Research and Applications

Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results

Brian J Zikmund-Fisher,[1,2,3] **Aaron M Scherer,**[3,4] **Holly O Witteman,**[5,6,7] **Jacob B Solomon,**[3] **Nicole L Exe,**[3] **Beth A Tarini,**[8] **and Angela Fagerlin**[9]

[1]Department of Health Behavior and Health Education, University of Michigan, Ann Arbor, MI, USA, [2]Department of Internal Medicine, University of Michigan, [3]Center for Bioethics and Social Sciences in Medicine, University of Michigan, [4]Department of Internal Medicine, University of Iowa, Iowa City, IA, USA, [5]Department of Family and Emergency Medicine, Laval University, Quebec City, Quebec, Canada, [6]Office of Education and Professional Development, Faculty of Medicine, Laval University, [7]Population Health and Optimal Health Practices Research Unit, Research Centre of the CHU de Québec-Université Laval, [8]Department of Pediatrics, University of Iowa and [9]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA

Corresponding Author: Brian J Zikmund-Fisher, Department of Health Behavior and Health Education, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA. Phone: 734-936-9179. Fax: 734-763-7379. E-mail: bzikmund-d@umich.edu

Received 1 September 2016; Revised 2 November 2016; Accepted 21 November 2016

ABSTRACT

Objective: Most electronic health record systems provide laboratory test results to patients in table format. We tested whether presenting such results in visual displays (number lines) could improve understanding.

Materials and Methods: We presented 1620 adults recruited from a demographically diverse Internet panel with hypothetical results from several common laboratory tests, first showing near-normal results and then more extreme values. Participants viewed results in either table format (with a "standard range" provided) or one of 3 number line formats: a simple 2-color format, a format with diagnostic categories such as "borderline high" indicated by colored blocks, and a gradient format that used color gradients to smoothly represent increasing risk as values deviated from standard ranges. We measured respondents' subjective sense of urgency about each test result, their behavioral intentions, and their perceptions of the display format.

Results: Visual displays reduced respondents' perceived urgency and desire to contact health care providers immediately for near-normal test results compared to tables but did not affect their perceptions of extreme values. In regression analyses controlling for respondent health literacy, numeracy, and graphical literacy, gradient line displays resulted in the greatest sensitivity to changes in test results.

Discussion: Unlike tables, which only tell patients whether test results are normal or not, visual displays can increase the meaningfulness of test results by clearly defining possible values and leveraging color cues and evaluative labels.

Conclusion: Patient-facing displays of laboratory test results should use visual displays rather than tables to increase people's sensitivity to variations in their results.

Key words: decision making, education of patients, electronic health record, computer graphics, clinical laboratory information systems

BACKGROUND AND SIGNIFICANCE

Patient access to electronic health record (EHR) systems has increased dramatically in recent years,[1] and most of these allow patients to view their laboratory test results outside of clinical consultations. As a result, increasing numbers of patients are now viewing medical test result data that they either have never seen before or have only seen in the context of a clinical visit with a health care provider who could explain and interpret it. Patients value this

access to information,[2–4] which is one component of a larger trend to encourage and facilitate greater patient involvement in medical decision making and health self-management.[5–7] Yet simply having access to test results is insufficient to ensure that patients can use that information to improve their health or their care.

Currently, most EHR patient portals present test results in tabular formats that are similar to those used to report results to clinicians.[8] These tables are difficult to interpret, and patients (especially those with lower numeracy or literacy skills) face significant barriers when trying to identify whether test results are outside of the standard reference range.[9]

Furthermore, the clinical importance of having results outside of the standard range is generally undefined in tabular formats. While patients do not have primary responsibility for identifying and acting upon laboratory findings (clinicians are generally expected to review and act upon urgent results), they will of course attempt to interpret and draw conclusions from whatever information is made available to them. The more ambiguous test result displays are in terms of clinical meaning, the more patients will wonder how bad their results are, which could lead them to draw erroneous inferences. Test values represent patients' short-term risk (eg, of bleeding if platelet counts are low) and long-term risk (eg, of diabetes-related complications if hemoglobin A1c remains high). If patients view every out-of-range test value similarly, they are likely to overreact to slightly elevated or reduced values that are not urgent (especially if the ordering clinician does not provide an interpretive note) and fail to recognize results that are urgent.

However, the fact that patients currently have difficulty interpreting test results does not imply that they would not if the data were presented in more effective formats.[10] Visual displays often help people understand data such as risk statistics,[11–14] and visual displays that show health care providers multiple test results appear to speed their interpretation,[15] although different displays appear to work better for different use cases.[16] When only one result is being communicated at a time, number line graph displays are used in many applications to enable users to see where a single value falls in the range of possible values (eg, the page position within the total pages of an e-book). Thus, presenting laboratory test results in number line graphs rather than tables may increase the usability of these data.[17] Such designs need to be studied rigorously to evaluate their effects on patient comprehension, risk perception, and activation. To our knowledge, no experimental research has tested whether presenting test results in visual number line graph formats instead of tables alters how people interpret test values.

We conducted an online survey in which respondents imagined receiving multiple laboratory test results through an EHR patient portal. Our primary research question was: To what extent do different visual displays help people discriminate between test results that do or do not require urgent action? By varying whether participants viewed their test results in either number line graph or tabular formats as well as the test values themselves, we provide empirical evidence of how display format influences people's understanding of what their laboratory test results mean.

MATERIALS AND METHODS

Participants

We recruited a stratified random sample of US adults from a panel of Internet users administered by Survey Sampling International (SSI), which recruits panel members through various opt-in methods. SSI uses a probability-weighted random process to identify which panel members should receive which surveys based on sample requirements. To ensure demographic diversity and offset variations in response rates, we established quotas based on respondent age and race (thereby approximating the distributions of these characteristics in the US population). The sampling algorithm continued to route SSI participants to the survey until all quotas were achieved. We recruited over a one-week period in May 2016. Upon completion, participants were entered into instant-win contests and regular drawings administered by SSI for modest prizes.

Design and procedure

Participants were asked to imagine that they had recently visited their doctor's office to discuss medications and had undergone a set of blood tests shortly thereafter. They were then asked to imagine that they were viewing the results of these tests on an online EHR portal.

Our experimental design focused on participants' reactions to 3 specific test results (platelet count, alanine aminotransferase [ALT], and serum creatinine) selected because they varied in the size of established standard ranges (relative to possible test values) and in the level of patient risk associated with varying degrees of deviation from standard ranges. Each test was reported initially as slightly outside the standard range and then as a more extreme result (ie, further from the standard range). Specifically, all participants first viewed a platelet count of 135×10^9/L and then 25×10^9/L, an ALT value of 80 U/L and then 360 U/L, and a creatinine value of 2.2 mg/dL and then 3.4 mg/dL.

Our primary between-subject experimental manipulation was to vary which of 4 formats (Figure 1) was used to display the test results. One group (Table condition) viewed results presented in a table format that included the test result value and standard range in number form, as well as the appropriate units. This is similar to what is typically presented in EHRs such as Epic. All remaining participants viewed visual displays (described in detail below) that placed the test value on a horizontal number line graph, with the numerical values corresponding to the ends of the number line and the ends of the standard range clearly marked and labeled, as shown in Figure 2. These ranges were chosen in consultation with clinicians to represent a set of values that would include all but the most unusual outlier values.

We tested 3 different designs in order to explore to what degree any differences in patient reactions were attributable to the use of number line visual displays in general vs specific design features (eg, colors or labels) of the displays (Figure 2). Participants in the Simple Line condition viewed displays that were colored solid gray except for a green range labeled "standard range." Participants in the Block Line condition viewed number lines that had areas outside the standard range divided into solid-color blocks (using a stoplight red-yellow-green color palette) that indicated diagnostic categories such as "very low" (red) and "borderline high" (yellow). This design provided additional reference points to increase the evaluability of (likely unfamiliar) results[18,19] and supported increased gist (vs verbatim) processing to facilitate meaning derivation.[20] The cutoffs for these categories were determined in consultation with clinician team members as representing plausible descriptions of the level of risk associated with different degrees of deviation from the standard range. Lastly, participants in the Gradient Line condition viewed number lines similar to those in the Block Line condition, except that the

Table:

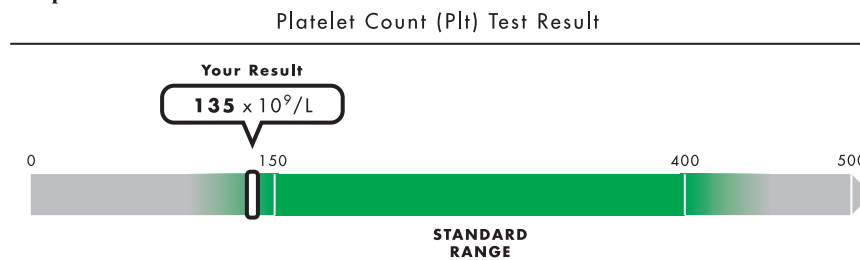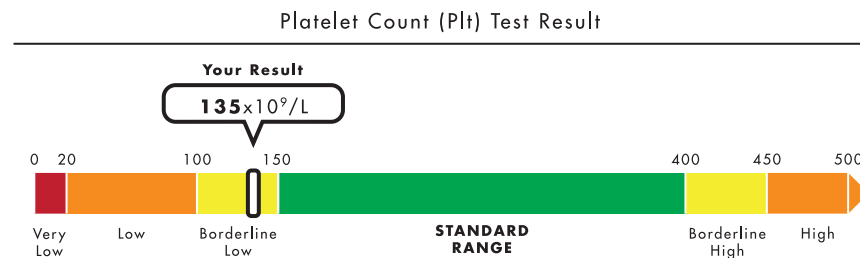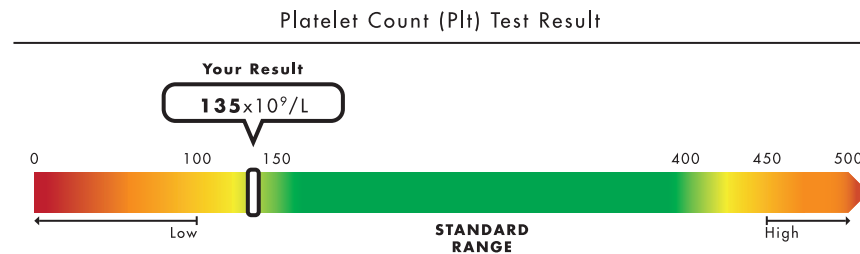| Test	Your Result	Standard Range	Units
Platelet Count (PLT)	135	150-400	$x10^9/L$

Simple Line:

Block Line:

Gradient Line:

Figure 1. Visual display formats used in this study

transitions between categories were graphically smoothed by using color gradients to more accurately represent the continuous nature of the underlying value-to-risk relationship. In addition, category labels were replaced by 2 small arrows labeled "low" and "high." This design therefore represents a compromise between simplicity and the volume of information provided in the Block Line design.

Outcome measures

Perceived test urgency

Because our primary research question focused on increasing patients' ability to discriminate between urgent and nonurgent test results, our primary outcome measure was the difference between participants' subjective sense of urgency when given test results that were slightly and extremely outside the standard range. For each of the 3 focal tests, we asked 2 questions: "How alarming does this re-

sult feel to you?" and "How urgent of an issue is this result?" Both questions utilized a 6-point Likert scale response option, with 1 = not at all and 6 = very.

Behavioral intentions

We also included a question about participants' specific behavioral intentions in response to the test. Specifically, we asked, "Which of the following best describes what you would do in response to your [test name] test result?" Response options were "Nothing," "Talk to your doctor about this test result at your next regular appointment," "Ask to see your doctor at the first available appointment," "Go to a hospital or your doctor's office tomorrow," "Go to a hospital as soon as you can get free later today," and "Go to a hospital immediately." The test result display remained visible during all
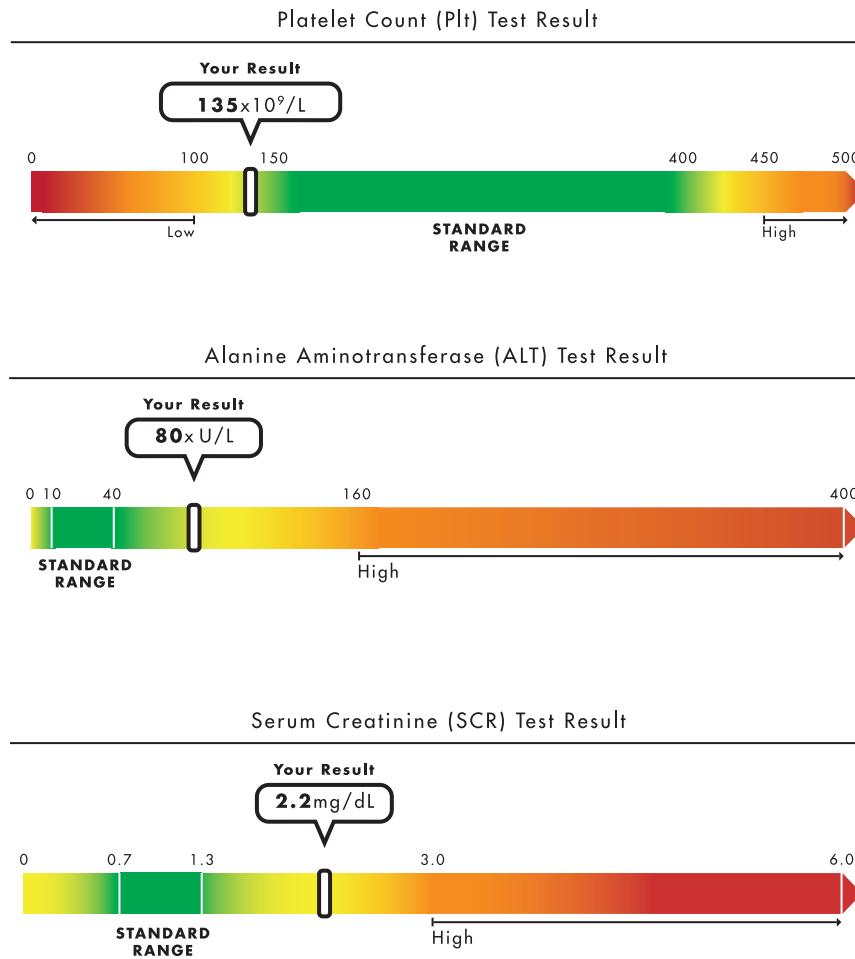
Figure 2. Visual displays of the 3 near-normal test values on Gradient Lines

questioning so that all questions measured data interpretation, not recall.

Display format preferences

At the end of the survey, we asked 4 questions to measure user perceptions of the data formats (table or line graph). Participants rated "In your opinion, how well did these images describe the test results?" on a 5-point scale from "not at all well" to "extremely well"; "How helpful were these images in helping you to understand the test results?" on a 5-point scale from "not at all helpful" to "extremely helpful"; "If you were receiving laboratory test results in real life, would you like to see the test results presented using this type of image?" on a 5-point scale from "definitely no" to "definitely yes"; and "How much would you trust what these images are telling you about your health?" on a 5-point scale from "do not trust at all" to "trust completely."

Individual difference measures

In addition to completing standard demographic questions, participants also completed 4 individual difference measures that we hypothesized might affect their ability to interpret test result tables or graphs. First, because ample evidence exists that even highly educated adults can have poor numeracy skills,[21–23] all study participants completed the Subjective Numeracy Scale.[24] The Subjective Numeracy Scale measures perceived quantitative ability and preference for receiving information in numerical form (range: 1–6; higher values indicate greater subjective numeracy) and has previously been shown to correlate with the ability to recall and comprehend textual and graphical risk communications.[25,26] Second, participants completed Chew's screening question for health literacy ("How confident are you filling out forms by yourself?," where 1 = not at all confident and 5 = extremely confident), which has been validated and shown to be highly correlated with the Rapid Estimate of Adult Literacy in Medicine and Short Test of Functional Health Literacy Assessment.[27] Third, participants also completed 6 questions (numbers 5–9, 11, and 13) from Galesic and Garcia-Retamero's graphical literacy scale.[28] We used the total number of correct answers (0–6) as an abbreviated measure of graphical literacy (reduced due to this scale's significant time demand on respondents). Fourth, respondents answered "How familiar are you with the medical test results like the ones discussed in this survey?" on a scale from 1 = not at all familiar to 5 = extremely familiar.

Data management

All data were collected anonymously using the Qualtrics® online survey platform. Participants were identified and prevented from taking the survey multiple times via unique identification numbers provided by SSI within the redirected URL. The design, sampling

process, data management procedures, and outcome measures received exempt status approval from the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board.

Statistical analyses

Perceptions of alarm and urgency for each test were highly correlated ($r = 0.83$–0.90). As a result, we created an aggregate scale of perceived urgency using the average of both questions for each test (Cronbach's $\alpha = 0.91$–0.95). To quantify sensitivity between values that are near normal and do not require immediate medical attention and values that are extreme and do require immediate medical attention, we then created an urgency difference score equal to perceived urgency (extreme value) minus perceived urgency (near-normal value) as the primary outcome variable of our analysis.

In addition, because the categories included in the behavioral intentions variable are not necessarily mutually exclusive (eg, one might describe waiting until a future appointment as doing "nothing"), for analysis purposes we divided that variable into 2 groups corresponding to willingness to wait (first 2 categories) and taking some form of immediate action (last 4 categories). We also combined the 4 questions assessing user perceptions of the display formats into a single, highly reliable measure of display format preferences (Cronbach's $\alpha = 0.88$).

We report descriptive statistics for the urgency difference score, willingness to wait, and user preferences across the 4 display formats. We also report the results of linear regression analyses predicting the urgency difference score, using display format as well as respondent numeracy, health literacy, graphical literacy, and familiarity with medical tests as continuous variables. All analyses were performed using Stata 14[29] and all tests of significance were 2-sided and used $\alpha = 0.05$.

RESULTS

Sample characteristics

Out of 1822 people who began the survey, 1621 completed it (an 89% completion rate). In addition, 1 response was dropped due to a reported age <18 years old. Table 1 reports sample demographic characteristics and the distribution of scores for the health literacy, numeracy, and graphical literacy measures among the remaining 1620 respondents.

Perceived test urgency

As shown in Table 2, perceptions of perceived urgency for the extreme values were universally high, regardless of which display format was shown. However, perceptions of the near-normal values varied substantially across formats, especially for the ALT and serum creatinine tests. The pattern is consistent: participants who saw their near-normal values in a table display rated those results as most urgent, while those who saw a gradient line display perceived the results as least urgent.

The within-subject urgency difference score is similarly smallest for the table conditions and largest for the gradient line conditions. Full distributions (violin plots) of this variable for each of the 3 tests are presented in Figure 3. The notable finding is that the modal urgency difference score for table displays is 0 for the displays of ALT and serum creatinine results. As shown in Table 2, over 56% of people who saw tables gave precisely the same ratings of urgency when ALT = 80 U/L vs ALT = 360 U/L, and 44% gave identical responses

when serum creatinine = 2.2 mg/dL vs = 3.4 mg/dL. By contrast, when those same test results were presented in any of the line graph formats, most participants had positive urgency difference scores, suggesting a more nuanced understanding of the result and its meaning for their health.

To assess the significance of these patterns, we conducted linear regression analyses of the urgency difference scores, controlling for individual differences in health literacy, subjective numeracy (Cronbach's $\alpha = 0.84$), and graphical literacy (mean correct answers = 3.76 out of 6). Table 3 shows that both the gradient line and simple line displays resulted in significantly greater sensitivity to the change in test result for all 3 laboratory tests. The effect of display format was notably larger for the ALT test than for the other 2 tests. In addition, the coefficients for the Gradient Line condition were significantly larger than those for the Block Line condition for all 3 tests (all $P \leq .002$) and the coefficient for the Simple Line condition for the ALT test (F [1,1595] = 23.23, $P < .001$).

When comparing the effects of the individual difference measures using a forward selection process, increased graphical literacy had the strongest relationship with degree of sensitivity, adding substantially to model fit when compared to a model with only display format (platelets $\Delta R^2 = 0.146$, $P < .001$; ALT $\Delta R^2 = 0.080$, $P < .001$; serum creatinine $\Delta R^2 = 0.040$, $P < .001$). It was also the strongest predictor in the final models shown in Table 3. Higher scores on the health literacy screening question also significantly predicted higher urgency difference scores (ΔR^2 vs model with display format and graphical literacy only: platelets $\Delta R^2 = 0.016$, $P < .001$; ALT $\Delta R^2 = 0.013$, $P < .001$; serum creatinine $\Delta R^2 = 0.012$, $P < .001$). Adding subjective numeracy to graphical literacy and health literacy had little effect ($\Delta R^2 < 0.004$ in all regressions), and familiarity with medical tests was not correlated with urgency difference scores.

Behavioral intentions

When respondents were presented with the near-normal platelet test result of 135×10^9/L, there were no significant differences across display formats in their stated willingness to wait (either by doing nothing or waiting until the next scheduled appointment). However, participants who viewed their near-normal ALT and serum creatinine test results in table format were significantly less willing to wait than participants who received the same results in any of the number line displays: ALT (table = 26.9% vs simple line = 44.2%, blocks line = 42.0%, gradient line = 51.8%; χ^2 (3) = 55.25, $P < .001$; serum creatinine (table = 35.5% vs simple line = 43.3%, blocks line = 46.5%, gradient line = 47.7%; χ^2 (3) = 15.20, $P < .001$).

Display format preferences

Mean preference ratings for tables ($M = 3.51$, SD = 1.09) were substantially lower than for the number line displays (simple line: $M = 3.62$, SD = 0.91; blocks line: $M = 3.76$, SD = 0.92; gradient line: $M = 3.68$, SD = 0.91). The overall pattern of variation was significant by one-way analysis of variance (F [3, 1614] = 5.02, $P = .002$), but pairwise comparisons identified only the table vs blocks line comparison as individually significant ($P < .001$).

DISCUSSION

Interpreting laboratory test results is difficult for most people.[9] Furthermore, our results show that when such results are presented in

Table 1. Sample demographics (*n* = 1620)

Characteristic	Category	Frequency (%)	Mean (SD)
Age			48.9 (15.7)
Gender	Male	768 (47.5)	
	Female	846 (52.3)	
	Transgender	4 (0.3)	
Ethnicity	Hispanic (any race)	205 (12.7)	
Race[a]	White	1252 (77.4)	
	African-American	211 (13.0)	
	All other	91 (7.0)	
Education	<High school	30 (1.9)	
	High school only	260 (16.1)	
	Some college/trade	547 (33.8)	
	Bachelor's degree	502 (31.0)	
	Master's/doctorate	279 (17.2)	
Health Literacy	1	29 (1.8)	4.24 (0.96)
	2	68 (4.2)	
	3	213 (13.2)	
	4	482 (30.0)	
	5	817 (50.8)	
Subjective Numeracy Scale	1.00–1.99	26 (1.6)	4.47 (0.97)
	2.00–2.99	104 (6.4)	
	3.00–3.99	288 (17.8)	
	4.00–4.99	622 (38.5)	
	5.00–5.99	528 (32.7)	
	6.00	47 (2.9)	
Graphical Literacy Scale	0	91 (5.7)	3.76 (1.83)
	1	137 (8.5)	
	2	212 (13.2)	
	3	226 (14.0)	
	4	249 (15.5)	
	5	354 (22.0)	
	6	342 (21.2)	
Familiarity with Medical Tests	1	167 (10.4)	3.31 (1.18)
	2	197 (12.3)	
	3	469 (29.2)	
	4	519 (32.3)	
	5	254 (15.8)	

Note: Reports results only for those respondents who completed each question or measure.

[a]Respondents could mark more than one race.

table formats, many people perceive no difference between values that represent minor nonurgent deviations and those that are more clinically concerning. But our results also demonstrate that using even the simplest number line graphics instead of tables to visually represent test results can decrease perceptions of urgency about values near the standard range and therefore increase sensitivity to variations in test values.

The most powerful visual cue that our number line displays provide patients is the concrete placement of test results in a visual space of possible values. Patients thus literally see whether their values are high, low, or in the middle, and they construct cognitive and emotional responses to the test results accordingly. Hence, visual formats are powerful tools that can be designed to shape patient perceptions in ways that facilitate appropriate response, either action (in the case of clinically urgent values) or inaction (in the case of values that are appropriately handled through existing interactions or processes).

Differences in perception of test urgency and behavioral motivation were particularly apparent for ALT and serum creatinine tests, for which values can deviate substantially from the standard range

(to varying degrees) before the patient faces immediate medical consequences or risks. Thus, these are the situations in which patients viewing test results without clinician guidance may be most likely to become unnecessarily concerned near-normal values. While the specificity of the observed effects implies that the choice of display format may be less critical for some types of laboratory tests (ie, those for which deviations quickly become medically concerning) vs others, visual displays are preferred over tables and were at least as effective as table formats for all tests. Given that a simple substitution of well-designed visual display formats appears to reduce such problems at minimal cost, we believe our results provide justification for their adoption.

Several of our designs focused on increasing the evaluability of test result data for patients by providing meaning-rich cues and reference standards to anchor the "how does this compare to X" comparative process that people naturally use to derive meaning from data.[18,30,31] In 2 of the formats we tested in this study, the displays included additional contextual cues in the form of either a color gradient and clearly marked "low" and "high" regions in the Gradient Line condition or distinct colored blocks labeled with specific terms

Table 2. Means (standard deviations) of urgency score and urgency difference score ratings, by display format

	Near-Normal Test Value	Extreme Test Value	Within-Subject Difference Score	Difference Equals 0 (%)
Platelet Count				
Table	3.95 (1.32)	5.20 (1.22)	1.24 (1.44)	26.5
Simple Line	3.72 (1.38)	5.26 (1.17)	1.54 (1.41)	17.5
Blocks Line	3.94 (1.18)	5.20 (1.09)	1.26 (1.29)	19.0
Gradient Line	3.73 (1.39)	5.30 (1.05)	1.57 (1.62)	15.8
ALT				
Table	4.90 (1.22)	5.26 (1.19)	0.37 (.91)	56.3
Simple Line	4.00 (1.36)	5.44 (1.06)	1.44 (1.57)	21.3
Blocks Line	3.97 (1.13)	5.35 (1.02)	1.38 (1.34)	20.2
Gradient Line	3.56 (1.30)	5.39 (1.03)	1.83 (1.59)	14.8
Serum Creatinine				
Table	4.36 (1.27)	4.74 (1.30)	0.39 (1.03)	43.7
Simple Line	4.09 (1.22)	4.81 (1.08)	0.72 (1.12)	27.7
Blocks Line	3.99 (1.11)	4.58 (1.06)	0.59 (.94)	28.7
Gradient Line	3.91 (1.20)	4.74 (1.08)	0.84 (1.19)	24.0

Note: Urgency scores are the average value of 2 questions asked on a 1–6 scale, where greater numbers represent higher perceived urgency of the test result. Urgency difference score = urgency score (extreme value) − urgency score (near-normal value) for each individual study participant.

such as "very low" and "borderline high" in the Block Line condition. Such evaluative labels have previously been shown to facilitate interpretation of unfamiliar health data.[32] However, our results found consistently lower perceived urgency about near-normal values (and hence greater overall sensitivity) with the Gradient Line design, which had fewer distinct evaluative categories, than with the Blocks Line design. Further research is needed to determine the optimal tradeoff between providing additional contextual information and the need for simplicity (since visual complexity inhibits understanding and use of data[33,34]) in patient-facing displays.

However, perhaps the most important, yet difficult, challenge in implementing visual displays of laboratory test results will be determining the range of values to be shown. For example, should a visual display of ALT include values from 0 to 400 U/L, a wider range (eg, 0–800 U/L), or a narrower range (eg, 0–100 U/L)? Use of a narrower range would make small changes of 10 or fewer units visually larger, implying that such changes are more medically significant. Doing so might be particularly valuable in situations where a patient's attention to (seemingly) small variations in test result values is important for self-management. Yet, narrowing the range of values shown increases the possibility of outlier results that fall off the end of the display. Conversely, extending the range to minimize outliers ensures that the vast majority of test results will fall within a very small range of the display. Nor is dynamic scaling of displays a good solution either, as there is clear evidence that people fail to adjust appropriately for scale changes.[35] The most likely solution will be to select default display ranges that include most, but not all, test values but that nonetheless make clinically significant shifts in test values visually salient.

The generalizability of these findings is primarily limited by our use of a hypothetical scenario. Participants did not receive actual test results relevant to true personal medical conditions. This meant that they lacked the personal relevance that such test results
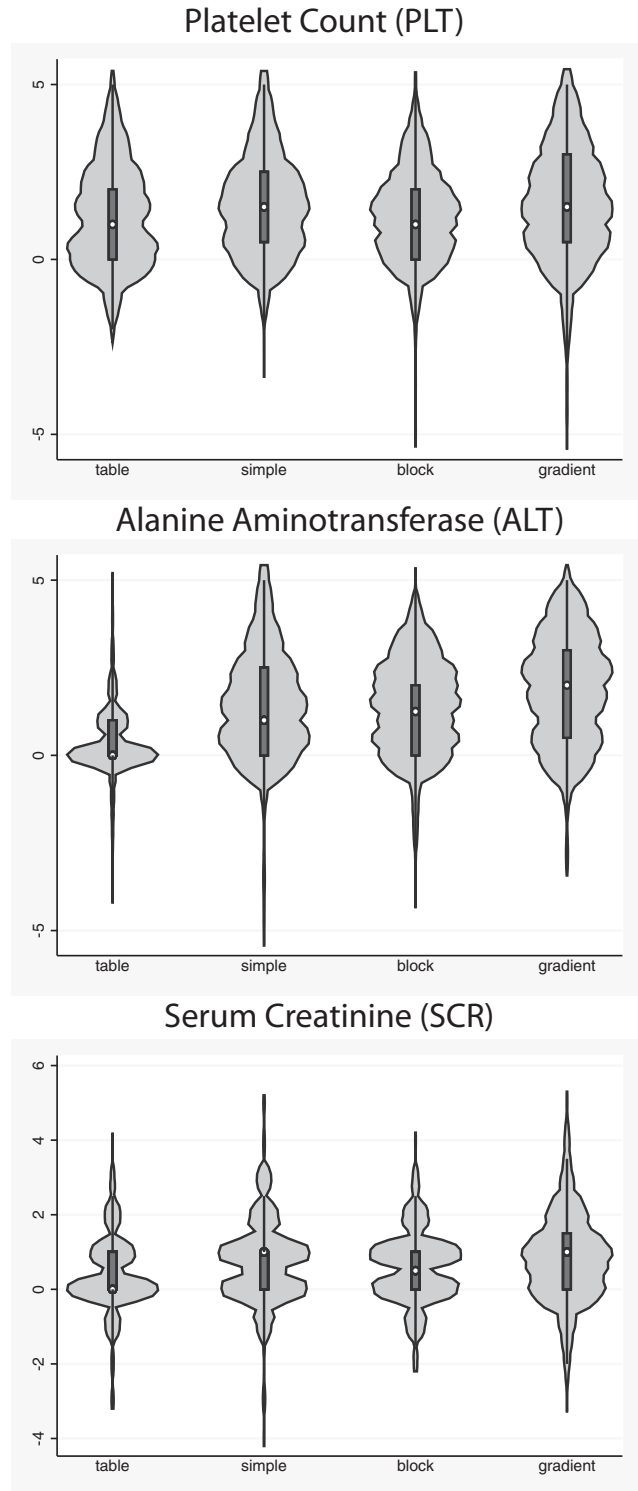
Figure 3. Violin plots of urgency difference scores, by display format

have for patients actually attempting to manage chronic diseases or track illnesses. Nor did we measure the respondents' level of familiarity with the specific tests presented. Nonetheless, our randomized experimental design allowed us to carefully disentangle the effects of test type, test result level, and display format on people's reactions in a way that would be impossible to do with

Table 3. Linear regression results showing predictors of urgency difference score

	Platelets		ALT		Creatinine	
	Coef.	P-value	Coef.	P-value	Coef.	P-value
Display Format						
Table	—	—	—	—	—	—
Simple Line	0.20	.03	1.01	<.001	0.29	<.001
Block Line	−0.00	.99	1.02	<.001	0.18	.02
Gradient Line	0.28	.003	1.44	<.001	0.43	<.001
Graphical Literacy (0–6)	0.29	<.001	0.21	<.001	0.10	<.001
Health Literacy (1–5)	0.23	<.001	0.21	<.001	0.13	<.001
Subjective Numeracy (1–6)	−0.10	.01	−0.07	.06	0.01	.76
Familiarity with Medical Tests	0.00	.94	−0.03	.30	−0.02	.39
Constant	−0.34	.08	−0.84	<.001	−0.47	.002

Notes: Urgency difference score = urgency score (extreme value) – urgency score (near-normal value).

Higher scores represent greater graphical literacy, health literacy, subjective numeracy, and familiarity with medical tests.

patients of varying backgrounds and diagnoses receiving independently varying test results. Furthermore, there are many scenarios in which patients with no prior knowledge of a given laboratory test might view laboratory results in a patient portal. We believe that these are the scenarios that could most easily lead to unnecessarily alarmed patients urgently contacting their health care professionals, and thus we designed the study to identify ways this negative outcome could be avoided.

Governments and hospital systems have recently invested enormous resources to support the adoption of EHR systems. According to a 2015 survey of hospitals, 92% offered patients the ability to view medical records, compared to 43% of hospitals in 2013.[1] These initiatives are increasing patients' ability to directly access and view laboratory test results and have been justified by the concept that patients will be able to translate access to test results into better disease self-management and better preparation for clinic visits. Yet all these potential benefits depend on patients being able to interpret the results they are given. They depend on patients' gist interpretations being appropriately sensitive to variations in what the test data show. The present practice of providing numerical test result values in EHRs in simple table formats is patently insufficient to enable many patients to achieve meaningful use. Systematic, user-centered design that draws upon existing knowledge of the psychology of information evaluability is needed before our investment in patient access to test results yields the health outcome returns of which they are potentially capable.

Acknowledgments

REFERENCES

1. American Hospital Association. *Individuals' Ability to Electronically Access Their Hospital Medical Records, Perform Key Tasks is Growing [Internet].* Washington, DC: American Hospital Association; 2016. http://www.aha.org/research/reports/tw/16jul-tw-healthIT.pdf. Accessed October 29, 2016.
2. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. *Int J Med Inf.* 2002;68(1–3):113–27.
3. Keselman A, Slaughter L, Arnott-Smith C, *et al.* Toward consumer-friendly PHRs: patients' experience with reviewing their health records. *AMIA Annu Symp Proc.* 2007;2007:399–403.
4. Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record: a quantitative study. *Fam Pract.* 2011;28(2):188–94.
5. Laine C, Davidoff F. Patient-centered medicine: a professional evolution. *J Am Med Assoc.* 1996;275(2):152–56.
6. Bernat JL, Peterson LM. Patient-centered informed consent in surgical practice. *Arch Surg.* 2006;141:86–92.
7. Carman KL, Dardess P, Maurer M, *et al.* Patient and family engagement: a framework for understanding the elements and developing interventions and policies. *Health Aff (Millwood).* 2013;32(2):223–31.
8. Krist AH, Woolf SH. A vision for patient-centered health information systems. *JAMA* 2011;305(3):300–01.
9. 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.
10. O'Kane M, Freedman D, Zikmund-Fisher BJ. Can patients use test results effectively if they have direct access? *BMJ.* 2015;350:h673.
11. Tufte ER. The Visual Display of Quantitative Information, *vol. 2.* Cheshire, CT: Graphics Press; 2001.
12. Garcia-Retamero R, Dhami MK. Pictures speak louder than numbers: on communicating medical risks to immigrants with limited non-native language proficiency. *Health Expect.* 2011;14 (Suppl 1):46–57.
13. 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.

14. Hawley ST, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Educ Couns.* 2008;73(3): 448–55.

15. Bauer DT, Guerlain S, Brown PJ. The design and evaluation of a graphical display for laboratory data. J Am Med Inform Assoc. 2010;17(4):416–24.

16. Torsvik T, Lillebo B, Mikkelsen G. Presentation of clinical laboratory results: an experimental comparison of four visualization techniques. *J Am Med Inform Assoc.* 2013;20(2):325–31.

17. Leckart S. The Blood Test Gets a Makeover [Internet]. *Wired.* 2010. http://www.wired.com/magazine/2010/11/ff_bloodwork/. Accessed January 5, 2013.

18. Hsee CK, Blount S, Loewenstein GF, Bazerman MH. Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychol Bull.* 1999;125(5):576–90.

19. Hsee CK, Zhang J. General evaluability theory. *Perspect Psychol Sci.* 2010;5(4):343–55.

20. Reyna VF. How people make decisions that involve risk: a dual-processes approach. *Curr Dir Psychol Sci.* 2004;13(2):60–66.

21. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. *Med Decis Making.* 2001;21(1): 37–44.

22. Nelson W, Reyna VF, Fagerlin A, Lipkus IM, Peters E. Clinical implications of numeracy: theory and practice. *Ann Behav Med.* 2008;35(3):261–74.

23. Peters E, Hibbard J, Slovic P, Dieckmann N. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Aff (Millwood).* 2007;26(3):741–48.

24. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry H, Smith DM. Measuring numeracy without a math test: Development of the subjective numeracy scale (SNS). *Med Decis Making.* 2007;27(5): 672–80.

25. Zikmund-Fisher BJ, Angott AM, Ubel PA. The benefits of discussing adjuvant therapies one at a time instead of all at once. *Breast Cancer Res Treat.* 2011;129(1):79–87.

26. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the subjective numeracy scale (SNS): effects of low numeracy on comprehension of risk communications and utility elicitations. *Med Decis Making.* 2007;27(5):663–71.

27. Chew LD, Griffin JM, Partin MR, *et al.* Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med.* 2008;23(5):561–66.

28. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. *Med Decis Mak.* 2011;31(3):444–57.

29. Stata Corporation. *Stata Statistical Software.* College Station, TX: Stata Corp.; 2015.

30. Hsee CK. Less is better: when low-value options are valued more highly than high-value options. *J Behav Decis Mak.* 1998;11(2):107–21.

31. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. "Is 28% good or bad?" evaluability and preference reversals in health care decisions. *Med Decis Mak.* 2004;24(2):142–48.

32. Peters E, Dieckmann NF, Västfjäll D, Mertz CK, Slovic P, Hibbard JH. Bringing meaning to numbers: the impact of evaluative categories on decisions. *J Exp Psychol Appl.* 2009;15(3):213–27.

33. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Improving understanding of adjuvant therapy options by using simpler risk graphics. *Cancer.* 2008;113(12):3382–90.

34. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. A demonstration of "less can be more" in risk graphics. *Med Decis Mak.* 2010;30(6):661–71.

35. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. What's time got to do with it? Inattention to duration in interpretation of survival graphs. *Risk Anal.* 2005;25(3):589–95.