Design Features of Graphs in Health Risk Communication: A Systematic Review

JESSICA S. ANCKER, MPH, YALINI SENATHIRAJAH, RITA KUKAFKA, DRPH, MA
JUSTIN B. STARREN, MD, PHD

Abstract This review describes recent experimental and focus group research on graphics as a method of communication about quantitative health risks. Some of the studies discussed in this review assessed effect of graphs on quantitative reasoning, others assessed effects on behavior or behavioral intentions, and still others assessed viewers’ likes and dislikes. Graphical features that improve the accuracy of quantitative reasoning appear to differ from the features most likely to alter behavior or intentions. For example, graphs that make part-to-whole relationships available visually may help people attend to the relationship between the numerator (the number of people affected by a hazard) and the denominator (the entire population at risk), whereas graphs that show only the numerator appear to inflate the perceived risk and may induce risk-averse behavior. Viewers often preferred design features such as visual simplicity and familiarity that were not associated with accurate quantitative judgments. Communicators should not assume that all graphics are more intuitive than text; many of the studies found that patients’ interpretations of the graphics were dependent upon expertise or instruction. Potentially useful directions for continuing research include interactions with educational level and numeracy and successful ways to communicate uncertainty about risk.

Introduction
Quantitative risk communication is a critical component of informatics applications that support such activities as shared medical decision-making, informed consent, health risk appraisal, and counseling about difficult decisions pertaining to cancer or genetic screening. Effective risk communication can improve awareness of health risks and promote risk-reducing behavior in support of health promotion and disease prevention. One of the many challenges to risk communication with the public is the difficulty in expressing quantitative information in an easily comprehensible form. Universal cognitive limitations cause biases in interpreting numerical probabilities. Small probabilities are particularly difficult to interpret; under some conditions people overestimate them, and under others they ‘round down’ to zero. For many consumers, these difficulties in interpreting probabilities are compounded by limited numeracy skills and by discomfort with numerical expressions of risk. Understanding numerical information can be even more difficult when analytic reasoning processes are impaired by age, stress, or other factors.

Graphs are an appealing alternative to numbers because they are visually interesting and exploit rapid, automatic visual perception skills. Some aspects of graph interpretation also require more effortful cognitive skills such as interpretation and calculation, which may rely upon learned strategies. Nevertheless, a well designed visual display can reduce the amount of mental computation by replacing it with automatic visual perception.

Graphs are often used in print and electronic materials for patient education, decision support, and health risk communication. Many studies from fields such as psychophysics, human factors, and marketing did not involve health risks; the studies of health graphs were often atheoretical, and results were sometimes inconsistent. In recent years, a number of relevant studies have appeared in the medical, psychological, and patient counseling literature. Here, we update Lipkus and Hollands’ article in a systematic review of experimental or focus group research on graphs of quantitative health risks.
Search Methods
We searched for evaluation studies of graphs describing probabilities, frequencies, or chances of health events that had not been covered in Lipkus and Hollands’ review. We excluded commentaries and instructions (e.g.,16), studies of pain scales, utility measures, or illustrations that communicated threat or causal relationships (e.g.,17), and studies in which graphics were not used as an independent variable (e.g.,14,18,19). We also excluded dissertations. We searched 3 bibliographic databases (PsycInfo, MEDLINE, CINAHL) and one portal (ACM Portal) for 1998–2005 inclusive using topic headings chosen from the controlled terminology of each database and additional key words.3 We read all titles, then read the abstract and full text of potentially relevant articles. Upon identifying eligible articles, we used the ‘find citing research’ and ‘find similar’ tools and searched reference lists. We also did key word searches on the websites of selected journals (Health Psychology, Risk Analysis and Medical Decision Making). In this review, we describe 24 studies. The searches produced: 969 unique articles on MEDLINE of which 14 met our criteria; 245 on PsycINFO of which 3 met our criteria and had not already been identified; and 54 on the Medical Decision Making web site of which 2 were eligible and had not been identified. The other searches produced no articles that had not been previously identified; 5 articles were identified through bibliography search. Details about the searches are available upon request.

Outcome Measures in Graphics Studies
Synthesizing the findings of these studies is challenging because of the variety of outcome measures used by the researchers.20 As others have noted,21 risk communication can be undertaken with different goals: (a) to increase understanding; (b) to change risky behavior; or (c) for cooperative conflict resolution. Also, some risk communication designs are designed to increase concern, while others are designed to calm fears.22 Researchers who pursue the first goal above—improving understanding—use outcome measures such as the accuracy or consistency of quantitative reasoning or perceptions (e.g.,10,23–25). Examples from this review include whether users can estimate a proportion represented in a graphic,25 whether they can read the number of survivors at a particular time point from a survival curve26 and whether they produce the same ranking of risks when expressing their perceptions in different graphical formats.27,28 From this perspective, graphical elements that cause perceptions of the risk to deviate from the probability of the outcome (including framing, axis distortion, or relative comparisons) may be denounced as unethical.23,24,29–32

Design Features of Risk Graphs
In our summary of the findings, we describe each type of graphic (bar chart, pie chart, survival curve, etc.) and its data scale (ordinal, discrete, continuous). We call attention to 3 design features that have not always been highlighted in the original studies but that help shed light on the results: (1) whether part-to-whole relationships can be assessed visually; (2) the graphical perception abilities exploited; and (3) the format of numbers in the graphics.

1 Part-to-whole relationships: The ability to estimate what proportion object A represents of a larger object B appears to be an automatic perceptual skill that can be invoked when a graphic displays the entire object B.31 An example is a stacked bar chart, a bar that extends from 0% to 100% with a segment in a different color to indicate the proportion affected by the disease. The part-to-whole relationship is available visually: a segment 10 units high in a bar 100 units high represents a risk of exactly 10%. Usually, in part-to-whole graphics, the size of the graphic element is proportional to the quantity it depicts, so a segment 10 units high depicts a risk exactly twice as large as one 5 units high. This property is considered desirable for data integrity.23,29 In other cases, such as a bar chart with a y axis truncated at 50%, the sizes of the graphic elements are proportional to the quantities, but the part-to-whole relationship for each risk is not available visually. It is also possible for graphics to be non-proportionate and to not contain part-to-whole visual information (e.g., a risk scale ranging from 1 in a million to 1 in 10). In these, the size difference between elements is not directly related to the difference between quantities.

2 Features that exploit basic graphical perception abilities: The classic psychophysical research of Cleveland and
McGill ranked visual perception tasks by their accuracy.\textsuperscript{10,42} Accuracy was excellent when judging positions or lengths against a common scale (such as heights of bars of a bar graph); good when judging angles (such as size of slices in a pie chart) and slopes (such as slopes of a line graph); fair when judging areas (such as circles); and poor when judging volumes or color and gray-scale densities.\textsuperscript{10}

3 Numerical format: Performing mathematical calculations such as converting from ratios to percentages is a learned skill; ability to perform such tasks varies with education, health literacy, and numeracy\textsuperscript{6,7,43}. A probability of 6 in 100 is formally equivalent to both 6% and 0.06, but the different formats strongly affect reasoning. For example, with ratios, problem-solving ability and comprehension are worse when the denominators are different than when they are the same: it is harder to compare and calculate with the pair of numbers “1 in 250” and “1 in 1000” than it is with “4 in 1000” and “1 in 1000”.\textsuperscript{5,44,45} Ratios with the same denominator have been called “natural frequencies.”\textsuperscript{5,44,45} In a study in outpatient clinics, only 56% could identify the larger of two risks when they were written in the “1-in-x” format.\textsuperscript{46} Complex-looking ratios such as 513/570 are more demanding to process than equivalent but simpler ones (such as 9/10) or decimals (e.g., 0.90), as shown by preference reversals with different formats.\textsuperscript{47}

A discussion of more complex graphical perception tasks, such as integrating information from multiple sources, would require attention to more complex theories.\textsuperscript{12} However, most risk graphics involve relatively simple tasks such as providing information about an individual risk, comparing several risks, or judging trends in risk over time.

Research Review

Icon Arrays

An icon array portrays a risk at the discrete level of measurement as a group of individual icons, such as dots or stick figures. In numerical reasoning, people tend to perform better on probability problems when the data are presented at the discrete level rather than as percentages or proportions.\textsuperscript{44,45,48} Slovic et al. review evidence that presenting information in terms of individuals can produce mental imagery with strong affective elements.\textsuperscript{9}

An icon display reduced the influence of vivid text anecdotes in a study of choices of medical treatment (Fig. 1).\textsuperscript{49} In this study, people were asked to imagine having angina and being offered more successful (75% success rate) but more arduous bypass surgery, or less successful (50% success rate) but less arduous balloon angioplasty. They also read anecdotes about patients who had had the procedures. The number of anecdotes describing success strongly affected participants’ choices. When the proportion of successes in the anecdotes was the same as the treatments’ success rates (for example, when 3 of the 4 bypass stories described a treatment success), respondents became more likely to choose the more successful alternative (bypass). When one anecdote described success and one a failure, most respondents chose the less arduous treatment (angioplasty). The anecdote effect was significantly smaller when respondents saw icon displays depicting the two treatments’ success rates.\textsuperscript{49} The icon array showed the part-to-whole relationship and the square icons were touching, so the display might have been visually processed as areas rather than as discrete icons.

In a focus group of women, participants preferred icon arrays with smaller denominators because they seemed simpler but also tended to think that graphics with larger denominators portrayed risks as smaller.\textsuperscript{37} The findings are not consistent with the common ratio-bias effect, in which risks described as ratios of small numbers are considered smaller than numerically equivalent risks described with large numbers (e.g., 1 in 20 is considered less likely than 1 in 200).\textsuperscript{50} In another focus group study with low-income women, participants preferred seeing an individualized risk estimate depicted as a bar chart with an ordinal scale (low, average, or high risk) rather than as an icon array or a percentage, and rather than a bar chart showing a series of relative risks for women in different risk categories.\textsuperscript{39}

Fuller et al. used several tasks to assess how elderly patients interpreted discrete icon displays.\textsuperscript{51} The patients could match percentages to icon arrays displaying different proportions (70% to 98% accuracy for different tasks). They were less accurate when marking the graph to show probabilities (either ratios with different denominators [38% to 79% accuracy] or percentages [51% to 98% accuracy]). The authors did not assess whether the graphs were successful in conveying the personal applicability of the risk. A short
Part-to-Whole Relationships in Icon Arrays

The importance of the part-to-whole relationship is suggested by a series of studies by Stone and colleagues\textsuperscript{34,53} to follow up a 1997 study.\textsuperscript{54} In these studies, undergraduates received information about pairs of fictitious products, each carrying a small probability of a harmful effect (e.g., tire blowouts with tires). Participants estimated how much the safer product would be worth. The graph showed the number harmed but the at-risk group was provided as a number, so the part-to-whole relationship was not available visually (Fig. 2). People were willing to pay more (i.e., were more risk-averse) when the number harmed was depicted as an array of stick figures, asterisks, or faces, or as a bar graph than when it was number. Asterisks and faces led to similar results, suggesting that human figures had no ‘humanizing effect.’ In one study,\textsuperscript{34} the graphs were compared to graphs that did portray the part-to-whole relationships visually, for example, a bar graph showing only the number affected compared to a stacked bar graph of those affected as a proportion of the entire group (Fig. 3). Participants were willing to pay more for the safer product with a graph that did not show the part-to-whole relationship than when given a number. However, when they saw part-to-whole relationship in the graph, they were not willing to pay more than when they saw numerical probabilities. Stone et al. suggest that risk aversion for rare events is the result of graphs that fail to show part-to-whole relationships, not of all graphs, and label the effect a ‘foreground effect’.\textsuperscript{34,53}

These studies provide a new perspective on an older pair of studies that used dots to depict only the denominator of the probability; viewers were told that the risk was 1 against the number of dots. An icon display of the risk of rare side-effects from a vaccine increased the number of subjects who said they would get vaccinated, presumably by focusing their attention on the denominator.\textsuperscript{55} However, a similar study could not replicate the effect.\textsuperscript{56}
Human Figures Versus Other Icons

Although Stone et al found no differences in behavior after viewing asterisk and face displays, Schapira et al’s qualitative study found that women considered human figure icons (like those in Fig. 2) to be more meaningful, easier to understand, and easier to identify with than bar charts. Women in one of the focus groups, which had a lower mean age and educational level, perceived risk of breast cancer as larger when it was shown as a part-to-whole human icon display than when it was shown as a part-to-whole bar graph; however, no quantitative results were collected. Some participants said the icon display suggested population risk, while a continuous scale suggested personal risk. In direct contrast, Royak-Schaler et al. found that focus groups of low-income women preferred a part-to-whole bar chart (similar to Fig. 3) to a part-to-whole icon array; however, in this study the bar chart had evaluative labels (‘high risk,’ ‘low risk’, or ‘average risk’), but the icon array did not, which may have made it difficult for viewers to place the risk in context.

Random Arrangement in Icon Arrays

In the Stone, Fagerlin, and Schapira studies, the icons depicting people affected by disease were arranged in a row or a block, so risks could be estimated by judging block length or area (Figs. 2 and 3). The Royak-Schaler icon array also did so, but the human figures were staggered rather than arranged in neat rows, which could have made it somewhat more difficult to use visual area judgment. Other arrays show those affected scattered throughout the array (Fig. 4). In genetic risk counseling, such a random arrangement has been described as helpful in promoting understanding of chance. With random icon arrays, area judgment is not available to the viewer. It is possible that comparisons are made through a gray-scale judgment (one of the least accurate of the visual perception tasks), through mental summation of areas (also relatively inaccurate), or by counting or computing. Viewers were less accurate at estimating proportions in random arrays than in sequentially arranged ones. In the Schapira et al. study, women disliked the random arrangement because they could determine the probability only by counting. However, some women said it better conveyed the idea of randomness.

Risk Tables, Ladders, and Scales

A risk table, ladder, or scale depicts a range of risks from very low to very high as context for an individual risk. When risks are ranked vertically in a table, a graphic is called a risk ladder, and when they are horizontal, the graphic is called a risk scale (Fig. 5 and 6) or visual analog scale. Because position on a risk ladder or scale is evaluated as distance from a baseline, Lipkus and Hollands propose that they exploit the most efficient of the Cleveland and McGill basic visual perception skills.

Risk ladders and scales often provide information about other risks for comparison. A horizontal scale was used to compare the risks from a blood transfusion (such as contracting HIV) with other hazards such as the annual chance.
**Figure 5.** Risk scale with comparative risks. This scale was used for demonstrating risk magnitudes; how viewers interpret the logarithmic scale is not known. Reprinted with permission from Blackwell Publishing from page 787 of Lee DH, Mehta MD. Evaluation of a visual risk communication tool: effects on knowledge and perception of blood transfusion risk. Transfusion 2003;43:779–87.

Place an x in either the magnifying glass or the lower part of the scale to describe the chance of each event happening to you.

**Figure 6.** Magnifier risk scale. This scale was used for eliciting risk perceptions. The magnifying lens at the low end allowed users to respond with smaller values for very low risks; a second study suggests that it reduces the magnitude of higher risks as well. From Woloshin S, Schwartz LM, Byram S, Fischhoff B, Welch HG. A new scale for assessing perceptions of chance: A validation study. Med Decis Making 2000;20(3):298–307. Copyright 2000 by Sage Publications. Reprinted by permission of Sage Publications, Inc.
of dying in a car accident (Fig. 5). The ladder was as effective as numbers alone in increasing knowledge and reducing dread about rare hazards of transfusion. This scale, the Paling Perspective Scale, depicts a range of probabilities on a logarithmic scale from 1 in 1 (certainty) to 1 in 1 trillion, centered on 1 in a million, described as “effective zero.” The rationale for choosing the comparative risks was not described; some were familiar, others unfamiliar, some cumulative, others one-time.

More systematic explorations of risk ladder design have been done in environmental health. When a person’s exposure to an environmental hazard was explained by referring to a location on a risk ladder, perceived risk was associated with location on the ladder rather than numerical magnitude of the risk. The ladders also illustrated unfamiliar concepts with text and graphics, such as icon arrays of the number of cigarettes needed to produce a cancer risk comparable to a given level of radon risk. Another way to place unfamiliar concepts in context was to place a pointer at the level of risk at which protective actions are recommended; this helped viewers determine that levels of risk below this threshold are not very serious. Johnson and Slovic compared numbers and a risk ladder for communicating uncertainty (confidence intervals) about risk estimates. The numbers were ratios with different denominators, and the graphic showed no part-to-whole information. When compared to numbers alone, the ladder did not affect perceived risk; it did decrease trust in the information but also improved peoples’ ability to notice the full range of possible risks.

**Survival and Mortality Curves**

Survival and mortality trends are cognitively complex because they involve changes over time. In a classic study, McNeil et al provided numeric information about a treatment with high short-term mortality but good long-term and median survival (surgery) and a treatment with good short-term but worse long-term and median survival (radiation). When information was framed in terms of mortality, more people chose radiation, perhaps because the short-term mortality appeared more salient in this frame. Framing effects were nearly as strong with physicians as with laypeople.

When survival or mortality data are presented graphically (Fig. 7), changes in risk are inferred from curve slopes. Part-to-whole relationships are available, though not very salient, when the y axis extends to 100%. Because line graphs portray data points as a single visual element, viewers do not need to integrate the information themselves; line graphs thus help experts perform complex tasks such as assessing rates of change. However, Armstrong et al. showed that only 74% of a sample recruited from a jury pool could interpret a survival curve well enough to determine the number of survivors at various time points, and only 55% could calculate the difference in survival between two time points. After a training exercise, ability to interpret survival at one time point improved but accuracy in calculating differences was unchanged. The effect of learning can also be inferred from older studies in which choice of treatments was strongly affected by the amount of instruction in interpreting survival curves. With minimal explanation, patients tended to choose the treatment with better long-term survival; patients given extensive explanations were more likely to take medium-range outcomes into account. Physicians were more likely to be influenced by middle parts of the curves than were patients, which could also be due to education. Survival curves may reduce a tendency to overweight immediate survival by drawing attention to longer-term outcomes. When patients were given a choice between treatments described with survival percentages, 59% chose the treatment with better immediate survival; this dropped to 34% when they viewed a pair of survival curves. The order in which patients viewed survival graphs significantly affected their preference for short-term versus long-term survival. More educated patients were less likely to choose the treatment with better short-term survival.

Participants answered comprehension questions more accurately with survival curves or both survival and mortality curves than with mortality curves alone; the effect was strongest in the lowest educational group and among nonwhites. In this study, participants were asked to imagine being at high risk for colon cancer and were given a choice between colectomy and an easier but less successful alternative (annual exam). They were less likely to choose colectomy when viewing mortality curves; the effect might...
have been due to the reduced understanding of the information with the mortality curves.26

In another study, people were influenced more by the distance between curves than the numerical differences (Fig. 7).69 If a pair of 15-year survival curves are displayed on an x-axis of a certain length, they will diverge more than if the first 5 years of data are stretched and displayed on an axis of the same length.70 This flattening effect markedly reduces the difference between peoples’ estimates of treatment effectiveness.69

In a study of women with BRCA1/2 mutations who were deciding on possible prophylactic measures against breast cancer, those who received a set of personalized survival curves were more satisfied with their decisions than those who received a similar educational booklet without survival curves.40 However, the survival curves did not change their actual decisions.

**Persuasiveness and Comprehension**

In one study, text descriptions of statistical data about interactions between disease and genetics were better understood and perceived as higher quality evidence than bar charts of the same data.41 Poor comprehension was associated with impressions that the evidence was of poor quality and was not persuasive. This study should be interpreted in light of a difference between the graphs and texts that may have made the graphs harder to understand: the graphs displayed the “relative mortality rate,” explained as “the actual number of deaths divided by the expected number of deaths,” whereas the text described one group of people as “about 20% more likely to die early deaths” than the other.

**Data Scale**

Patients’ descriptions of risk may differ if elicited with discrete scales (e.g., a number affected out of 1000 people, or icons), ordinal ones (low, medium, and high), or continuous ones (e.g., on a scale from 0% to 100%). Women estimating their risk of breast cancer provided different estimates when using an icon display (a grid of 100 female figures) and a continuous scale (a horizontal line anchored at 0% and 100%).28 Icons elicited risk estimates that were higher and farther from epidemiological risk as assessed by the Gail model.71 Woloshin et al. asked participants to rank the likelihood of several health events, then asked them to describe each event’s likelihood with words (an ordinal scale ranging from “not at all likely” to “extremely likely”), numbers (a “1-in-x chance”), a horizontal risk scale ranging from 0 in 100 to 100 in 100, and another scale supplemented with an image of a magnifying glass to illustrate probabilities smaller than 1% (Fig. 6).27 Rankings with the verbal scale were the most reliable, usable, and strongly correlated with participants’ rankings, and rankings with the “1-in-x” numbers had the worst performance. This result is consistent with other studies of the reliability and usability of word scales.72,73 The magnifier scale had slightly lower correlations and was perceived as less usable than the verbal scale but permitted people to make lower estimates for very rare events. A subsequent study by another research group compared the magnifier with the standard horizontal risk scale.74 This work confirmed that the magnifier scale enabled appropriately low estimates for very rare events but also showed that it substantially lowered risk estimates for more common events.24 This effect was seen when participants estimated risks of various health events without being given numeric information about the magnitudes of those risks. Licensed anglers were shown risk ladders describing the hazards of eating contaminated fish in discrete numbers or ordinal categories (“higher risk”, “moderate risk”, and “lower risk”).75; 57% preferred the quantitative ladder. In the study of willingness to pay for safer products, icon arrays and bar charts produced similar results, suggesting that level of measurement made no difference.53

A qualitative study of 40 women found that simple bar charts depicting absolute lifetime risk of various events were preferred over line graphs, thermometer graphs, icon arrays, and survival curves.98 Participants wanted graphics to be supplemented with text.

**Multiple Quantitative Endpoints**

Two series of experiments using quantitative reasoning endpoints confirm the applicability of the basic graphic perception findings to health settings.10,76 Feldman-Stewart et al. assessed speed and accuracy of students’ and patients’ judgments with 6 data formats: vertical part-to-whole bar chart, horizontal part-to-whole bar chart, pair of numbers, part-to-whole icon graphic with random arrangement, icon graphic with the icons arranged in a block, and pie chart.26 Participants were slowest and least accurate at judging the larger of two quantities with the pie chart and the random-arrangement icons. Estimates of the differences between quantities were best with number pairs and sequentially arranged icons. Participants performed no better with their preferred formats.25 Patients took longer than students.

One trial compared graphics for conveying risks to physicians.24 Physicians saw data from a fictitious clinical trial in which one treatment had a high failure rate. Clinical trials may be halted midcourse if results in one group are much worse than in the other; the physicians were asked if the data warranted halting the trial. Five formats were given (tables of success rates, tables of failure rates, pie charts, stacked bar charts, and icon arrays). Most noticed the high failure rate in icon arrays and deals; fewer did with pie charts or stacked bar graphs.24 However, most liked the bar graphs and disliked the icon array. These results are consistent with research showing that proportions are difficult to judge when mental summation is required,11 although not with a finding that pie charts were superior when mental summation of slices was required.76 The authors suggest that the icons’ success was due to the framing effect of drawing attention to the failures, which would be consistent with the Stone et al. ‘foreground effect’.24 Another explanation is that the discrete icons could be counted, but the other graphs required area estimation; however, earlier findings that results with bar charts were the same as results with icon arrays54 argues that viewers may not be using counting as a strategy.

**Discussion**

The best design for a graphic depends upon the purpose of the risk communication. Some communications are intended to enhance quantitative understanding or promote good arithmetic judgments, whereas others are intended to promote behavior change.
For good quantitative judgments, the size of a graphic element should be proportional to the number it portrays. When the size diverges from the number, people are more influenced by the size than by the number. For example, quantitative accuracy is best when numerator and denominator of ratios are both visually salient. Part-to-whole bar charts and part-to-whole sequentially arranged icon arrays probably invoke automatic visual area processing and proportion judgments and can be used to help viewers attend to the mathematical proportion. This may help them de-emphasize the emotional content of accompanying text. With experts and lay users given some instruction, survival curves can help draw attention to information that is otherwise ignored, such as middle-term outcomes.

Patients can recognize proportions fairly successfully with part-to-whole sequential icon arrays. By contrast, proportions are difficult to assess in randomly arranged icon arrays and possibly also when the icons are jittered. This could account for the dislike of random-arrangement arrays found in qualitative studies. Thus, sequentially arranged icon arrays may be better than random ones in any situation that requires the viewer to estimate a proportion or compare two proportions. Additional work may be needed to confirm the hint in some studies that randomly arranged icon arrays help convey the difficult concept of chance or uncertainty.

Relatively few studies have attempted to express the even more difficult concept of uncertainty around a probability estimate (confidence intervals). Communicating uncertainty in risks should be a topic for continuing study, given older findings that laypeople are often unfamiliar with the concept of scientific uncertainty.

Graphs emphasizing the numerator of a risk ratio are more likely to promote risk behavior changes. When part-to-whole information is not available, as in the Stone et al. studies, icon arrays and bar charts apparently draw attention to the numerator; when this numerator depicts adverse events, viewers make more risk-averse choices than they do with the numbers alone. A graphic that shows only the numerators of two risks is analogous to the epidemiologic quantity of the relative risk. The part-to-whole graphs depict numerator and denominator with equal salience and are analogous to an incidence or rate measure. Providing relative risks without absolute risks has long been known to inflate the apparent magnitude of the risk difference, even with educated audiences.

Bar charts, risk ladders, scales and sequentially arranged icons have been used successfully to help viewers place individual risks in context of other risks or make specific comparisons between risks. Perceptions are strongly influenced by the design of graphics. Magnifying the low end of a risk scale to call attention to very small probabilities reduces the perceived size of low risks as well as higher risks. If the scale of a ladder is altered so that a particular risk is closer to the high end of the ladder, this inflates viewers' perception of that risk.

Studies that ask patients to express information graphically produce somewhat different results from studies of comprehension. For example, viewers could match a numeric proportion to an icon array with that proportion colored in, but were relatively inaccurate when asked to mark the proportion on a blank icon array. When patients express perceived risks on different types of graphic devices, they give inconsistent results. The types of graphic best for demonstrating information to patients may be different from the types best for eliciting patient perceptions.

Qualitative research is important to learn more about how patients interpret graphs, but relying too heavily on patients' likes and dislikes may pose a problem because they sometimes like graphics that lead to poor quantitative judgments. For example, viewers appear to like graphics that are simpler, with fewer visual elements (for example, arrays with few icons rather than arrays with many icons and bar charts with verbal ordinal categories rather than icon arrays). However, it is not clear that such simple graphics are successful at conveying complex information. In fact, Schapira et al. found that women viewing the simpler array appeared to have an inflated perception of the risk it portrayed. Elting et al. found that doctors performed worst with the format they liked best, and best with the one they strongly disliked, and Feldman-Stewart found that laypeople made similar judgments whether they used a format they liked or one they disliked. Focus groups liked human figures in graphics even though a different research group found that replacing human figures with asterisks in an icon array produced no difference in participants' decisions. Personalized survival curves improved satisfaction with choices but did not affect the choices. Parrott et al. suggest that people process familiar graphic forms through learned heuristics rather than through comprehension of the information, resulting in gaps between the intended meaning and the meaning constructed by the viewer; such heuristics could also produce a preference for a particular graphic form whether or not it results in good comprehension. Parrott et al. also point out that risk graphics often reduce complex multivariate relationships between hazards and multiple risk factors to simplistic unidimensional relationships; if viewers believe the true relationship is multivariate, they may dismiss the graph as noncredible.

Future research might help develop graphics that are both acceptable and successful in promoting quantitative judgments or behavioral outcomes.

Interactions with education level, literacy, numeracy, and culture are also likely to be fruitful continuing areas of research. Although graphs often seem to be more intuitive than words, the literature shows that graphical literacy is strongly affected by expertise and familiarity with specific graphical formats. Patients may require instruction to be able to interpret certain formats, such as survival graphs, and they want textual explanations for illustrations. Instruction may also improve comprehension of many other formats, even familiar ones such as bar charts; speed and accuracy of judgments are worse among novices than among experts. Future research should also integrate the literature on comprehension of different number formats (e.g., percentages versus rates) to avoid confusing from the use of hard-to-understand numbers in graphs. If patients do not fully understand what they are seeing, they may not find the information credible or persuasive. Visual framing and order effects may be stronger.
with less educated viewers, and lower educational level may be associated with mistrust of depictions of scientific uncertainty. Parrott et al have suggested that the historical use of statistics to support discriminatory theories might lead African-Americans to be suspicious of statistics regardless of how they are presented. Such issues must be explored in continuing research among culturally and educationally diverse participant groups. Better methods for communicating risk can help patients integrate risk data into genuinely informed decisions about health care.

References

70. Lau EW, Ng GA. Visual illusions created by survival curves and the need to avoid potential misinterpretation. Medical Decision Making. 2002;22:238–44.