MEDICINE

Risk literacy in medical decision-making

How can we better represent the statistical structure of risk?

By **Joachim T. Operskalski**
and **Aron K. Barbey**

Imagine that you have received a positive result on a routine cancer screening test. Follow-up biopsies were inconclusive, and the decision to treat aggressively or monitor conservatively is yours. Consider the following information: 0.1% of the population has a terminal version of this form of cancer, 99% of those people will appear positive on the test you have been administered, and 5% of those without terminal cancer will still have a benign condition that tests positive. Given your test result, what is the probability that you have terminal cancer and should treat it aggressively? When judging risks and trying to predict the future, how should you decide? We need to better understand the structure of risks and how the human mind creates representations of risk and probability.

In the scenario above, if your rapid impression of the information at hand supports the intuition that the probability of terminal cancer is close to 95%, then your line of thinking is not surprising, but you would be wrong. The true probability is 2%, with a complementary 98% risk of having undergone unnecessary (and even dangerous) treatment (*1*). This is hard to grasp intuitively without using Bayes' theorem of conditional probability (*2*); even physicians and medical students are prone to this error without having been explicitly instructed on the statistics of rare events (*3*). However, it might be understood more readily if you considered the situation in another way: The small number of false positives in the large population without cancer is a greater number than the majority of true positives in the smaller subpopulation that actually have the disease.

Far from being a statistical curiosity, this is the exact dilemma faced by patients and their physicians every day, and it will become more common as we learn more about the genetic factors that affect health. A "precision medicine" approach formalizes the insight that no two patients are alike (*4*, *5*). Improving medical diagnostic technology will improve many people's lives, but to

Decision Neuroscience Laboratory, Beckman Institute, Urbana, IL 61801, USA. Email: barbey@illinois.edu

prevent doing harm, the medical field and society at large need to understand that technological advances will not remove uncertainty. By understanding the way we think about risk, we can bring insight from the pages of statistical textbooks into legislative sessions, hospital exam rooms, and family discussions alike.

Contemporary research on medical decision-making originates from two traditions in the psychological sciences. The "heuristics and biases" view claims that errors in reasoning are an unavoidable consequence of our mental architecture as humans. Under this view, we each possess a fast and impression-based system of decision-making, in addition to a slower, reflective one capable of complex calculations that checks and verifies our quick impressions (*6*). Errors in reasoning may result when time pressure or situational complexity cause the reflective system to accept as fact the incorrect output of the impression-based system. Thus, the primary solution to such errors is education in statistics and awareness of our own biases. The "ecological rationality"

> *"The mind embodies a natural capacity to perform elementary set operations..."*

---

view of decision-making offers another way out, however. By considering reasoning errors as the consequence of having evolved to mentally process information in specific formats, we can better understand risk and uncertainty by communicating with one another using the right types of information (*3*). For example, abstract concepts, such as single-event probabilities (e.g., 1% chance), would not have appeared naturally in environments inhabited by primitive humans, so we never evolved to think about single-event probability with ease (*7*). Instead, thinking about long-run frequencies relative to a reference class (e.g., 1 out of 100 times) is easier because it conveys the same information as percents or probabilities in a way that reflects how we experience the natural world. Indeed, even experts in statistics have difficulty making judgments of single-event probabilities without explicitly calculating them. However, when the same

problems are represented as event frequencies, even the statistical layperson can generate a response in accordance with Bayes' theorem. For example, when a sample of college undergraduates was asked to solve problems on the basis of single-event probabilities, only 12% generated a Bayesian response, whereas 56% of another sample did so when given numbers in a natural frequency format (*8*). Using probing questions and visual aids depicting frequency information raised the accuracy to 76% and 92%, respectively.

Few believe that numerical formats automatically endow us with the ability to mentally calculate the probability of some event by counting observations and comparing them to our prior beliefs. Which cognitive processes, then, account for this remarkable facilitation of Bayesian inference under "natural" or event frequencies? The mind embodies a natural capacity to perform elementary set operations, such as taking the intersection of sets (e.g., "A and B") and union of sets (e.g., "A or B") (*9*). Such set operations can be induced by frequency formats, which provide cues to the set structure of the problem and therefore facilitate Bayesian inference (*9*). Indeed, the key variable that predicts accurate inference is not the statistical format of the problem but the transparency of nested set relations (*8*, *9*). As long as the nested set structure of events in a larger reference class is made accessible, one can accurately understand and reason from single-event probabilities (*10*). However, when the set structure of the problem is obscured by the use of unusual (small or large) reference class sizes, even frequency formats are difficult to understand (*11*).

Where does contemporary research on human judgment and decision-making leave us as patients, physicians, and policy-makers? Even among professionals, there remains a lack of consensus on screening guidelines for diseases such as breast or prostate cancer when their prognostic values are ambiguous (*12*). In 2014, the U.S. Preventive Services Task Force recommended against using prostate-specific antigen (PSA) test results in diagnosing and treating prostate cancer because the false positive rate and ambiguity of prognosis meant that men with benign or slow-growing tumors were undergoing unnecessary

Visual representations of risk structure

As medical science reveals more about genetic health risk factors, patients will need tools to understand the uncertainty inherent in those risk factors.

Question: 0.1% of the population has a disease, and a test detects it 99% of the time but falsely identifies 5% of healthy people as sick. What is the likelihood of a positive test result being accurate?
Each format below conveys the same fundamental information about risk structure.

A Single-event probability format
Bayes' theorem is necessary (and difficult) when using single-event probabilities to calculate the probability of a hypothesis (having the disease) given the evidence for it (a positive test result).
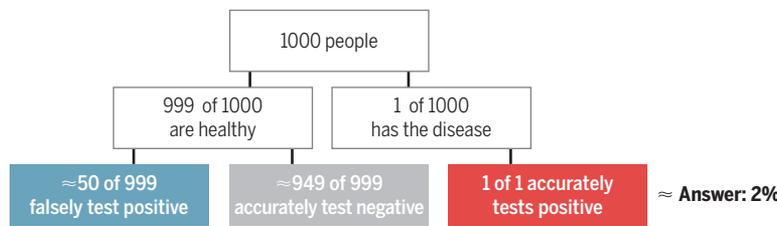
P(disease) = 0.1% prevalence of disease
P(positive test|disease) = 99% true positive rate
P(positive test|no disease) = 5% false positive rate

$$P(\text{disease}|\text{positive test}) = \frac{0.1\% \times 99\%}{(0.1\% \times 99\%) + (99.9\% \times 5\%)} = 1.94\% \approx \textbf{Answer: 2\%}$$
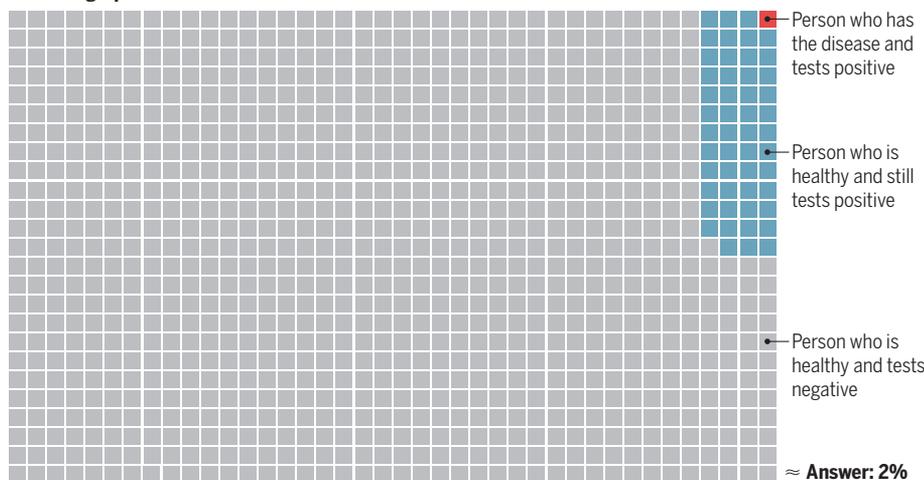
B Nested set format
This view facilitates accurate judgment because it represents base rates (prevalence) and reference class size (1 of 1000) without having to multiply a conditional probability by the base rate.

$$P(\text{disease}|\text{positive test}) = \frac{P(\text{disease AND positive test})}{P(\text{positive test})} = \frac{1}{50+1} = 1.96\% \approx \textbf{Answer: 2\%}$$

1000 people

999 of 1000 are healthy

1 of 1000 has the disease

≈50 of 999 falsely test positive

≈949 of 999 accurately test negative

1 of 1 accurately tests positive   ≈ **Answer: 2%**

C Pictograph format

Person who has the disease and tests positive

Person who is healthy and still tests positive

Person who is healthy and tests negative

≈ **Answer: 2%**

and risky treatment procedures (*13*). How did policy-makers arrive at this decision? They understood the statistical structure of risk in medical health problems. Health care systems that screened healthy men for PSA regularly had an approximately 98% 5-year survival rate for those diagnosed with cancer, whereas those that tested only men with symptoms had a 71% survival rate. Why, then, was there no significant difference in cancer mortality (26 versus 27 out of 100,000) between the systems across the entire population? Screening people aggressively when they are young ends up diagnosing extra cases who do not truly have cancer, or whose cancer is so slow-growing that they would not have been at risk for far beyond 5 years anyway. This may seem like an innocuous difference, but the extra diagnoses are associated with treatments that have their own risks. Using natural frequencies and drawing attention to the appropriate reference class (the whole population, not just people who are diagnosed before dying) draws attention to the fact that a higher survival rate is associated with a higher overall rate of positive test results and no difference in base rates of cancer (as indicated by mortality in the whole population) (*14*). Our instincts may tell us that 5-year survival rates are important, but only if placed in the appropriate context—natural frequencies or nested sets.

What measures can be applied to improve the ability to draw accurate statistical inferences and to fully appreciate the risks posed by medical decisions? It is important to consider conditions that lead to errors, such as risks that could be framed equivalently as gains or losses. Considering all possible outcomes (instead of only survival or mortality) can prevent inappropriate use of heuristics. Representing the problem in natural frequency formats or highlighting its nested set structure is helpful, and depicting outcomes using visual diagrams can facilitate fully informed decision-making via the creation of mental simulations (see the figure).

In the era of precision medicine, we are using a data-driven approach to understand patients as individuals rather than group averages. We are moving away from doctors telling patients with certainty what is wrong and what to do, toward a model in which doctors empower patients to take charge of their own care. To facilitate safe self-stewardship, the biomedical field needs to equip patients with the tools to make well-informed decisions in the face of uncertainty. ■

REFERENCES
1. D. M. Eddy, in *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, A. Tversky, Eds. (Cambridge Univ. Press, 1982), pp. 249–267.
2. G. Gigerenzer, U. Hoffrage, *Psychol. Rev.* **102**, 684 (1995).
3. G. Gigerenzer, *Med. Decis. Making* **16**, 273 (1996).
4. www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative
5. F. S. Collins, H. Varmus, *N. Engl. J. Med.* **372**, 793 (2015).
6. K. Stanovich, R. West, *Behav. Brain. Sci.* **23**, 645 (2000).
7. P. M. Todd, G. Gigerenzer, *Curr. Dir. Psychol. Sci.* **16**, 167 (2007).
8. L. Cosmides, J. Tooby, *Cognition* **58**, 1 (1996).
9. A. K. Barbey, S. A. Sloman, *Behav. Brain Sci.* **30**, 241 (2007).
10. W. P. Neace, S. Michaud, L. Bolling, K. Deer, L. Zecevic, *Judgment Decis. Making* **3**, 140 (2008).
11. S. A. Sloman, D. Over, L. Slovak, J. M. Stibel, *Organ. Behav. Hum. Decis. Process.* **91**, 296 (2003).
12. K. Jorgensen, P. Gotzche, *JAMA* **315**, 1402 (2016).
13. U.S. Preventive Services Task Force, "Final Recommendation Statement: Prostate Cancer: Screening" (2012); www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/prostate-cancer-screening.
14. G. Gigerenzer, *Psychol. Sci. Public Interest* **8**, 2 (2010).

10.1126/science.aaf7966

Editor's Summary

Article Tools	Visit the online version of this article to access the personalization and article tools: http://science.sciencemag.org/content/352/6284/413
Permissions	Obtain information about reproducing this article: http://www.sciencemag.org/about/permissions.dtl