This paper was originally published by Sage as:
Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). **Do icon arrays help reduce denominator neglect?** *Medical Decision Making*, *30*(6), 672–684. https://doi.org/10.1177/0272989X10369000

This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence (funded by the DFG, German Research Foundation) respectively.

# Do Icon Arrays Help Reduce Denominator Neglect?

*Rocio Garcia-Retamero, PhD, Mirta Galesic, PhD, Gerd Gigerenzer, PhD*

***Background and Objective***. *Denominator neglect is the focus on the number of times a target event has happened (e.g., the number of treated and nontreated patients who die) without considering the overall number of opportunities for it to happen (e.g., the overall number of treated and nontreated patients). In 2 studies, we addressed the effect of denominator neglect in problems involving treatment risk reduction where samples of treated and non-treated patients and the relative risk reduction were of different sizes. We also tested whether using icon arrays helps people take these different sample sizes into account. We especially focused on older adults, who are often more disadvantaged when making decisions about their health.* ***Design***. *Study 1 was conducted on a laboratory sample using a within-subjects design; study 2 was conducted on a nonstudent sample interviewed through the Web using a between-subjects design.* ***Outcome Measures***. *Accuracy of understanding risk reduction.* ***Results***. *Participants often paid too much attention to numerators and insufficient attention to denominators when numerical information about treatment risk reduction was provided. Adding icon arrays to the numerical information, however, drew participants' attention to the denominators and helped them make more accurate assessments of treatment risk reduction. Icon arrays were equally helpful to younger and older adults.* ***Conclusions***. *Building on previous research showing that problems with understanding numerical information often do not reside in the mind but in the representation of the problem, the results show that icon arrays are an effective method of eliminating denominator neglect.* ***Key words***: *decision aids; risk communication or risk perception; health literacy; numeracy; bounded rationality; heuristics and biases.* **(Med Decis Making 2010;30:672–684)**

Increased emphasis on patient-centered decision making and disease prevention has shifted the responsibility of decision making to patients,[1] who now more than ever need to understand health-related numerical information. For instance, informed consent laws mandate that patients must be informed about risks and benefits and give their consent before any treatment can be implemented. This information is often provided in a numerical format, and its interpretation can require rather advanced knowledge of statistical concepts.[2]

Recent research in health decision making has shown that many patients—and their physicians—have difficulties grasping a host of numerical concepts that are prerequisites for understanding health-relevant risk communications.[3–8] These communications often take the form of baseline risk estimates, risk reduction with one or more treatments, and increased risk of side effects of treatments.[9] Cross-cultural comparisons suggest that these difficulties are widespread, extending even to highly educated populations[10–12] and possibly increasing with age.[13,14] Is it possible to help patients make accurate and effective decisions by presenting risks about health in appropriate formats? This is the central question of this article.

Ratio concepts—of which risks and probabilities are examples—are particularly challenging and prone to biases that undermine judgment and decision making.[15–17] For instance, people often pay too

much attention to numerators and insufficient attention to denominators.[18–21] That is, they focus on the number of times a target event has happened (e.g., 6 patients who died from the surgery) and pay less attention to the overall number of opportunities for it to happen (e.g., 10 or 1000 patients who underwent the surgery). In fact, 6 patients who died from surgery may or may not be significant depending on the overall number of patients who underwent the procedure (10 or 1000). This denominator neglect effect, described by Reyna[22] and Reyna and Brainerd,[23] has been studied extensively.

In a series of studies by Epstein and colleagues[18,24–26] (see also Miller and others[27] and Stanovich and West[28]), participants had to choose 1 of 2 samples from which to draw a winning (e.g., red) jelly bean. A typical task involved choosing between a sample containing 1 red out of 10 beans and a sample containing 7 red out of 100 beans. Participants often chose the sample containing 7 red beans: they favored the option with a greater absolute number but smaller proportion of red jelly beans over the option with fewer red beans but better odds. That is, they preferred the option with the larger numerator and did not pay much attention to the denominator, even when the probabilities were smaller in their chosen sample. Similar results were obtained by Denes-Raj and others[18] and Kirkpatrick and Epstein[29] when presenting pairwise choices of gambles, by Acredolo and others[30] and Kokis and others[31] using the same procedure with children, and by Yamagishi[32] in a medical context.

In this article, we address denominator neglect in problems involving treatment risk reduction. To the best of our knowledge, all existing studies of formats for presenting risk reduction assumed that the samples of treated and nontreated patients were of the same size,[33–35] and even experts in the field recommended this arrangement.[36–38] For instance, in a study by Galesic, Garcia-Retamero and Gigerenzer,[35] participants were provided with information about the usefulness of screening in reducing the risk of several cancers. The information involved the number of patients who die within groups who do or do not go to the screening. The group who received the screening and the group who did not receive the screening were set to be of equal size (either 100 or 1000 patients).

In medical practice, however, the number of patients who receive a certain treatment is often smaller than those who do not.[39–42] Similarly, in daily life, patients—and their physicians—might be able to think of more people who did not go to

a certain screening or take a novel drug than those who did. Denominator neglect could then have important consequences when estimating treatment risk reduction in such conditions. Imagine, for instance, a person who would like to estimate the effect of a treatment, and the groups of treated and nontreated patients that she can recall are unequally sized (e.g., 100 and 500, respectively). If this person disregards the overall number of treated and nontreated people, she might perceive the treatment to be more effective than it actually is. Thus, she might underestimate the proportion of patients who died after receiving the treatment (e.g., 2 of 100) while overestimating the proportion of people who died and did not receive the treatment (e.g., 50 of 500). Our first goal, then, is to test the hypothesis that people will show denominator neglect when judging the effectiveness of a treatment using information from unequally sized groups of treated and nontreated patients.

Second, we test the hypothesis that using visual displays will help people attend to and take differential sample sizes into account when judging the effectiveness of a treatment. As Reyna[22] and Reyna and Brainerd[23] suggested, visual displays can help people represent superordinate classes such as the overall number of patients who did and did not receive a treatment. Reyna and others[16,43–46] provided experimental support for this hypothesis by using line drawings of familiar objects (i.e., pictograms), visual grids with squares representing individuals, and icon arrays. Experimental studies have confirmed as well that icon arrays are a promising method for communicating medical risk reduction.[35,47–50] Icon arrays might then help draw people's attention to the overall number of unaffected patients and thus reduce denominator neglect in problems involving treatment risk reduction. Whether this is, in fact, the case is still an open question.

Finally, we test the hypothesis that icon arrays are useful even for people who are otherwise disadvantaged when making decisions about health. As we mentioned above, older people have problems understanding complex medical information, not only numerically presented data,[13,51–53] and they have severe difficulties when performing numerical calculations—especially when task complexity increases[53]—due to age-related cognitive decline.[54,55] Older adults also show inconsistent preferences and search for health-related information less often than young adults.[56,57] Finally, older adults focus more on interpersonal and experiential components of a problem rather than the propositional content.[58–60] At the

same time, older adults face more medical decisions than younger people and are disproportionately more likely to encounter and be required to use quantitative information when making their health-related choices.[52] Older adults, therefore, require special assistance. Our final hypothesis is then that icon arrays will be useful for an older population.

To test these hypotheses, we conducted 2 studies. The first was conducted in a laboratory setting using a within-subject design. To test the generalizability of the results, the second study was conducted via a survey with a more diverse sample of the general population and employing a between-subjects design. In both studies, we analyzed participants' understanding of medical risk reduction after treatment. To test whether people pay more attention to numerators than to denominators and whether icon arrays increase accuracy of understanding of treatment risk reduction, participants in our studies were presented with scenarios involving equally effective treatments but differing in the overall number of treated and nontreated patients. In some conditions, the number of patients who did receive a treatment was equal to those who did not; in other conditions, it was smaller or larger. In both studies, some participants were provided with visual aids (i.e., icon arrays) in addition to numerical information about risk reduction, whereas others received numerical information only.

## STUDY 1

### Method

*Participants* The participants were 117 individuals: 57 older adults (median age of 68 years, range 51–75; 49% men and 51% women; 49% with high school or lower education and 51% with college or university education) and 60 university students from the Free University of Berlin (Germany; median age of 25 years, range 19–33; 45% men and 55% women). The questionnaire was administered on computers in our lab at the Max Planck Institute for Human Development in Berlin (Germany) and took from 18 to 22 minutes to complete. Participants were invited to our lab and received 10 euros. They were randomly assigned to the experimental groups.

*Stimuli and procedure* Participants were presented with 8 medical scenarios of the usefulness of hypothetical new drugs for reducing cholesterol that also decrease the risk of dying from a heart attack. The drugs were modeled after the drug simvastatin.[61]

The order of the 8 scenarios was randomized. In one scenario, for instance, participants got the following information:

> A new drug for reducing cholesterol, Estatin, decreases the risk of dying from a heart attack for patients with high cholesterol. Here are the results of a study of 600 such patients: 50 out of 500 of those who did not take the drug died of a heart attack, compared with 2 out of 100 of those who took the drug.

The rest of the drugs were named Astatin, Bestatin, Cestatin, Destatin, Festatin, Gestatin, and Istatin, respectively.

Three independent variables were manipulated in the study. First, each drug's relative risk reduction was manipulated within-subjects with 2 levels. Relative risk reduction was set at either 40% (from 10% to 6%) or 80% (from 10% to 2%), with 4 scenarios at each level. Second, the overall numbers of treated and nontreated patients (i.e., the sizes of the denominators) were manipulated within-subjects to be either 500–500, 100–500, 500–100, or 100–100, where the first and second quantities reflect the overall number of patients who did and did not take the drug, respectively (see Table 1). The sizes of the numerators (i.e., the number of treated and nontreated patients who died) varied within conditions depending on the drug's relative risk reduction and the sizes of the denominators.

Independently of these manipulations, half of the participants received, in addition to the numerical information about risk reduction, 2 icon arrays presenting the risk of dying of a heart attack when the drug was and was not taken, respectively. All icon arrays contained either 500 or 100 circles depending on the overall number of patients who did and did not take the drug. Deceased patients were shown as black circles at the end of the array. We used circles because previous research[62] did not find differences in effects of arrays with faces compared to more abstract symbols. An example of the condition involving icon arrays is shown in Figure 1 (original material was in German).

As a dependent variable, accuracy of risk understanding after reading the information about each medical scenario was measured. First, following the procedure used by Schwartz and others,[8] participants were asked how many of 1000 patients with high cholesterol might die of a heart attack if they *do not* take the drug. Second, they were asked how many of 1000 patients with high cholesterol might die of a heart attack if they *do* take the drug. By

**Table 1**   Stimuli Used in Studies 1 and 2 (in Bold)

| Denominator Sizes | Treatment Risk Reduction, % | Treated Patients | | Nontreated Patients | |
| --- | --- | --- | --- | --- | --- |
| | | Dead Patients | Population Size | Dead Patients | Population Size |
| 500-500 | 40 | 30 | 500 | 50 | 500 |
| | **80** | **10** | **500** | **50** | **500** |
| 100-500 | 40 | 6 | 100 | 50 | 500 |
| | **80** | **2** | **100** | **50** | **500** |
| 500-100 | 40 | 30 | 500 | 10 | 100 |
| | **80** | **10** | **500** | **10** | **100** |
| 100-100 | 40 | 6 | 100 | 10 | 100 |
| | **80** | **2** | **100** | **10** | **100** |

Number of treated and nontreated patients who died from a heart attack for all denominator size and risk reduction conditions in studies 1 and 2.
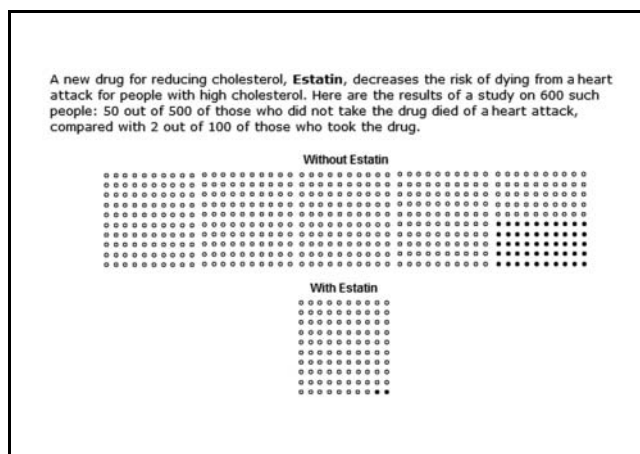


*Figure 1   Numerical information about risk reduction and icon arrays that participants received in study 1. This example presents the information provided in the 100–500 condition with relative risk reduction of 80%.*

deducting the second from the first answer and dividing it by the first, we calculated the estimated relative risk reduction. Participants were classified depending on whether their estimates were accurate, lower, or higher than the exact value (i.e., .40 and .80 when risk reduction was set at 40% and 80%, respectively). Participants' estimates were treated as correct only when they were exactly right.

In sum, the experimental design had 4 factors: participants' age (older v. young adults; between-subjects), the drugs' relative risk reduction (40% v. 80%; within-subjects), the sizes of the denominators (500–500, 100–500, 500–100, and 100–100; within-subjects), and icon arrays (presence v. absence; between-subjects). To assess the effect of these factors on accuracy of risk understanding, we conducted analysis of variance (ANOVA; for a justification, see

Lunney[63] and Cleary and Angel[64]). Degrees of freedom for the analyses containing repeated-measures factors were corrected by using the Greenhouse–Geisser[65] technique. We conducted Tukey's HSD (honest significant difference) test in post hoc analyses.

## Results

*Do participants show denominator neglect in their estimates of risk reduction?* Figure 2 shows the percentage of participants whose estimates of risk reduction were accurate, lower, or higher than the exact value in each of the sizes of the denominator conditions. Figure 3 shows average deviations of inaccurate estimates from the exact value. When information about treatment risk reduction was provided numerically and the sizes of the denominators were equal (i.e., in the 500–500 and 100–100 conditions when no icons were provided), participants' estimates were fairly accurate. In contrast, when the denominators were different, the estimates were rather inaccurate. These results held for both the 40% and the 80% relative risk reduction conditions. Consistent with these findings, an ANOVA with sizes of the denominators and relative risk reduction as within-subjects factors on the percentage of participants whose estimates of risk reduction were accurate when information about the drug was provided only numerically showed a main effect of sizes of the denominators, $F_{2, 126} = 24.67$, $P = 0.001$.

As Figure 2 shows, when the number of treated patients was lower than the number of those who did not receive the treatment (i.e., in the 100–500 conditions), 53% of participants' estimates of treatment risk reduction were *higher* than the exact value compared to 47% that were accurate. Note that in such cases, the number of patients who received the treatment and died is much lower than the number
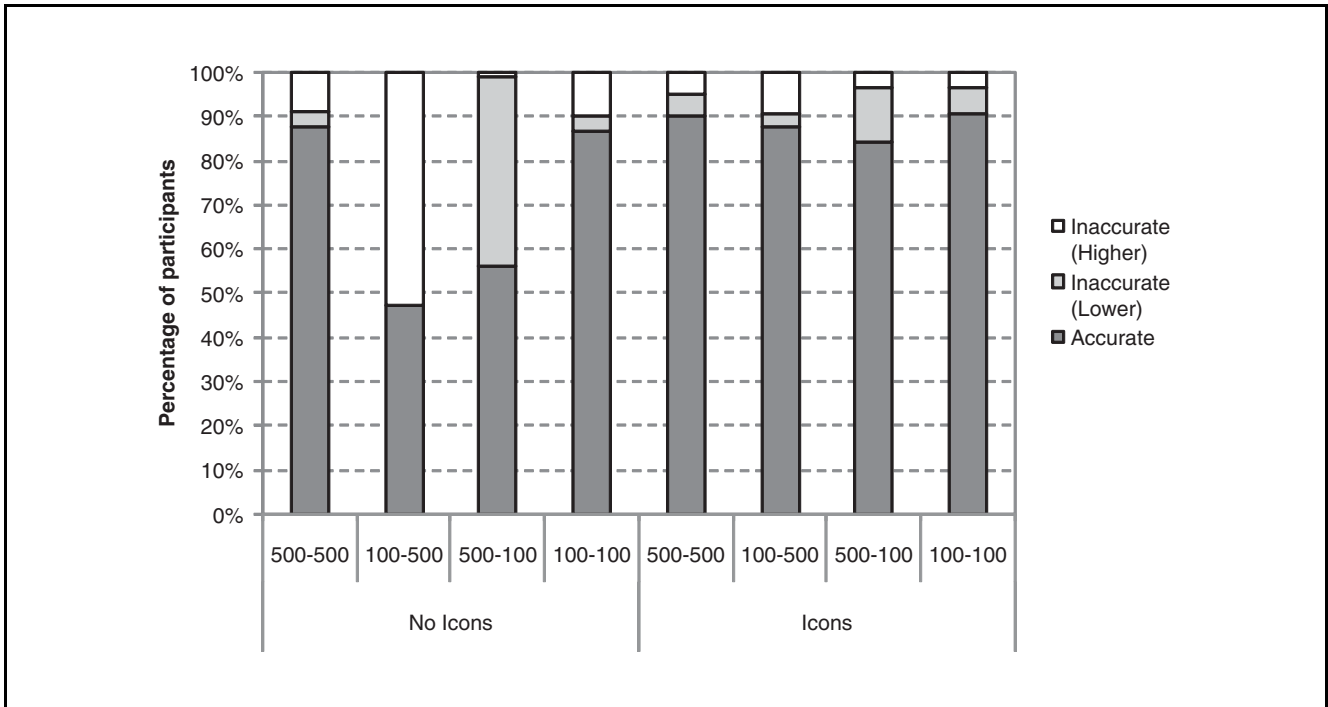
*Figure 2   Percentage of participants whose estimates of risk reduction were accurate, lower, or higher than the exact value as a function of the sizes of the denominators and the provision of icon arrays in study 1. Note: The first and second quantities in the notation of the sizes of the denominators reflect the overall number of patients who did and did not take the drug, respectively. Results are collapsed across the 2 levels of the treatment relative risk reduction.*

of patients who did not receive the treatment and died (e.g., 2 and 50, respectively, when risk reduction is 80%). It is possible that many participants in these conditions paid more attention to absolute numbers in the numerators than in the denominators. This might have led them to believe that the treatment had a larger effect than it actually did.

In contrast, when the number of treated patients was higher than the number of patients who did not receive treatment (i.e., in the 500–100 conditions), 43% of participants' estimates of treatment risk reduction were *lower* than the exact value compared to 1% and 56% that were higher and accurate, respectively. In these conditions, the number of patients who received the treatment and died is equal to or even higher than the number of patients who did not receive the treatment and died (e.g., 30 and 10, respectively, when risk reduction is 40%). This might have led participants to believe that the treatment had a smaller effect than it actually did.

Finally, when the overall number of treated and nontreated patients was equal (in the 500–500 and 100–100 conditions), 87% of the participants'

estimates of treatment risk reduction were *accurate* compared to 4% and 9% that were lower and higher than the exact value, respectively. In these conditions, participants did not necessarily have to take proportions into account to make accurate estimates but could focus on the absolute numbers in the numerators (e.g., 30 and 50, when risk reduction is 40% in the 500–500 condition).

Deviations of inaccurate estimates from the exact value are in line with these findings (see Figure 3). Participants whose estimates about treatment risk reduction were inaccurate often believed that the treatment had a *greater* effect than it actually did in both the 500–500 and 100–500 conditions (i.e., in the conditions with a *larger* number of nontreated patients who died). In contrast, when the number of nontreated patients who died was *smaller* (i.e., in the 100–100 and 500–100 conditions), participants whose estimates about treatment risk reduction were inaccurate often believed that the treatment had a *smaller* effect than it actually did. In sum, the results support the hypothesis that many participants would pay more attention to the number of
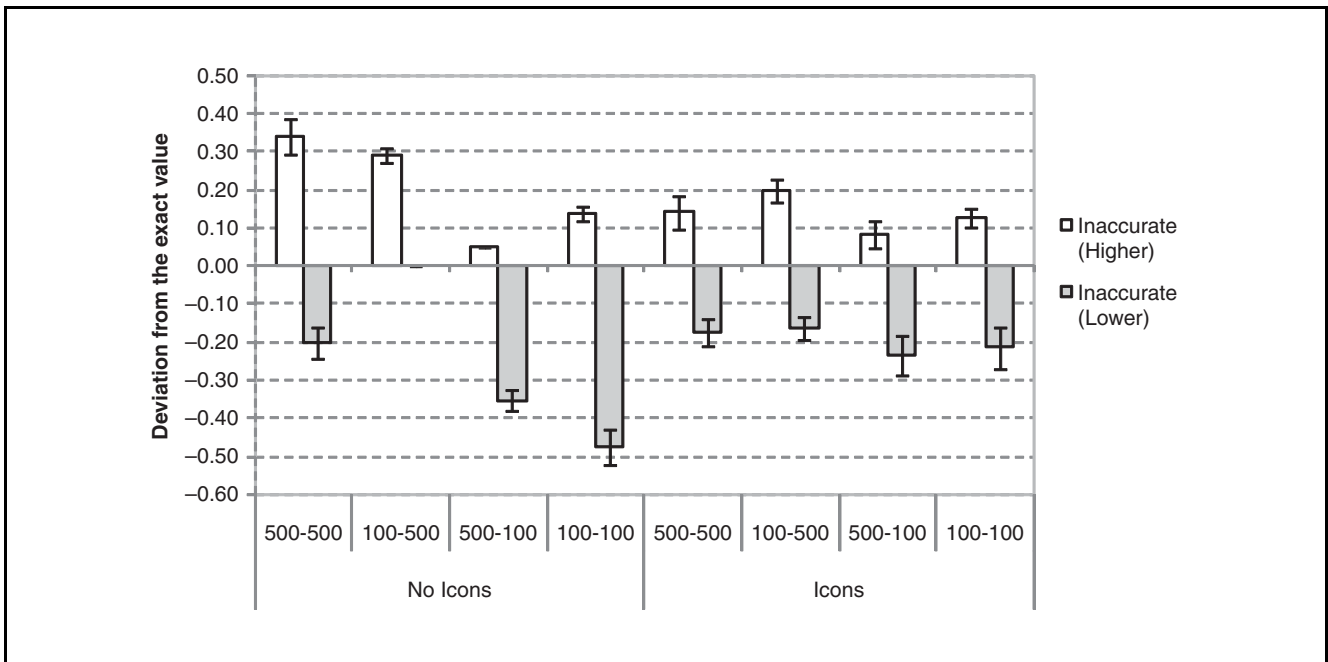
*Figure 3   Average deviation of inaccurate estimates from the exact value as a function of the sizes of denominators and the provision of icon arrays in study 1. Note: The first and second quantities in the notation of the sizes of the denominators reflect the overall number of patients who did and did not take the drug, respectively. Results are collapsed across the 2 levels of the treatment relative risk reduction. Error bars indicate 1 standard error.*

treated and nontreated patients who died (i.e., numerators) than to the overall number of treated and nontreated patients (i.e., denominators).

*Do icon arrays help reduce denominator neglect?*
As Figures 2 and 3 show, when icon arrays were added to the numerical information, participants paid attention to both numerators and denominators. The ANOVA with relative risk reduction and sizes of the denominators as within-subjects factors and icon arrays as a between-subjects factor showed an interaction between sizes of denominators and icon arrays on percentages of participants whose estimates of risk reduction were accurate, $F_{2, 274} = 14.54$, $P = 0.001$.

When the sizes of the denominators were different and icon arrays were added to the numerical information, the percentage of participants who estimated the treatment risk reduction correctly increased from 52% to 86% ($P = 0.0001$). The latter percentage is very similar to that when the sizes of the denominators were equal (i.e., 90%). Moreover, deviations of inaccurate estimates from the exact value were rather small in all conditions. These results held for both the 40% and the 80% relative risk reduction conditions. Icon arrays, therefore,

helped the participants understand that it is important to take the overall number of treated and nontreated patients into account for an accurate evaluation of effectiveness of the treatments.

*Do icon arrays improve accuracy of risk understanding in both younger and older adults?* Understanding medical information presented numerically was more difficult for older than younger adults: the percentage of older adults who estimated the risk reduction accurately was lower (73%, $s_{\bar{x}} = 2.1$) than that of younger adults (85%, $s_{\bar{x}} = 1.6$). This result is consistent with the main effect of age in participants' estimates of risk reduction in the ANOVA with relative risk reduction and sizes of the denominators as within-subjects factors and age and icon arrays as between-subjects factors, $F_{1, 113} = 8.99$, $P = 0.003$. The effect of icon arrays, however, was independent of age, as icon arrays had a positive effect on accuracy and did not interact with age: the percentage of older adults who made accurate estimates increased from 63% ($s_{\bar{x}} = 3.3$) to 83% ($s_{\bar{x}} = 2.5$) when icon arrays were added to the numerical information. Similarly, this percentage increased from 76% ($s_{\bar{x}} = 2.8$) to 94% ($s_{\bar{x}} = 1.6$) in younger adults.

## Discussion

When the information was presented numerically, participants paid too much attention to numerators (i.e., the number of patients who died with and without treatment) and insufficient attention to denominators (i.e., the overall number of patients who did and did not receive the treatment). Interestingly, icon arrays helped participants take both numerators and denominators into account: when icon arrays were added to the numerical information, the denominator neglect disappeared. Icon arrays were helpful additions for both younger and older adults. These results held even when level of education was controlled in the analyses.

A critic, however, could call the generalizability of our findings into question because study 1 was conducted in a laboratory setting using a convenient sample of participants. The critic could further argue that our results may well have been inflated because a within-subject design was used. The within-subjects manipulation might make salient the differences between numerators (e.g., 2 treated patients died and 50 nontreated in the 100–500 condition, compared to 10 and 10 in the 500–100 condition, when risk reduction is 80%). As a consequence, participants might have paid more attention to numerators than to denominators. A between-subjects test of our results would represent a much stricter criterion and perhaps a more appropriate one because "much of life resembles a between-subjects experiment."[66,67] In line with this hypothesis, Stanovich and West[28] (see also LeBoeuf and Shafir[68]) argued that several biases and framing effects display different relationships with individual difference variables in within-subjects compared to between-subjects designs. Whether icon arrays help people take the sizes of the denominators into account when a less transparent design is used remains untested. In study 2, we test the generalizability of our findings by using a more diverse sample of participants and a between-subjects design.

## STUDY 2

### Method

*Participants* The participants were 354 individuals selected from an online pool maintained by Survey Sampling, Inc. The structure of the sample in terms of gender and age was designed to reflect that of the total population of Germany: thus, the sample consisted of 49.2% men and 50.8% women with a median age of 44 years (range 20–99). We split the participants into 3 age groups according to the overall group's 33rd and 66th percentile, which are 37 and 50 years, respectively. The median age in the 3 groups was 30, 44, and 57 years, respectively. Male and female participants were evenly distributed in the 3 groups. Fifty-five percent of the participants had a secondary level of schooling (high school or lower), and 45% had at least some college or university education. All participants were Internet users. The questionnaire was administered using the online survey software Unipark and took from 7 to 15 minutes to complete. Participants in the study had the opportunity to win a prize or donate a small amount to charity. They were randomly assigned to the groups.

*Stimuli and procedure* The instructions and procedure of study 2 were identical to those of study 1, except that each participant was presented with only one medical scenario describing Estatin, a hypothetical new drug with a relative risk reduction of 80%. The overall number of treated and nontreated patients was manipulated between subjects to be 500–500, 100–500, 500–100, or 100–100 (see Table 1).

The design of the study had 3 factors: participants' age (older v. middle-aged v. younger adults; between-subjects), sizes of the denominators (500–500, 100–500, 500–100, and 100–100; between-subjects), and icon arrays (presence v. absence; between-subjects). As in study 1, we measured estimates of relative risk reduction.

### Results

*Do participants show denominator neglect in a between-subjects design?* Figures 4 and 5 mirror Figures 2 and 3. As in study 1, participants' average estimates of relative risk reduction were fairly accurate when information about treatment risk reduction was provided only numerically and the denominators were equal, whereas they were rather inaccurate when the denominators were different. The ANOVA with sizes of the denominators as a between-subjects factor showed a main effect of this variable on percentages of participants whose estimates of risk reduction were accurate, $F_{3, 177} = 4.64$, $P = 0.004$, when information about the drug was provided only numerically.

Similar to study 1, 42% of the participants *overestimated* risk reduction when the number of treated patients was lower than the number of those who did not receive the treatment (i.e., in the 500–
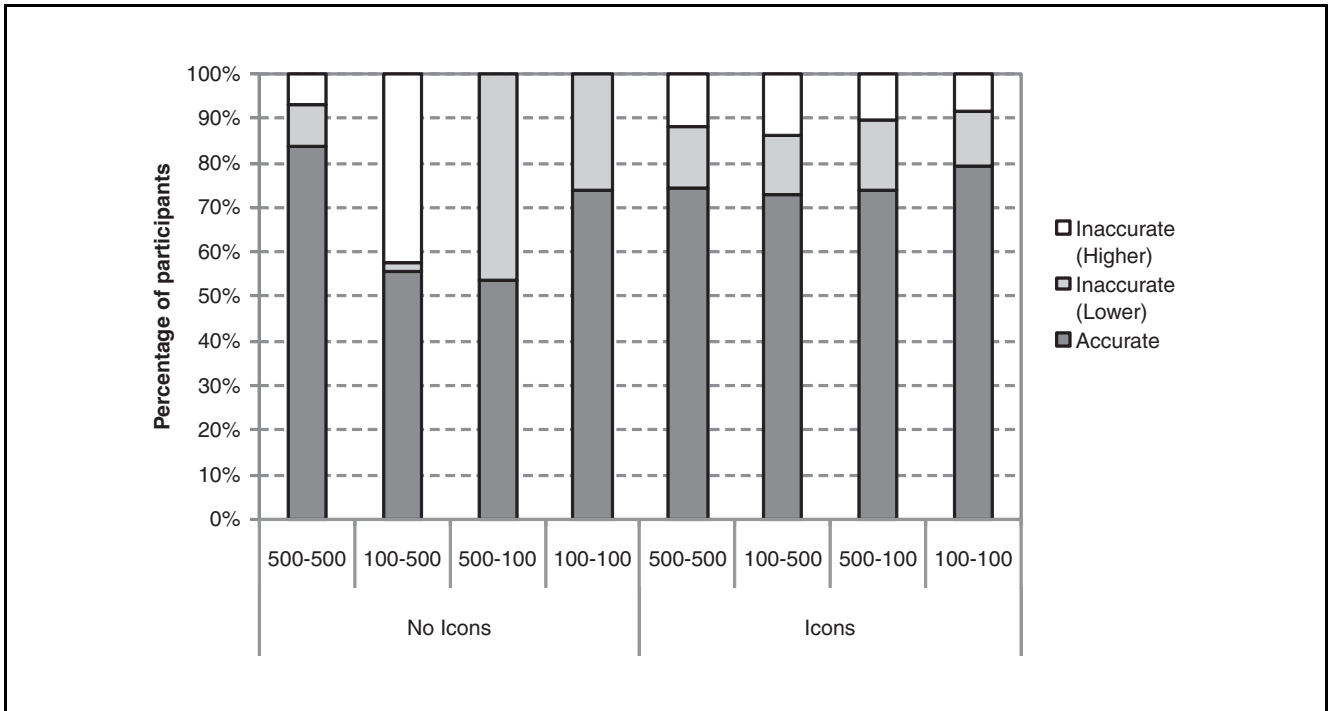
*Figure 4   Percentage of participants whose estimates of risk reduction are accurate, lower, or higher than the exact value as a function of the sizes of the denominators and the provision of icon arrays in study 2. Note: The first and second quantities in the notation of the sizes of the denominators reflect the overall number of patients who did and did not take the drug, respectively.*

100 condition) compared to 2% and 56% of the participants who underestimated risk reduction and provided an accurate estimate, respectively. In contrast, when the number of treated patients was higher than the number of patients who did not receive treatment (i.e., in the 500–100 conditions), 46% of the participants *underestimated* risk reduction compared to 54% of the participants who provided an accurate estimate. Similarly, participants whose estimates of treatment risk reduction were inaccurate often believed that the treatment had a *greater* effect than it actually did in the 500–500 and 100–500 conditions (i.e., when the number of nontreated patients who died was *larger*). In contrast, in the 100–100 and 500–100 conditions (i.e., when the number of nontreated patients who died was *smaller*), participants whose estimates about treatment risk reduction were inaccurate often believed that the treatment had a *smaller* effect than it actually did (see Figure 5). In sum, when the sizes of the denominators were manipulated between-subjects, participants also paid more attention to the number of treated and nontreated patients who died (i.e., numerators) than to the

overall number of treated and nontreated patients (i.e., denominators).

*Do icon arrays help reduce denominator neglect in the between-subjects design?* As in study 1, when icon arrays were added to the numerical information, participants paid attention to both numerators and denominators (see Figures 4 and 5). The ANOVA with sizes of the denominators and icon arrays as between-subjects factors showed an interaction between the 2 variables on percentages of participants whose estimates of risk reduction were accurate, $F_{3, 346} = 2.87$, $P = 0.036$.

When the icon arrays were added to the numerical information, the percentage of participants who estimated the treatment risk reduction correctly was similar in all conditions (75% on average, compared to 11% and 12% of the participants who overestimated and underestimated risk reduction, respectively). Moreover, average deviations of inaccurate estimates from the exact value are rather small in all conditions. In sum, when the sizes of the denominators were manipulated between-subjects, icon arrays again helped the participants
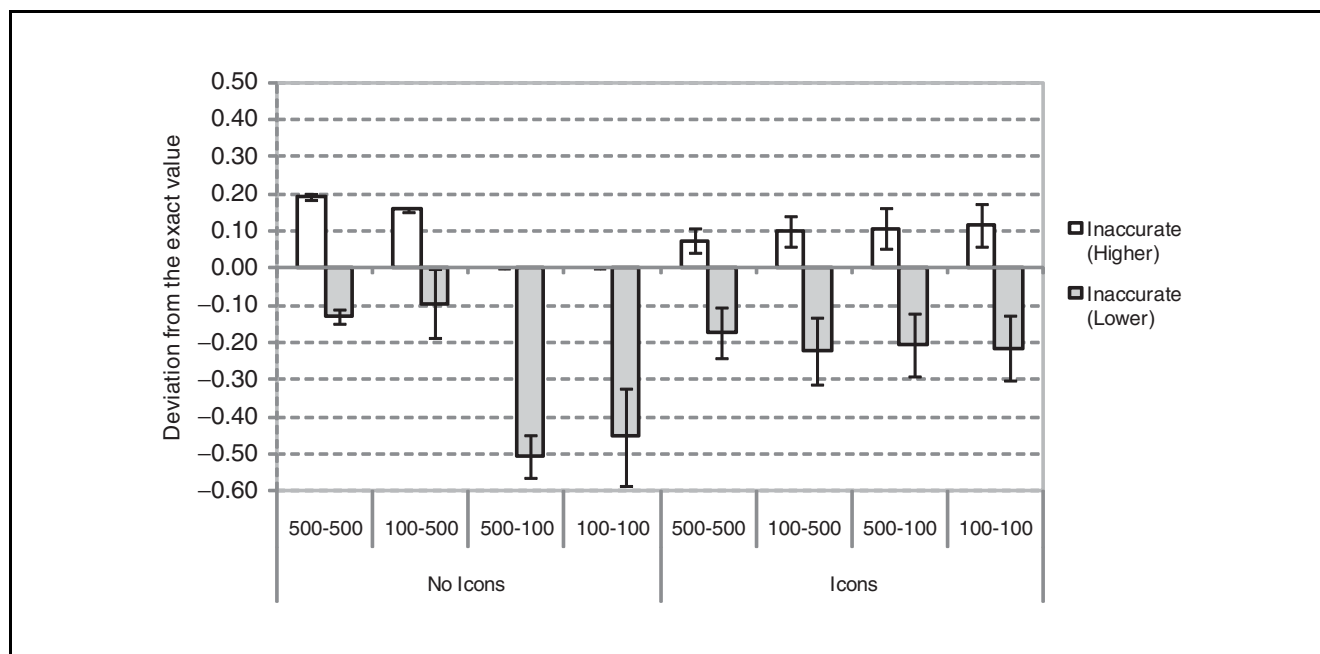
*Figure 5  Average deviation of inaccurate estimates from the exact value as a function of the sizes of denominators and the provision of icon arrays in study 2. Note: The first and second quantities in the notation of the sizes of the denominators reflect the overall number of patients who did and did not take the drug, respectively. Error bars indicate 1 standard error.*

understand that it is important to take the overall number of treated and nontreated patients into account for an accurate evaluation of the effectiveness of the treatments.

*Do icon arrays improve accuracy of risk understanding in both younger and older adults in a between-subjects design?* Understanding medical information presented numerically was more difficult for older than middle-aged and younger adults. Consistent with this result, the ANOVA with sizes of the denominators, icon arrays, and age as between-subjects factors showed a main effect of age on percentages of participants whose estimates of risk reduction were accurate, $F_{2, 329} = 2.79$, $P = 0.041$. In fact, the percentage of older adults who estimated the risk reduction accurately was lower (55%, $s_{\bar{x}} = 4.7$) than that of middle-aged and younger adults (71%, $s_{\bar{x}} = 4.1$, and 83%, $s_{\bar{x}} = 3.5$, respectively).

As in study 1, the effect of icon arrays was positive and independent of age: the percentage of older adults who made accurate estimates increased from 47% ($s_{\bar{x}} = 7.0$) to 63% ($s_{\bar{x}} = 6.4$) when icon arrays were added to the numerical information. Similarly, this percentage increased from 65% ($s_{\bar{x}} = 5.3$) to 77% ($s_{\bar{x}} = 6.5$) and from 77% ($s_{\bar{x}} = 4.9$) to 88%

($s_{\bar{x}} = 4.9$) in middle-aged and younger adults, respectively.

**GENERAL DISCUSSION**

Patients need to understand numerical information such as ratio concepts to make effective decisions about their health. Understanding a treatment risk reduction, for instance, implies taking into account the number of treated and nontreated patients who die out of the overall number of patients who do and do not receive the treatment. In this article, we addressed the problem of whether people estimate risk reduction accurately when the number of treated and nontreated patients differs substantially.

Our results show that participants often pay too much attention to numerators and insufficient attention to denominators. This denominator neglect[19,23] is compatible with previous evidence found by Epstein and colleagues in choices of pairs of samples from which to draw a winning bean or in gambles.[18,24–26] Our results are also in line with the research by Chapman[69] (see also Hoemann and Ross[70] and Surber and Haines[71]), who showed that

problems in which a denominator is shared (1-sample problems) or equal (2-sample equal sample size problems) are easier to solve than problems in which denominators differ across options. Finally, Yamagishi[32] has similarly shown that causes of death with greater absolute numbers are perceived as more risky even if they have smaller proportions than others with smaller absolute numbers.

Our studies, however, are unique in their efforts to address the effect of denominator neglect in problems involving assessments of risk reduction, not only assessments of single probabilities.[32] Furthermore, in contrast to previous research, which has focused exclusively on people's perceptions when they are explicitly told that the probabilities expressed with different denominators are the same, our studies focus on *estimates* when the actual treatment risk reduction is not mentioned. Our task, therefore, reproduces the situation we often encounter when making health-related choices. Furthermore, our results show denominator neglect both in a relatively limited laboratory sample and in a more diverse sample interviewed through the Web; the results hold in both a within-subject and a between-subjects design and in both younger and older adults.

Moreover, older adults are particularly disadvantaged as they showed denominator neglect more often than younger adults. These results are consistent with previous literature showing that older adults have more difficulties understanding complex medical information.[72] In fact, the result that people—especially those who are older—could disregard crucial information when making important decisions about their health is a troubling finding. In recent years, people have been living longer but are not always able to rely on younger family members for support because of the increasing geographical dispersion of families and the modern emphasis on independence. Older adults, therefore, need to maintain strong decision-making capabilities as they age.[52,53] In addition, patients today often receive information not only from their physicians but also from a number of other sources such as the media, health insurance companies, government health pamphlets, the Internet, and their friends and relatives.[73] These alternative sources do not always follow the convention of presenting information about risk reduction using ratios with the same sizes of denominators.[41]

Interestingly, we show an effective method to eliminate the denominator neglect: providing icon arrays in addition to numerical information draws participants' attention to both the numerators and the denominators and helps them make more accurate assessments of risk reduction. Icon arrays were equally helpful to younger and older adults. These results support and extend our own and others' previous findings about the usefulness of icon arrays in communicating medical risks.[10,12,35,37,44,49,50] Specifically, they provide experimental support of Ancker and others'[36] hypothesis that graphs that make part-to-whole relations visually available help people attend to the relationship between the numerator (i.e., the number of treated and non-treated patients who are affected) and the denominator (i.e., the entire population at risk; see also Lipkus[50]). These findings also extend the literature on denominator neglect in judgment and decision making because they add to the existing experimental support of Reyna and Brainerd's[19,23,43] hypothesis that visual displays can help people represent superordinate classes (e.g., the overall number of patients who did and did not receive a treatment). Finally, our results have implications for medical practice as they suggest suitable ways to communicate quantitative medical data, especially to people who are more disadvantaged when making decisions about health. In fact, our findings support the medical convention of reporting risks using ratios with the same denominator (e.g., when communicating drug benefits and harms with a drug facts box[74]). When this common practice is not feasible, using visual displays would be an effective method of preventing difficulties in understanding health-relevant risk communications.

In sum, our results support the notion that problems in communicating risks occur because inappropriate information formats are often used and not because of biases in people's minds. That is, problems with understanding numerical information often do not reside in the mind but in the representation of the problem.[5,75] Similar reductions in what superficially looked like biased thinking were observed in the case of conditional probabilities (when expressed as natural frequencies),[76] relative risk reductions (when expressed as absolute risk reductions),[77] and single-event probabilities.[78] In the same vein, we show that denominator neglect disappears when both the numerator and the denominator are presented in a transparent way.

To the best of our knowledge, this article is the first to illustrate that people show denominator neglect when estimating treatment risk reduction and that icon arrays are a helpful method to eliminate such bias. At the same time, it leaves several

questions open for future research. For instance, in this study, we have focused on icon arrays because they seem to be particularly promising for communicating risk reductions as a result of medical screenings and treatments.[35,37,48] In fact, it was Otto Neurath (1882–1945), a philosopher, economist, and prominent member of the Vienna Circle, who first proposed the use of arrays of symbols to explain complex statistics—social and economic facts—to uneducated Viennese.[79] Since then, arrays of icons symbolizing different types of people and objects have been widely used in many areas, including medicine.[36,37,47] Icon arrays, however, have the disadvantage of taking a lot physical space compared to other visual formats (e.g., bars and pies).[36,47,50] It would then be interesting to explore the effectiveness of these alternative visual formats in reducing denominator neglect. If they are useful, there might be some conditions in which they are more convenient than icon arrays (e.g., when reporting complex information under space constraints).

Another limitation of our studies is that they involved information about the number of treated and nontreated people who *die* out of the number of those who do and do not receive the treatment. Previous research showed that denominator neglect varies depending on whether the information is framed in a positive or a negative way (e.g., choosing a sample from which to draw a *winning* v. a *losing* bean).[24,26] The matching task in a medical context would be estimating risk reduction when reporting the number of people who *survive* or *die* after receiving a treatment. Whether icon arrays also help reduce denominator neglect when the task is framed in a positive way is an avenue for future research.

Another open question relates to the fact that our studies did not involve real patient-doctor interactions.[73] Although our studies enabled us to draw clear conclusions and demonstrate the generalizability of our findings, it is possible that icon arrays would show additional benefits in clinical settings. Whether icon arrays eliminate denominator neglect in real patient-doctor interactions still needs to be studied.

## REFERENCES

1. Deber RB. Physicians in health care management: the patient–-physician partnership: changing roles and the desire for information. CMAJ. 1994;151:171–6.

2. Golbeck AL, Ahlers-Schmidt CR, Paschal AM, Dismuke SE. A definition and operational framework for health numeracy. Am J Prev Med. 2005;29:375–6.

3. Galesic M, Garcia-Retamero R. Statistical numeracy for health in United States and Germany. Arch Intern Med. 2010;170:462–468.

4. Garcia-Retamero R, Galesic M. Who profits from visual aids? The impact of numeracy and graphical literacy on understanding of medical risks. Soc Sci Med. 2010;70:1019–1025.

5. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest. 2008;8:53–96.

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

7. Peters E, Vastfjall D, Slovic P, Mertz CK, Mozzocco K, Dickert S. Numeracy and decision making. Psychol Sci. 2006;17:406–13.

8. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966–72.

9. Montori VM, Rothman RL. Weakness in numbers: the challenge of numeracy in healthcare. J Gen Intern Med. 2005;11:1071–2.

10. Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. Am J Public Health. 2009;99:2196–202.

11. Garcia-Retamero R, Galesic M. Heuristics. In: Kattan MW, ed. The Encyclopedia of Medical Decision Making. Thousand Oaks, CA: Sage; 2009.

12. Garcia-Retamero R, Galesic M. Risk Communication and Medical Decision Making: A Cross-Cultural Perspective. Springer. In press.

13. Kutner M, Greenberg E, Jin Y, Paulsen C. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy (NCES 2006–483). US Department of Education. Washington, DC: National Center for Education Statistics; 2006.

14. Peters E, Finucane ML, MacGregor DG, Slovic P. The bearable lightness of aging: judgment and decision processes in older adults. In: National Research Council, Stern PC, Carstensen LL, eds. The Aging Mind: Opportunities in Cognitive Research. Washington, DC: National Academies Press; 2000.

15. Hecht SA, Close L, Santisi M. Sources of individual differences in fraction skills. J Exp Child Psychol. 2003;86:277–302.

16. Reyna VF, Brainerd CJ. The origins of probability judgment: a review of data and theories. In: Wright G, Ayton P, eds. Subjective Probability. New York: John Wiley; 1994.

17. Siegler RS. Children's Thinking. Englewood Cliffs, NJ: Prentice Hall; 1991.

18. Denes-Raj V, Epstein S, Cole J. The generality of the ratio-bias phenomenon. Pers Soc Psychol Bull. 1995;21:1083–92.

19. Reyna VF. How people make decisions that involve risk: a dual-processes approach. Curr Dir Psychol. 2004;13:60–6.

20. Slovic P, Monahan J, MacGregor DG. Violence risk assessment and risk communication: the effects of using actual cases, providing instructions, and employing probability vs. frequency formats. Law Hum Behav. 2000;24:271–96.

21. Wolfe CR. Information seeking on Bayesian conditional probability problems: a fuzzy-trace theory account. J Behav Decis Making. 1995;8:85–108.

22. Reyna VF. Class inclusion, the conjunction fallacy, and other cognitive illusions. Dev Rev. 1991;11:317–36.

23. Reyna VF, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learn Individ Differ. 2008;18:89–107.

24. Denes-Raj V, Epstein S. Conflict between intuitive and rational processing: when people behave against their better judgment. J Pers Soc Psychol. 1994;66:819–29.

25. Pacini R, Epstein S. The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. J Pers Soc Psychol. 1999;76:972–87.

26. Pacini R, Epstein S. The interaction of three facets of concrete thinking in a game of chance. Think Reasoning. 1999;5:303–25.

27. Miller DT, Turnbull W, McFarland C. When a coincidence is suspicious: the role of mental simulation. J Pers Soc Psychol. 1989;57:581–9.

28. Stanovich KE, West RF. On the relative independence of thinking biases and cognitive ability. J Pers Soc Psychol. 2008;94:672–95.

29. Kirkpatrick L, Epstein S. Cognitive-experiential self-theory and subjective probability: evidence for two conceptual systems. J Pers Soc Psychol. 1992;63:534–44.

30. Acredolo C, O'Connor J, Banks L, Horobin K. Children's ability to make probability estimates: skills revealed through application of Anderson's functional measurement methodology. Child Dev. 1989;60:933–45.

31. Kokis J, Macpherson R, Toplak M, West RF, Stanovich KE. Heuristic and analytic processing: age trends and associations with cognitive ability and cognitive styles. J Exp Child Psychol. 2002;83:26–52.

32. Yamagishi K. When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. Appl Cogn Psychol. 1997;11:495–506.

33. Fagerlin A, Ubel PA, Smith DM, Zikmund-Fisher BJ. Making numbers matter: present and future research in risk communication. Am J Health Behav. 2007;31:47–56.

34. Zikmund-Fisher BJ, Ubel PA, Smith DM, et al. Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. Patient Educ Couns. 2008;73:209–14.

35. Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks: overcoming low numeracy. Health Psychol. 2009;28:210–6.

36. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. JAMIA. 2006;13:608–18.

37. Paling J. Strategies to help patients understand risks. BMJ. 2003;327:745–8.

38. International Patient Decision Aids Standards (IPDAS). Criteria for judging the quality of patient decision aids, 2005. Available from: http://www.ipdas.ohri.ca/IPDAS_checklist.pdf

39. Grossarth-Maticek R, Ziegler R. Randomized and non-randomized prospective controlled cohort studies in matched pair design for the long-term therapy of corpus uteri cancer patients with a mistletoe preparation. Eur J Med Res. 2008;13:107–20.

40. Lichtenberg P, Levinson D, Sharshevsky Y, Feldman D, Lachman M. Clinical case management of revolving door patients: a semi-randomized study. Acta Psychiatr Scand. 2008;117:449–54.

41. Sedrakyan A, Shih C. Improving depiction of benefits and harms: analyses of studies of well-known therapeutics and review of high-impact medical journals. Med Care. 2007;45:S23–8.

42. Walitza S, Werner B, Romanos M, Warnke A, Gerlach M, Stopper H. Does methylphenidate cause a cytogenetic effect in children with attention deficit hyperactivity disorder? EHP. 2007;115:936–40.

43. Brainerd CJ, Reyna VF. Inclusion illusions: fuzzy-trace theory and perceptual salience effects in cognitive development. Dev Rev. 1990;10:365–403.

44. Lloyd FJ, Reyna VF. A Web exercise in evidence-based medicine using cognitive theory. J Gen Intern Med. 2001;16:94–9.

45. Lloyd FJ, Reyna VF, Whalen P. Accuracy and ambiguity in counseling patients about genetic risk. Arch Intern Med. 2001;161:2411–3.

46. Brainerd CJ, Reyna VF. Autosuggestibility in memory development. Cogn Psychol. 1995;28:65–101.

47. Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. BMJ. 2002;324:827–30.

48. Fagerlin A, Wang C, Ubel PA. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? Med Decis Making. 2005;25:398–405.

49. Stone ER, Sieck WR, Bull BE, Yates JF, Parks SC, Rush CJ. Foreground:background salience: explaining the effects of graphical displays on risk avoidance. Organ Behav Hum Decis Process. 2003;90:19–36.

50. Lipkus IM. Numeric, verbal, and visual formats of convening health risks: suggested best practices and future recommendations. Med Decis Making. 2007;27:695–713.

51. Tuijnman A. Benchmarking Adult Literacy in America: An International Comparative Study (International Adult Literacy Survey). Washington, DC: Department of Education, OVAE; 2000.

52. Finucane ML, Slovic P, Hibbard JH, Peters E, Mertz CK, MacGregor DG. Aging and decision-making competence: an analysis of comprehension and consistency skills in older versus younger adults considering health-plan options. J Behav Dec Making. 2002;15:141–64.

53. Finucane ML, Mertz CK, Slovic P, Scholze E. Task complexity and older adults' decision-making competence. Psychol Aging. 2005;20:71–84.

54. Baltes PB, Lindenberger U. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? Psychol Aging. 1997;12:12–21.

55. Baltes PB, Staudinger UM, Lindenberger U. Lifespan psychology: theory and application to intellectual functioning. Annu Rev Psychol. 1999;50:471–507.

56. Löckenhoff C, Carstensen LL. Socioemotional selectivity theory, aging, and health: the increasingly delicate balance between regulating emotions and making tough choices. J Pers. 2004;72:1395–424.

57. Johnson MMS. Age differences in decision making: a process methodology for examining strategic information processing. J Gerontol Psychol Sci. 1990;45:75–8.

58. Blanchard-Fields F. Emotion and everyday problem solving in adult development. In: Magai C, McFadden SH, eds. Handbook

of Emotion, Adult Development, and Aging. London: Academic Press; 1996.

59. Labouvie-Vief G. Dynamic development and mature autonomy: a theoretical prologue. Hum Dev. 1982;25:161–91.

60. Sinnot JD. Everyday Problem Solving: Theory and Application. New York: Praeger; 1989.

61. Skolbekken JA. Communicating the risk reduction achieved by cholesterol reducing drugs. BMJ. 1998;316:1956–8.

62. Stone ER, Yates JF, Parker AM. Effects of numerical and graphical displays on professed risk-taking behavior. J Exp Psychol Anim Behav Process Appl. 1997;3:243–56.

63. Lunney GH. Using analysis of variance with a dichotomous dependent variable: empirical study. J Educ Meas. 1971;7:263–9.

64. Cleary PD, Angel R. The analysis of relationships involving dichotomous dependent-variables. J Health Soc Behav. 1984;25:334–48.

65. Greenhouse SW, Geisser S. On methods in the analysis of profile data. Psychometrika. 1959;24:95–112.

66. Kahneman D. A psychological point of view: violations of rational rules as a diagnostic of mental processes. Behav Brain Sci. 2000;23:681–3.

67. Garcia-Retamero R, Hoffrage U, Dieckmann A. When one cue is not enough: combining fast and frugal heuristics with compound cue processing. Q J Exp Psychol. 2007;60:1197–215.

68. LeBoeuf RA, Shafir E. Deep thoughts and shallow frames: on the susceptibility to framing effects. J Behav Decis Making. 2003;16:77–92.

69. Chapman RH. The development of children's understanding of proportions. Child Dev. 1975;46:141–8.

70. Hoemann HW, Ross BM. Children's concepts of chance and probability. In: Brainerd C, ed. Children's Logical and Mathematical Cognition. New York: Springer; 1982.

71. Surber CF, Haines BA. The growth of proportional reasoning. Ann Child Dev. 1987;4:35–87.

72. Galesic M, Gigerenzer G, Straubinger N. Natural frequencies facilitate accurate judgments about medical screenings for elderly and people with lower numeracy skills. Med Decis Making. 29:368–71.

73. Garcia-Retamero R, Galesic M. Physician trust. In: Kattan MW, ed. The Encyclopedia of Medical Decision Making. Thousand Oaks, CA: Sage; 2009.

74. Schwartz LM, Woloshin S, Black WC, Welch HG. Using a drug facts box to communicate drug benefits and harms. Ann Intern Med. 2009;150:516–27.

75. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. BMJ. 2003;327:741–4.

76. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. Psychol Rev. 1995;102:684–704.

77. Covey JA. Meta-analysis of the effects of presenting treatment benefits in different formats. Med Decis Making. 2007;27:638–54.

78. Gigerenzer G, Hertwig R, van den Broek E, Fasolo B, Katsikopoulos K. ''A 30% chance of rain tomorrow'': how does the public understand probabilistic weather forecasts? Risk Anal. 2005;25:623–9.

79. Neurath O. International Picture Language. London: Kegan Paul, Trench, Trubner; 1936. Available from: http://www.vknn.at/neurath/isotype.html