Published in: *Academic Medicine, 73* (5), 1998, 538–540.

Using Natural Frequencies to Improve Diagnostic Inferences

Ulrich Hoffrage and **Gerd Gigerenzer**
Max Planck Institute for Human Development, Berlin, Germany

Abstract

Purpose. To test whether physicians' diagnostic inferences can be improved by communicating information using natural frequencies instead of probabilities. Whereas probabilities and relative frequencies are normalized with respect to disease base rates, natural frequencies are not normalized.
Method. The authors asked 48 physicians in Munich and Düsseldorf to determine the positive predictive values (PPVs) of four diagnostic tests. Information presented in the four problems appeared either as probabilities (the traditional way) or as natural frequencies.
Results. When the information was presented as probabilities, the physicians correctly estimated the PPVs in only 10% of cases. When the same information was presented as natural frequencies, that percentage increased to 46%.
Conclusion. Representing information in natural frequencies is a fast and effective way of facilitating diagnostic insight, which in turn helps physicians to better communicate risks to patients, and patients to better understand these risks.

What does a positive medical test result mean? Physicians often have difficulty inferring the probability of a disease from statistical information relevant to positive test results.[1–5] In one study, David Eddy[1] provided physicians with information that can be summarized as follows (numbers rounded): For a woman at age 40 who participates in routine screening, the probability of breast cancer is 1%. If a woman has breast cancer, the probability is 80% that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 10% that she will still have a positive mammogram.

Imagine a woman from this age group with a positive mammogram. What is the probability that she actually has breast cancer? This probability, also called the positive predictive value (PPV) of a test, can be calculated from Bayes' rule by using 1 (below), where p(disease) is the prevalence of the disease (here .01); p(pos | disease) is the sensitivity of the test, that is, the proportion of positive results among people suffering from the disease (here .8); and p(pos | no disease) is the false-positive rate of the test, that is, the proportion of positive results among people not suffering from the disease (the complement of the specificity of the test; here .l). Inserting this information into Bayes' rule results in a PPV of .075. Yet most of the physicians (95 of 100) in Eddy's study estimated the predictive value of the test to be between .7 and .8.[1] Eddy concluded that they had confused the sensitivity of the test with the PPV.

How can physicians' diagnostic inferences based on prevalence, sensitivity, and specificity be improved? Our empirical claim is that merely changing the representation of information can improve diagnostic inferences without any training or instruction. Eddy and other authors[1–3] communicated information in terms of probabilities or percentages—the information format commonly used in medical textbooks and curricula. However, this is only one way of representing numerical information. Mathematical probability is in fact only a few hundred years old,[6] which makes one doubt whether human information processing has actually evolved to work with probabilities. Indeed, there is evidence suggesting that minds are evolutionarily adapted to process information in *natural frequencies,* that is, absolute frequencies as they result from observ-

$$\text{PPV} = \frac{p(\text{disease})p(\text{pos} \mid \text{disease})}{p(\text{disease})p(\text{pos} \mid \text{disease}) + [1 - (\text{disease})]p(\text{pos} \mid \text{no disease})} \quad (1)$$

ing cases that have been representatively sampled from a population.[7, 8]

Natural frequencies are not normalized with respect to the base rates of disease or no disease. Using natural frequencies, one can communicate the information in Eddy's mammography problem as follows: Ten out of every 1,000 women at age 40 who participate in routine screening have breast cancer. Of these ten women with breast cancer, eight will have a positive mammogram. Of the remaining 990 women without breast cancer, 99 will still have a positive mammogram.

Imagine a group of 40-year-old women with positive mammograms. How many of them actually have breast cancer? Now the answer can easily be „seen." There are 107 women with positive test results (8 + 99), but only eight of them will have breast cancer. More generally, for

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

natural frequencies the computation of PPV can be reduced to:
where TP denotes the number of true positives (here 8) and FP denotes the number of false positives (here 99). Communicating information in natural frequencies means specifying the absolute numbers of TPs and FPs, which are not directly provided when frequencies are normalized with respect to the base rates of disease or no disease (i.e., the true-positive and false-positive rates).[7]

Both the evolutionary argument that reasoning processes are adapted to natural frequencies— rather than to normalized frequencies, probabilities, or percentages—and the observation that Bayesian computations are easier when the information is communicated in natural frequencies led us to predict that natural frequencies can help physicians to correctly infer the PPV of a diagnostic test. We carried out the following study to test this prediction.

Method

We tested 48 physicians in Munich and Düsseldorf: 18 in university hospitals, 16 in private or public hospitals, and 14 in private practice. The mean number of years of professional experience was 14, with a range from one month to 30 years. The physicians worked on four diagnostic problems in which they had to infer the presence of (1) breast cancer from a positive mammography test (as described above), (2) colorectal cancer from a positive Hemoccult test, (3) phenylketonuria from a positive Guthrie test, and (4) ankylosing spondylitis (Bekhterev's disease) from a positive HL-antigen-B27 test.

The participants received a booklet containing all four problems, two of which presented information in probabilities and two in natural frequencies. We systematically varied the formats and orders of the problems among the physicians. We invited them to make notes, calculations, or drawings while working on the problems, which we later analyzed for evidence of their reasoning strategies. Also, after the physicians had filled out the booklets, taking an average of 30 minutes, we interviewed them about their reasoning strategies. We coded answers to the problems as being in accord with Bayes' rule only when (1) the numerical estimate was within five percentage points of the correct one, and (2) the physician's notes, calculations, or drawings and the interview confirmed that the answer was neither a guess nor the result of another strategy.

Results

When the information in the problems was communicated in probabilities, the physicians were able to reason according to Bayes' rule in only 10% of the cases. This poor performance is consistent with that found in earlier studies.[1–5] However, when the same information was communicated in natural frequencies, the percentage of correct estimates increased to 46%. For the individual problems, the improvements of the estimates when the information was given in natural frequencies were: breast cancer, from 8% to 46%; colorectal cancer, from 4% to 67%; phenylketonuria, from 21% to 42%; and ankylosing spondylitis, from 8% to 29%.

The physicians spent an average of about 25% more time on the problems involving probabilities than they did on those involving natural frequencies, indicating that they found them more difficult to solve. Many of the physicians appeared nervous and uncertain when the information was communicated in probabilities, but appeared relaxed when it was in natural frequencies. For instance, when working on probability problems, the physicians made complaints such as: „I simply can't do that. Mathematics is not my forte.“ However, with natural frequencies, a typical remark was: „Now it's different. It's quite easy to imagine. There's a frequency; that's more visual.“ Moreover, the physicians were less skeptical about the relevance of statistical information to medical diagnosis when it was communicated in frequencies. Despite the usefulness of the frequency format, the physicians did not spontaneously translate probabilities into natural frequencies.

What strategies did the physicians use when they did not reason according to Bayes' rule? When the information was presented as probabilities, the two most frequent strategies for estimating the PPV—either simply reporting the sensitivity of the test or subtracting the false-positive rate from the sensitivity—relied on the diagnostic information of the test alone and ignored the prevalence of the disease. However, when information was presented in terms of natural frequencies, the two most frequent non-Bayesian strategies—reporting either the prevalence of the disease or the probability of a positive test result—ignored the diagnostic information of the test and focused exclusively on base rates. Notice that either of these two strategies usually provides better estimates of the PPV than do the two most dominant strategies for the probability problems.

Discussion

The results of our study and others show that natural frequencies can serve as an effective tool in inferring the predictive value of a test. This result has two implications. First, because information in medical texts is routinely communicated in probabilities or percentages, medical students as well as physicians ought to be taught how to translate these numbers into natural frequencies.

Sedlmeier[9] designed a computerized tutorial system that teaches people how to do so. People who were taught to translate probabilities into natural frequencies performed twice as well on Bayesian inference problems as did people who were taught the standard method of inserting probabilities into Bayes' rule. Even more striking, performance in the group that translated information into natural frequencies remained stable in a five-week follow-up test (median performance 90% correct), whereas that in the standard group showed the usual deterioration due to forgetting (15% correct). Training in how to translate probabilities into natural frequencies should be part of medical education.

The second, equally important, implication concerns communication of risks, not only within medical textbooks, but between physicians and patients as well. For instance, before consenting to medical treatment on the basis of a diagnosis, patients should understand the uncertainties involved, such as the chances of actually having the disease. Natural frequencies have been shown to work with lay people as well as with physicians.[7, 10] Physicians should use this simple tool to communicate with patients about risk, thereby helping them to become better decision makers.

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (Ho 1847/1-1). The authors thank Maria Zumbeel for conducting the interviews and Valerie M. Chase, Anita Todd, and Angelika Weber for helpful comments.

References

1. Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, UK: Cambridge University Press.
2. Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New English Journal of Medicine, 299,* 999–1001.
3. Berwick, D. M., Fineberg, H. V., & Weinstein, M. C. (1981). When doctors meet numbers. *American Journal of Medicine, 71,* 991–998.
4. Windeler, J., & Köbberling, J. (1986). Empirische Untersuchung zur Einschätzung diagnostischer Verfahren am Beispiel des Haemoccult-Tests [An empirical study of the judgments about diagnostic procedures using the example ot the hemoccult test]. *Klinische Wochenschrift, 64,* 1106–1112.
5. Gigerenzer, G., Hoffrage, U., & Ebert, A. (1998). AIDS counselling for low-risk clients. *AIDS Care, 10,* 197–211.
6. Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge, UK: Cambridge University Press.
7. Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704.
8. Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73.
9. Sedlmeier, P. (1997). BasicBayes: A tutor svstem for simple Bayesian inference. *Behavior Research Methods Instruments & Computers, 27,* 328–336.
10. Gigerenzer, G. (1996). The psychology of good judgment: Frequency formats and simple algorithms. *Journal of Medical Decision Making, 3,* 273–280.