Positive Mammografie = Brustkrebs?

Von den Schwierigkeiten im Umgang mit statistischen Informationen

Bei der Interpretation medizinischer Testbefunde ist das statistische Denken des Arztes gefordert. Jedoch haben Ärzte oft Schwierigkeiten im Umgang mit statistischen Informationen. Diese Schwierigkeiten können verhältnismässig leicht überwunden werden, wenn die relevante nummerische Information auf eine Weise dargestellt wird, die der menschlichen Informationsverarbeitung gut zugänglich ist.

Ulrich Hoffrage[1], Stephanie Kurzenhäuser[1] und Gerd Gigerenzer[1]

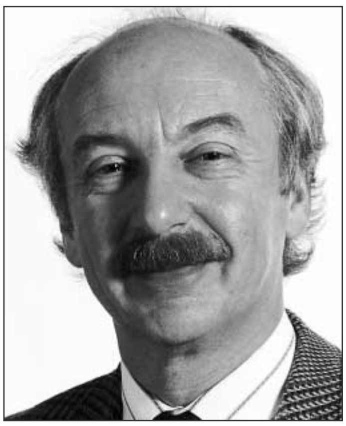
Ärzte können mit statistischen Informationen nicht immer so umgehen, wie man dies als Patient erwarten würde [1]. Dabei müssen Ärzte und PatientInnen im Alltag häufig Urteile unter Unsicherheit treffen, die einen angemessenen Umgang mit Zahlen erfordern. Ein Beispiel: Eine 55-jährige Frau, ohne einschlägige Symptome, ist dem Rat ihres Arztes gefolgt, im Rahmen der Brustkrebsfrüherkennung jedes Jahr eine Mammografie durchführen zu lassen. Bei einer solchen Untersuchung erhält sie einen positiven Befund. Schockiert über das Ergebnis, fragt sie ihren Arzt: «Heisst das, ich habe Brustkrebs?» «Nein, das kann man noch nicht sicher sagen.» Sie möchte es genauer wissen: «Wie hoch ist die Wahrscheinlichkeit, dass ich tatsächlich Brustkrebs habe?» Dem Arzt liegen folgende Informationen (a) zur Krankheit und (b, c) zum Testverfahren vor [2]:

(a) Die Wahrscheinlichkeit (p), dass eine symptomfreie Frau im Alter von 55 Jahren Brustkrebs (K) hat, beträgt *0,6 Prozent* (Prävalenz).
(b) Wenn eine dieser Frauen Brustkrebs hat, dann beträgt die Wahrscheinlichkeit, dass sie einen positiven Mammografie-Befund (M+) erhält, *94 Prozent* (Sensitivität des Tests).
(c) Wenn eine dieser Frauen jedoch *keinen* Brustkrebs (-K) hat, dann beträgt die Wahrscheinlichkeit, dass sie dennoch einen positiven Mammografie-Befund erhält, *7 Prozent* (Falsch-Alarm-Rate des Tests = 1 minus Spezifität)[2].

Um zu bestimmen, mit welcher Wahrscheinlichkeit die Frau mit dem positiven Mammogramm tatsächlich Brustkrebs hat, müssen obige Informationen in die so genannte Bayes'sche Regel eingesetzt werden (linke Seite der *Abbildung*). Die Rechnung ergibt 7,5 Prozent. In ei-


Ulrich Hoffrage

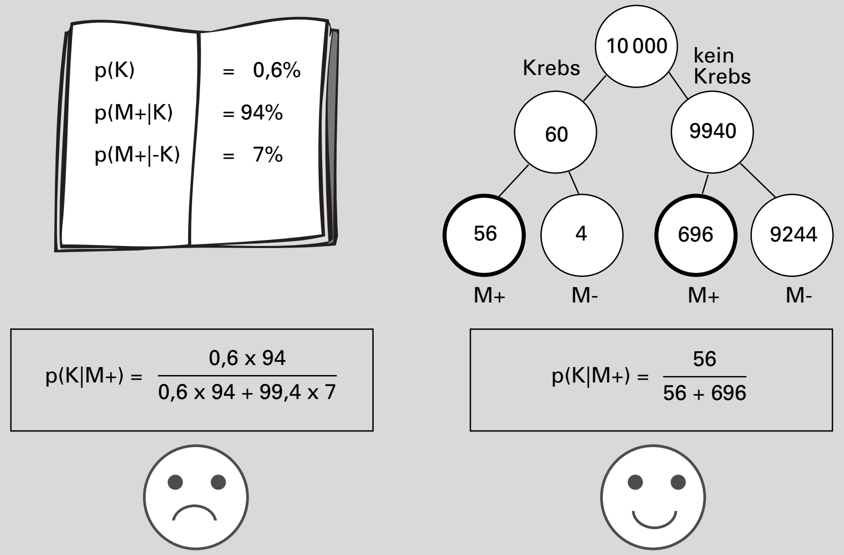

Stephanie Kurzenhäuser


Gerd Gigerenzer

1 Max-Planck-Institut für Bildungsforschung, Berlin
2 Die Werte sind entnommen aus Kerlikowske K et al.: Likelihood Ratios for Modern Screening Mammography – Risk of Breast Cancer Based on Age and Mammographic Interpretation (siehe Literaturangabe). Es sei hier ausdrücklich betont, dass sowohl die Sensitivität als auch die Spezifität der Mammografie stark abhängig sind von den Rahmenbedingungen, unter denen die Mammografie durchgeführt wird. In einem qualitätsgesicherten, systematisch durchgeführten Screening mit vielen Befundungen von speziell ausgebildeten und erfahrenen Radiologen wird es deutlich weniger übersehene Brustkrebse und falsch-positive Befunde geben als in kleinen gynäkologischen Praxen.

Abbildung:

Dieselben Informationen, dargestellt in Form von Wahrscheinlichkeiten und in Form von natürlichen Häufigkeiten



ner in den USA durchgeführten Untersuchung von David Eddy [3] schlossen 95 von 100 befragten Ärzten aus ähnlichen Informationen (p[K] = 1%, p[M+|K] = 80%, und p[M+|-K] = 9,6%), dass die Wahrscheinlichkeit für das Vorliegen von Brustkrebs zwischen 70 und 80 Prozent liegen würde; in Wirklichkeit betrug sie gemäss den Zahlen von Eddy 8 Prozent. Die Ärzte in Eddy's Studie sind nicht die einzigen, die hier Schwierigkeiten hatten; derart grosse Diskrepanzen zwischen menschlichem Urteil und der Wahrscheinlichkeitstheorie konnten auch bei deutschen Ärzten (im Kontext von Darmkrebs-Screening mit dem Haemoccult-Test) [4], bei Studenten an der Harvard Medical School [5] und bei Studenten in Deutschland und Österreich [6] nachgewiesen werden. Diese Fehleinschätzungen, die im Einzelfall zu gravierenden Fehlentscheidungen bezüglich weiterer Diagnostik oder Therapie führen können, lassen sich jedoch leicht vermeiden, wenn man die obigen Informationen in natürliche Häufigkeiten übersetzt.

Natürliche Häufigkeiten

Natürliche Häufigkeiten ergeben sich, wenn man in einer repräsentativen Stichprobe auszählt, welche Fälle dort wie häufig auftreten. In der Regel sind es diese Häufigkeiten, die den heute in Lehrbüchern üblicherweise verwendeten Wahrscheinlichkeiten zu Grunde liegen. Ist diese Information dort nun einmal in Form von Wahrscheinlichkeiten gegeben, so lässt sie sich leicht in natürliche Häufigkeiten zurückübersetzen: Im ersten Schritt wird die Prävalenz auf eine fiktive Anzahl von Personen bezogen, um zu bestimmen, wie viele von ihnen die Krankheit haben. Im zweiten Schritt wird die Sensitivität des Tests benutzt, um zu ermitteln, wie viele der kranken Personen ein positives Ergebnis erhalten, und im dritten Schritt wird mit Hilfe der Falsch-Alarm-Rate des Tests festgestellt, wie viele der gesunden Personen ein positives Ergebnis erhalten. Für das oben gewählte Beispiel ergibt sich mit dieser Methode:

(a) Von 10 000 Frauen haben 60 Brustkrebs.
(b) Von diesen 60 Frauen erhalten 56 ein positives Mammogramm.
(c) Von den 9940 nicht an Brustkrebs erkrankten Frauen erhalten allerdings 696 ebenfalls ein positives Mammogramm.

Aus den nun als natürliche Häufigkeiten dargestellten Informationen wird leicht ersichtlich, dass es unter 10 000 Frauen[3] 752 mit einem positiven Mammogramm geben wird (56+696), dass von diesen aber nur 56 tatsächlich Brustkrebs haben (rechte Seite der *Abbildung*). (Dieser Quotient, 56/752, entspricht der oben genannten Wahrscheinlichkeit von 7,5 Prozent, mit der eine Frau mit positivem Mammogramm tatsächlich Brustkrebs hat.)

Ausserdem wird nun leicht ersichtlich, was ein negatives Mammogramm bedeutet. Insgesamt sind 9248 negative Ergebnisse zu erwarten: 4 unter den 60 erkrankten Frauen, bei denen der Brustkrebs übersehen wird (4/60 = 6% = 1 minus Sensitivität), und 9244 unter den 9940 gesunden Frauen, die völlig zu Recht ein negatives Mammogramm erhalten (93% = Spezifität). Für eine Frau, die ohne Mammografie mit 99,4-prozentiger Wahrscheinlichkeit sicher sein konnte, keinen Brustkrebs zu haben (1 minus Prävalenz), ist diese Wahrscheinlichkeit nach Erhalt eines negativen Ergebnisses also «immerhin» um 0,56 Prozent auf 99,96 Prozent gestiegen (dies sind die 9244/9248).

Natürliche Häufigkeiten verbessern die Einschätzung von Risiken

In einer Serie von Experimenten mit Ärzten [7], Medizinstudenten [8] sowie Studenten verschiedenster Fachrichtungen [6] haben wir den Versuchsteilnehmern diagnostische Aufgaben gestellt. Jedesmal sollten sie die Bedeutung einer diagnostischen Information, wie zum Beispiel eines positiven Mammogramms, abschätzen. Dazu wurden ihnen die statistischen Informationen über die Krankheit und über das Testverfahren entweder in Form von Wahrscheinlichkeiten oder in Form von natürlichen Häufigkeiten gegeben. In jedem dieser Experimente schnitten die Versuchsteilnehmer deutlich besser ab, wenn die Informationen in Form von natürlichen Häufigkeiten präsentiert wurden. Auch für Jurastudenten und fertig ausgebildete Juristen erwiesen sich natürliche

3 Da für praktische Zwecke eine grobe Abschätzung zumeist völlig ausreichend ist und Rundungsfehler von daher vernachlässigt werden können, empfiehlt es sich im Allgemeinen, kleinere Bezugsgrössen zu wählen, z.B. 1000 statt 10 000.

Häufigkeiten als überaus hilfreich, um zu verstehen, was eine Übereinstimmung zwischen dem DNA-Profil einer am Tatort eines Verbrechens aufgefundenen biologischen Spur und dem DNA-Profil eines Verdächtigen bedeutet [8]. Die Methode, statistische Informationen in natürliche Häufigkeiten zu übersetzen, um die Bedeutung von Testresultaten besser zu verstehen, konnte auch erfolgreich für die Lösung komplexerer Aufgaben eingesetzt werden (mehrere hintereinander durchgeführte Tests, Tests mit unklarem Befund sowie Tests für mehrere Krankheiten) [9]. Des Weiteren war diese Methode in einer Lehreinheit für Medizinstudenten weit effizienter als die traditionelle Methode, nach der den Studenten die Bayes'sche Regel beigebracht und ihnen gezeigt wird, wie dort die Wahrscheinlichkeiten einzusetzen sind [10].

Relative und absolute Risikoreduktion

Informationen in Form von natürlichen Häufigkeiten sind nicht nur hilfreich, wenn es darum geht, die Bedeutung eines positiven Mammogramms einzuschätzen. Sie können auch der Frau helfen, die vor der Entscheidung steht, ob sie an einem Brustkrebs-Screening teilnehmen soll oder nicht. Das Risiko, an Brustkrebs zu sterben, kann durch regelmässige Teilnahme an der Mammografie reduziert werden. Diese Reduktion kann absolut (das heisst bezogen auf alle Frauen) oder relativ (bezogen auf alle Frauen, die ohne Mammografie in einem definierten Zeitraum an Brustkrebs sterben würden) ausgedrückt werden. Wird der Nutzen der Mammografie als absolute Risikoreduktion dargestellt, erscheint er geringer, als wenn man ihn – wie sonst üblich – als relative Risikoreduktion ausdrückt. Aber es ist die absolute Reduktion, die den Nutzen aus der Perspektive der einzelnen Frau darstellt, die sich als eine von vielen Screeningteilnehmerinnen sehen muss. So entspricht die absolute Risikoreduktion den oben vorgestellten natürlichen Häufigkeiten: Sie bezieht die Anzahl der durch das Screening geretteten Leben nicht auf eine Untergruppe, sondern auf die Gesamtmenge der am Screening teilnehmenden Frauen (zur Relevanz dieses Unterschiedes siehe den Artikel «Individuelle Entscheidungsfindung am Beispiel der Brustkrebs-Früherkennung» von Klazien Matter-Walstra und Ulrich Hoffrage im vorliegenden Heft, sowie [11, 12]).

Schlussfolgerungen

Die Darstellung von statistischer Information in Form von natürlichen Häufigkeiten ist eine einfache und praktikable Methode, von der Ärzte wie auch Laien profitieren können, wenn es darum geht, Unsicherheiten und Risiken zu interpretieren und/oder zu kommunizieren. Dies ist auch ein Schritt auf dem Weg hin zum Idealbild des informierten, mündigen Patienten, der im Dialog mit dem Arzt die in Diagnostik und Therapie involvierten Unsicherheiten und Risiken verstehen und gegeneinander abwägen kann, um so zu einer für ihn richtigen Entscheidung zu gelangen. ■

Korrespondenzadresse:
PD Dr. Ulrich Hoffrage
Max-Planck-Institut
für Bildungsforschung
Lentzeallee 94, D-14195 Berlin
Tel. 0049-30-82406 273,
Fax: 0049-30-82406 394
E-Mail:
hoffrage@mpib-berlin.mpg.de

Literatur

1. Berwick DM, Fineberg HV, Weinstein, MC (1981): When doctors meet numbers. Am J Med 71: 991–998.

2. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V (1996): Likelihood Ratios for Modern Screening Mammography – Risk of Breast Cancer Based on Age and Mammographic Interpretation. JAMA 3: 39–43.

3. Eddy DM (1982): Probabilistic reasoning in clinical medicine: Problems and opportunities. In: Kahneman D, Slovic P, Tversky A (Eds): Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press, pp. 249–267.

4. Windeler J, Köbberling J (1986): Empirische Untersuchung zur Einschätzung diagnostischer Verfahren am Beispiel des Haemoccult-Tests. Klin Wschr 64: 1106–1112.

5. Casscells W, Schoenberger A, Grayboys T (1978): Interpretation by physicians of clinical laboratory results. N Engl J Med 299: 999–1001.

6. Gigerenzer G, Hoffrage U (1995): How to improve Bayesian reasoning without instruction: Frequency formats. Psychol Rev 102: 684–704.

7. Hoffrage U, Gigerenzer G (1998): Using natural frequencies to improve diagnostic inferences. Acad Med 73: 538–540.

8. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G (2000): Communicating Statistical Information. Science 290: 2261–2262.

9. Krauss S, Martignon L, Hoffrage U (1999): Simplifying Bayesian inference: The general case. In Magnani L, Nersessian NJ, & Thagard P, (Eds): Model-based reasoning in scientific discovery. New York: Kluwer Academic/Plenum Publishers, pp. 165–179.

10. Kurzenhäuser S, Hoffrage U (2000): Teaching Bayesian Reasoning: An evaluation of a classroom tutorial. Manuskript zur Veröffentlichung eingereicht. Für eine zusammenfassende Darstellung der Befunde siehe auch Hoffrage U, Kurzenhäuser S, Gigerenzer G (2000): Wie kann man die Bedeutung medizinischer Testbefunde besser verstehen und kommunizieren? Zeitschr f ärztl Fortb u Qual.sicherung 94: 713–720.

11. Forrow L, Taylor WC, Arnold RM (1992): Absolutely relative: how research results are summarized can effect treatment decisions. Am J Med 92: 121–124.

12. Malenka DJ, Baron JA, Johansen S, Wahrenberger JW, Ross JM (1993): The framing effect of relative and absolute risk. J Gen Intern Med 8: 543–548.