

Numbers Can Be Worth a Thousand Pictures: Individual Differences in Understanding Graphical and Numerical Representations of Health-Related Information

Wolfgang Gaissmaier and Odette Wegwarth
Max Planck Institute for Human Development

David Skopec, Ann-Sophie Müller, and
Sebastian Broschinski
Zurich University of the Arts

Mary C. Politi
Washington University School of Medicine

Objective: Informed medical decision making requires comprehending statistical information. We aimed to improve the understanding of conveying health-related statistical information with graphical representations compared with numerical representations. First, we investigated whether the iconicity of representations (i.e., their abstractness vs. concreteness) affected comprehension and recall of statistical information. Second, we investigated whether graph literacy helps to identify individuals who comprehend graphical representations better than numerical representations. **Method:** Participants ($N = 275$) were randomly assigned to receive different representations of health-related statistical information, ranging from very low iconicity (numbers) to very high iconicity (icon arrays including photographs). Comprehension and recall of the information were assessed. Additionally, participants rated the accessibility of the information and the attractiveness of the representation. Graph literacy was assessed by means of a recently developed scale. **Results:** The only difference between representations that affected comprehension and recall was the difference between graphics and numbers; the actual level of iconicity of graphics did not matter. Individuals with high graph literacy had better comprehension and recall when presented with graphics instead of numbers, and they rated graphical information as more accessible than numerical information, whereas the reverse was true for individuals with low graph literacy, $F(4, 185) = 2.60$, $p = .04$, $\eta_p^2 = .05$, and $F(4, 245) = 2.71$, $p = .03$, $\eta_p^2 = .04$, respectively. Both groups judged graphical representations as more attractive than numerical representations. **Conclusion:** An assessment of graph literacy distinguished individuals who are best informed with graphical representations of statistical information from those who are better informed with numerical representations.

Keywords: graph literacy, health literacy, icon arrays, medical decision making, risk communication

Supplemental materials: <http://dx.doi.org/10.1037/a0024850.supp>

Increasing efforts have been made to involve patients in medical decisions (Barry, 1999; Gigerenzer & Gray, 2011; O'Connor et al., 2007). To engage in informed and shared decision making, both

physicians and patients must evaluate and discuss the benefits and harms of treatment options, which requires comprehending statistical information. However, many people, including experts, have difficulty understanding health statistics (see Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007, and Reyna, Nelson, Han, & Dieckmann, 2009, for recent reviews).

Consequently, recent research has investigated how to better inform individuals about risks and benefits of screening or treatment options and how best to present numbers to represent risks (Fagerlin, Ubel, Smith, & Zikmund-Fisher, 2007; Gigerenzer et al., 2007; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). In addition, a variety of graphical representations can improve understanding of risks (Ancker, Senathirajah, Kulkafka, & Starren, 2006; Kurz-Milcke, Gigerenzer, & Martignon, 2008; Zikmund-Fisher, Ubel, et al., 2008). However, in a recent review of best practices for risk communication, Lipkus (2007) criticized the lack of theory with regard to the impact of graphical displays.

In this study, we aimed to improve the understanding of conveying health-related statistical information with graphical representations compared with numerical representations in two

This article was published Online First August 15, 2011.

Wolfgang Gaissmaier and Odette Wegwarth, Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany; David Skopec, Ann-Sophie Müller, and Sebastian Broschinski, Department of Design, Zurich University of the Arts, Zurich, Switzerland; Mary C. Politi, Department of Surgery, Washington University School of Medicine, St. Louis, MO.

This research was funded by the Max Planck Institute for Human Development, Berlin, Germany. Our thanks go to Angela Neumeyer-Gromen for helping us to gather the underlying clinical evidence for the topic medication, to Birgit Silberhorn for helping us with the data collection, and to Mary Louise Grossman for editing the manuscript. Mary C. Politi is a member of the U.S. Prescription Medicine Adherence Advisory Board for Merck.

Correspondence concerning this article should be addressed to Wolfgang Gaissmaier, Harding Center for Risk Literacy, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: gaissmaier@mpib-berlin.mpg.de

ways. First, we investigated the impact of the iconicity of information representations (i.e., their abstractness vs. concreteness) on comprehension and recall. Second, we investigated whether there are individual differences in comprehension of and preferences for graphical versus numerical information.

Impact of Iconicity on Comprehension and Recall

Iconicity refers to how much a representation resembles what it is supposed to represent versus the extent to which it is an abstraction (e.g., Moles, 1958/1968; Morris, 1946/1955). We believe that iconicity could be a useful concept in the domain of risk communication to understand differences between numerical and graphical representations and between different kinds of graphical representations. Numbers represent information with low iconicity, because they are strong abstractions; graphics are of higher iconicity, and increasingly so the more realistically they depict what they represent (e.g., using actual photographs).

One advantage of representations with low iconicity, such as mere numbers, could be that the information is reduced to its essential elements, eliminating unnecessary and potentially distracting features. Gathering precise numerical knowledge from more highly iconic graphical representations requires additional steps such as counting, whereas this information can simply be read off from numbers. For precise, verbatim knowledge, numbers alone could therefore be better representations. In line with this hypothesis, Feldman-Stewart, Kocovski, McConnell, Brundage, and Mackillop (2000) showed that numbers allowed a better assessment than graphical representations of what they called “detailed-level information,” which is similar to what we refer to as *verbatim knowledge*.

However, Feldman-Stewart et al. (2000) showed that for what they called “gross-level information,” graphical representations were better than numbers. *Gross-level information* refers to ordinal relations between quantities—for instance, that one quantity is bigger than another—without specifying exactly how much bigger. Building on those results, we tested whether high iconicity better enables people to acquire such knowledge, which we refer to as *gist knowledge* (i.e., qualitative bottom-line meaning; Reyna, 2008). Some evidence has shown, for instance, that symbols are easier to learn if they are of higher iconicity and thus are closer to what they actually represent (Mirenda, 2003) and that charts are better suited to convey gist knowledge than are numbers (Hawley et al., 2008).

A second advantage of high iconicity could be that it makes it easier to recall the information. A classic finding in cognitive psychology is that compared with words, pictorial information has a memory advantage (e.g., Paivio, Rogers, & Smythe, 1968; Park, Puglisi, & Sovacool, 1983; Shepard, 1967). Therefore, it could also be that highly iconic graphics are recalled more easily than numbers. Recall is an important aspect of making informed medical decisions, because one will often make or discuss such decisions when no information is laid out conveniently in front of oneself.

Individual Differences in Comprehension and Recall of Graphical Versus Numerical Information

The key assumption here is that there is no one-size-fits-all way of communicating risks; rather, different individuals can best be informed by different representations. Although using graphical information is generally recommended for facilitating risk comprehension (e.g., Paling, 2003), interpreting graphics requires additional skills beyond understanding numerical risks. Thus, one cannot assume that everyone intuitively understands graphics (Galesic & Garcia-Retamero, in press).

Instruments designed to measure the ability to understand quantitative information in health have thus far largely focused on *numeracy*—the ability to understand numerical information (e.g., Lipkus, Samsa, & Rimer, 2001; Reyna & Brainerd, 2007; Schwartz, Woloshin, Black, & Welch, 1997). Moreover, no standard instrument measuring the more general construct of health literacy (or aspects thereof) includes a systematic measurement of graph comprehension, even though they include quantitative skills as an important component (e.g., Baker, 2006; Parker, Baker, Williams, & Nurss, 1995).

Therefore, Galesic and Garcia-Retamero (in press) recently developed a graph literacy scale that assesses the ability to understand graphical information about health. They demonstrated that only those people who had high graph literacy benefitted from graphical information that complemented numbers (Garcia-Retamero & Galesic, 2010). We tested whether graph literacy therefore predicts individual differences in comprehension of and preference for graphical versus numerical information.

Overview of the Study and Hypotheses

The general task was designed to carefully assess how well participants could comprehend and recall health-related information that we provided to them, in relation to two independent variables, (a) how the information was represented in terms of iconicity, ranging from numbers to highly iconic graphics and (b) individual differences in graph literacy. In line with Edwards and Elwyn’s (1999) call for using more than one way to assess the success of different kinds of risk communication, outcome measures included both more objective knowledge questions (i.e., gist and verbatim) about health-related information and more subjective assessments of accessibility of the information and attractiveness of the representations. Moreover, we compared the comprehension of information when materials were laid directly in front of participants to recall at two later time points (i.e., after approximately 30 min and after approximately 2 weeks).

Our first hypothesis was that highly iconic representations would result in better gist knowledge of the information than lesser iconic representations, whereas the reverse should hold true for verbatim knowledge. In addition, we hypothesized that higher iconicity would particularly benefit recall. Our second hypothesis is that people with high graph literacy can particularly acquire and recall knowledge better with graphics than with numbers. It is an open question whether the comprehension and recall of people with low graph literacy do not improve with graphics alone rather than with numbers or whether they are actually better off with

numbers.¹ Both hypotheses were primarily concerned with the objective knowledge questions, and whether the results would be mirrored on the more subjective assessments was an exploratory question.

Method

Design

We used two health-related topics in this study. The first topic was a comparison of the frequency of benefits and side effects of three painkilling medications (i.e., aspirin, ibuprofen, and paracetamol [acetaminophen]) in comparison to placebo (hereinafter referred to as *medication*). The clinical evidence was taken from three Cochrane reviews (Derry, Derry, Moore, & McQuay, 2009; Edwards et al., 1999; Toms, Derry, Moore, & McQuay, 2009). The second topic was the impact of smoking on the risk of dying from lung cancer, colon cancer, prostate cancer, heart disease, stroke, and chronic obstructive pulmonary disease (hereinafter referred to as *smoking*). The clinical evidence was taken from Woloshin, Schwartz, and Welch (2008). Note that all numbers were simplified by rounding to improve understanding. For an overview of the clinical evidence, see Table S1 in the supplemental online materials.

On the basis of the topics medication and smoking, we developed three kinds of visualizations in collaboration with the Zurich University of the Arts under the direction of David Skopec (see Table 1). For each kind of visualization, there were five levels of iconicity (Table 1). Note that for all visualizations, numbers alone correspond to the lowest level of iconicity (Level 1, in column 2), and the highest level of iconicity included photographs in all three kinds of visualizations (Level 5, in column 6). The representations in between (i.e., Levels 2–4, columns 3–5) are graphical abstractions of the photographs, which become more and more similar, but not identical, to the photographs with increasing iconicity.

All graphics represented icon arrays at various levels of abstraction, with the exception of the most abstract graphics (i.e., Table 1, column 3), which are probably more similar to horizontal bar charts (medication–stick figures, smoking–stick figures) or pie charts (medication–pills). However, even those graphics shared an important feature with the icon arrays, namely that they had clearly separated and countable units. Two different kinds of visualization were developed by the same designer (Ann-Sophie Müller) for the first topic (medication). The first kind of visualization (medication–stick figures) showed horizontally aligned units that appeared increasingly humanlike with increasing iconicity. The second kind of visualization (medication–pills) showed circularly aligned units that looked increasingly pill-like with increasing iconicity. Using these two different kinds of visualizations based on the same topic allowed us to check whether our findings were the same in both cases. To test whether these results would also extend to a different topic, smoking, a separate designer (Sebastian Broschinski) developed a third kind of visualization (smoking–stick figures), which was similar to the first one (medication–stick figures) in that it also showed horizontally aligned units that looked increasingly humanlike with increasing iconicity. Importantly, all graphics were designed to allow the numbers to be precisely assessed and included small quantities that made this assessment comparatively easy.

In total, there were 3 (kinds of visualizations) \times 5 (levels of iconicity) between-subjects conditions. The lowest level of iconicity (numbers) was the same for both visualizations of medication and was therefore treated as one condition, resulting in 14 distinct between-participants conditions. The independent variables were the level of iconicity of the representation provided to participants and their graph literacy as assessed by the scale developed by Galesic and Garcia-Retamero (in press). The dependent variables were gist knowledge, verbatim knowledge, accessibility of the information, and attractiveness of the representation (see Measures section).

Procedure

The study was approved by the ethics committee of the Max Planck Institute for Human Development, and each participant provided informed consent. After assessing demographic characteristics, participants were randomly assigned to one of the 14 conditions. Participants were asked a series of questions concerning one of the two health-related topics on a computerized questionnaire. They were provided with an actual printout of one of the 14 different representations, and they had to answer various questions with the materials directly in front of them (Time 1 [T1]). These questions concerned objective dimensions (gist knowledge and verbatim knowledge) and subjective dimensions (accessibility of the information and attractiveness of the representation). Example items for each dimension are listed in Table S2 in the supplemental online materials. Next, participants returned the printout with the representation to the experimenter, filled out the graph literacy scale, and worked on another, unrelated experiment,² which took on average 34 min ($SD = 12$ min). Then, participants received a task in which they had to recall both gist and verbatim knowledge that they had learned previously (Time 2 [T2]). Finally, numeracy was assessed as a possible confounding variable. After exactly 2 weeks, participants received a previously announced e-mail inviting them to participate in another recall test of both gist and verbatim knowledge (Time 3 [T3]).

Measures

In this section, we describe measures in order of assessment.

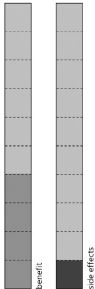




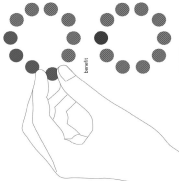
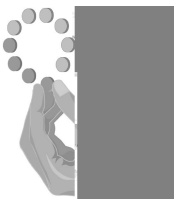


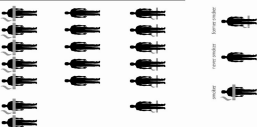


Demographics. Participants were asked to indicate their sex, age, and highest level of education (high school or less vs. college degree or more).

Gist knowledge. We assessed gist knowledge with five non-numerical questions asking for ordinal comparisons between quan-

¹ We did not have any specific hypothesis as to how the actual level of iconicity—beyond the difference between graphics and numbers—interacts with graph literacy, because the graph literacy scale assesses only the ability to handle graphical information in general and because too little is known as of now about the impact of iconicity.

² This experiment is part of another, yet unpublished, study in which participants were asked to evaluate a range of expert statements with regard to their credibility. These statements concerned different topics, including finance, environmental issues, and health, and the central manipulation was the level of uncertainty that the expert revealed. This manipulation was done within participants and was independent of the condition in the experiment presented in this article, so this other study did not introduce systematic bias.

Table 1
Excerpts of the Information Designs for the Different Kinds of Visualizations and Levels of Iconicity

Kind of visualization	Iconicity level				
	1	2	3	4	5
Medication—stick figures ^a	Benefit: 4 out of 10 Side effects: 1 out of 10				
Medication—pills ^a	Benefit: 4 out of 10 Side effects: 1 out of 10				
Smoking—stick figures ^b	Smoker: 7 out of 100 Nonsmoker: 5 out of 100 Former smoker: 6 out of 100				

^a It was explained to participants that *benefit* and *side effects* referred to the number of patients with a benefit (i.e., pain reduction) and side effects (e.g., drowsiness or gastric disorders), respectively. ^b It was explained to participants that numbers indicated how many out of every 100 people are expected to die of the disease within the next 10 years.

ties only, which were presented to participants in two consecutive blocks that varied in their difficulty. The first block consisted of three questions, for which participants needed to consider data on one dimension at a time only (e.g., “Which medication has the fewest side effects?”). In the second block, consisting of two questions, participants needed to consider the information on several dimensions (e.g., considering benefits and side effects together to answer “Which medication is worst overall?”). The gist knowledge score represents the average proportion of precisely correct answers achieved on the two blocks. The questions assessing gist knowledge were asked at T1 as well as at T2 and T3 to assess the recall thereof.

Verbatim knowledge. We assessed verbatim knowledge with 12 numerical questions, which were presented to participants in three consecutive blocks that varied in their difficulty and consisted of four questions each. In the first block, participants needed to read off frequencies from the information chart (e.g., “How many patients experience side effects with ibuprofen?”). In the second block, participants needed to compute absolute differences between two frequencies from the information chart (e.g., “How many patients experience a benefit of ibuprofen that they would not have had with a placebo?”). In the third block, participants needed to compute relative differences (i.e., percentage changes) between two frequencies from the information chart (e.g., “People who take ibuprofen have a % lower risk of experiencing a side effect compared with people who take paracetamol”). The verbatim knowledge score represents the average proportion of precisely correct answers achieved on the three blocks. The questions to assess verbatim knowledge were asked at T1 as well as at T2 and T3 to assess the recall thereof.

Accessibility. Subjective accessibility of the information was assessed with five questions, each of which was answered on a 5-point scale ranging from 1 (*not at all*) to 5 (*very much*). The five questions assessed these five aspects of the information: comprehensibility, usefulness, seriousness, intuitive accessibility, and difficulty of answering the questions. Answers were averaged to generate one accessibility score for each participant. The internal consistency of this scale was good (Cronbach's $\alpha = .78$).

Attractiveness. Subjective attractiveness of the representation was assessed with eight questions, each of which was answered on a 5-point scale ranging from 1 (*not at all attractive*) to 5 (*very attractive*). They assessed these aspects of the representation: overall impression, attractiveness of colors, imagery, technical implementation, size, font size, font, and composition. Answers were averaged so that each participant had one attractiveness score. The internal consistency of this scale was good (Cronbach's $\alpha = .82$).

Graph literacy. We assessed graph literacy with a scale recently developed and validated on nationally representative samples in Germany and the United States by Galesic and Garcia-Retamero (in press). The scale assesses an individual's comprehension of health-related information on the basis of graphical representations on three levels of difficulty: reading the data, reading between the data, and reading beyond the data. For examples of items measuring each of the three levels, see Figure S1 in the supplemental online materials. The scale consists of 13 items. Here, we briefly report its key psychometric properties, based on Galesic and Garcia-Retamero's German sample because our study also took place in Germany. It took participants 9.2 min on average ($SD = 5.7$) to complete the scale. On average, German participants

answered 9.4 (72.31%) of 13 questions correctly ($SD = 2.6$). Cronbach's alpha was .74, and the average item-total correlation was .37, indicating a satisfactory level of internal consistency. The average correlation between individual items was .19, showing that each item assessed a somewhat different aspect of graph literacy. To assess the scale's validity, Galesic and Garcia-Retamero assessed its correlations with other variables. Graph literacy correlated with education (.29), numeracy (.47), and graph comprehension items from other literacy questionnaires (.32).

Numeracy. We assessed numeracy as a control variable, because it is correlated with graph literacy (Galesic & Garcia-Retamero, in press; in our sample, $r[273] = .40, p < .001$). This assessment was done using the 11 items from Lipkus et al. (2001) plus one additional item by Schwartz et al. (1997; the item involving a coin toss). The same 12 items have previously been used successfully by other authors (Galesic, Garcia-Retamero, & Gigerenzer, 2009), and more generally, the numeracy scale is a widely used and accepted measurement instrument (e.g., Galesic & Garcia-Retamero, 2010; Peters et al., 2006).

Participants

Two hundred eighty participants (20 in each of the 14 conditions) were included in the basic experiment that consisted of working with the materials (T1) and the first recall test (T2). All individuals were invited from the participant pool of the Max Planck Institute for Human Development. Most participants were Caucasian (96.8%), and a few were Asian (2.1%) or Hispanic (1.1%). Each participant was paid €15 (approximately \$18 at that time), which included payment for the unrelated study between T1 and T2. Five participants did not finish the experiment and were thus excluded from the sample, so the final sample consisted of 275 participants. Of these 275 participants, 215 (78.2% follow-up rate) also completed the recall test after 2 weeks (T3). Descriptive statistics for demographics, graph literacy, and numeracy for all 275 participants, for the subsample of 215 participants at T3, and for the 60 participants who dropped out after T1 and T2 and did not participate at T3, respectively, can be found in Table 2. In comparison to participants at T3, those who dropped out were older, less graph literate, and less numerate, and fewer of them had at least a college degree. Still, the subsample of participants at T3 was very similar to the overall sample, except that they were slightly younger.

We assessed whether there were differences between participants across the 14 distinct conditions in demographics, numeracy, and graph literacy. At T1 and T2, there were differences in numeracy, and at T3 there were differences in numeracy and gender.³ Note that including numeracy, gender, or both as control variables did not affect the results.

Data Analysis

We calculated descriptive statistics for the dependent variables gist knowledge, verbatim knowledge, accessibility, and attractiveness. There were no missing values because the computerized

³ At T1 and T2, the 14 conditions differed with regard to numeracy, $F(13, 261) = 1.85, p = .04, \eta_p^2 = .08$. At T3, the 14 conditions differed with regard to gender composition, $\chi^2(13, N = 215) = 23.00, p = .04$, and numeracy, $F(13, 201) = 2.24, p = .01, \eta_p^2 = .13$.

Table 2

Characteristics of Participants at T1–T2 and T3; Characteristics of Those Who Dropped Out After T1–T2; and a Comparison of Participants at T3 and Those Who Dropped Out

Characteristic	T1–T2	T3	Dropouts	T3 vs. Dropouts (<i>p</i>)
<i>N</i>	275	215	60	
Mean age [95% CI]	31 [30, 32]	29 [28, 30]	39 [34, 44]	<.001
% Female [95% CI]	53.5 [47.6, 59.4]	55.3 [48.7, 62.0]	46.7 [34.1, 59.3]	=.245
% College degree or more [95% CI]	70.2 [64.8, 75.6]	74.9 [69.1, 80.7]	53.3 [40.7, 66.0]	=.002
Mean graph literacy [95% CI]	.82 [.80, .84]	.84 [.82, .86]	.77 [.72, .82]	=.017
Mean numeracy [95% CI]	.88 [.86, .90]	.89 [.87, .91]	.82 [.78, .86]	=.002

Note. T1 = Time 1; T2 = Time 2; T3 = Time 3; CI = confidence interval.

questionnaire did not allow for item nonresponse. For the main analyses, participants' graph literacy was split at the median to obtain one group with relatively low graph literacy and one group with relatively high graph literacy.⁴

To analyze gist and verbatim knowledge as a function of iconicity and graph literacy, we ran a repeated-measures ANOVA with the within-subjects factors time (T1, T2, T3) and type of knowledge (gist vs. verbatim), as well as the between-subjects factors graph literacy (high vs. low), iconicity (1–5), and kind of visualization (medication–stick figures, medication–pills, smoking–stick figures). This analysis could only be performed for the subsample of 215 participants who also participated in the follow-up because it required data for T1, T2, and T3. To analyze accessibility and attractiveness as a function of iconicity and graph literacy, we ran a multivariate ANOVA, which again included the between-subjects factors graph literacy (high vs. low), iconicity (1–5), and kind of visualization (medication–stick figures, medication–pills, smoking–stick figures). We performed this analysis on the entire sample of 275 participants. Results remained identical when only the subsample of 215 participants was included. Note that because the condition with the lowest iconicity (numbers) was identical in both representations of medication, half of these participants were randomly assigned to medication–stick figures and the other half were assigned to medication–pills in both of the ANOVAs.

Results

Descriptive Statistics

The mean accuracy scores on gist knowledge for T1, T2, and T3 were .83 (*SD* = .14), .76 (*SD* = .21), and .70 (*SD* = .24), respectively. The mean accuracy scores on verbatim knowledge for T1, T2, and T3 were .54 (*SD* = .24), .35 (*SD* = .20), and .23 (*SD* = .16), respectively. The mean scores for accessibility and attractiveness were 3.38 (*SD* = 0.78) and 3.38 (*SD* = 0.73), respectively. The group of participants with low graph literacy, as defined by the median split, had a mean accuracy of .75 (*SD* = .13) on this scale, whereas participants with high graph literacy had a mean accuracy of .94 (*SD* = .03).

Gist and Verbatim Knowledge

Gist and verbatim knowledge as a function of iconicity and graph literacy are illustrated in Figure 1 for T1, T2, and T3,

averaged across the three kinds of visualization. We first checked whether iconicity had a systematic impact on both gist and verbatim knowledge as hypothesized. This was not the case: Knowledge did not generally increase with iconicity, $F(4, 185) = 0.94$, $p = .44$, $\eta_p^2 = .02$. It was also not the case that higher iconicity benefitted gist but not verbatim knowledge, $\text{Iconicity} \times \text{Type of Knowledge}$ $F(4, 185) = 0.52$, $p = .72$, $\eta_p^2 = .01$. Finally, higher iconicity did not benefit knowledge more strongly in recall than in working with the materials, $\text{Iconicity} \times \text{Time}$ $F(8, 370) = 0.38$, $p = .93$, $\eta_p^2 = .01$.

Next, we looked at graph literacy. Participants with high graph literacy had higher gist and verbatim knowledge scores in the graphical conditions compared with the numbers-only condition, and the reverse was true for participants with low graph literacy, $\text{Iconicity} \times \text{Graph Literacy}$ $F(4, 85) = 2.60$, $p = .04$, $\eta_p^2 = .05$, and this interaction did not differ between the different kinds of visualizations, $\text{Iconicity} \times \text{Graph Literacy} \times \text{Kind of Visualization}$ $F(8, 185) = 0.64$, $p = .75$, $\eta_p^2 = .03$.

The interaction between iconicity and graph literacy stemmed purely from the difference between graphics and numbers, whereas the actual level of iconicity of the graphics had no effect. When the numbers-only condition (i.e., Table 1, column 2) was excluded from the analyses, the $\text{Iconicity} \times \text{Graph Literacy}$ interaction disappeared, $F(3, 161) = 0.41$, $p = .74$, $\eta_p^2 = .01$. Instead, a main effect of graph literacy remained, indicating that individuals with high graph literacy performed better on all levels of iconicity beyond numbers, $F(1, 161) = 14.67$, $p < .001$, $\eta_p^2 = .08$. Surprisingly, in the numbers-only conditions, the reverse effect was true: People with high graph literacy actually had lower gist and verbatim knowledge scores than people with low graph literacy, $F(1, 24) = 3.50$, $p = .07$, $\eta_p^2 = .13$.

In other words, people with high graph literacy generally achieved higher gist and verbatim knowledge scores with graphics

⁴ The continuous graph literacy score was not well suited as a between-subjects factor in the subsequent analyses of variance (ANOVAs) because the distribution of scores was skewed and there were too few observations of each particular level of graph literacy in each of the 14 conditions. The median split ensured a sufficient number of participants with either low or high graph literacy in each of the 14 conditions. Additionally, this is how the scale was used by its developers (e.g., Garcia-Retamero & Galesic, 2010), and median splits are also typically used in research using the related construct of numeracy for the same reason as that of a skewed distribution of scores (e.g., Peters et al., 2006).

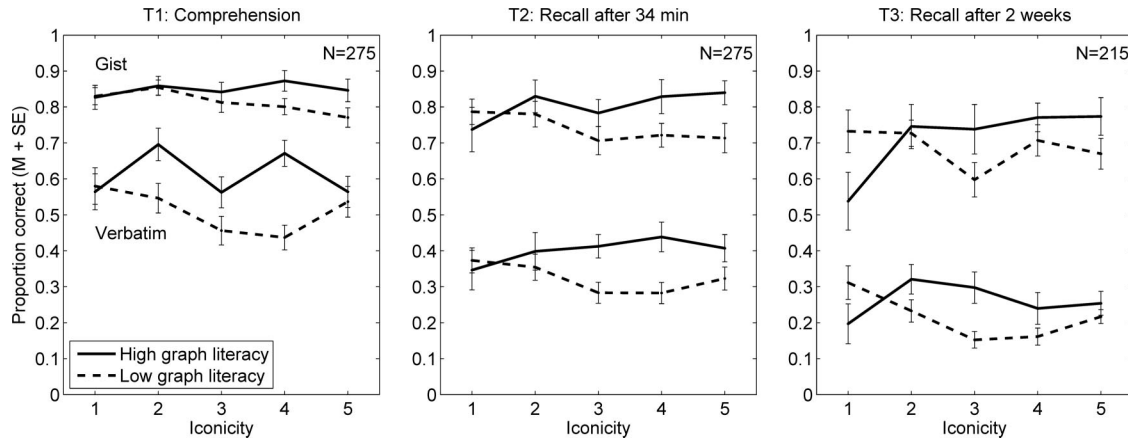


Figure 1. The only difference between representations that affected comprehension and recall was the difference between numbers (i.e., iconicity = 1) and graphics (i.e., iconicity ≥ 2). Individuals with high graph literacy had better comprehension and recall when presented with graphics instead of numbers, and the reverse was true for individuals with low graph literacy. T1 = Time 1; T2 = Time 2; T3 = Time 3.

than with numbers, independent of the actual level of iconicity of the graphics. For people with low graph literacy, the reverse was true. Figure 2 shows the difference between numbers (Table 1, column 2) and graphics (i.e., pooled across all graphical conditions shown in Table 1, columns 3–6) for all three kinds of visualizations, separately for people with high and low graph literacy and separately for gist versus verbatim knowledge and points of measurement.

Accessibility and Attractiveness

Figure 3 illustrates that the results for subjective accessibility were similar to objective gist and verbatim knowledge: Accessibility did not generally increase with iconicity, $F(4, 245) = 1.68$, $p = .16$, $\eta_p^2 = .03$. For participants with high graph literacy, however, accessibility did increase with iconicity, whereas it decreased with iconicity for participants with low graph literacy, Iconicity \times Graph Literacy $F(4, 245) = 2.71$, $p = .03$, $\eta_p^2 = .04$, and this interaction did not differ between the different kinds of visualizations, Iconicity \times Graph Literacy \times Kind of Visualization $F(8, 245) = 0.71$, $p = .68$, $\eta_p^2 = .02$. Similar to the results on gist and verbatim knowledge, in the numbers-only condition (Table 1, column 2) accessibility was rated higher by participants with low graph literacy than by participants with high graph literacy, $F(1, 32) = 4.90$, $p = .03$, $\eta_p^2 = .13$.

The picture was very different for attractiveness, which was generally rated higher for graphics than for numbers, reflected in a main effect of iconicity, $F(4, 245) = 7.61$, $p < .001$, $\eta_p^2 = .11$ (see Figure 3). This effect did not depend on graph literacy, Iconicity \times Graph Literacy $F(4, 245) = 0.06$, $p = .99$, $\eta_p^2 = .00$, which held true across the different kinds of visualization, Iconicity \times Graph Literacy \times Kind of Visualization $F(8, 245) = 0.80$, $p = .60$, $\eta_p^2 = .03$.

Discussion

We investigated the impact of iconicity of representations and individuals' graph literacy on gist and verbatim knowledge and

recall of health information, as well as on preferences for different representations. Iconicity ranged from numbers to icon arrays with photographs, with less iconic graphics in between. To our knowledge, this study is the first to systematically explore the concept of iconicity of representations in the context of communicating health-related statistical information. In addition, it adds to the understanding of the recently developed concept of graph literacy.

The most important result was that neither graphics nor numbers were superior for conveying gist or verbatim knowledge per se. Rather, only participants with high graph literacy achieved better gist and verbatim comprehension and recall with graphics than with numbers, and they also rated graphics as more subjectively accessible. For participants with low graph literacy, in contrast, the opposite held true. This interaction is consistent with our second hypothesis, and it was robust in the sense that it did not depend on how exactly the graphics were designed, at least with regard to the variations in graphics studied here. In line with these results, Stone, Yates, and Parker (1997) showed differences in risk perception between graphical and numerical formats, but no differences between different graphical formats.

Counter to our first hypothesis, however, higher iconicity did not result in improved gist knowledge, and less iconic information (particularly numbers) did not result in better verbatim knowledge, in contrast to Feldman-Stewart et al. (2000) and Hawley et al. (2008). Our study differed in that numbers represented relatively small quantities and could be read off precisely, which was also true of the graphical representations, which could result in the loss of a potential advantage of numbers for verbatim knowledge. Moreover, higher iconicity did not lead to improved recall, contrary to findings in the memory literature that pictures are better remembered than words (Paivio et al., 1968; Park et al., 1983; Shepard, 1967). The only main effect of iconicity was that on attractiveness: Graphical representations were generally judged to be more attractive than numbers, regardless of participants' graph literacy.

In the numbers-only condition, participants with low graph literacy actually achieved higher gist and verbatim knowledge

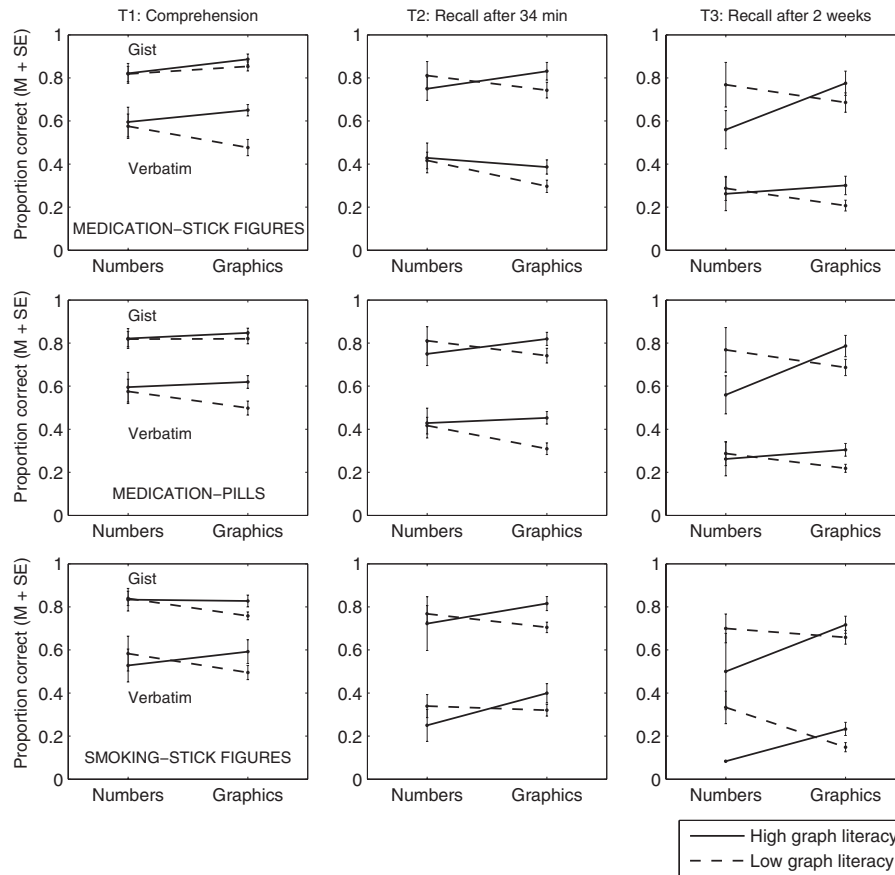


Figure 2. For all three kinds of visualizations (see Table 1), individuals with high graph literacy had better comprehension and recall with graphics than with numbers. Individuals with low graph literacy, in contrast, had better comprehension and recall with numbers than with graphics. T1 = Time 1; T2 = Time 2; T3 = Time 3.

scores than participants with high graph literacy. This finding was surprising, because graph literacy is correlated with numeracy, and thus participants with high graph literacy also had higher numeracy on average (.91 vs. .86). At the same time, our sample was relatively highly educated in general, so that the numeracy scores of even participants with low graph literacy were high in comparison to the general population. If numeracy is sufficiently high, it could be the case that subjective preferences for different representational formats become more important than skill. In line with this idea, participants with low graph literacy also subjectively evaluated numerical information to be more accessible than did participants with high graph literacy.

Limitations

The high level of education in our sample is one of the limitations of our study. In fact, even those participants who were classified by the median split as having low graph literacy had about the same or even slightly higher graph literacy scores than the average score found in a nationally representative sample of people ages 25–69 in Germany (Galesic & Garcia-Retamero, in press). Therefore, it is important to be clear that the statements we make about differences between low and high graph literacy are interpreted not in an absolute manner but relative to this sample.

Note that research on skills required to understand quantitative information has commonly defined groups of low versus high skill relative to the sample and not by comparison to absolute standards (e.g., Peters et al., 2006), or even relative to subsamples such as younger and older adults (e.g., Galesic et al., 2009). Still, future research needs to investigate whether the results reported here also hold true for more representative samples. It is encouraging that our results are consistent with the results of a similar study that tested representative samples (Garcia-Retamero & Galesic, 2010). Additionally, one could argue that it is particularly surprising that we found substantial differences between participants with relatively low and relatively high graph literacy, given that the general level of graph literacy in our sample was high.

A second limitation is that we do not know whether our manipulation of iconicity was perceived by our participants in the way we intended, because we did not include a manipulation check. That is, we cannot know for sure that participants would agree with us as to which graphical representation was of higher or lower iconicity in comparison to the other representations. Because the graphical representations were developed by designers who are experts on the concept of iconicity, we believe that participants would agree with us on the order of iconicity of the graphical representations within each of the topics. However, it is less clear

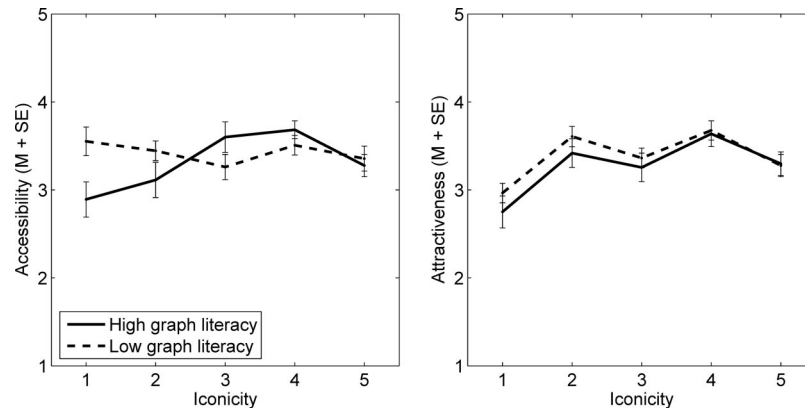


Figure 3. Individuals with high graph literacy evaluated graphics (i.e., iconicity ≥ 2) to be more accessible than numbers (i.e., iconicity = 1), whereas the reverse was true for individuals with low graph literacy (left panel). Graphics were generally rated as more attractive than numbers, even by individuals with low graph literacy (right panel).

whether they would perceive the distance between two levels of iconicity similarly for different topics, for instance whether the difference between iconicities of 4 and 5 (Table 1, columns 5 and 6) would be perceived similarly for the topics smoking-stick figures and medication-stick figures. In this regard, however, it is comforting that for the subjective evaluations of attractiveness—which is probably the measure that comes closest to a manipulation check of the perception of iconicity—the level of iconicity had a comparable effect across all three different topics.

Implications

The findings of this study have important practical implications, because they clearly demonstrate that not everyone can be successfully informed using the same mode of representation. Past research has identified and developed important tools for informing people about risk and uncertainty (e.g., Fagerlin, Ubel et al., 2007; Gigerenzer et al., 2007). To date, when informing patients with patient decision aids, decision aid developers typically include a wide range of representations all at once to account for individual differences in preference for and understanding of risk representations. For instance, one decision aid quality criterion listed by the International Patient Decision Aids Standards Collaboration stated, “The patient decision aid provides more than one way of explaining the probabilities (e.g., words, numbers, diagrams)” (Elwyn et al., 2006, Table 2, p. 2, in Additional Details section). Given an ever-increasing range of methods for communicating probabilities (see Bunge, Mühlhauser, & Steckelberg, 2010, for an overview), including multiple formats could result in an information overload for patients. In fact, evidence already exists that people sometimes prefer simplified and reduced information (e.g., Peters et al., 2007; Zikmund-Fisher, Fagerlin, & Ubel, 2008). It is possible that assessing an individual’s level of graph literacy could help determine which risk representation format to include during decision communication.

However, assessing patients’ skills regarding graph literacy or numeracy before providing information to them in clinical practice might be too time consuming. With regard to numeracy, some have recommended assessing subjective (i.e., self-assessed) rather

than objective numeracy (e.g., Fagerlin, Zikmund-Fisher et al., 2007; Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007). The advantages of subjective numeracy are that it can be assessed more quickly and that it is less aversive for patients in clinical settings. Future research should carefully develop subjective or shorter measures of graph literacy to explore whether they could be reliably used in practice.

One way to minimize information overload and circumvent the problem of assessing graph literacy in clinical settings at the same time could be to allow patients to choose how they prefer to receive risk information. This could best be achieved with computer kiosks. However, because kiosks will not often be available in a busy doctor’s office, one could alternatively provide patients with drug facts boxes, simple tabular presentations of clinical data that were developed and tested by Schwartz, Woloshin, and Welch (2007). These drug facts boxes could be accompanied by graphics on the back side, allowing patients to choose which kind of representation to focus on.

Our findings suggest that in principle, patients could choose the representation that allows them to comprehend the information best: Participants with high graph literacy evaluated graphics as more accessible than numbers and achieved higher gist and verbatim knowledge and recall with graphics than with numbers, and the reverse was true for participants with low graph literacy. However, even those participants with low graph literacy thought that graphical information was more attractive, which suggests that they might choose graphical representations although their gist and verbatim knowledge was lower with graphics than with numbers. Future research could examine experimentally the impact of allowing individuals to choose their preferred risk representations on understanding, compared with a situation in which they are randomly assigned to different risk representations.

Conclusion

In conclusion, these findings suggest that health-related information should be conveyed differently to different individuals. An assessment of graph literacy distinguished individuals who are best informed with graphical representations of statistical information

from those who are better informed with numerical representations. However, to successfully articulate strategies of presentation in clinical settings, more research is needed on the role of graph literacy in risk communication and how best to tailor information to individual characteristics to improve understanding of health statistics.

References

- Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association*, 13, 608–618.
- Baker, D. W. (2006). The meaning and the measure of health literacy. *Journal of General Internal Medicine*, 21, 878–883.
- Barry, M. D. (1999). Involving patients in medical decisions. *JAMA*, 282, 2356–2357.
- Bunge, M., Mühlhauser, I., & Steckelberg, A. (2010). What constitutes evidence-based patient information? Overview of discussed criteria. *Patient Education and Counseling*, 78, 316–328.
- Derry, C., Derry, S., Moore, R. A., & McQuay, H. J. (2009). Single dose oral ibuprofen for acute postoperative pain in adults. *Cochrane Database of Systematic Reviews*, 3, Art. No. CD001548. doi:10.1002/14651858.CD001548.pub2
- Edwards, A., & Elwyn, G. (1999). How should effectiveness of risk communication to aid patients' decision be judged? A review of the literature. *Medical Decision Making*, 19, 428–434.
- Edwards, J., Oldman, A., Smith, L. A., Collins, S., Carroll, D., Wiffen, P. J., . . . Moore, R. A. (1999). Single dose oral aspirin for acute pain. *Cochrane Database of Systematic Reviews*, 4, Art. No. CD002067. doi:10.1002/14651858.CD002067
- Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., . . . Whelan, T. (2006). Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *British Medical Journal*, 333, 417–419.
- Fagerlin, A., Ubel, P. A., Smith, D. M., & Zikmund-Fisher, B. J. (2007). Making numbers matter: Present and future research in risk communication. *American Journal of Health Behavior*, 31(Suppl. 1), S47–S56.
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. *Medical Decision Making*, 27, 672–680.
- Feldman-Stewart, D., Kocovski, N., McConnell, B. A., Brundage, M. D., & Mackillop, W. J. (2000). Perception of quantitative information for treatment decisions. *Medical Decision Making*, 20, 228–238.
- Galesic, M., & Garcia-Retamero, R. (2010). Statistical numeracy for health: A cross-cultural comparison with probabilistic national samples. *Archives of Internal Medicine*, 170, 462–468.
- Galesic, M., & Garcia-Retamero, R. (in press). Graph literacy: A cross-cultural comparison. *Medical Decision Making*. doi:10.1177/0272989X10373805
- Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, 28, 210–216.
- Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science and Medicine*, 70, 1019–1025.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients to make sense of health statistics. *Psychological Science in the Public Interest*, 8, 53–96.
- Gigerenzer, G., & Gray, J. A. M. (Eds.). (2011). *Better doctors, better patients, better decisions: Envisioning health care 2020*. Cambridge, MA: MIT Press.
- Hawley, S. T., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Lucas, T., & Fagerlin, A. (2008). The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, 73, 448–455.
- Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, 290, 2261–2262.
- Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: Graphical and analog tools. In W. T. Tucker et al. (Eds.), *Annals of the New York Academy of Sciences: Vol. 1128. Strategies for risk communication: Evolution, evidence, experience* (pp. 18–28). New York, NY: Blackwell.
- Lipkus, I. M. (2007). Numeric, verbal, and visual formats of conveying health risks: Suggested best practices and future recommendations. *Medical Decision Making*, 27, 696–713.
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21, 37–44.
- Mirenda, P. (2003). Toward functional augmentative and alternative communication for students with autism: Manual signs, graphic symbols, and voice output communication aids. *Language, Speech, and Hearing Services in Schools*, 34, 203–216.
- Moles, A. (1968). *Information theory and esthetic perception*. Urbana, IL: University of Illinois Press. (Original work published 1958)
- Morris, C. (1955). *Signs, language, and behavior*. New York: Braziller. (Original work published 1946)
- O'Connor, A. M., Wennberg, J. E., Legare, F., Llewellyn-Thomas, H. A., Moulton, B. W., Sepucha, K. R., . . . King, J. S. (2007). Towards the "tipping point": Decision aids and informed patient choice. *Health Affairs*, 26, 716–725.
- Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11, 137–138.
- Paling, J. (2003). Strategies to help patients understand risks. *British Medical Journal*, 327, 745–748.
- Park, D. C., Puglisi, J. T., & Sovacool, M. (1983). Memory for pictures, words, and spatial location in older adults: Evidence for pictorial superiority. *Journal of Gerontology*, 38, 582–588.
- Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The Test of Functional Health Literacy in Adults: A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine*, 10, 537–541.
- Peters, E., Dieckmann, N. F., Dixon, A., Hibbard, J. H., Mertz, C. K., & Slovic, P. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, 64, 169–190.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17, 407–413.
- Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy-trace theory. *Medical Decision Making*, 28, 829–833.
- Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, 17, 147–159.
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, 135, 943–973.
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966–972.
- Schwartz, L. M., Woloshin, S., & Welch, H. G. (2007). The drug facts box: Providing consumers with simple tabular data on drug benefit and harm. *Medical Decision Making*, 27, 655–662.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6, 156–163.

- Stone, E. R., Yates, J. F., & Parker, A. M. (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied*, 3, 243–256.
- Toms, L., Derry, S., Moore, R. A., & McQuay, H. J. (2009). Single dose oral paracetamol (acetaminophen) with codeine for postoperative pain in adults. *Cochrane Database of Systematic Reviews*, 1, Art. No. CD001547. doi:10.1002/14651858.CD001547.pub2
- Woloshin, S., Schwartz, L. M., & Welch, H. G. (2008). The risk of death by age, sex, and smoking status in the United States: Putting health risks in context. *Journal of the National Cancer Institute*, 100, 845–853.
- Zikmund-Fisher, B. J., Fagerlin, A., & Ubel, P. A. (2008). Improving understanding of adjuvant therapy options by using simpler risk graphics. *Cancer*, 113, 3382–3390.
- Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale (SNS): Effects of low numeracy on comprehension of risk communications and utility elicitation. *Medical Decision Making*, 27, 663–671.
- Zikmund-Fisher, B. J., Ubel, P. A., Smith, D. M., Derry, H. A., McClure, J. B., . . . Fagerlin, A. (2008). Communicating side effect risks in a tamoxifen prophylaxis decision aid: The debiasing influence of pictographs. *Patient Education and Counseling*, 73, 209–214.

Received August 17, 2010

Revision received June 15, 2011

Accepted June 17, 2011 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.