

POLICY FORUM: MEDICINE

Communicating Statistical Information

Ulrich Hoffrage,* Samuel Lindsey, Ralph Hertwig, Gerd Gigerenzer

Decisions based on statistical information can mean the difference between life and death—for instance, when a cancer patient has to decide whether to undergo a painful medical procedure based on the likelihood that it will succeed, or when a jury has to decide whether to convict someone based on DNA evidence. Unfortunately, most of us, experts included, have difficulty understanding and combining statistical information effectively.

For example, faculty, staff, and students at Harvard Medical School were asked to estimate the probability of a disease given the following information (1): “If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person’s symptoms or signs?” The estimates varied wildly, ranging from the most frequent estimate, 95% (given by 27 out of 60 participants), to the correct answer, 2% (given by 11 out of 60 participants) (2). In a study requiring interpretation of mammography outcomes (3), almost all physicians confused the sensitivity of the test (the proportion of positive test results among people with the disease) with its positive predictive value (the proportion of people with the disease among those who receive a positive test result). This is a common confusion that even crops up in scholarly articles (3) and statistical textbooks (4) and certainly affects the ability of lay people (5) to understand the statistical information. Recent discussions of genetic testing have indicated that genetic counselors are experiencing the same difficulty (6).

It makes little mathematical difference whether statistics are expressed as probabilities, percentages, or absolute frequen-

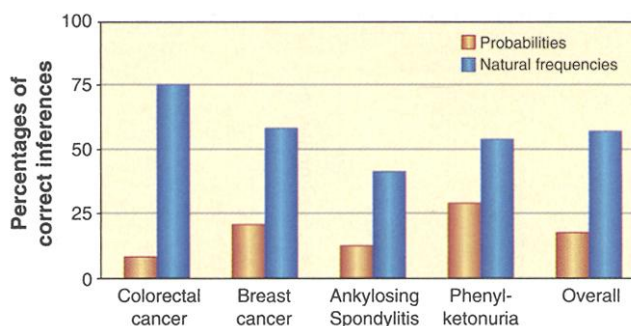


Fig. 1. Interpreting statistics. Medical students' percentage of correct inferences in four realistic diagnostic tasks.

cies. It does, however, make a psychological difference. More specifically, statistics expressed as natural frequencies improve the statistical thinking of experts and nonexperts alike.

Natural Frequencies

To illustrate how natural frequencies differ from probabilities, we use the example of a cancer screening test. The probability of colorectal cancer can be given as 0.3% [base rate]. If a person has colorectal cancer, the probability that the hemoccult test is positive is 50% [sensitivity]. If a person does not have colorectal cancer, the probability that he still tests positive is 3% [false-positive rate]. What is the probability that a person who tests positive actually has colorectal cancer? A restatement of the same problem in terms of natural frequencies would be that out of every 10,000 people, 30 have colorectal cancer. Of these, 15 will have a positive hemoccult test. Out of the remaining 9970 people without colorectal cancer, 300 will still test positive. How many of those who test positive actually have colorectal cancer?

Only 1 out of 24 physicians gave the correct answer when the statistical information was expressed in probabilities (7). When it was presented in natural frequencies, 16 out of 24 other physicians gave the

correct answer: 15 out of 315 (i.e., 5%). Whereas natural frequencies seem to help people make statistical inferences, probabilities apparently hinder them. Unfortunately, in contexts in which the positive predictive value of a test is at issue, statistics are typically expressed and communicated in the form of probabilities, although they can easily be translated into natural frequencies, as follows:

1. Select a population and use the base rate to determine how many people in the population have the disease.
2. Take that result and use the test's sensitivity to determine how many people have the disease and a positive test.
3. Take the remaining number of healthy people and use the test's false-positive rate to determine how many people do not have the disease but still test positive.
4. Compare the number obtained in step 2 with the sum of those obtained in steps 2 and 3 to determine how many people with a positive test actually have the disease.

Natural frequencies facilitate inferences because they carry implicit information about base rates and reduce the number of computations required to determine the positive predictive value of a test (8, 9, 10). They also correspond to the way in which humans have experienced statistical information over most of their history.

Applications in Medicine

To illustrate the effect of natural frequencies, we asked 96 advanced medical students to solve four realistic diagnostic tasks. Each participant worked on two probability and two frequency versions; the order of representation format and which task was in which format was balanced (11). For each of the tasks, more participants correctly inferred the likelihood of having the disease given a positive test when the statistics were communicated as natural frequencies (Fig. 1).

Other medical practitioners could also profit from representing statistical information in terms of natural frequencies.

"...statistics expressed as natural frequencies improve the statistical thinking of experts and nonexperts alike."

U. Hoffrage, R. Hertwig, and G. Gigerenzer are at the Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. S. Lindsey is at the Department of Psychology, 102 Gilmer Hall, University of Virginia, Charlottesville, VA 22903, USA.

*To whom correspondence should be addressed. E-mail: hoffrage@mpib-berlin.mpg.de

This Policy Forum was collaboratively written to combine work submitted to *Science* independently by the first and second authors.

Consider the statistics AIDS counselors must understand and communicate. In Germany, the prevalence of HIV in heterosexual men who are not in any known risk group is around 0.01%. The false-positive rate of the HIV test (in which one blood sample is subjected to multiple tests) is around 0.01%, and its sensitivity is around 99.9% [exact estimates vary (12)]. To explore how counselors actually communicate these risks, we sent a male, low-risk client to 20 German public health centers to have 20 HIV tests. During the mandatory pretest counseling, the client asked the counselor about the prevalence, sensitivity, false-positive rate, and the chance that he actually had the virus if the test were positive (13). Not a single counselor communicated the risks to the client in natural frequencies. Instead, they used probabilities and percentages, and, in the majority of the counseling sessions, the information was either inconsistent or wrong. For instance, one counselor estimated the base rate and the false-positive rate to be around 0.1%, and the sensitivity to be 99.9%, and then stated that the client's probability of infection given a positive test is also 99.9% (applying steps 1 to 4 above to his estimates yields a probability of 50%). In fact, 15 out of the 20 counselors told this low-risk client that it is 99.9% or 100% certain that he has HIV if he tests positive (applying steps 1 to 4 to the numbers found in the literature yields an actual probability of 50%).

Percentages can mislead in other ways. For example, it may sound impressive to learn that mammography screening can reduce the risk of breast cancer fatality in women by 25% [for 50- to 74-year-old women (14)]. However, this percentage does not say anything about the actual frequencies. If 4 out of 1000 women without symptoms die of breast cancer within the next 10 years (15), the relative risk reduction of 25% means that 1 woman in 1000 women who undergo screening would be saved. A woman without symptoms is most likely not one of the 4 to whom the risk reduction applies, but one of the other 996 instead—and many of these women may suffer as a result from the screening. For instance, false-positives occur and, moreover, cancers that grow so slowly that they present little risk will be diagnosed and unnecessarily treated. As long

as health organizations inform women in terms of probabilities and relative risk reduction about the benefits and harms of screening, a truly informed decision is unlikely.

Applications in Law

Determinations of facts and verdicts in legal proceedings often depend on scientific evidence. The communication of statistics

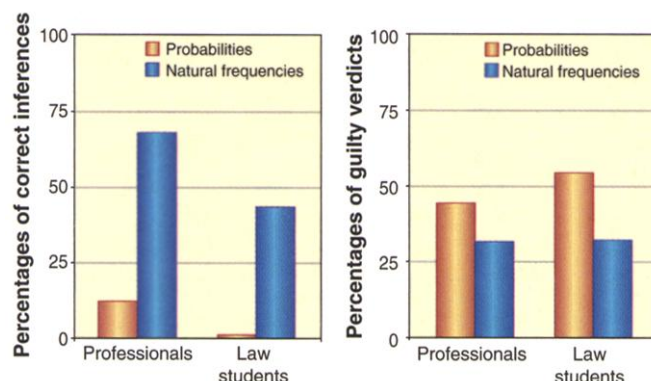


Fig. 2. Interpreting statistics. Legal experts' percentage of correct inferences (left) and of guilty verdicts (right) in two criminal court case files.

is as important to the making of legal decisions by judges, attorneys, forensic experts, and jurors as it is to medical decision-makers (16, 17). In considering the admissibility standards for scientific evidence, the U.S. Supreme Court has specifically indicated that courts need to consider "known or potential rate of error, and the existence and maintenance of standards controlling the technique's operation" (18).

In a study conducted in Germany, we asked 27 professionals who would soon qualify as judges and 127 advanced law students to evaluate two criminal-court case files involving rape (19). In both cases, a DNA match was reported between a DNA sample from the defendant and one recovered from the victim. Aside from this evidence, there was little reason to suspect that the defendant was the perpetrator. Expert testimony reported the frequency of the recovered DNA profile as 1 in 1,000,000 and then stated that it was practically certain that the analysis would show a match for a person who indeed had the DNA profile (in other words, sensitivity = 100%). The expert also reported the rates of technical and human mishaps leading to false-positive results in laboratory tests to be about 0.003 (20).

When these statistics were expressed as probabilities, only 13% of the professionals and under 1% of the law students correctly inferred the probability that the defendant was actually the source of the trace. But when the identical statistics were stated as natural frequencies, 68% and 44% of these same par-

ticipants made the correct inference (Fig. 2, left). The different ways of expressing the same statistical information altered the verdicts in each case. When the information was presented as probabilities, 45% of the professionals and 55% of the students rendered a verdict of guilty, but only 32% and 33% did so when the same statistics were expressed as natural frequencies (Fig. 2, right). When verdicts hinge on statistical evidence, understanding that evidence is crucial, and pursuing this simple method of fostering statistical insight could contribute to that goal (21, 22).

Implications for Teaching

The beneficial effects of natural frequencies on statistical reasoning in the studies reported above occurred without training or instruction. Systematic training in the use of natural frequencies can even help people to reason with probabilities. The key is to teach representations rather than rules—that is to teach people how to translate probabilities into natural frequencies, as shown in steps 1 to 4. Traditionally, however, students are instead taught how to plug probabilities into mathematical formulas such as Bayes's rule.

Teaching representations rather than rules—and expressing statistical information in natural frequencies where appropriate—can help to foster the statistical reasoning needed to make sound decisions.

References and Notes

- W. Casscells, A. Schoenberger, T. Grayboys, *N. Engl. J. Med.* **299**, 999 (1978).
- The correct solution can be obtained by applying Bayes's rule.
- D. M. Eddy, in *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, A. Slovic, A. Tversky, Eds. (Cambridge Univ. Press, Cambridge, 1982), pp. 249–267.
- G. Gigerenzer, in *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, G. Keren, C. Lewis, Eds. (Erlbaum, Hillsdale, NJ, 1993), pp. 313–339.
- J. J. Koehler, *Behav. Brain Sci.* **19**, 1 (1996).
- R. Weiss, *Washington Post*, 2 December 2000, p. A10.
- U. Hoffrage, G. Gigerenzer, *Acad. Med.* **73**, 538 (1998).
- G. Gigerenzer, U. Hoffrage, *Psychol. Rev.* **102**, 684 (1995).
- _____, *Psychol. Rev.* **106**, 425 (1999).
- B. Mellers, A. McGraw, *Psychol. Rev.* **106**, 417 (1999).
- The tasks are displayed at <http://www-abc.mpib-berlin.mpg.de/users/hoffrage/papers/4tasks.html>.
- G. J. Stine, *Acquired Immune Deficiency Syndrome: Biological, Medical, Social, and Legal Issues* (Prentice-Hall, Englewood Cliffs, NJ, 1996).
- G. Gigerenzer, U. Hoffrage, A. Ebert, *AIDS Care* **10**, 197 (1998).
- K. Kerlikowske et al., *JAMA* **273**, 149 (1995).
- L. Nystöm, et al., *J. Med. Screen.* **3**, 85 (1997).
- See, e.g., J. J. Koehler, *Jurimetrics* **34**, 21 (1993).
- _____, *Univ. Colorado Law Rev.* **67**, 859 (1996).
- Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), pp. 593–594.
- S. Lindsey, R. Hertwig, G. Gigerenzer, in preparation.
- An average of estimates based on laboratory proficiency tests. See J. J. Koehler, A. Chia, S. Lindsey, *Jurimetrics* **35**, 201 (1995).
- S. Breyer, *Science* **280**, 537 (1998).
- R. Hertwig, U. Hoffrage, in *Frequency Processing and Cognition*, P. Sedlmeier, T. Betsch, Eds. (Oxford Univ. Press, New York, in press).
- We thank the German Research Foundation for financial support (Ho 1847/1 and He 2768/6-1).