

Using Icon Arrays to Communicate Medical Risks: Overcoming Low Numeracy

Mirta Galesic

Max Planck Institute for Human Development

Rocio Garcia-Retamero

University of Granada

Gerd Gigerenzer

Max Planck Institute for Human Development

Objective: Icon arrays have been suggested as a potentially promising format for communicating risks to patients—especially those with low numeracy skills—but experimental studies are lacking. This study investigates whether icon arrays increase accuracy of understanding medical risks, and whether they affect perceived seriousness of risks and helpfulness of treatments. **Design:** Two experiments were conducted on samples of older adults ($n = 59, 62$ to 77 years of age) and university students ($n = 112, 26$ to 35 years of age). **Main Outcome Measures:** Accuracy of understanding risk reduction; perceived seriousness of risks; perceived helpfulness of treatments. **Results:** Icon arrays increased accuracy of both low- and high-numeracy people, even when transparent numerical representations were used. Risks presented via icon arrays were perceived as less serious than those presented numerically. With larger icon arrays (1,000 instead of 100 icons) risks were perceived more serious, and risk reduction larger. **Conclusions:** Icon arrays are a promising way of communicating medical risks to a wide range of patient groups, including older adults with lower numeracy skills.

Keywords: icon arrays, risk communication, older population, numeracy skills

Icon arrays are graphical representations consisting of a number of stick figures, faces, circles, or other icons symbolizing individuals who are affected by some risk (for an example, see Figure 1). Such displays seem to be gaining in popularity both in medical practice and in public media (Ancker, Senathirajah, Kukafka, & Starren, 2006; Edwards, Elwyn, & Mulley, 2002; Elmore & Gigerenzer, 2005; Paling, 2003). However, experimental research on icon arrays is still scarce, and there are many open questions related to their use regarding effectiveness and impact.

Because many people have difficulties using numbers and processing elementary probability expressions (Kutner, Greenberg, Jin, & Paulsen, 2006; Lipkus, Samsa, & Rimer, 2001; Peters et al., 2006; Schwartz, Woloshin, Black, & Welch, 1997; Tuijnman,

2000), visual displays have been proposed as a potentially promising method for communicating medical risks to such low-numeracy people (Ancker et al., 2006; Lipkus & Hollands, 1999; Montori & Rothman, 2005). In particular, icon arrays seem to be promising for communicating risk reduction as a result of medical screenings and treatments (e.g., Fagerlin, Wang, Ubel, 2005; Feldman-Stewart, Kocovski, McConnell, Brundage, & Mackillop, 2000; Paling, 2003). However, to the best of our knowledge, no study so far has explicitly tested this promise. As numeracy skills are correlated with education and possibly decline with age (e.g., Davids, Schapira, McAuliffe, & Nattinger, 2004; Hibbard, Peters, Slovic, & Finucane, 2001), such people may be disadvantaged in general understanding of complex medical information, not only numerically presented data. Therefore, the first question we address in this study is whether icon arrays improve accuracy of understanding of numerically presented risks for both high- and low-numeracy people.

Another issue is whether icon arrays can help even when information is presented in numerical formats that are relatively easy to understand. It has been shown that large improvements in understanding can be gained even with numerical representations, simply by using absolute risk reduction (ARR) instead of relative (RRR) risk reduction (e.g., Bucher, Weinbacher, & Gyr, 1994; Forrow, Taylor, & Arnold, 1992; Hux & Naylor, 1995; Sarfati, Howden-Chapman, Woodward, & Salmond, 1998). Additional visual representations might be ineffective, or even make the presentation too complex and lead to lower accuracy. Therefore, our second question is whether icon arrays lead to additional improvements in accuracy even when transparent numerical representations are used.

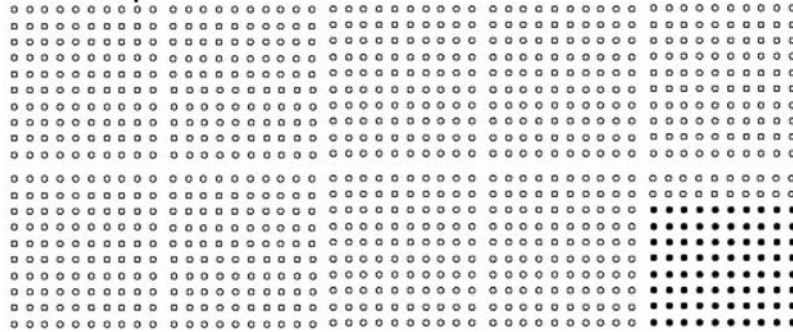
Mirta Galesic, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany; Rocio Garcia-Retamero, Department of Experimental Psychology, University of Granada; Gerd Gigerenzer, Center for Adaptive Behavior and Cognition and Harding Center for Risk Literacy, Max Planck Institute for Human Development, Berlin, Germany.

This study is part of the project “Helping People With Low Numeracy to Understand Medical Information” funded by the Foundation for Informed Medical Decision Making and the Max Planck Society. We are grateful to Gregor Caregnato for supervising the lab experiments and Anita Todd for editing the manuscript.

Correspondence concerning this article should be addressed to Mirta Galesic, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: galesic@mpib-berlin.mpg.de

For people with symptoms of arterial disease, aspirin can reduce the risk of having a stroke or heart attack by 13%.

Without aspirin



With aspirin

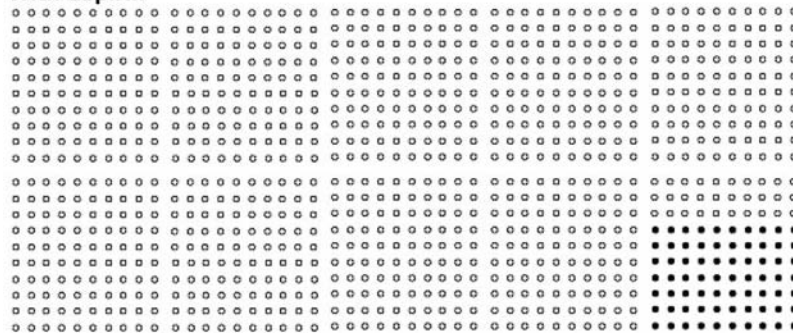


Figure 1. Example of a condition from Experiment 1: Numerical information about relative risk reduction and additional visual information (icon array).

Some studies suggest that presenting risks by means of icon arrays instead of numbers might increase perceived seriousness of risk. Presenting information about risks on the level of individuals might make it easier to imagine scary outcomes (Slovic, Monahan, & MacGregor, 2000; Slovic, Peters, Finucane, & MacGregor, 2004). On the other hand, icon arrays could also draw attention to a large number of unaffected people and thus make risks seem less serious than when they are in numerical form. For instance, when icon arrays were used to present only the number of drivers who experienced traffic accidents due to bad tires, but not the much larger number of drivers who did not have an accident (Schirillo & Stone, 2005; Stone, Yates, & Parker, 1997), an increase in willingness to pay for better tires was found. However, when pie charts and bar graphs were used (Stone et al., 2003) to present both people with and people without traffic accidents (i.e., both the numerator and the denominator of the risk ratio), the opposite effect was found: Numerically presented risks were now perceived as more serious than the equivalent risks presented visually. To the best of our knowledge, a similar test was not done with icon arrays. Therefore, within our third research question, we test two opposing hypotheses about the effect of icon arrays: One proposes that they will increase and the second that they will decrease the perceived seriousness of risk compared to numerical representations. We examine two types of risk perception relevant in a medical context: seriousness of baseline risks and perceived risk reduction, the latter operationalized as helpfulness of treatments.

A qualitative study (Schapira, Nattinger, & McHorney, 2001) found that otherwise identical risk was perceived to be lower when it was presented with a larger rather than with a smaller number of icons. This result is in contrast with the so called *ratio-bias effect*—the finding that equivalent likelihoods can be perceived as larger when presented with larger rather than smaller denominators (cf., Denes-Raj, Epstein, & Cole, 1995; Reyna & Brainerd, 1994). For instance, participants in a study conducted by Yamagishi (1997) rated the likelihood of a cancer described as killing 1,286 out of 10,000 people as higher than when they were told the cancer kills 24.14 out of 100 people. Rudski and Volksdorf (2002) tested this phenomenon using icon arrays and found similar effects, in line with the ratio-bias phenomenon. However, they did not use health-related stimuli: Their participants were asked to judge the likelihood that they will pick a black jelly bean from a tray containing different numbers of black and white beans. To the best of our knowledge there is no experimental research on ratio-bias effects in icon arrays presenting medical information. In addition, most of the existing studies of ratio-bias effect did not involve health-related problems. Therefore, our fourth question is whether icon arrays with different total number of icons lead to changes in perception of otherwise equivalent risks. We study both changes in baseline risks and changes in risk reduction. The same baseline risk presented with a larger overall number of icons might suggest higher risk because more people seem to be affected. Similarly, the same risk reduction might appear larger when presented with a

larger number of icons because more people “become healthy” as a result of the treatment.

In sum, we investigate four questions about icon arrays. First, can icon arrays help people with low numeracy skills understand medical risks more accurately? Second, can icon arrays increase accuracy beyond the improvement provided by transparent numerical representations such as absolute (instead of relative) risk reduction? Third, do icon arrays increase perceived seriousness of risks and helpfulness of treatments relative to numerical representations? Fourth, does overall number of icons affect the perception of risks and risk reduction? In this paper, we describe two experiments designed to shed light on those questions.

We focus particularly on studying how useful icon arrays are for presenting quantitative risks to low numeracy people. Numeracy skills are part of a more general concept of quantitative or mathematic literacy. For instance, PISA (Programme for International Student Assessment) studies assess four broad areas of mathematical literacy: having a basic number sense, recognizing geometric patterns, understanding functional relationships, and understanding the concept of uncertainty (PISA, 2003). The National Assessment of Adult Literacy Survey assesses various aspects of quantitative literacy that one needs for everyday tasks like shopping or calculating postage (Kutner, Greenberg, & Baer, 2006). In the area of health, different aspects of numeracy—such as computational, analytical, and statistical numeracy—are considered to be important for making effective medical decisions (Golbeck, Ahlers-Schmidt, Paschal, & Dismuke, 2005; Speros, 2005). As we study understanding of risks and risk reductions, we focus primarily on statistical numeracy, including understanding the concept of a random toss and knowing how to perform elementary calculations with percentages (Lipkus et al., 2001; Schwartz et al., 1997).

Numeracy, and literacy skills in general, are often lower in older adults (e.g., Kutner et al., 2006; Tuijnman, 2000). To assess both ends of the spectrum of numeracy skills in the population within the limits of our laboratory, we conducted our experiments on two very different groups of participants: students and older adults. If numeracy skills influence people’s understanding of risks even in such disparate groups, this would highlight the importance of tailoring risk communication to the level of patients’ numeracy skills.

Method

We conducted two experiments, described in more detail in the sections below. Both were pretested on a sample of 135 students from the University of Granada, Spain. The results of the pretest (available from the authors) encouraged us to proceed with the main study, conducted in our lab at the Max Planck Institute for Human Development in Berlin, Germany, with two groups of participants: a group of 59 older adults (49% age 62 to 69 years, 51% age 70 to 77 years; 49% women and 51% men; 57% with high school or lower education, 43% with college or university education) and a group of 112 university students (63% age 18 to 25 years, 37% age 26 to 35 years; 57% women and 43% men). They received 10 euros for their participation. The questionnaire was administered on computers in our lab. Besides the two experiments described here, it included other questions about medical risks and numeracy skills, and took 1 hr to complete.

Measurement of Numeracy

All participants completed the numeracy scale consisting of 11 items from one scale (Lipkus et al., 2001) and one additional item from another (Schwartz et al., 1997; the item involving a coin toss), for a maximum score of 12. Examples of items are “Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?”; “Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%?”; and “If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1,000?”

The students had notably higher numeracy scores than the older participants (M_s 10.3 vs. 8.6; SD_s 1.78 vs. 2.51). Differences still appeared when the students were compared only with the highly educated older participants, whose mean score was 9.4 (SD 2.08; lower educated older person: M 8.0, SD 2.65). This finding was expected, as we know from previous research that even higher educated older persons have relatively low numeracy skills (Lipkus et al., 2001) and lower quantitative literacy in general (e.g., Kutner et al., 2006; Tuijnman, 2000). In the analyses that follow, we split the participants in each age group according to their group’s median numeracy scores (see Peters et al., 2006, for similar procedure). Among students, the low-numeracy group includes participants with 10 or fewer correct answers, whereas the high-numeracy group includes those with 11 or 12 correct answers. Among the older adults, the low-numeracy group has up to 8 correct answers, and the high-numeracy groups 9 or more correct answers. This means that we treat each participant’s numeracy skills as relative to their particular reference group (students or older adults). Had we split the respondents according to the median of all participants, we would have had disproportionately more students in the high numeracy group.

Role of the Funding Sponsor

This study is part of the project “Helping People With Low Numeracy To Understand Medical Information” funded by the Foundation for Informed Medical Decision Making and the Max Planck Society. There were no conflicts of interest regarding this study.

Experiment 1

Goals

This experiment was designed to give answers to the first two questions described in the Introduction: Whether icon arrays improve accuracy of risk understanding in low-numeracy people, and whether they lead to additional improvements in accuracy even if we use transparent numerical representations such as absolute risk reduction.

Procedure

The participants were presented with three medical scenarios: one involving the usefulness of aspirin in reducing the risk of stroke or heart attack for people with symptoms of arterial disease (ARR from 8% to 7%, or RRR of 13%); another about a hypothetical new drug (modeled after the drug Simvastatin; cf. Skol-

bekken, 1998) for reducing cholesterol that also decreases the risk of dying from a heart attack (ARR from 8% to 5%, or RRR of 38%); and the third about a fictitious case of screenings for early signs of acute appendicitis (ARR from 8% to 3%, or RRR of 63%). Order of tasks was randomized.

There were two between-subjects independent variables. First, half of each group of participants randomly got the numerical information in the form of either absolute or relative risk reduction. For instance, in the case of aspirin, participants in the RRR condition got the following text: "For people with symptoms of arterial disease, aspirin can reduce the risk of having a stroke or heart attack by 13%." Those in the ARR condition received "For people with symptoms of arterial disease, aspirin can reduce the risk of having a stroke or heart attack: 8% of such people who did not take aspirin had a stroke or heart attack, compared to 7% of such people who did take aspirin."

Independently of this manipulation, a random half got—in addition to the numerical information—two icon arrays, one presenting baseline risk of the disease without treatment, and the other presenting the risk with treatment. All icon arrays contained 1,000 circles. We used circles because previous research (Stone et al., 1997) did not find differences in effects of arrays with faces compared to more abstract symbols; in addition, faces looked messy when decreased to fit in a large array. Affected individuals were shown as black circles at the end of the array. An example of the condition involving icon arrays is shown in Figure 1 (original material was in German).

Accuracy of risk understanding was assessed with two questions, following the procedure used by Schwartz et al. (1997). In the case of aspirin, participants were first asked how many of 1,000 people with symptoms of arterial disease might have a stroke or heart attack if they do not take aspirin. Second, they were asked how many of 1,000 people with symptoms of arterial disease might have a stroke or heart attack if they do take aspirin. The first answer divided by 1,000 represents the estimated absolute risk. By

deducting the second from the first answer and dividing it by the first, we transformed the estimates into a relative risk reduction and coded it as correct if it was within $\pm 3\%$ of the right answer. This criterion was determined based on the distribution of errors, which peaked within this interval and became flat quickly after passing its limits. It appeared as if the participants either knew how to calculate the correct answer and did it correctly or with minor rounding errors, or they did not know how to calculate it at all.

As there were no differences in effects of presentation format between different tasks (about aspirin, cholesterol drug, and appendicitis screening), we analyzed them together. To assess the joint effects of presence versus absence of icon arrays, numerical format, age group, and relative numeracy (lower vs. higher than the age group median) on answering at least one of the three tasks correctly, we used analysis of variance (ANOVA), following Snedecor and Cochran (1967) who showed that this method can be used to get conservative results for large samples of a dichotomous dependent variable.

Results

Do icon arrays improve accuracy of estimated risk reduction in low-numeracy people? Figure 2 shows the percentage of participants who gave accurate estimate of risk reductions in at least one of the tasks. The figure suggests that icon arrays were especially useful for those students and older adults whose numeracy was low compared to the average of their groups. This is consistent with the results of an ANOVA, interaction Icon Array \times Numeracy: $F(1, 161) = 4.1, p = .04, \eta^2 = .03$. As Figure 2 shows, this effect was of similar size for students and older adults, interaction Icon Array \times Numeracy \times Age, $F(1, 161) = 1.2, p = .27, \eta^2 = .01$. Numeracy, $F(1, 161) = 5.7, p = .02, \eta^2 = .03$; but not age, $F(1, 161) = 0.4$, had an effect on accuracy: Participants with higher relative numeracy were more accurate in estimating relative risk reductions and this effect was independent of age.

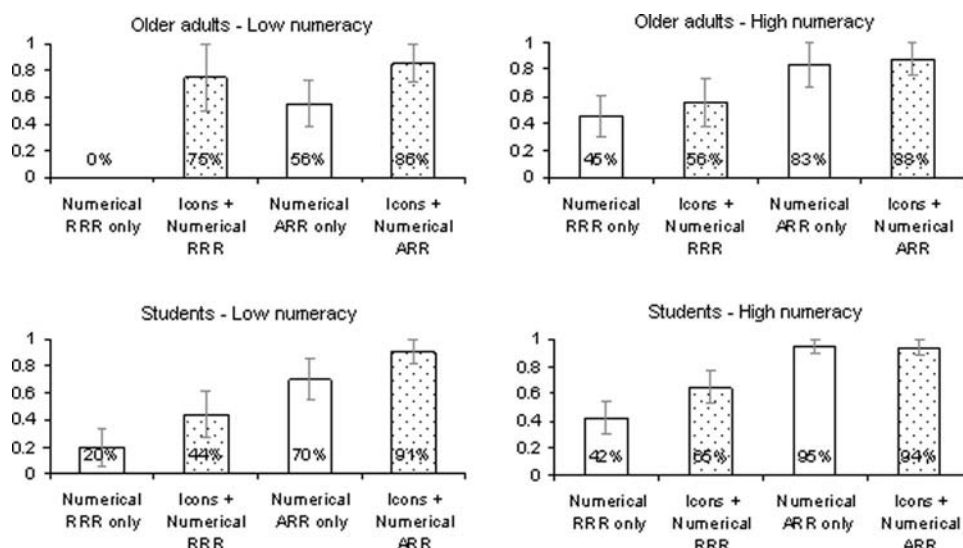


Figure 2. Percentage of older adults and students, with relatively low and high numeracy skills, who gave correct estimates of risk reduction for at least one task, by experimental condition. Error bars indicate \pm one standard error. RRR = relative risk reduction; ARR = absolute risk reduction.

Do icon arrays lead to additional improvements in accuracy even when transparent numerical representations are used? As shown in Figure 2, large improvements in accuracy were achieved when numerical information was presented in terms of absolute instead of relative risk reduction (ARR vs. RRR). However, icon arrays were a useful addition to both types of numerical representations. An ANOVA indicates that both numerical format, $F(1, 161) = 39.9, p = .001, \eta^2 = .20$; and icon arrays, $F(1, 161) = 10.3, p = .002, \eta^2 = .06$; had independently a positive effect on the accuracy, interaction Icon Array \times Numerical format: $F(1, 161) = 2.0, p = .16, \eta^2 = .01$. The relatively strong effect of icon array compared to that of the interaction of icon arrays and numerical format indicates that icon arrays were helpful not only for the condition with relative risk reductions, but also for the condition involving absolute risk reductions.

For both the students and the older adults, proportion of women was higher in the low compared to the high numeracy groups (73% vs. 47% for students, and 60% vs. 35% for older adults). Therefore we rechecked our results by including sex and its interaction with numeracy skills in the analysis. All conclusions remained unchanged. Women were somewhat less accurate on average (58% of them gave correct estimate for at least one task, compared with 70% of men) $F(1, 159) = 3.2, p = .08$, but this difference was independent of numeracy.

For the older adults, low and high numeracy groups also differed in education: 30% of those in the low numeracy group had higher education, compared to 53% of those in the high numeracy group. Therefore we rechecked our results by adding education and its interactions with sex and numeracy in the model for the older adults. Higher education led to increased accuracy (correct estimates were given by 52% vs. 79% of participants with lower vs. higher education, respectively) $F(1, 48) = 4.63, p = .04$, but showed no interactions with sex or numeracy.

Experiment 2

Goals

In this experiment we conducted a more systematic investigation of the effects of icon arrays on perception of risk and risk reduction: We studied whether icon arrays increase perceived seriousness of risks and helpfulness of screenings relative to numerical representations, and whether overall number of icons affects these perceptions.

Procedure

We used four tasks, each about a different fictitious type of cancer, denoted only by letters of the alphabet (S, T, U, W). This allowed us to investigate how different properties of icon arrays affect perception of risk without the contamination from possible preexisting knowledge about diseases and treatments. Two tasks had the baseline risk of cancer set to 10%, and two to 30%. Independently of this, two tasks had a 20% risk reduction (e.g., from 10% to 8%), and two a 60% reduction (e.g., from 10% to 4%).

There were two between-subjects independent variables. First, presentation format was either exclusively numerical (ratios of whole numbers) or exclusively visual (icon arrays). Second, the denominator (overall number of icons) was set to either 100 or

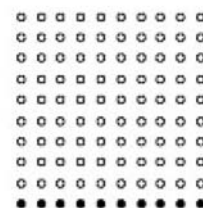
1,000 people. An example of the numerical condition with 100 as the denominator is: "Ten of 100 people who do not get screened die of cancer S. Eight of 100 people who do get screened die of cancer S." An example of the visual condition for the same task is shown in Figure 3 (original material was in German). Assignments to experimental conditions were independent for each of the four tasks, and the order of tasks was randomized.

Each task was followed by two questions. First, perception of risk was assessed with the question "Without screening, how serious does the risk of cancer S (T, U, V) seem to you?" Second, perception of risk reduction due to screening was evaluated by asking "How helpful do you think the screening is in reducing the risk of dying of cancer S (T, U, V)?" The latter question examines perceived helpfulness of treatment, which is likely to be closely related to the perception of risk reductions, and is at the same time more natural to ask than a direct question about the perceived size of a certain reduction (e.g., how large does the reduction of 20% seem). Each question was answered on a 15-point bipolar scale with endpoints labeled (for perception of risk, 1 = not at all serious–15 = very serious; for helpfulness of treatment, 1 = not at all helpful–15 = very helpful).

As mentioned, this experiment included four tasks, each for a different (fictitious) type of cancer. As the between-subjects independent variables affected answers to all tasks in similar ways, below we present averages across the tasks. To analyze the effects of visual versus numerical format and large versus small denominators, we used mixed linear models that accounted for the clustering due to repeated measurements on the same subjects. We tested for the interactions of all findings with age group (older adults vs. students), relative numeracy skills, sex, size of baseline risk, and size of risk reduction; as none were found, we do not present those results here. We present the results of simple mixed

In the following picture, black circles show people who will die of cancer S - with and without screening.

Without screening



With screening

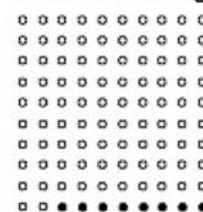


Figure 3. Example of a condition from Experiment 2: Icon arrays used to present baseline risks and risk reductions, based on 100 people.

linear models including a single factor (format or denominator size).

Results

Do icon arrays increase perceived seriousness of risk and helpfulness of treatment compared to numerical representations? Baseline risks of different cancers were perceived as more serious when presented in numerical rather than in visual form: average evaluations were 10.3 and 9.1 on a 15-point scale, respectively, Cohen's $d = .21$; $t(676.1) = 4.38$, $p = .001$. Similarly, risk reductions (i.e., helpfulness of screenings) were perceived to be higher when presented numerically instead of via icon arrays: average evaluations were 10.1 and 9.2, respectively, $d = .16$; $t(647.2) = 2.79$, $p = .005$. This finding is in line with earlier results (Stone et al., 2003): when people can see both numerator and denominator of a risk ratio (i.e., icons showing both the affected and the unaffected individuals), visual displays do not necessarily produce the affect-laden mental imagery that is thought to cause higher perceived risk (cf., Slovic et al., 2005).

Does the overall number of icons affect risk perception? For both the numerical and the visual format, perceived seriousness was higher when larger denominators were used. Here we focus on the conditions involving visual presentations only. Perceived seriousness of baseline risks was higher when arrays contained 1,000 rather than 100 icons: average evaluations were 9.3 versus 8.8 on a 15-point scale, respectively, $d = .13$; $t(669.0) = 1.45$, $p = .15$. The effect was more pronounced for the assessment of helpfulness of screenings: They were perceived to be more helpful with the larger overall number of icons. Average evaluations were 9.7 and 8.9, respectively, $d = .21$; $t(645.2) = 2.96$, $p = .003$. These results are consistent with studies using stimuli unrelated to health (Denes-Raj et al., 1995; Rudski & Volksdorf, 2002).

Discussion

In this study, we conducted experimental tests of the claims that icon arrays help low-numeracy people understand risk reductions accurately (cf., Lipkus & Hollands, 1999; Montori & Rothman, 2005; Paling, 2003). In our first experiment, we found that icon arrays improved accuracy of estimated relative risk reductions when added to numerical descriptions. They were helpful even when the reductions were presented in transparent numerical formats (absolute instead of relative risk reduction). Icon arrays lead to additional improvements in accuracy for both students and older adults, even after accounting for their numeracy skills, sex, and education. They were particularly useful for participants whose numeracy skills were lower than the median of their respective group.

Ideally, visual aids would help patients to make more informed medical decision without affecting their perceptions of otherwise equal risks presented in the same format. In our second experiment we tested, therefore, whether icon arrays affected these perceptions. In contrast to some concerns that visual displays may increase perceived seriousness of risks compared to numerical presentations (cf., Slovic et al., 2005; Stone et al., 1997), we found that icon arrays actually made the risks look less serious. As icon arrays display both people who are and people who are not affected by risks, they may discourage focusing on just the nu-

merators of risk ratios, which has been suggested to be responsible for higher seriousness of risks observed in some numerical and visual formats (cf. Stone et al., 2003; Reyna & Brainerd, 1994). The finding that icon arrays decrease perceived risks should be taken into account when considering their use for purposes of encouraging people to change their behavior. At the same time, icon arrays might be useful when the purpose of communication is to alleviate false fears or debunk exaggerated statements about the successfulness of various treatments.

We have also found that icon arrays with larger overall number of icons make otherwise equivalent risks seem somewhat more serious. These results are comparable to those found in studies on the ratio-bias effect (e.g., Denes-Ray et al., 1995; Rudski & Volksdorf, 2002). Participants in these studies perceived the same risks seem as larger when presented with larger denominators. Our results are also in accord with studies on human judgment (cf. Fiedler, 2000), which show that people tend to put more weight on information obtained from larger samples. Our findings are different than those of the only other study that has, to our awareness, examined this effect in health context (Schapira et al., 2001). However, as that study was based on focus groups, it is somewhat difficult to compare it with our experimental results.

A limitation of this study is that it was conducted in a laboratory setting and did not involve real patient-doctor interaction. Although this controlled setting enabled us to draw clearer conclusions about the effects of various properties of icon arrays (e.g., number of icons), it is possible that these displays have additional benefits or drawbacks in clinical settings. On the other hand, people are often exposed to information about health-related risks outside the doctors' office. In particular, the media and advertisers often use relative risk reductions to report information on effectiveness of medical drugs and other treatments (cf. Skolbekken, 1998). This can lead to mistaken hopes in medical procedures whose usefulness is actually not very impressive (cf. Gigerenzer & Edwards, 2003). Presenting risk reductions in absolute rather than relative numerical format certainly helps, but visual displays might be even more effective in attracting people's attention and affecting their judgment in today's world over flooded with information. Future research might address this hypothesis. In sum, our experiments suggest that icon arrays are a promising way of communicating medical risks to a wide range of patient groups.

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.
- Bucher, H., Weinbacher, M., & Gyr, K. (1994). Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *British Medical Journal*, 309, 761-764.
- Davids, S. L., Schapira, M. M., McAuliffe, T. L., & Nattinger, A. B. (2004). Predictors of pessimistic breast cancer risk perceptions in a primary care population. *Journal of General Internal Medicine*, 19, 310-315.
- Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 21, 1083-1092.
- Edwards, A., Elwyn, G., & Mulley, A. (2002). Explaining risks: Turning numerical data into meaningful pictures. *British Medical Journal*, 324, 827-830.

- Elmore, J. G., & Gigerenzer, G. (2005). Benign breast disease—The risks of communicating risk. *New England Journal of Medicine*, 353, 297–299.
- Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making*, 25, 398–405.
- Feldman-Stewart, D., Kocovski, N., McConnell, B. A., Brundage, M. A., & Mackillop, W. J. (2000). Perception of quantitative information for treatment decisions. *Medical Decision Making*, 20, 228–238.
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107, 659–676.
- Forrow, L., Taylor, W., & Arnold, R. (1992). Absolutely relative: How research results are summarized can affect treatment decisions. *American Journal of Medicine*, 92, 121–124.
- Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *British Medical Journal*, 327, 741–744.
- Golbeck, A. L., Ahlers-Schmidt, C. R., Paschal, A. M., Dismuke, S. E. (2005). A definition and operational framework for health numeracy. *American Journal of Preventive Medicine*, 29, 375–376.
- Hibbard, J. H., Peters, E. M., Slovic, P., & Finucane, M. L. (2001). Making health care report cards easier to use. *The Joint Commission Journal on Quality Improvement*, 27, 591–604.
- Hux, J. E., & Naylor, C. D. (1995). Communicating the benefits of chronic preventive therapy: Does the format of efficacy data determine patients' acceptance of treatment? *Medical Decision Making*, 15, 152–157.
- Kutner, M., Greenberg, E., & Baer, E. (2006). *A first look at the literacy of America's adults in the 21st century* (NCES 2006–470; U.S. Department of Education). Washington, DC: Institute of Education Sciences.
- Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy* (NCES 2006–483; U.S. Department of Education). Washington, DC: National Center for Education Statistics.
- Lipkus, I. M., & Hollands, J. G. (1999). The visual communication of risk. *Journal of the National Cancer Institute Monographs*, 25, 149–163.
- 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.
- Montori, V. M., & Rothman, R. L. (2005). Weakness in numbers—The challenge of numeracy in healthcare. *Journal of General Internal Medicine*, 11, 1071–1072.
- Paling, J. (2003). Strategies to help patients understand risks. *British Medical Journal*, 327, 745–748.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17, 407–413.
- Programme for International Student Assessment (PISA). (2003). *The PISA 2003 assessment framework—Mathematics, reading, science, and problem solving knowledge and skills*. Paris: OECD.
- Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective Probability*, pp. 239–272. New York: Wiley.
- Rudski, J. M., & Volksdorf, J. (2002). Pictorial versus textual information and the ratio-bias effect. *Perceptual and Motor Skills*, 95, 547–554.
- Sarfati, D., Howden-Chapman, P., Woodward, A., & Salmond, C. (1998). Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefits are expressed. *Journal of Medical Screening*, 5, 137–140.
- Schapiro, M. M., Nattinger, A. B., & McHorney, C. A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making*, 21, 459–467.
- Schirillo, J. A., & Stone, E. R. (2005). The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. *Risk Analysis*, 25, 555–566.
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, G. H. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127, 966–971.
- Skolbekken, J. A. (1998). Communicating the risk reduction achieved by cholesterol reducing drugs. *British Medical Journal*, 316, 1956–1958.
- Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: The effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, 24, 271–296.
- Slovic, P., Peters, E., Finucane, M. L., & MacGregor, D. G. (2004). Affect, risk, and decision making. *Health Psychology*, 24, S35–S40.
- Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods* (6th ed.). Ames: Iowa State University Press.
- Speros, C. (2005). Health literacy: Concept analysis. *Journal of Advanced Nursing*, 50, 633–640.
- Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. (2003). Foreground: Background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, 90, 19–36.
- 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.
- Tuijnman, A. (2000). *Benchmarking adult literacy in America: An international comparative study*. Washington, DC: International Adult Literacy Survey, Department of Education, OVAE.
- Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: Implications for risk communication. *Applied Cognitive Psychology*, 11, 495–506.