8): e187.
42. Morrow D, Hasegawa-Johnson M, Huang T, *et al*. A multidisciplinary approach to designing and evaluating electronic medical record portal messages that support patient self-care. *J Biomed Inform* 2017; 69: 63–74.
43. Zikmund-Fisher BJ, Ubel PA, Smith DM, *et al*. Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. *Patient Educ Couns* 2008; 73 (2): 209–14.
44. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Improving understanding of adjuvant therapy options by using simpler risk graphics. *Cancer* 2008; 113 (12): 3382–90.
45. Vogt F, Marteau TM. Perceived effectiveness of stop smoking interventions: Impact of presenting evidence using numbers, visual displays, and different timeframes. *Nicotine Tob Res* 2012; 14 (2): 200–8.
46. Snyder CF, Smith KC, Bantug ET, *et al*. What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. *Cancer* 2017; 123 (10): 1848–59.
47. Brundage MD, Smith KC, Little EA, *et al*. Communicating patient-reported outcome scores using graphic formats: results from a mixed-methods evaluation. *Qual Life Res* 2015; 24 (10): 2457–72.
48. Smith KC, Brundage MD, Tolbert E, *et al*. Engaging stakeholders to improve presentation of patient-reported outcomes data in clinical practice. *Support Care Cancer* 2016; 24 (10): 4149–57.
49. Arcia A, Woollen J, Bakken S. A systematic method for exploring data attributes in preparation for designing tailored infographics of patient reported outcomes. *EGEMS (Wash DC)* 2018; 6 (1): 2.
50. Tolbert E, Brundage M, Bantug E, *et al*. Picture this: presenting longitudinal patient-reported outcome research study results to patients. *Med Decis Making* 2018; 38 (8): 994–1005.
51. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. *Med Decis Making* 2011; 31 (3): 444–57.
52. Sun X, May A. a comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices. *Adv Hum-Comput Interact* 2013; 2013: 1–9.
53. Ambrosy AP, Fonarow GC, Butler J, *et al*. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol* 2014; 63 (12): 1123–33.

54. Benjamin EJ, Blaha MJ, Chiuve SE, *et al*. Heart disease and stroke statistics-2017 update: a report from the American Heart Association. *Circulation* 2017; 135 (10): e146–603.
55. Goldberg RJ, Spencer FA, Szklo-Coxe M, *et al*. Symptom presentation in patients hospitalized with acute heart failure. *Clin Cardiol* 2010; 33 (6): e73–80.
56. MacKenzie IS. Designing HCI experiments. In: *Human-Computer Interaction: An Empirical Research Perspective*. New York, NY: Elsevier; 2013: 157–89.
57. International Organization for Standardization. *International Standard ISO 9186-1:2014(E) Methods for Testing Comprehensibility*. London, United Kingdom: International Organization for Standardization; 2014.
58. Arcia A, Suero-Tejeda N, Bales ME, *et al*. Sometimes more is more: Iterative participatory design of infographics for engagement of community members with varying levels of health literacy. *J Am Med Inform Assoc* 2016; 23 (1): 174–83.
59. Ancker J, Begg MD. Using visual analogies to teach introductory statistical concepts. *Numeracy* 2017; 10 (2): 7.
60. Duit R. On the role of analogies and metaphors in learning science. *Sci Educ* 1991; 75 (6): 649–72.
61. Martin MA. It's like… you know": the use of analogies and heuristics in teaching introductory statistical methods. *J Stat Educ* 2017; 11: 2.
62. Glynn S. Making science concepts meaningful to students: teaching with analogies. In: Mikelskis-Seifert S, ed. *Four Decades of Research in Science Education: From Curriculum Development to Quality Improvement*. Munster, Germany: Waxmann Verlag; 2008: 113–25.
63. Arcia A, Bales ME, William Brown III MC, *et al*. Method for the development of data visualizations for community members with varying levels of health literacy. *AMIA Annu Symp Proc* 2013; 2013: 51–60.
64. Flynn KE, Dew MA, Lin L, *et al*. Reliability and construct validity of PROMIS® measures for patients with heart failure who undergo heart transplant. *Qual Life Res* 2015; 24 (11): 2591–9.
65. Cella D, Yount S, Rothrock N, *et al*. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* 2007; 45 (Suppl 1): S3–11.
66. HealthMeasures. Interpret scores: PROMIS®: meaningful change; 2019. http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/meaningful-change. Accessed May 27, 2019.
67. Lapin B, Thompson NR, Schuster A, Katzan IL. Clinical utility of patient-reported outcome measurement information system domain scales. *Circ Cardiovasc Qual Outcomes* 2019; 12 (1): e004753.
68. Baik D, Reading M, Jia H, Grossman LV, Masterson Creber R. Measuring health status and symptom burden using a web-based mHealth application in patients with heart failure. *Eur J Cardiovasc Nurs* 2019; 18 (4): 325–31.
69. Reading M, Masterson Creber R. Quantifying non-cardiac symptoms in patients with heart failure using PROMIS® measures collected using a mHealth application. *Circulation* 2018; 138: A14199.
70. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007; 8 (2): 53–96.
71. Garcia-Retamero R, Cokely ET. Designing visual AIDS that promote risk literacy: a systematic review of health research and evidence-based design heuristics. *Hum Factors* 2017; 59 (4): 582–627.
72. Raghubir P. Is 1/10 > 10/100? The effect of denominator salience on perceptions of base rates of health risk. *Int J Res Mark* 2008; 25 (4): 327–34.
73. Pew Research Center. Generations and age; 2019. http://www.pewresearch.org/topics/generations-and-age/ Accessed May 27, 2019.
74. Masterson Creber R, Prey J, Ryan B, *et al*. Engaging hospitalized patients in clinical care: study protocol for a pragmatic randomized controlled trial. *Contemp Clin Trials* 2016; 47: 165–71.
75. Nasreddine ZS, Phillips NA, Bédirian V, *et al*. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005; 53 (4): 695–9.
76. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004; 36 (8): 588–94.
77. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, *et al*. Measuring numeracy without a math test: development of the subjective numeracy scale. *Med Decis Making* 2007; 27 (5): 672–80.
78. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the subjective numeracy scale: effects of low numeracy on comprehension of risk communications and utility elicitations. *Med Decis Making* 2007; 27 (5): 663–71.
79. McNaughton C, Wallston KA, Rothman RL, Marcovitz DE, Storrow AB. Short, subjective measures of numeracy and general health literacy in an adult emergency department. *Acad Emerg Med* 2011; 18 (11): 1148–55.
80. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a short, 3-item version of the subjective numeracy scale. *Med Decis Making* 2015; 35 (8): 932–6.
81. 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.
82. Kubo SH, Schulman S, Starling RC, *et al*. Development and validation of a patient questionnaire to determine New York heart association classification. *J Card Fail* 2004; 10 (3): 228–35.
83. Arcia A, Grossman LV, George M, *et al*. Modifications to the ISO 9186 method for testing comprehension: successes and lessons learned. In: 2019 IEEE Workshop on Visual Analytics in Healthcare; October 20, 2019; Vancouver, Canada.
84. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med* 1997; 127 (11): 966–72.
85. Zikmund-Fisher BJ, Witteman HO, Dickson M, *et al*. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. *Med Decis Making* 2014; 34 (4): 443–53.
86. Garcia-Retamero R, Cokely ET. Effective communication of risks to young adults: using message framing and visual aids to increase condom use and STD screening. *J Exp Psychol Appl* 2011; 17 (3): 270–87.
87. Brewer NT, Gilkey MB, Lillie SE, Hesse BW, Sheridan SL. Tables or bar graphs? presenting test results in electronic medical records. *Med Decis Making* 2012; 32 (4): 545–53.
88. Lakoff G, Johnson M. *Metaphors We Live by*. Chicago, IL: University of Chicago Press; 1980.
89. Donnelly CM, McDaniel MA. Use of analogy in learning scientific concepts. *J Exp Psychol Learn Mem Cogn* 1993; 19 (4): 975–87.
90. Braasch JLG, Goldman SR. The role of prior knowledge in learning from analogies in science texts. *Discourse Process* 2010; 47 (6): 447–79.
91. Elting LS, Martin CG, Cantor SB, Rubenstein EB. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. *BMJ* 1999; 318 (7197): 1527–31.
92. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. *J Am Med Inform Assoc* 2006; 13 (6): 608–18.
93. Tractinsky N. Visual aesthetics. *The Encyclopedia of Human-Computer Interaction, 2nd ed*. Aarhus, Denmark: The Interaction Design Foundation; 2013. https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/visual-aesthetics Accessed April 7, 2019.
94. Tuch AN, Presslaber EE, Stocklin M, Opwis K, Bargas-Avila JA. The role of visual complexity and prototypicality regarding first impression of websites: working towards understanding aesthetic judgments. *Int J Hum Comput Stud* 2012; 70 (11): 794–811.
95. Cameron J, Worrall-Carter L, Page K, Stewart S, Ski CF. Screening for mild cognitive impairment in patients with heart failure: Montreal cognitive assessment versus mini mental state exam. *Eur J Cardiovasc Nurs* 2013; 12 (3): 252–60.
96. Sampson EL, Blanchard MR, Jones L, Tookman A, King M. Dementia in the acute hospital: prospective cohort study of prevalence and mortality. *Br J Psychiatry* 2009; 195 (1): 61–6.

97. Foster R, Walker S, Brar R, *et al*. Cognitive impairment in advanced chronic kidney disease: The Canadian Frailty Observation and Interventions Trial. *Am J Nephrol* 2016; 44 (6): 473–80.
98. Harkness K, Heckman GA, Akhtar-Danesh N, *et al*. Cognitive function and self-care management in older patients with heart failure. *Eur J Cardiovasc Nurs* 2014; 13 (3): 277–84.
99. Ouellet M-C, Sirois M-J, Beaulieu-Bonneau S, *et al*. Is cognitive function a concern in independent elderly adults discharged home from the emergency department in Canada after a minor injury? *J Am Geriatr Soc* 2014; 62 (11): 2130–5.
100. Leto L, Feola M. Cognitive impairment in heart failure patients. *J Geriatr Cardiol* 2014; 11: 316–28.
101. Feinberg LF, Whitlatch CJ. Decision-making for persons with cognitive impairment and their family caregivers. *Am J Alzheimers Dis Other Demen* 2002; 17 (4): 237–44.
102. Tao D, Yuan J, Qu X. Presenting self-monitoring test results for consumers: the effects of graphical formats and age. *J Am Med Inform Assoc* 2018; 25 (8): 1036–46.
103. Aigner W, Miksch S, Müller W, Schumann H, Tominski C. Visualizing time-oriented data-a systematic view. *Comput Graph* 2007; 31 (3): 401–9.