## Natural Frequencies Help Older Adults and People with Low Numeracy to Evaluate Medical Screening Tests

Mirta Galesic, Gerd Gigerenzer and Nils Straubinger

*Med Decis Making* 2009; 29; 368 originally published online Jan 6, 2009;
DOI: 10.1177/0272989X08329463

# Natural Frequencies Help Older Adults and People with Low Numeracy to Evaluate Medical Screening Tests

*Mirta Galesic, PhD, Gerd Gigerenzer, PhD, Nils Straubinger*

***Background.*** *Understanding information about medical screening tests often requires estimating positive predictive values (i.e., posterior probabilities), which is a notoriously difficult task. Previous studies have shown that representation of information in terms of natural frequencies (i.e., counts of occurrences that preserve base rates) facilitates judgments of positive predictive values. The objective of this study was to investigate whether natural frequencies facilitate accurate estimates in elderly people and whether performance depends on numeracy skills. Elderly people are more often than younger people required to use such information to make informed choices regarding medical procedures (e.g., screenings).* ***Method.*** *This was an experimental study in which information about 2 medical screening tests was presented either as conditional probabilities or natural frequencies. Participants were 47 older adults (62–77 years of age; average numeracy score 8.6) and 115 younger adults (18–35 years of age; average numeracy score 10.3).* ***Results.*** *When the screening information was presented in terms of conditional probabilities, only 15% of the younger adults and 18% of the older adults provided accurate estimates in at least 1 of the tasks. When information was presented in terms of natural frequencies, 55% of the younger adults and 58% of the elderly participants gave correct estimates. This effect occurred without explicit training. Furthermore, participants with higher numeracy scores performed better in the estimation tasks than those with lower numeracy scores.* ***Conclusions.*** *Natural frequencies help elderly and young patients—including those with lower numeracy skills—to understand positive predictive values of medical screening tests.* ***Key words:*** *natural frequencies; conditional probabilities; medical screening tests; older adults; numeracy.* ***(Med Decis Making 2009;29:368–371)***

To make informed decisions about their health care, patients need the ability to process, interpret, and act on quantitative information. This ability is an important aspect of health-related numeracy.[1] Understanding and using concepts such as proportions, percentages, probabilities, and risks may require especially high levels of numeracy.[1] Probability expressions in particular seem to cause many people difficulties.[2,3]

Received 31 March 2008 from the Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany. Revision accepted for publication 17 September 2008.

Address correspondence to Mirta Galesic, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Königin-Luise-Strasse 5, 14195 Berlin, Germany; e-mail: galesic@mpib-berlin.mpg.de.

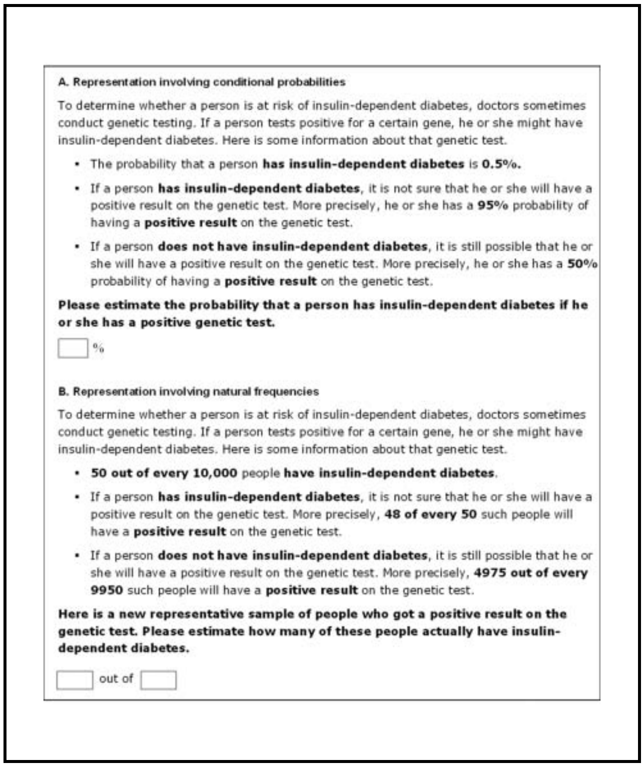
DOI: 10.1177/0272989X08329463

Bayesian inference, which is used to estimate the posterior probability that a person has a certain disease given a positive test result (i.e., positive predictive value), is a very demanding task. In medical diagnosis, Bayesian inference requires assessing the positive predictive value of medical tests given the prevalence of the disease (base rate) as well as the sensitivity (hit rate) and specificity (1 – false alarm rate) of the test.

Because Bayesian inference has proven to be a notoriously difficult task, some researchers have concluded that the human mind is just not built to reason in a Bayesian fashion.[4] An alternative view, which takes into account how the mind interacts with the environment (i.e., task representation), suggests that the commonly used representations of the information do not fit with the way our minds have evolved.[5–7] Indeed, previous studies have shown that a representation of information in terms of natural frequencies (i.e., counts of occurrences that preserve base rates) vastly improves performance in

Bayesian inferences compared with commonly used representations involving conditional probabilities.[5] This is primarily because natural frequencies preserve information about base rates and therefore simplify mental computations.[5] Natural frequencies have been successfully used to improve Bayesian reasoning in various groups, including medical and law students, physicians and judges, and even children.[8–10]

However, to the best of our knowledge, no studies have investigated whether natural frequencies can improve inferences in older adults or in people with lower numeracy skills. Elderly people may be required to assess statistical health-related information (e.g., information about medical screening tests and treatments) more often than younger people because the risk for many diseases increases with age. At the same time, elderly people may be at particular risk of misunderstanding probability expressions.[11] Furthermore, older adults are more strongly affected by the normal decline of those cognitive abilities that are more biologically determined (e.g., information processing) than are younger adults.[12] The normal loss of cognitive abilities may have a negative impact on performance in tasks such as Bayesian inferences, which require mental integration of information. The objective of this study is to examine whether natural frequencies can improve posterior probability judgments of older adults and of people with lower numeracy skills. Our hypothesis is that natural frequencies, which simplify mental computations, should facilitate reasoning in these groups of people.


**A. Representation involving conditional probabilities**

To determine whether a person is at risk of insulin-dependent diabetes, doctors sometimes conduct genetic testing. If a person tests positive for a certain gene, he or she might have insulin-dependent diabetes. Here is some information about that genetic test.

- The probability that a person **has insulin-dependent diabetes** is **0.5%.**
- If a person **has insulin-dependent diabetes**, it is not sure that he or she will have a positive result on the genetic test. More precisely, he or she has a **95%** probability of having a **positive result** on the genetic test.
- If a person **does not have insulin-dependent diabetes**, it is still possible that he or she will have a positive result on the genetic test. More precisely, he or she has a **50%** probability of having a **positive result** on the genetic test.

**Please estimate the probability that a person has insulin-dependent diabetes if he or she has a positive genetic test.**

[ ] %

**B. Representation involving natural frequencies**

To determine whether a person is at risk of insulin-dependent diabetes, doctors sometimes conduct genetic testing. If a person tests positive for a certain gene, he or she might have insulin-dependent diabetes. Here is some information about that genetic test.

- **50 out of every 10,000** people **have insulin-dependent diabetes**.
- If a person **has insulin-dependent diabetes**, it is not sure that he or she will have a positive result on the genetic test. More precisely, **48 of every 50** such people will have a **positive result** on the genetic test.
- If a person **does not have insulin-dependent diabetes**, it is still possible that he or she will have a positive result on the genetic test. More precisely, **4975 out of every 9950** such people will have a **positive result** on the genetic test.

**Here is a new representative sample of people who got a positive result on the genetic test. Please estimate how many of these people actually have insulin-dependent diabetes.**

[ ] out of [ ]

*Figure 1 Example of a task requiring calculation of positive predictive value of a medical screening procedure, formulated in terms of conditional probabilities or natural frequencies.*

## METHOD

The experiment was conducted at the Max Planck Institute for Human Development in Berlin, Germany. There were 2 groups of participants: one group of 47 older adults (49% aged 62–69 years, 51% aged 70–77 years of age; 49% women and 51% men; 57% with high school or lower education, 43% with college or university education) and a 2nd group of 115 younger adults (63% aged 18–25 years, 37% aged 26–35 years; 57% women and 43% men; all university students). Participants received 10 Euros for each of 2 sessions. The experiment was part of a larger computerized questionnaire comprising various questions about medical risks.

We used 2 modified Bayesian inference problems involving information about medical screening procedures from Hoffrage and Gigerenzer[13]: 1) the diabetes problem (i.e., genetic testing for early detection of diabetes) and 2) the trisomy 21 problem (i.e., neckfold skin test for detecting Down syndrome). The problems resembled those typically used in studies of Bayesian inference.[5] The problem information included the prevalence of disease and the sensitivity and the false alarm rate of the test. Participants were required to estimate the procedures' positive predictive value.

Participants were randomly assigned to 1 of 2 experimental conditions that represented the information as either conditional probabilities or natural frequencies. In the natural frequency condition, participants were free to enter any combination of numerators and denominators they wished. Figure 1 illustrates the representations using the diabetes problem. The correct answer for both problems was about 1%. We coded all answers that were below 5% as accurate estimates, as our focus was on whether participants could give estimates that were functional for health-related decisions.

Participants completed the numeracy scale consisting of 11 items from Lipkus and others[14] and an additional item (involving a coin toss) from Schwartz

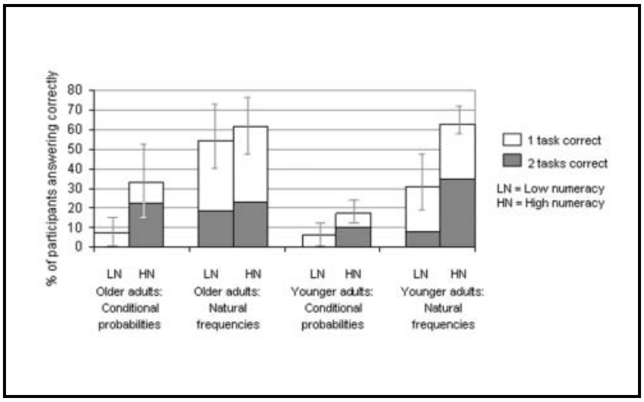



*Figure 2 Proportion of participants who solved 1 or both tasks correctly, by task representation, age group, and numeracy level. Error bars denote ±1 standard error for proportion of participants answering at least 1 task correctly.*

and others,[15] allowing for a maximum score of 12. The average of participants' numeracy scores was 9.7 (median = 11, $s = 2.2$). Younger adults ($\bar{x} = 10.3$, $s = 1.8$) had notably higher numeracy scores than older adults ($\bar{x} = 8.6$, $s = 2.8$), even when compared with only the highly educated elderly ($\bar{x} = 9.4$, $s = 2.1$). The median scores for elderly and younger adults were 9 and 11, respectively. Participants were split into numeracy groups (lower v. higher) using the average median of older and younger adults (median = 10).

## RESULTS

Each participant was assigned a score ranging from 0 to 2 (corresponding to the number of accurate estimates). Figure 2 shows the proportion of participants who gave accurate estimates in 1 or 2 tasks, respectively, depending on the representation, age group, and numeracy level. To analyze performance differences, we conducted a 3-factorial (Representation × Numeracy × Age group) analysis of variance for fixed factors and independent samples, using participants' scores as the dependent measure.

Although the number of accurate estimates was low, with representations in terms of conditional probabilities, it vastly increased when natural frequencies were provided, regardless of the age group and the numeracy level, $F(1, 154) = 15.38$, MS = 7.41, $P = 0.001$, $\eta^2 = 0.09$.[a] Furthermore, high-numeracy participants gave substantially more accurate estimates than low-numeracy participants, across age groups and representations, $F(1, 154) = 7.62$, MS = 3.67, $P = 0.01$, $\eta^2 = 0.05$. Performance did not differ substantially across age groups, $F(1, 154) = 1.05$, MS = 0.50, $P = 0.31$, $\eta^2 = 0.01$, and there were no interactions between the examined factors.

It is possible that our participants (students and elderly) had higher numeracy skills than an average member of the German population of similar age. As a result, we could expect that in the overall population, the performance with both formats—natural frequencies and percentages—would be lower. However, we do expect that the relative improvement due to the natural frequencies would be similar.

Note also that some of the participants had difficulties with correctly solving the problems even when natural frequencies were provided. Therefore, a safe solution for medical practice might seem to be to encourage doctors to communicate end results to patients (here, positive predictive values of the screening tests) instead of leaving it to patients to perform the calculation themselves. However, in our other studies,[6] we found that even doctors have difficulties performing the calculations and that they were in fact misinforming their patients. Therefore, some skill in understanding problems involving conditional probabilities is beneficial for informed medical decision making.

## CONCLUSIONS

The current study demonstrated for the 1st time that elderly and low-numeracy people benefit from natural frequencies. Unlike what might be expected considering the normal decline of cognitive abilities in old age, the effect was the same for younger and older adults. Based on the results of the present study and previous findings, we recommend representing medical test results in terms of natural frequencies.[8,9] Taken together, these results show that various groups involved in communication of medical information—physicians and medical students as well as younger and older patients with different numeracy skills—can profit from natural frequencies. Understanding of quantitative health-related information, which is a prerequisite of informed decision making in medicine, can be fostered by providing good representations.

[a]MS is short for "mean squares," computed by dividing the sum of squares for a given factor (here: Representation) with the appropriate degrees of freedom (here, $df = 1$). $\eta^2$ is a descriptive measure of the amount of variance explained, calculated by dividing the sum of squares for a factor by its sum with the error sum of squares.

## REFERENCES

1. 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.

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

3. Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. Psychol Sci. 2006;17:407–13.

4. Kahneman D, Tversky A. Subjective probability: a judgment of representativeness. Cogn Psychol. 1972;3:430–54.

5. Cosmides L, Tooby J. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition. 1996;58:1–73.

6. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. Psychological Science in the Public Interest. 2008;2:53–96.

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

8. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. Acad Med. 1998;73:538–40.

9. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical information. Science. 2000;290:2261–2.

10. Zhu L, Gigerenzer G. Children can solve Bayesian problems: the role of representation in mental computation. Cognition. 2006;98:287–308.

11. Hibbard JH, Peters EM, Slovic P, Finucane ML. Making health care report cards easier to use. Jt Comm J Qual Improv. 2001;27: 591–604.

12. Li S-C, Lindenberger U, Hommel B, Aschersleben G, Prinz W, Baltes PB. Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. Psychol Sci. 2004;15:155–63.

13. Hoffrage U, Gigerenzer G. How to improve the diagnostic inferences of medical experts. In: Kurz-Milcke E, Gigerenzer, G, eds. Experts in Science and Society. New York: Kluwer Academic/Plenum; 2004. p 249–68.

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

15. Schwartz LM, Woloshin S, Black WC, Welch GH. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997;127:966–71.