Graph Literacy: A Cross-Cultural Comparison

Mirta Galesic, PhD, Rocio Garcia-Retamero, PhD

Background. *Visual displays are often used to communicate important medical information to patients. However, even the simplest graphs are not understood by everyone.* ***Objective***. *To develop and test a scale to measure health-related graph literacy and investigate the level of graph literacy in the United States and Germany.* ***Design***. *Experimental and questionnaire studies.* ***Setting***. *Computerized studies in the laboratory and on probabilistic national samples in the United States and Germany.* ***Participants***. *Nationally representative samples of people 25 to 69 years of age in Germany (*n *= 495) and the United States (*n *= 492). Laboratory pretest on 60 younger and 60 older people.* ***Measurements***. *Psychometric properties of the scale (i.e., reliability, validity, discriminability) and level of graph literacy in the two countries.* ***Results***. *The new graph literacy scale predicted which patients can benefit from visual aids and had promising measurement properties. Participants in both countries completed approximately 9 of 13 items correctly (in Germany, $\bar{x}$ = 9.4, s = 2.6; in the United States, $\bar{x}$ = 9.3, s = 2.9). Approximately one third of the population in both countries had both low graph literacy and low numeracy skills.* ***Limitations***. *The authors focused on basic graph literacy only. They used a computerized scale; comparability with paper-and-pencil versions should be checked.* ***Conclusions***. *The new graph literacy scale seems to be a suitable tool for assessing whether patients understand common graphical formats and shows that not everyone profits from standard visual displays. Research is needed on communication formats that can overcome the barriers of both low numeracy and graph literacy.*

Key words: *patient decision making; risk communication; risk perception; shared decision making; education.* ***(Med Decis Making 2011;31:444–457)***

Graph literacy, or the ability to understand graphically presented information, is essential in everyday life: graphs are ubiquitous in newspapers and magazines, on television, and on the Internet. Graphs often provide important information for medical, financial, nutritional, and political choices. Recent studies have shown that graphical displays—bar charts, pie charts, line plots, and icon arrays—can improve understanding of the risks and benefits associated with medical treatments, screenings, and lifestyles.[1–3] For example, icon arrays help people with low numeracy skills to understand treatment-related risk reductions.[4–6] They can also promote consideration of beneficial treatments that have side effects[7] and limit the biases induced by anecdotal narratives.[8]

However, even the simplest graphs may be difficult to understand for many people. Bar charts, pie charts, and line plots were first used in the late 18th and early 19th centuries. William Playfair, an economist and author of *Commercial and Political Atlas* (1786) and *Statistical Breviary* (1801), first used those graphical formats.[9,10] Icon arrays are even more recent: they began to be widely used only in the early 20th century, when Otto Neurath (1882–1945), a philosopher, economist, and a prominent member of the Vienna Circle, used them to explain complex social and economic statistics to uneducated Viennese.[11] In other words, in most of human history, there were no graphical representations of statistical

---

Received 8 March 2009 from the Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany (MG, RG-R), and the Department of Experimental Psychology, University of Granada, Spain (RG-R). This study is part of two projects: "Helping People With Low Numeracy to Understand Medical Information," funded by the Foundation for Informed Medical Decision Making (United States) and the Max Planck Society (Germany), and "How to Improve Understanding of Risks About Health (PSI2008-02019)," funded by the Ministerio Ciencia e Innovación (Spain). The authors declare independence of the funding agencies. The authors thank Julia Schooler and Anita Todd for editing the manuscript. Revision accepted for publication 10 April 2010.

Address correspondence to Mirta Galesic, PhD, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany; telephone: +49 30 8240 6354; fax: +49 30 824 9939; e-mail: galesic@mpib-berlin.mpg.de.

DOI: 10.1177/0272989X10373805

information—at least not in the formats that are ubiquitous today. Therefore, there is no immediate reason that people should understand such graphs intuitively. For example, although pie charts are used very frequently to communicate various statistical facts, the scientific evidence about their usefulness is equivocal.[10,12,13]

The first aim of this article is to develop a scale that can be used to assess the graphical literacy skills needed to understand risks in the health domain. To date, graph understanding has not been assessed by any health literacy instrument.[14] Within national assessments of literacy,[15] only a few document literacy questions investigate selected aspects of graph comprehension, but most of these items are relatively complex and require an advanced understanding of graphs. In a similar vein, Kramarski and Mevarech developed a 36-item Graph Interpretation Test[16] to investigate the effects of different instructional methods on the ability of eighth-grade students to interpret graphs in general. However, their test is not embedded in the health domain, is too long to be used in clinical practice, is focused mostly on line graphs, and involves questions that require relatively advanced graph interpretation skills. Therefore, we have constructed a new graphical literacy scale that a) investigates both basic graph-reading skills and more advanced graph comprehension, b) involves examples of different types of graphs, c) is embedded in the context of medical decisions, and d) is brief enough for use in everyday clinical practice.

The second aim of our article is to investigate the extent and distribution of graph (il)literacy on probabilistic national samples in the United States and Germany—two countries with very different educational and medical systems. It is known that a significant part of the general population has problems understanding numerically presented statistical data, particularly lower educated people.[17,18] The same may hold for understanding of graphs. Indeed, a portion of the population may have problems with understanding *both* numerically and graphically presented information. To promote informed medical decision making, it is important to identify these people and either train them to understand existing forms of graphs or offer them representations that can be understood without training.

In what follows, we first describe the development and evaluation of the new graph literacy scale. We then report on the level of graph literacy in the United States and Germany.

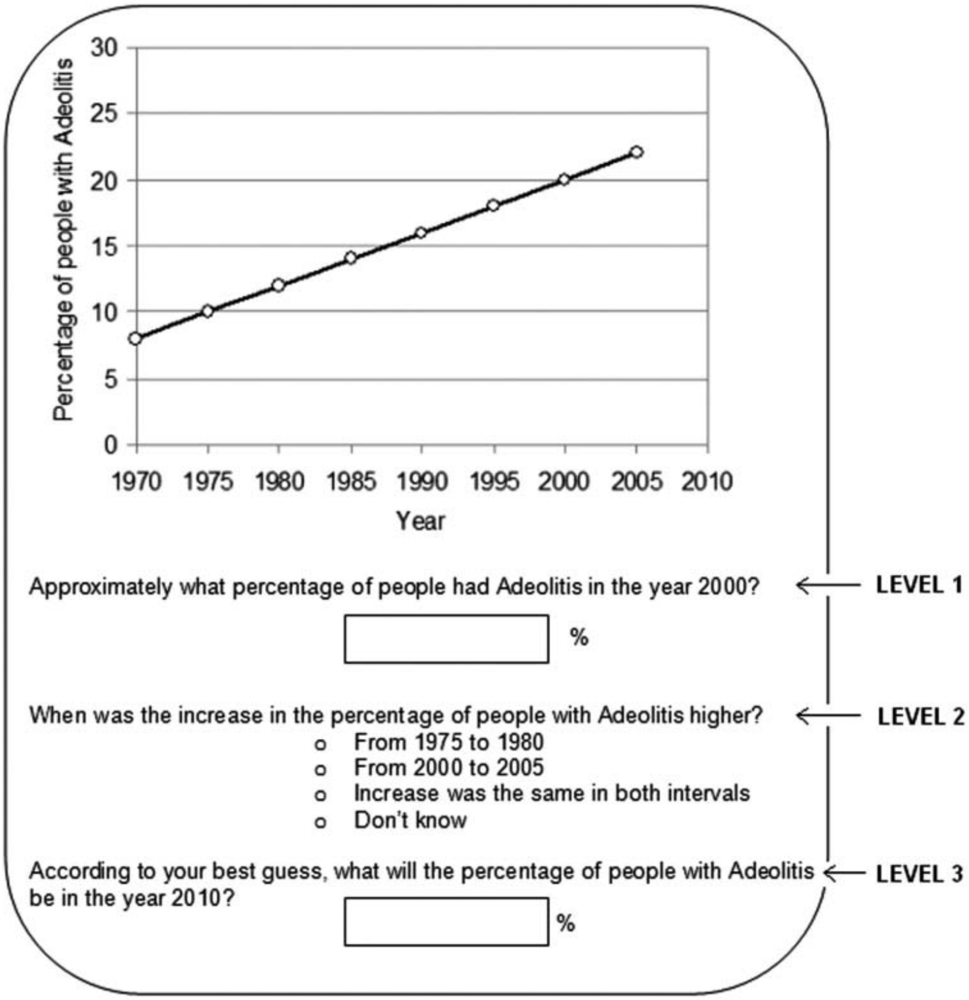
DEVELOPMENT OF THE GRAPH LITERACY SCALE

To determine which items to include in our graph literacy scale, we started from the traditional division of graph comprehension skills on 3 levels.[9] On the first level, one should have the ability to *read the data*, that is, to find specific information in a graph. For example, one should be able to read the height of a particular bar within a bar chart or the number of icons of a particular type in an icon array. On the second level, one should be able to *read between the data*, that is, to find relationships in the data as shown on a graph. For instance, one should be able to read the difference between 2 bars or sets of icons or sum up several slices on a pie chart. The highest level of graph comprehension is reflected in the ability to *read beyond the data*, or make inferences and predictions from the data. For example, one should be able to project a future trend from a line chart, understand the importance of attending to scale ranges and scale labels when comparing 2 charts, and use the existing labels to interpolate scale labels that are missing. For examples of items measuring each of the 3 skills, see Figure 1.

Following this classification, we developed the 42 items included in the initial scale. In creating these items, we were guided by several principles. First, we embedded all graphs in a medical context—each presented data that patients could realistically encounter when making health-related decisions. For example, we included tasks dealing with the communication of medical risks, treatment efficiencies, prevalence of diseases, and so on. Second, we designed items to cover 4 frequently used graph types—line plots, bar charts, pies, and icon arrays.[1–3,12] Third, we varied the complexity of graphs by changing the number of data series displayed on the same graph (1, 2, or 3) and whether the data were unidimensional or bidimensional.

Participants

We pretested the initial version of the scale on convenience samples of 60 German students (33 women, mean age 24.8 years) and 60 German older people (31 women, mean age 67.0 years, 31 with high school and 29 with college education), recruited from the pool of participants maintained by the Max Planck Institute for Human Development in Berlin. They were compensated at 10 euros per hour.



Approximately what percentage of people had Adeolitis in the year 2000?

☐ %

When was the increase in the percentage of people with Adeolitis higher?

- ○ From 1975 to 1980
- ○ From 2000 to 2005
- ○ Increase was the same in both intervals
- ○ Don't know

According to your best guess, what will the percentage of people with Adeolitis be in the year 2010?

☐ %

Figure 1 Examples of tasks measuring 3 levels of graph comprehension. Level 1 is the ability to read the data. Level 2 is the ability to read between the data. Level 3 is the ability to read beyond the data.

Procedure

The scale was administered on computers in our laboratory. Besides the newly developed, 42-item graphical literacy scale, we also administered several previously developed items to evaluate convergent validity. These items investigated several aspects of graph comprehension, including reading the data, reading between the data, and reading beyond the data.[9] We selected items from the International Adult Literacy Survey (IALS),[19] the National Assessment of Adult Literacy (NAAL),[20] and the Program for International Student Assessment (PISA).[21] We also included 2 items from the Kramarski and Mevarech Graph Interpretation Test[16] and an additional unpublished item kindly shared with us by those authors. This last item measured the ability to recognize which of several graphs depicts the relationship between time and distance of a car traveling from one place to another and back.

Participants completed 2 additional measures that served to establish the divergent validity of the new graphical literacy scale. First, they completed 3 of the 4 items from the numeracy part of the short form of the Test of Functional Health Literacy.[22] The excluded item (understanding the information on an appointment slip for a diabetic clinic) was judged to be too culturally and content specific. Second, participants completed 4 numeracy items selected from Schwartz and others[18] and Lipkus and others.[23] Both scales included items designed to measure the basic numerical skills needed to understand statistical information.

Results

We evaluated the scale on several criteria: duration, discriminability (i.e., the ability to differentiate between those taking the test[24]), reliability, and validity.

Duration. The initial version of the graphical literacy scale took on average 21 minutes to complete (s = 8.0; median: 19 minutes). Older people took significantly longer to complete the scale compared to the students ($\bar{x}$ = 27, s = 7.0 v. $\bar{x}$ = 16, s = 4.1 minutes, respectively).

Discriminability. Participants completed from 10 to 41 items correctly, with an overall mean of 34 correct items (students: 36 items; older adults: 31 items). The probability of answering individual items correctly ranged from 10% to 99%, with a mean of 80%. The discriminability of items was higher among the older adults than among the students.

Reliability. The correlations between individual items and the total score ranged from .07 to .63, with a mean of .38. Cronbach's alpha was .85, indicating a satisfactory level of internal consistency.

Validity. The average correlation of the total score with the graph comprehension items taken from the existing literacy questionnaires was .44, indicating a satisfactory convergent validity. As for the divergent validity, the correlation with the test of functional literacy was .19, suggesting that it measures a different type of skill. The correlation with the numeracy scale was relatively high at .51, suggesting that the same meta-cognitive abilities that lead to high numeracy scores also foster good graphical literacy skills.

EVALUATION OF THE GRAPH LITERACY SCALE

On the basis of the pretest results, we selected 13 items to be included in the refined version of the scale. The items were chosen according to the following criteria: 1) discriminability (percent correct lower than 90%), 2) item-total correlation of at least .3, 3) correlation with existing graph comprehension items of at least .3, 4) representation of the 3 levels of graph comprehension (reading the data, reading between the data, and reading beyond the data) and different types of graphs (bar, pie, and line charts, as well as icon arrays), and 5) the scale had to be short—ideally not longer than 10 minutes—and efficient, with each item measuring a somewhat different aspect of graph literacy. The complete scale is given in Appendix A.

Participants

The final version of the scale was administered on probabilistic national samples of German and US populations from 25 to 69 years of age, within the project "Helping People with Low Numeracy to Understand Medical Information," funded by the Foundation for Informed Medical Decision Making. The samples were chosen from large panels of households selected through probabilistic telephone surveys and equipped with Internet access in case they did not have it. These panels, maintained by the companies Forsa (Germany, 20,000 or 11% of initially contacted households) and Knowledge Networks (United States, 43,000 or 16% of initially contacted households), enable computerized surveying over the Internet while facilitating generalizations from the data to the general population. Such panels have been used successfully in a number of studies in the areas of health and medicine, political and social sciences, and economics and public policy.[25–29] Methodological studies have shown that data from such panels are comparable to the results obtained through traditional probabilistic sampling methods.[30]

Of the panel members who were invited to participate in our study, 52% in Germany and 54% in the United States completed the questionnaire within the designated survey period (3 weeks). The survey was completed by 1001 participants in Germany and 1009 participants in the United States. Of those, a random half of the participants completed the graph literacy scale, resulting in 495 participants in Germany and 492 participants in the United States, and are included in the analyses that follow. The other half completed different tasks related to understanding of numerical information. These results will be reported elsewhere.

Table 1 Characteristics of the Sample Used for Evaluation of the Graph Literacy Scale

	Germany (n = 495)			United States (n = 492)		
	Sample Size (Unweighted)	**Sample % (Weighted)**	**Population %**[a]	**Sample Size (Unweighted)**	**Sample % (Weighted)**	**Population %**[b]
Male sex	254	50.3	49.9	236	48.4	49.2
Age (years)						
25–39	125	31.4	32.5	120	31.2	35.7
40–54	210	39.0	39.9	194	40.6	38.3
55–69	160	29.6	27.7	178	28.2	26.1
Education						
High school or less	393	74.1	72.3	356	44.5	44.6
Some college or more[c]	102	25.9	27.7	136	55.5	55.4

a. Source: Statistisches Bundesamt Deutschland, Microcensus, 2007 (https://www-genesis.destatis.de/genesis/online; accessed September 15, 2008).
b. Source: US Census Bureau, Current Population Survey, 2008 Annual Social and Economic Supplement (http://pubdb3.census.gov/macro/032008/perinc/toc.htm; accessed September 15, 2008).
c. In Germany, this category includes people with Abitur.

Characteristics of the 2 samples, compared with the official population estimates, are shown in Table 1. According to official statistics, there are more highly educated people in the United States than in Germany. Consequently, lower educated people in the United States were oversampled to make sure that the subsamples of lower educated people, which were the focus of this project, were comparable between the countries. For all between-country comparisons, we used weights to achieve realistic proportions of education groups. In the analyses that follow, unless otherwise stated, all results are based on the weighted data.

Procedure

The questionnaire was administered through the Web. Some respondents (64% in Germany and 62% in the United States) completed the questionnaire via personal computers, whereas the rest used WebTV with infrared keyboards. We checked whether this variable affected the results but did not find any differences between the 2 groups in either country.

We put special effort into making the German and English versions of the questionnaire comparable. All questions were developed in English and edited by a native English speaker, translated into German by a native German speaker with excellent knowledge of English, back-translated into English by another person of equivalent language skills, and compared with the original English version. Any inconsistencies were resolved by a native German speaker and an excellent English speaker familiar with the research objectives. Finally, the English and the German versions were compared and edited by a bilingual German and English speaker who was raised simultaneously in both languages and who lived in both Germany and the United States for significant periods of time. When programming the questionnaire, special care was taken that the interface looked the same in both German and the US versions, including background color, number of questions on the screen, font color and size, and the design of the plausibility checks.

We also conducted 2 experiments to assess the predictive validity of the graph literacy scale. The experiments examined whether the graph literacy scale predicts usefulness of visual displays for people who have problems understanding numerical information, reflected in their low numeracy scores. First, we presented information about reducing the risks of stroke (task 1) and heart attack (task 2) for patients with symptoms of arterial disease. In both tasks, respondents were presented with data from a fictitious study, in which one randomly selected group of 100 patients took a placebo, whereas another group of 100 patients took a fictitious drug named Vitarilen. The information was presented to randomly selected subgroups of respondents either only numerically or numerically with additional graphical displays in the form of bar charts (see Appendix B). We asked the participants to answer

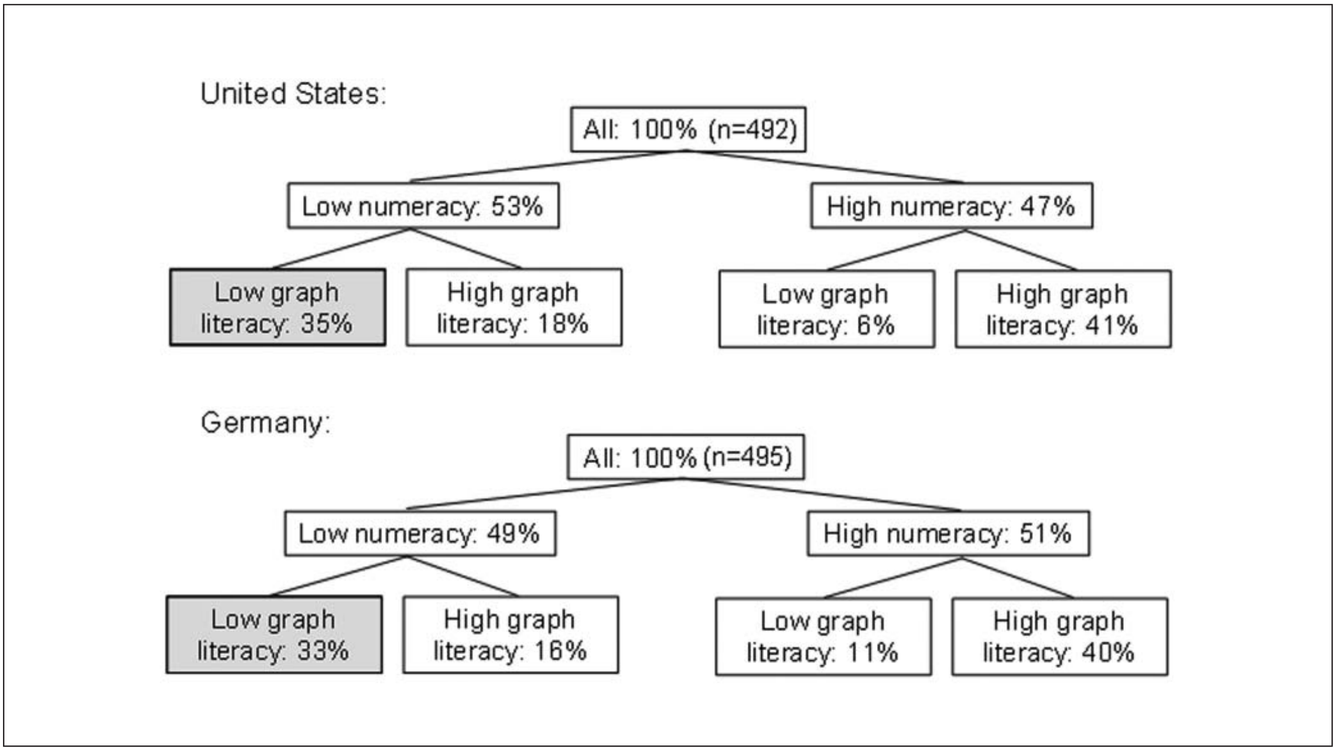


Figure 2 Graph literacy skills among people with low and high numeracy skills, in the United States and Germany. Groups are defined by median split (for numeracy: 6 correct answers; for graph literacy: 9 correct answers).

how many of 1000 patients would have a stroke (task 1) or a heart attack (task 2) with and without Vitarilen.

Second, we assessed whether graph literacy was associated with the ability to use icon arrays to reduce denominator neglect (the focus on the number of treated and nontreated patients who died, without sufficiently considering the overall number of patients). We presented information about the risk of dying of a heart attack for people with high cholesterol with or without taking different fictitious drugs. Depending on the experimental condition, the overall number of patients who took and who did not take the drugs was either the same (e.g., 100 for both groups) or different (e.g., 100 and 800), making it important to pay attention to both numbers to infer risk reductions accurately. The information was presented either only numerically or both numerically and visually using icon arrays (see Appendix B for the materials). Participants were asked what the percent reduction was in the number of deaths among people took the drug relative to those who did not.

The Ethics Committee of the Max Planck Institute for Human Development approved the methodology, and all participants consented to participation through an online consent form at the beginning of the survey.

Results

The final version of the graph literacy scale took 9 to 10 minutes to complete (in Germany, $\bar{x} = 9.2$, $s = 5.7$; in the United States, $\bar{x} = 10.1$, $s = 5.7$) and had good measurement properties. When calculating participants' results, we required exactly correct answers to all questions except for question Q7, where we allowed as correct all answers that fell between 23 and 25.

Reliability. Cronbach's alpha was .74 in Germany and .79 in the United States, and average item-total correlations were .37 and .42 in Germany and the United States, respectively, indicating a satisfactory level of internal consistency. Average correlations between individual items were .19 in Germany and .23 in the United States, indicating

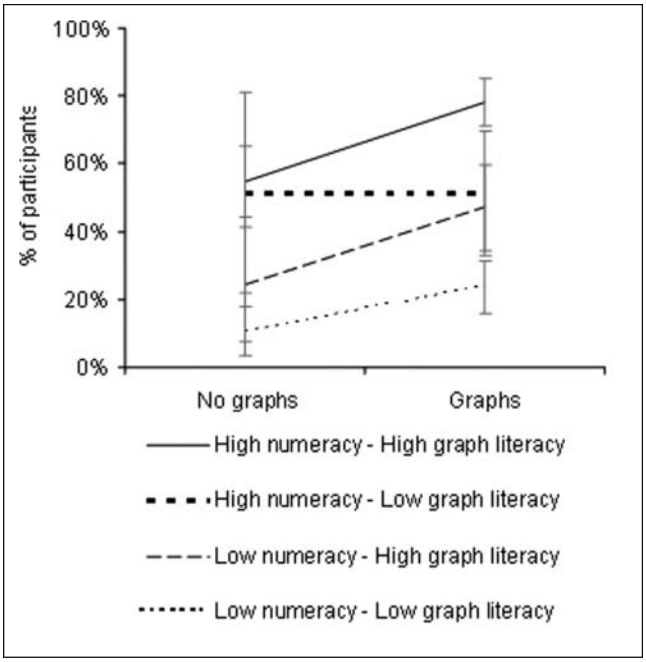


Figure 3 Percentage of participants who answered correctly at least one of the tasks in experiment 1, with and without graphs, by numeracy and graph literacy groups (error bars are 95% confidence interval [CI]).

that each item measured a somewhat distinctive aspect of graph literacy. This is reflected in relatively low internal consistencies of items testing each skill: .62, .48, and .45, for levels 1, 2, and 3, respectively.

Validity. The average correlation of the total score with education level was .29 in Germany and .54 in the United States. As shown in Figure 2, the correlation with numeracy was also substantial (.47 in Germany and .55 in the United States). Correlation with the graph comprehension items from the existing literacy questionnaires was .32 in Germany and .50 in the United States, indicating satisfactory convergent validity. The existing items correlated most highly with items testing basic and moderately advanced graph literacy skills ("reading the data" and "reading between the data"; average correlation .36). The correlation with items testing more advanced graph literacy skills ("reading beyond the data") was lower but nevertheless substantial (.33).

PREDICTIVE VALIDITY OF THE GRAPH LITERACY SCALE

We describe here a portion of the results to illustrate the validity of the graph literacy scale; for more details, see Garcia-Retamero and Galesic.[31]

In the first experiment, we tested whether graph literacy predicts the usefulness of simple bar charts in interpreting risk reductions. Figure 3 presents the percentage of respondents who answered at least 1 of the 2 tasks correctly. As expected, respondents whose numeracy was high compared to average were more likely to answer correctly than those with low numeracy. Critically, graphs were more helpful to the low-numeracy respondents who had high rather than low graph literacy. A similar trend is observed for the high-numeracy participants. Results were similar in both countries (in a logistic regression, the interaction of country, presence of graphs, and graph literacy was not significant).

In the second experiment, we tested the effect of graph literacy on the usefulness of another type of visual aid—icon arrays—in reducing denominator neglect when evaluating the benefits of medical treatments. Figure 4 shows the percentage of participants who answered correctly. Again, high-numeracy respondents were more likely to give the right answer. However, those low-numeracy respondents who had higher graph literacy profited significantly from the addition of icon arrays. Respondents with low numeracy and low graph literacy did not show improvement due to icons. As in experiment 1, these results were similar in both countries.

EXTENT OF GRAPH LITERACY IN THE UNITED STATES AND GERMANY

Participants in both countries completed approximately 9 of 13 items correctly (in Germany, $\bar{x} = 9.4$, $s = 2.6$; in the United States, $\bar{x} = 9.3$, $s = 2.9$). Table 2 shows percentage of correct responses to each of the items. The items testing the ability to "read the data" were answered correctly by a large majority of participants in both countries. The items testing the 2 more advanced skills—reading between the data and reading beyond the data—were more difficult. The most difficult item was the one that required noticing that it is not possible to compare the effectiveness of 2 different drug treatments when the data are displayed on different charts with unlabeled axes (Q11). Only 20% of participants in the United States and 16% in Germany knew this. A similar item (Q10), testing the ability to notice that 2 different graphs present the same data but use different scale ranges, produced a higher but still troubling

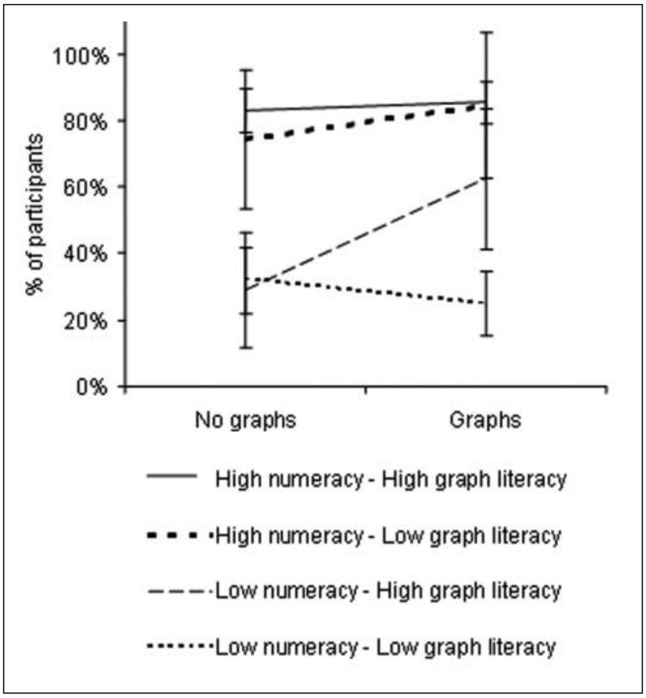


Figure 4 *Percentage of participants who answered correctly in experiment 2, with and without graphs, by numeracy and graph literacy groups (error bars are 95% confidence interval [CI]).*

level of inaccuracy, with 66% of participants in the United States and 63% in Germany giving the correct answer.

Of particular concern for health communicators is that a significant portion of both populations has both low numeracy and low graph literacy skills. As Figure 2 shows, approximately one third of people in both countries are likely to have problems understanding both numerically presented information and standard visual displays.

DISCUSSION

We developed and evaluated a graph literacy scale to identify people who have problems understanding graphically presented information, particularly information related to health issues. The scale has promising psychometric properties and may be suitable for use in many clinical and research circumstances. The scale successfully identified people for whom graphically presented information may be very useful but also those who are less likely to profit from visual aids. Among people with low numeracy skills, who are disadvantaged when it comes to grasping a host of numerical concepts that are prerequisites for understanding health-relevant risk communications,[32] a significant portion (approximately one third) can be aided by means of standard visual displays. However, a large percentage of low-numeracy people also have low graph literacy skills, and they may require either specially designed information formats that are undemanding in terms of both numeracy and graph literacy, such as analogies[33,34] or natural frequencies[35] and/or additional training in the use of standard graphs.

We administered the scale on probabilistic national samples in Germany and the United States. In both countries, the scores were highest on items designed to measure the most basic graph comprehension skill: reading the data. On average, 85% of people in Germany and 86% in the United States answered these questions correctly. The 2 more advanced skills had significantly lower average scores: about two thirds of people in each country were able to answer these questions. Although these percentages may seem high, it is important to note that there are still significant parts of the population that cannot perform elementary tasks involving very simple graphs. For example, 16% of Americans (12% of Germans) do not know what a quarter of a pie chart is in percentages (Q3). Similarly, 15% of people in the United States (17% in Germany) cannot read the height of a bar chart with fully labeled axes and gridlines as an additional help (Q1). These percentages translate into rather striking numbers when expanded to the total adult population 25 to 69 years of age in both countries. In addition, we found that graph literacy correlates with education in both Germany and the United States. This result suggests that understanding graphs is not entirely intuitive but requires a certain level of meta-knowledge about graphs acquired through formal education. The correlation of graph literacy and education was stronger in the United States than in Germany. This may be the result of differences in education systems, particularly the stronger focus on math and science education in Germany from an early age.[36]

By design, internal consistency and interitem correlations among graph items demonstrated considerable heterogeneity, and the internal consistencies of items testing each skill were low. To make the best use of the short time available for completing the scale, we designed an instrument that

Table 2 Percentage of Correct Responses to Items Included in the Final Scale

Items	Overall % Correct Responses	
	Germany (n = 495)	United States (n = 492)
Reading the data		
Q1. Reading off a point on a bar chart	82.7	84.6
Q3. Knowing what a quarter of a pie is in %	87.7	83.5
Q5. Reading off a point on a line chart	81.7	84.8
Q8. Reading off number of icons	88.6	90.3
Average	85.2	85.8
Reading between the data		
Q2. Determining difference between 2 bars	67.1	69.6
Q4. Summing slices within a quarter of a pie	74.2	77.6
Q6. Comparing slopes of a line at 2 intervals	82.1	61.6
Q9. Determining difference between 2 groups of icons	51.0	58.1
Average	68.6	66.7
Reading beyond the data		
Q13. Reading off a point on a bar chart when bar falls between 2 labels	80.1	75.2
Q7. Projecting future trend from a line chart	81.8	79.2
Q10. Comparing 2 bar charts: Attending to scale range	62.8	66.1
Q11. Comparing 2 line charts: Attending to scale labels	15.5	19.3
Q12. Differentiating slope and height of a line	86.1	77.5
Average	65.3	63.5
Mean number of correct answers ($s_{\bar{x}}$); max = 13	9.4 (.17)	9.3 (.18)

captured different aspects of graph literacy and contained no redundant items. On each skill level, we intentionally included items involving very different visual displays: bars, pies, lines, and icon arrays.

Although we designed items reflecting different levels of graph literacy, we did not aspire to design a Guttman scale[24] because we wanted to keep the scale short and broad in scope. Understanding graphs includes a number of loosely related processes, from perceptual and interpretative to integrative processes.[37,38] It would be difficult to systematically test these processes on each skill level and for different types of graphs in the time available in most clinical and research settings. Therefore, we used the framework of different skill levels to select a diverse set of items rather than to systematically test all processes involved in each skill. Nevertheless, the majority of participants who answered more difficult items correctly (level 3 skill) also answered the less difficult items well. For example, on average, of those who answered correctly an item on the third level of difficulty (read beyond the data), 90% answered correctly items on the first level (read the data), compared to only 73% of those who did not answer the level 3 items correctly. Similarly, of those who gave correct answers to items on the second level of difficulty (read between the data), 92% answered items on the first level correctly, compared to only 72% of those who did not answer the level 2 items correctly.

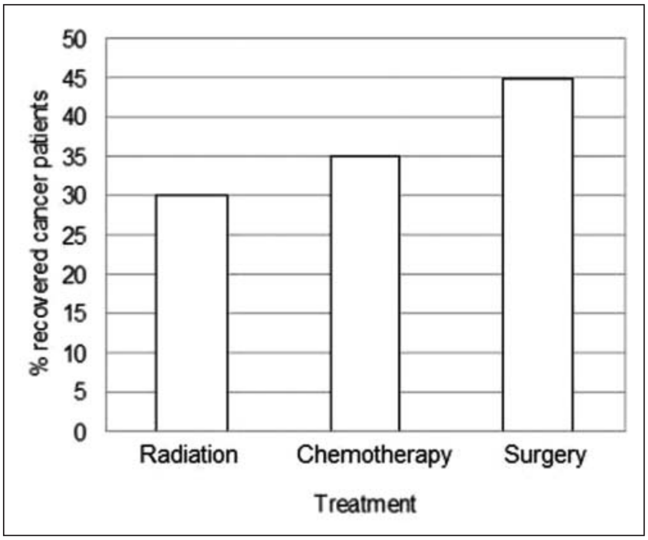
People with low graph literacy often have low levels of numeracy skills (Figure 2). In fact, elementary graph literacy measured by our test correlates more highly with elementary statistical numeracy than with more advanced graph comprehension items (see section Development of the Graph Literacy Scale). Nevertheless, as we showed in experiments testing predictive validity of the scale, graph literacy predicts how helpful graphs are to people independent of numeracy. Graphs help low-numeracy people with relatively high graph literacy, but they do not help to those with low graph literacy. Furthermore, as Figure 4 shows, about a third of people who are below the median of the population in numeracy have above-median values for graph literacy. This relatively large proportion is not surprising given that most of our items do not require any calculation, with the exception of 2 questions that require fairly simple deduction of 2 integers (45–30 and 60–40). It is more likely that both numeracy and graph literacy skills require a certain level of meta-knowledge about statistics and the meaning of statistical indicators. Our research shows that to some people, this knowledge is more accessible in visual rather than numeric formats but also that a large segment of the population simply does not know enough statistics to be helped by any of the standard formats.

To the best of our knowledge, this article is the first to develop a graph literacy scale to identify people who have problems understanding graphically presented medical information. At the same time, it leaves several questions open. For instance, one avenue for future research could be to test the scale on physicians. Recent research on numeracy in health decision making has shown that not only patients but also their physicians have difficulty grasping numerical concepts that are prerequisites for understanding health-relevant risk communications.[39] Another open question relates to the generalizability of our scale. As we mentioned above, our aim was to develop an instrument that could be used to assess graphical literacy in the health domain. To what extent is our scale useful to evaluate graph literacy in general or in other important domains such as finance, nutrition, or education? Although our studies enabled us to draw clear conclusions and demonstrate the generalizability of our results, it is possible that there are substantial differences between domains. Furthermore, we used a computerized questionnaire, and equivalence of results obtained using paper and pencil should be checked. Finally, the present version of the graph literacy scale focuses on understanding of simple bar, line, and pie charts and icon arrays. Further research on understanding more complex graphs, such as survival curves, is needed.

Our research suggests that understanding of both numerical and standard graphical representations of statistical information requires a certain level of statistical thinking. However, unlike reading and writing, statistical thinking is not routinely taught in schools. As a result, a large part of the population is insufficiently prepared to cope with many novel risks and uncertainties of the modern world. The goal of informed decision making hinges on educating the general public to understand statistical information about medical treatments and on finding alternative, more intuitive formats for communicating risks.

APPENDIX A
Graph Literacy Scale (Final Version)

Here is some information about cancer therapies.



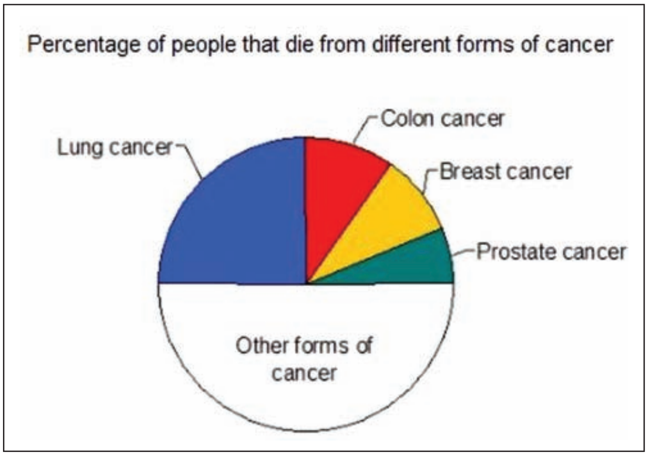
Q1. What percentage of patients recovered after chemotherapy?

_______ %

Q2. What is the difference between the percentage of patients who recovered after a surgery and the percentage of patients who recovered after radiation therapy?

_______ %

Here is some information about different forms of cancer.



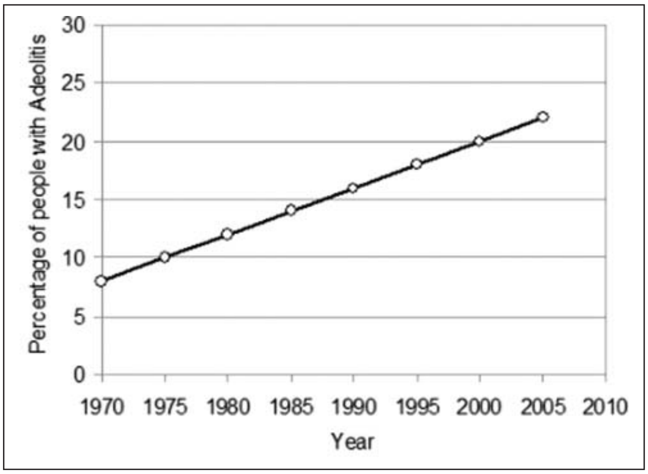
Q3. Of all the people who die from cancer, approximately what percentage dies from lung cancer?

________ %

Q4. Approximately what percentage of people who die from cancer die from colon cancer, breast cancer, and prostate cancer taken together?

________ %

Here is some information about an imaginary disease called Adeolitis.



Q5. Approximately what percentage of people had Adeolitis in the year 2000?

________ %

Q6. When was the increase in the percentage of people with Adeolitis higher?

From 1975 to 1980 1

From 2000 to 2005 2

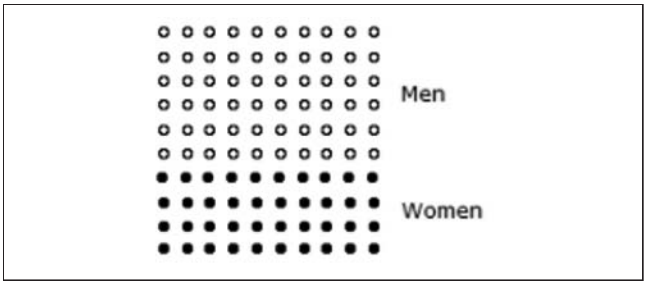
Increase was the same in both intervals 3

Don't know 4

Q7. According to your best guess, what will the percentage of people with Adeolitis be in the year 2010?

________ %

The following figure shows the number of men and women among patients with disease X. The total number of circles is 100.



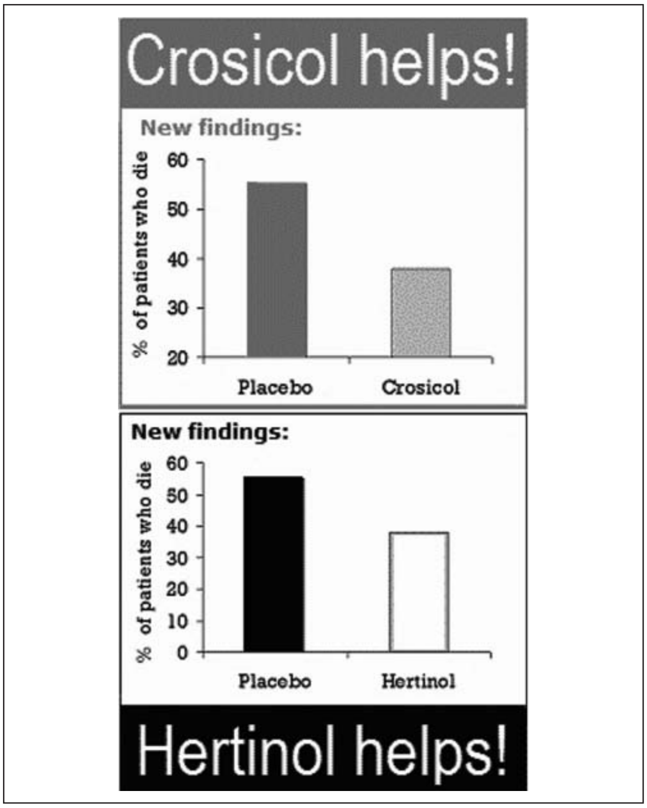
Q8. Of 100 patients with disease X, how many are women?

________ women

Q9. How many more men than women are there among 100 patients with disease X?

________ men

Q10. In a magazine you see two advertisements, one on page 5 and another on page 12. Each is for a different drug for treating heart disease, and each includes a graph showing the effectiveness of the drug compared to a placebo (sugar pill).



Compared to the placebo, which treatment leads to a larger decrease in the percentage of patients who die?

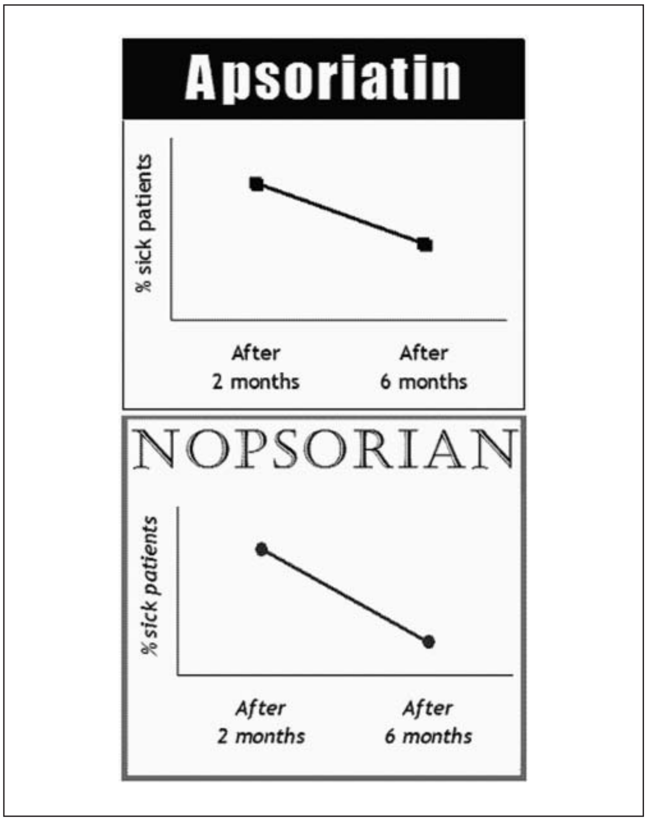
Crosicol 1

Hertinol 2

They are equal. 3

Can't say 4

Q11. In the newspaper you see two advertisements, one on page 15 and another on page 17. Each is for a different treatment of psoriasis, and each includes a graph showing the effectiveness of the treatment over time.



Which of the treatments contributes to a larger decrease in the percentage of sick patients?

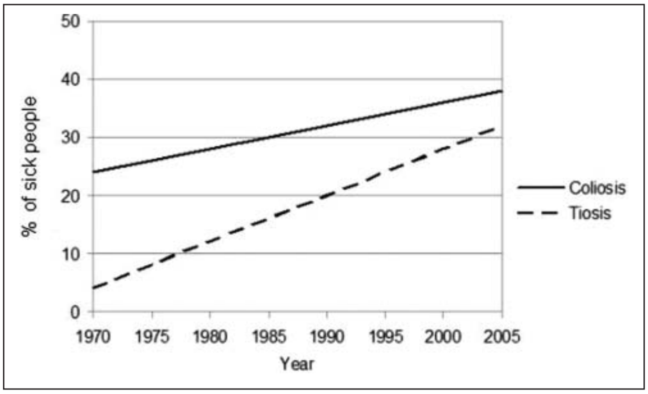
Apsoriatin 1

Nopsorian 2

They are equal. 3

Can't say 4

Q12. Here is some information about the imaginary diseases Coliosis and Tiosis.



Between 1980 and 1990, which disease had a higher increase in the percentage of people affected?

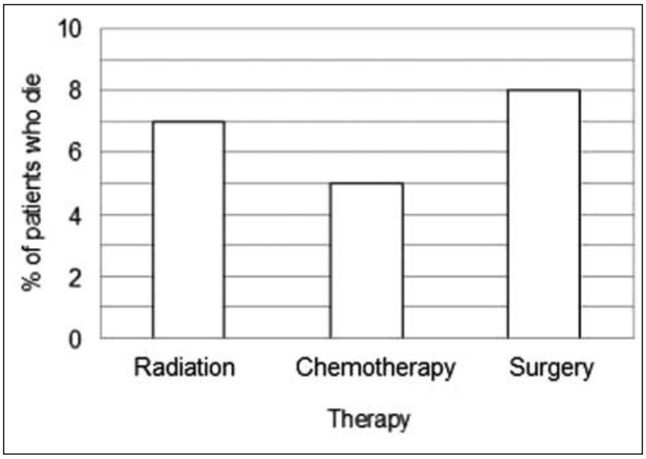
Coliosis 1

Tiosis 2

The increase was equal. 3

Can't say 4

Q13. Here is some information about cancer therapies.



What is the percentage of cancer patients who die after chemotherapy?

________ %

APPENDIX B
Materials Used in Experiments Testing the Predictive Validity of the Graph Literacy Scale

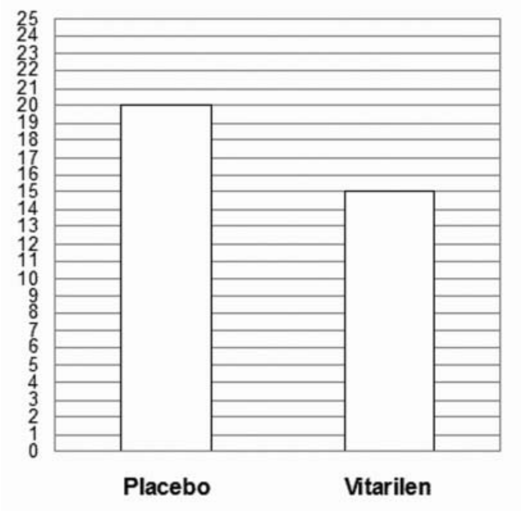
EXPERIMENT 1

[This example shows task 1. Task 2 had the same structure but involved heart attack and different numbers.]

A study was conducted to test the effectiveness of a new drug called Vitarilen. The drug is supposed to reduce the risk of stroke and heart attack for patients with symptoms of arterial disease. One randomly selected group of 100 patients took a placebo, while another group of 100 patients took Vitarilen. Compared to the group that took a placebo, the relative reduction in risk of having a stroke in the group that took Vitarilen was 25%.

[For conditions involving visual displays:]

This is also shown in the following figure:



Percentage of Patients Who Have a Stroke in Each Group

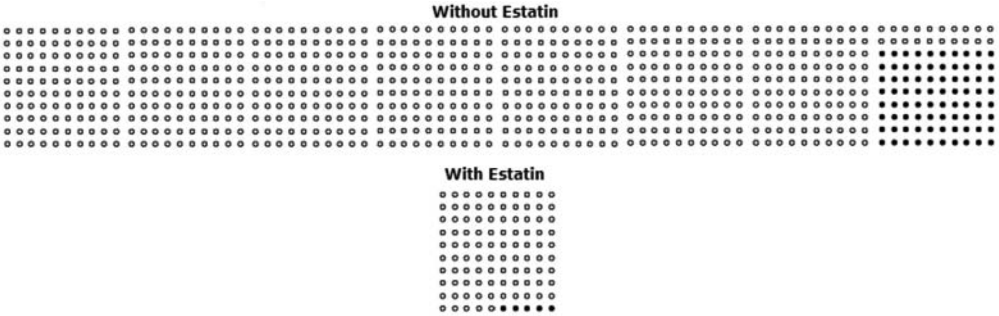
EXPERIMENT 2

[This example is an experimental condition in which groups of patients who took and who did not take the drug were of different sizes. Other conditions involved different numbers.]

Here is some information about a new (fictitious) drug that decreases cholesterol level.

A new drug that reduces cholesterol, **Estatin**, decreases the risk of dying of a heart attack for people with high cholesterol. Here are the results of a study of 900 people with high cholesterol: 80 out of 800 people who did not take the drug died of a heart attack, compared to 5 out of 100 people who took the drug.

[For conditions involving visual displays:]



REFERENCES

1. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. J Am Med Inform Assoc. 2006;13:608–18.

2. Lipkus IM, Hollands JG. The visual communication of risk. J Natl Cancer Inst Monogr. 1999;25:149–63.

3. Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. Med Decis Making. 2007;27:696–713.

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

5. Garcia-Retamero R, Galesic M, Gigerenzer G. Do icon arrays help reduce denominator neglect? Med Decis Making. In press.

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

7. Waters EA, Weinstein ND, Colditz GA, Emmons K. Formats for improving risk communication in medical tradeoff decisions. J Health Commun. 2006;11:167–82.

8. 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.

9. Friel SN, Curcio FR, Bright GW. Making sense of graphs: critical factors influencing comprehension and instructional implications. J Res Math Ed. 2001;32:124–58.

10. Spence I. No humble pie: the origins and usage of a statistical chart. J Educ Behav Stat. 2005;30:353–68.

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

12. Spence I, Lewandowsky S. Displaying proportions and percentages. Appl Cogn Psychol. 1991;5:61–77.

13. Feldman-Stewart D, Kocovski N, McConnell BA, Brundage MD, Mackillop WJ. Perception of quantitative information for treatment decisions. Med Decis Making. 2000;20:228–38.

14. Ancker JS, Kaufman D. Rethinking health numeracy: a multidisciplinary literature review. J Am Med Inform Assoc. 2007;14:713–21.

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

16. Kramarski B, Mevarech ZR. Enhancing mathematical reasoning in the classroom: effects of cooperative learning and metacognitive training. Am Educ Res J. 2003;40:239–80.

17. Galesic M, Garcia-Retamero R. Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. Arch Intern Med. 2010;8:462–8.

18. 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.

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

20. National Center for Education Statistics. National Assessment of Adult Literacy (NAAL): Sample Questions 1985. Available from: http://nces.ed.gov/NAAL/sample_items.asp. Accessed 15 December 2008.

21. Programme for International Student Assessment (PISA). The PISA 2003 Assessment Framework—Mathematics, Reading, Science, and Problem Solving Knowledge and Skills. Paris: OECD; 2003.

22. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns. 1999;38:33–42.

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

24. Kline P. The New Psychometrics: Science, Psychology, and Measurement. London: Routledge; 1998.

25. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and email for health care information: results from a national survey. JAMA. 2003;289:2400–6.

26. Jacoby WG. Value choices and American public opinion. Am J Pol Sci. 2006;50:706–23.

27. Lerner JS, Gonzalez RM, Small DA, Fischoff B. Effects of fear and anger on perceived risks of terrorism: a national field experiment. Psychol Sci. 2003;14:144–50.

28. Miller JD, Scott EC, Okamoto S. Public acceptance of evolution. Science. 2006;313:765–6.

29. Schlenger WE, Caddell JM, Ebert L, et al. Psychological reactions to terrorist attacks: findings from the national study of Americans' reactions to September 11. JAMA. 2002;288:581–8.

30. Chang L, Krosnick JA. National surveys via RDD telephone interviewing versus the Internet: comparing sample representativeness and response quality. Public Opin Q. 2009;73:641–8.

31. Garcia-Retamero R, Galesic M. Who profits from visual aids? Overcoming challenges in people's understanding of risk. Soc Sci Med. 2010;70:1019–25.

32. 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:S47-56.

33. Edwards A. Communicating risks through analogies. BMJ. 2003;327:749.

34. Galesic M, Barton A, Wegwarth O, Gaissmaier W, Gigerenzer G. Do analogies help elderly people understand medical information? Int J Psychol. 2008;43:32.

35. Galesic M, Gigerenzer G, Straubinger N. Natural frequencies help older adults and people with low numeracy to evaluate medical screening tests. Med Decis Making. 2009;29:368–71.

36. Stigler JW, Gonzales P, Kawanaka T, Knoll S, Serrano A. The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States. Washington, DC: US Department of Education, National Center for Education Statistics; 1999.

37. Carpenter PA, Shah P. A model of the perceptual and conceptual processes in graph comprehension. J Exp Psychol Appl. 1998;4:75–100.

38. Shah P, Hoeffner J. Review of graph comprehension research: implications for instruction. Educ Psychol Rev. 2002;23:47–69.

39. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest. 2007;8:53–96.