

Wie kann man die Bedeutung medizinischer Testbefunde besser verstehen und kommunizieren?

Ulrich Hoffrage, Stephanie Kurzenhäuser und Gerd Gigerenzer

Max-Planck-Institut für Bildungsforschung, Berlin

Zusammenfassung

Bei der Interpretation medizinischer Testbefunde ist das statistische Denken des Arztes gefordert: Wie groß ist die Wahrscheinlichkeit für das Vorliegen einer Krankheit bei einem positiven Testbefund? Diese Wahrscheinlichkeit lässt sich mithilfe der Regel von Bayes bestimmen. Mehrere Studien zeigen jedoch, dass Ärzte oft Schwierigkeiten mit derartigen bayesianischen Inferenzen haben. Wir zeigen, dass ihre Urteile deutlich verbessert werden können, wenn numerische Information auf eine Weise repräsentiert wird, die der menschlichen Informationsverarbeitung leicht zugänglich ist. Dies ist bei den üblicherweise verwendeten Wahrscheinlichkeiten nicht der Fall, wohl aber bei „natürlichen Häufigkeiten“, die sich durch die Auszählung von beobachteten Einzelfällen in einer natürlichen Umgebung ergeben. In einer Reihe von Studien haben wir das Repräsentationsformat der relevanten statistischen Information variiert. Wurde die Information nicht in Form von Wahrscheinlichkeiten oder Prozenten, sondern in natürlichen Häufigkeiten präsentiert, konnten sowohl medizinische Experten als auch Laien ihre Urteile deutlich verbessern. Und so lässt sich auch der Umgang mit Wahrscheinlichkeiten und Prozenten leicht erlernen: Zwei Trainingsprogramme, in denen gezeigt wurde, wie Wahrscheinlichkeiten in natürliche Häufigkeiten zu übersetzen sind, versetzten die Teilnehmer in die Lage, sehr gute Ergebnisse auch bei der Lösung dieser Aufgaben zu erzielen. Abschließend diskutieren wir die Bedeutung, die eine verständliche Risiko- und auch Nutzenkommunikation im Arzt-Patienten-Verhältnis hat.

Sachwörter: Risiko-Kommunikation, Bayes-Theorem, Repräsentation von Information, statistisches Denken, Mediziner aus- und -fortbildung

Hier ist eine alltägliche Situation, in der Ärzte und Patientinnen mit Zahlen konfrontiert werden und Urteile unter Unsicherheit treffen müssen: Eine Frau, ohne einschlägige Symptome, ist dem Rat ihres Arztes gefolgt, jedes Jahr im Rahmen der Brustkrebsfrüherkennung eine Mammographie durchführen zu lassen. Bei der letzten 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 zur Krankheit und zum Testverfahren vor (2):

- Die Wahrscheinlichkeit, dass eine symptomfreie Frau Brustkrebs (K) hat, beträgt 1%.
- Wenn eine Frau Brustkrebs hat, dann beträgt die Wahrscheinlichkeit, dass sie einen positiven Mammographie-Befund (M) erhält, 80%.
- Wenn eine Frau jedoch *keinen* Brustkrebs ($\neg K$) hat, dann beträgt die Wahrscheinlichkeit, dass sie dennoch einen positiven Mammographie-Befund erhält, 10%.

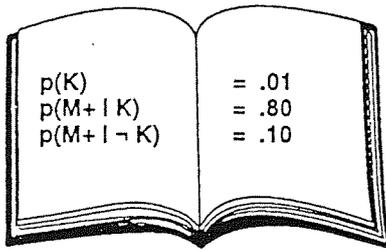
In einer in den USA durchgeführten Untersuchung von David Eddy schlossen 95 von 100 befragten Ärzten aus dieser Information, dass die Wahrscheinlichkeit für das Vorlie-

In einer amerikanischen Studie aus dem Jahr 1981 waren über die Hälfte von 151 befragten praktizierenden Ärzten nicht in der Lage, die Begriffe Prävalenz und Inzidenz korrekt zu definieren, und nur ein Fünftel fand die richtige Lösung bei einer Aufgabe zum Zusammenhang von Prävalenz und prädiktivem Wert eines positiven Tests (1). Dies sind nur zwei Beispiele, die zeigen, dass Ärzte mit statistischen Begriffen und Informationen nicht immer so umgehen können, wie man dies als Patient erwarten würde. Der Titel, unter dem die Ergebnisse dieser Studie veröf-

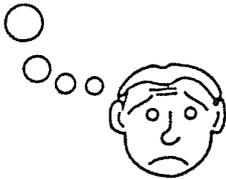
fentlicht wurden – „When doctors meet numbers“ – lässt die Schlussfolgerung der Autoren bereits erahnen: „Our results suggest that important problems exist in the consumption of quantitative information by medically trained individuals.“ (S. 996). In dem vorliegenden Artikel wollen wir uns nicht damit begnügen, einen Missstand aufzuzeigen, sondern wollen eine einfache und praktikable Methode vorstellen, mit der Ärzte das Ergebnis eines medizinischen Tests besser interpretieren und damit auch dessen Bedeutung besser kommunizieren können.



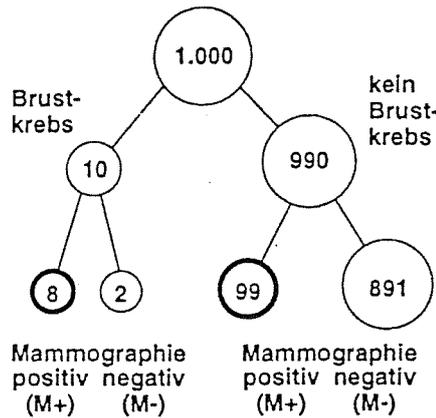
Wahrscheinlichkeiten



$$p(K | M+) = \frac{.01 \times .80}{.01 \times .80 + .99 \times .10}$$



Natürliche Häufigkeiten



$$p(K | M+) = \frac{8}{8 + 99}$$

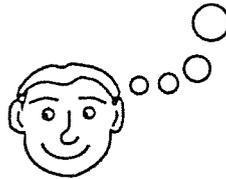


Abb. 1. Repräsentation ein und derselben Information in Wahrscheinlichkeiten und natürlichen Häufigkeiten. Die entsprechenden kognitiven Algorithmen sind unterschiedlich komplex.

gen von Brustkrebs hier zwischen 70% und 80% liegen würde (2). Zieht man jedoch die Regel von Bayes heran, um aus den oben angegebenen Informationen die gesuchte Wahrscheinlichkeit zu errechnen, so sieht man, dass diese lediglich 7,5% beträgt (Abb. 1, linke Seite). Eddy's Studie ist kein Einzelfall: Solche Diskrepanzen zwischen menschlichem Urteil und der Wahrscheinlichkeitstheorie konnten auch bei deutschen Ärzten im Kontext von Darmkrebs-Screening mit dem Haemocult-Test (3), bei Studenten an der Harvard Medical School (4) und bei Studenten in Deutschland und Österreich (5) nachgewiesen werden. Diese Fehleinschätzungen, die im Einzelfall zu gravierenden Fehlentscheidungen bezüglich weiterer Diagnostik oder Therapie führen können, wurden dahingehend interpretiert, dass die menschliche Informationsver-

arbeitung nicht den Gesetzen der Wahrscheinlichkeitstheorie folgt und somit „irrational“ sei. Aber kann man eine solch weit reichende Schlussfolgerung schon auf Grundlage dieser Befunde ziehen, oder müssen weitere wichtige Einflussgrößen beachtet werden?

Menschliche Informationsverarbeitung ist an ein spezifisches Informationsformat angepasst

Ein Taschenrechner hat einen Algorithmus für Multiplikation. Er verarbeitet dabei Zahlen – aber Zahlen können in verschiedenen Formaten repräsentiert werden, wie zum Beispiel arabisch, römisch oder binär. Die Algorithmen üblicher Taschenrechner sind auf das arabische System eingestellt. Werden in einen solchen Taschenrechner binäre Zah-

len eingetippt, produziert er unsinnige Ergebnisse. Daraus kann man jedoch nicht schließen, dass er keinen Algorithmus für Multiplikation hätte. Wie die Algorithmen eines Rechners sind auch kognitive Algorithmen auf bestimmte Formate eingestellt. (Versuchen Sie einmal eine Division mit römischen Zahlen!) Übertragen wir nun diese enge Verknüpfung von Algorithmus und Informationsformat auf Urteilen unter Unsicherheit. Wie in Eddy's Studie, ist auch in den meisten Lehrbüchern für Medizin und Statistik Information in Form von Wahrscheinlichkeiten und Prozenten repräsentiert. Stellen wir die oben erwähnte pessimistische Auffassung über menschliche Informationsverarbeitung einmal infrage und nehmen wir an, dass Menschen – wie auch andere Organismen – im Laufe der Evolution sehr wohl Algorithmen für den Umgang mit einer unsicheren Welt ausgebildet haben. Auf welches Format sind diese kognitiven Algorithmen eingestellt? Sicher nicht auf Wahrscheinlichkeiten und Prozentwerte, die erst seit wenigen Jahrhunderten existieren – das mathematische Konzept der Wahrscheinlichkeit kam erst in der Mitte des 17. Jahrhunderts auf (6), und Prozente wurden erst mit der Einführung des metrischen Systems nach der französischen Revolution eine gängige Repräsentationsform (7). Diese wenigen Jahrhunderte stellen im Hinblick auf die Evolution menschlicher Intelligenz eine vernachlässigbare Zeitspanne dar. So ist unsere These, dass kognitive Algorithmen nicht an Wahrscheinlichkeiten, sondern an *natürliche Häufigkeiten* angepasst sind. Natürliche Häufigkeiten ergeben sich aus der Aggregation von (repräsentativ gezogenen) Einzelfällen, wie Menschen sie auch bereits schon in einer Zeit ohne Bücher und Statistiken beobachtet haben. Dazu ein Beispiel: Stellen Sie sich einen Arzt in einer Gesellschaft vor, die keine Bücher und keine Statistiken kennt. Eine bisher unbekante



Krankheit (K) ist ausgebrochen, von der lediglich ein Symptom (M) bekannt ist. Der Arzt hat 1.000 Personen untersucht.

- Von diesen 1.000 Personen hatten 10 die Krankheit.
- Von diesen 10 Personen zeigten 8 das Symptom.
- Allerdings gab es auch unter den 990 gesunden Personen 99, die das kritische Symptom zeigten.

Nun erscheint ein neuer Patient, er hat das Symptom. Wie hoch ist die Wahrscheinlichkeit, dass er auch die Krankheit hat? Der Arzt, der alle Beobachtungen selbst gemacht hat, benötigt keine komplizierten Berechnungen, um diese Wahrscheinlichkeit zu schätzen. Er braucht lediglich die beobachtete Anzahl der Fälle mit Symptom und Krankheit (8) durch jene der Symptomfälle (8 + 99) zu dividieren. Man sieht sofort, dass das Risiko nicht bei 70% bis 80%, sondern bei weniger als 8% liegt (Abb. 1, rechte Seite). Und selbst wenn der Arzt nicht Strichliste geführt hätte, um die genauen Zahlen zu erhalten: Eine Vielzahl von Studien zeigt, dass sowohl Menschen als auch Tiere Häufigkeitsinformationen in ihrer natürlichen Umgebung automatisch und ziemlich genau registrieren (z.B. 8, 9, 10, 11).

Vergleicht man die beiden Seiten von Abbildung 1, so sieht man, dass die statistischen Informationen dieselben sind: Die Wahrscheinlichkeiten können in natürliche Häufigkeiten übersetzt werden und umgekehrt. Ferner sind beide Gleichungen Versionen der Regel von Bayes. Und doch gibt es einen Unterschied: Kognitive Algorithmen für die Regel von Bayes sind einfacher, wenn die relevante Information in natürlichen Häufigkeiten statt in Wahrscheinlichkeiten oder Prozentwerten vorgegeben wird (12, 13). Im Folgenden zeigen wir, dass sich mit dieser einfachen Methode – nämlich die relevanten statistischen Informationen in dem Format zu repräsentieren, in dem sie leichter verstanden werden – die Urteile so-

wohl von Laien als auch von Experten verbessern lassen.

Wie verbessert man die Einschätzung von Risiken durch Laien und Experten?

In der Fachliteratur und der durch sie geprägten öffentlichen Meinung ist aufgrund einer Reihe von Befunden wie den oben genannten das Bild des irrationalen Laien entstanden, der bei der Einschätzung von Risiken weitgehend versagt. Diese „negativen“ Befunde basieren auf Studien, in denen die Information in Form von Wahrscheinlichkeiten vorgegeben wurde. Lassen sich intuitive Urteile von Laien verbessern, indem die Information in natürlichen Häufigkeiten statt in Wahrscheinlichkeiten präsentiert wird? In einer Studie mit Laien, das heißt Personen ohne einschlägiges Statistik-Training, haben wir mit einer kombinierten Ergebnis- und Prozessanalyse bei 1.800 individuellen Aufgabenlösungen die jeweils benutzten Strategien identifiziert (5). Verwendet wurden 15 verschiedene Probleme (z.B.: Was ist die Wahrscheinlichkeit von Brustkrebs nach einem positiven Mammogramm? Was ist die Wahrscheinlichkeit einer HIV-Infektion nach einem positiven HIV-Test? Was ist die Wahrscheinlichkeit eines Autounfalls, wenn man mit einem betrunkenen Fahrer mitfährt?) Die Teilnehmer erhielten die Information entweder als Wahrscheinlichkeiten oder als natürliche Häufigkeiten. Wurde die Information in Wahrscheinlichkeiten gegeben, wie es für Textbücher und viele psychologische Experimente üblich ist, so wurde in nur 16% der Fälle im Einklang mit der Regel von Bayes geurteilt. Wurde die Information hingegen in natürlichen Häufigkeiten gegeben, stieg der Anteil der bayesianischen Lösungen auf 46%. Die Beurteilung von Risiken durch Laien kann also deutlich verbessert werden, ohne dass diese in statisti-

schem Denken unterrichtet werden: Wird die relevante Information in einem intuitiv zugänglichen Format vorgegeben, übernimmt diese externe Repräsentation einen Teil der Denkarbeit.

Können auch Experten von dieser Methode profitieren? Wie oben bereits erwähnt, haben die meisten Ärzte große Schwierigkeiten, die Wahrscheinlichkeit für das Vorliegen einer Erkrankung bei gegebenem positiven Testausgang korrekt einzuschätzen. So erschreckend dieser Befund zunächst klingen mag, er ist kaum überraschend, wenn man bedenkt, dass statistisches Denken, inklusive der Anwendung der Regel von Bayes, in den meisten medizinischen Lehrbüchern in Form von Wahrscheinlichkeiten, Prozenten oder relativen Häufigkeiten gelehrt wird. Wir testeten 48 Ärzte mit einer mittleren Berufserfahrung von 14 Jahren, die an Universitätskliniken, in privaten Praxen und im öffentlichen Dienst arbeiten (14). Jeder Arzt erhielt vier medizinisch-diagnostische Aufgaben. Bei zwei Aufgaben wurde die Information in Form von Wahrscheinlichkeiten gegeben – hier stimmten ihre Urteile nur in 10% der Fälle mit der Regel von Bayes überein. Bei den anderen beiden Aufgaben wurden die Informationen in Form natürlicher Häufigkeiten gegeben, damit stieg dieser Anteil auf 46%. Mit den Worten eines Arztes: „Es ist jetzt anders, das kann ich mir ganz gut vorstellen. Es handelt sich um eine Anzahl, das ist visueller.“ Die Präsentation in Häufigkeiten bewirkte auch, dass die Ärzte schneller zu einer Einschätzung gelangten, und dass sie weniger nervös, weniger verunsichert und entspannter waren. Das Ergebnis der Ärzte entspricht also – trotz ihrer langjährigen Erfahrung mit diagnostischen Urteilen – dem der Laien: Auch sie können deutlich profitieren, wenn Informationen in natürlichen Häufigkeiten statt in Wahrscheinlichkeiten kommuniziert werden.



AIDS-Beratung

AIDS erhält in den Medien viel Aufmerksamkeit, und in den deutschen Gesundheitsämtern wurden entsprechende Beratungszentren eingerichtet. Deren Aufgabe ist unter anderem, Ratsuchenden zu erklären, was es bedeutet, wenn der HIV-Test positiv ausfallen würde – eine verantwortungsvolle Aufgabe, wenn man bedenkt, dass die Reaktionen der Betroffenen auf die Mitteilung eines positiven HIV-Tests bis hin zum Suizid reichen können (15). Wie erklären professionelle AIDS-Berater die Bedeutung eines eventuell positiven Ergebnisses ihren Klienten? Um diese Frage zu beantworten, hat einer unserer Mitarbeiter, ein 25-jähriger Mann, die öffentlichen Gesundheitsämter in 20 deutschen Städten besucht, um insgesamt 20 HIV-Tests durchführen zu lassen (16). Da vor jedem Test eine obligatorische Beratung stattfindet, konnte er dort die Berater in einem halbstandardisierten Interview verdeckt befragen.

Die Hauptergebnisse der Studie waren: (1) Die Hälfte der Berater erklärte dem Klienten fälschlicherweise, dass ein positives Testergebnis mit 100-prozentiger Sicherheit auf das Vorliegen einer HIV-Infektion hindeuten würde; weitere fünf erklärten, dass die Sicherheit 99,9% oder höher sei. (2) Die Mehrheit der Berater erklärte, dass es keine falsch-positiven Diagnosen gäbe (diese Aussage steht im Widerspruch zur Literatur, der zufolge es diese sehr wohl gibt – kontrovers ist lediglich die Frage, wie häufig sie sind, und damit auch die Frage, wie mit welcher Sicherheit ein positives Testergebnis auf das Vorliegen einer Infektion hindeutet). (3) Kein einziger der 20 professionellen AIDS-Berater hat die relevante Information in der Form von natürlichen Häufigkeiten erteilt, alle numerischen Informationen wurden in Form von Wahrscheinlichkeiten und Prozentwerten vermittelt – was Fehleinschätzung des Risikos und

Konfusion sicherlich begünstigt hat.

Diese Ergebnisse haben direkte praktische Konsequenzen für die Ausbildung von AIDS-Beratern und für die Aufklärung der Klienten, insbesondere dann, wenn Selbsttests oder die in verschiedenen Ländern diskutierte HIV-Testung großer Bevölkerungsteile eingeführt werden. Wir brauchen eine Ausbildung für statistisches Denken, die vermittelt, wie Risiken verständlich und effektiv kommuniziert werden können. Ein solches Training haben wir sowohl in Form eines Computer-Tutorials als auch eines Classroom-Tutorials erprobt.

Trainingsprogramme für Aus- und Fortbildung

Statistisches Denken, insbesondere die Regel von Bayes und andere Methoden zur Einschätzung von Risiken, nimmt im Mathematikunterricht in der Schule nur wenig Raum ein. Auch sind bislang nur wenige kontrollierte Trainingsstudien durchgeführt worden. Diese Studien, in denen hauptsächlich ergebnisorientiertes Feedback gegeben wurde, zeigten nur geringen oder keinen Erfolg der Trainingsmaßnahmen. Basierend auf der oben dargestellten These, dass menschliche Informationsverarbeitung an natürliche Häufigkeiten adaptiert ist, haben wir eine Methode entwickelt, statistisches Denken effektiv zu unterrichten (17, 18). Während traditionelle Methoden lehren, Wahrscheinlichkeiten in die Regel von Bayes einzusetzen („Regel-Lernen“; Abb. 1, linke Seite), lehrt hier ein computergestütztes Tutorsystem, Unsicherheiten und Risiken in Form natürlicher Häufigkeiten zu repräsentieren („Repräsentations-Lernen“; Abb. 1, rechte Seite). Das heißt, ausgehend von Informationen im gängigen Wahrscheinlichkeitsformat wird beim Repräsentations-Lernen gezeigt, wie diese in natürliche Häufigkeiten zu übersetzen und dann zur Aufgabenlösung

zu verrechnen sind. Im Unterschied zu den oben referierten Studien wird hier also nicht die Repräsentation der Aufgabeninformation selbst verändert, sondern die Repräsentation der gegebenen Information beim Lösungsprozess.

In einer Studie mit 54 Studierenden verschiedener Fachrichtungen haben wir die Effizienz beider Verfahren verglichen (17, 18). Der unmittelbare Lernerfolg beim Lernen von Repräsentationen war etwa doppelt so hoch wie beim traditionellen Lernen von Regeln. Doch der unmittelbare Lernerfolg ist wegen des bekannten rapiden Vergessens von statistischen Prinzipien allein nicht ausreichend (ein Arzt aus unserer oben genannten Studie (14): „Wir haben so eine Formel gelernt. Ich habe sie vergessen.“) Bei den Personen, welche durch Lernen von Regeln trainiert wurden, war nach fünf Wochen der Lernerfolg weitgehend verschwunden, nur noch 15% der Aufgaben konnten in einem nachfolgenden Test gelöst werden. Beim Lernen von Repräsentationen dagegen trat kaum Vergessen ein: die Leistung lag fünf Wochen nach dem Training immer noch bei 90%.

Diese Ergebnisse zeigen, wie hilfreich sich die hier skizzierten Erkenntnisse auf das Erlernen von statistischem Denken im Allgemeinen und die Regel von Bayes im Besonderen auswirken, vor allem in Kombination mit den modernen Methoden des computerunterstützten, interaktiven Lernens. Allerdings sieht die gängige Unterrichtspraxis an deutschen Universitäten und anderen Aus- und Weiterbildungsstätten oft (noch) ganz anders aus: Es findet traditioneller Frontalunterricht für Gruppen mit Tafel und Overheadprojektor statt. Lerncomputer stehen den Teilnehmern meist nicht zur Verfügung. Aus diesem Grunde haben wir zusätzlich zum Computer-Tutorial eine einstündige Trainingseinheit entwickelt, die den zuvor dargestellten Ansatz des Repräsentations-Lernens im typischen Seminarsetting umsetzt (19). Der



Ablauf sah wie folgt aus: Nach Vor-test und Einführung einiger grund-legendender Begriffe wurde an einem Beispiel erläutert, wie die Wahr-scheinlichkeitsinformation in natür-liche Häufigkeiten zu übersetzen ist. Dieser Lösungsweg wurde dann an drei weiteren Beispielen vertieft. Dem Tutor standen Overheadfolien zur Verfügung, die Teilnehmer er-hielten Handouts mit der Bitte, sich aktiv zu beteiligen und die Rech-nungen dort nachzuvollziehen. Er-gebnisse und offene Fragen wurden gemeinsam diskutiert.

Erprobt wurde diese Trainingsein-heit mit Medizinstudenten aus der Vorklinik. Ähnlich wie im Compu-ter-Tutorial wurde auch in dem Classroom-Tutorial die Effizienz des Repräsentations-Lernens mit der Kontrollbedingung Regel-Ler-nen verglichen, die – abgesehen vom Lösungsweg – weitestgehend gleich gestaltet war. Insgesamt wur-den 208 Studenten bei einer durch-schnittlichen Gruppengröße von 15 Teilnehmern unterrichtet, 109 er-hielten ein Repräsentations-Train-ing und 99 ein Regel-Training. In dem vor der jeweiligen Trainings-einheit durchgeführten Vortest er-reichten die Teilnehmer beider Gruppen einen vergleichbar niedri-gen Anteil bayesianischer Lösungen von 1% bzw. 3%. Der Nachtest zur Evaluation des Trainingserfolgs, der wie der Vortest aus einer Textaufga-be im Wahrscheinlichkeitsformat bestand, fand nicht direkt im An-schluss, sondern erst zwei Monate später statt, da wir auch hier primär am langfristigen Lernerfolg interes-siert waren. Von den 69 Teilnehmern des Regel-Trainings, die den Nach-test bearbeitet haben, vermochten nach diesen zwei Monaten nur 16% die Aufgabe richtig zu lösen, bei den 66 Teilnehmern an der Repräsentati-ons-Bedingung, die den Nachtest bearbeitet haben, waren es 47%. Diese Erfolgsquote von knapp 50% fiel hier nicht so hoch aus wie die beim computerunterstützten Einzel-lernen, was aufgrund der weniger optimalen Lernbedingungen in der

vorgefundenen Unterrichtssituation (geringerer Grad an individueller Aktivität, Interaktivität und Wiederholungen) nicht weiter überrascht. Dennoch war auch im traditionellen Seminarsetting das Repräsentations-Lernen dem Regel-Lernen einmal mehr deutlich überlegen. Festzuhal-ten ist, dass sich die hier skizzierten Erkenntnisse mit Erfolg in verschie-denen Kontexten der Ausbildung im statistischen Denken einsetzen las-sen.

Risikokommunikation in der Medizin: Zum Wohle des Patienten?

Eine Bemerkung, die auf die oben vorgestellten Arbeiten oft spontan gemacht wird, lautet: „Gut, natürli-che Häufigkeiten mögen helfen, die Bedeutung eines positiven Testbefunds besser zu erfassen, aber letzt-lich ist dies doch gar nicht wichtig. Bei einem positivem Befund wird ohnehin weiter diagnostiziert oder gleich mit der Therapie begonnen, und wenn es ein falsch-positiver Befund war und die Krankheit nicht vorliegt, so wird sich dies schon früher oder später herausstellen.“ Dazu möchten wir Folgendes zu be-denken geben. Elmore et al. errech-neten zum Beispiel, dass fast die Hälfte aller Frauen, die an zehn Screening-Mammographien teil-nehmen, mindestens einmal einen falsch-positiven Befund erhalten (20). Die Mitteilung eines positiven Befunds und die Interpretation „Ich habe wahrscheinlich Brustkrebs“ hat schwere psychische Folgen. Ist nach erfolgter Abklärung, in der sich ein solcher Befund als falsch-positiv herausstellt, alles wieder beim Alten? Lerman et al. berich-ten, dass selbst drei Monate nach er-folgter Entwarnung 26% aller Frau-en psychisch und 17% sogar in der Verrichtung ihrer Alltagsgeschäfte beeinträchtigt sind (21). Hätte man diesen Frauen von Anfang an gesagt „Nur eine von zehn Frauen hat nach einem solchen Ergebnis tatsächlich

Brustkrebs“ sollte die Entwarnung, so könnte man spekulieren, glaub-hafter sein.

Bisher war ausschließlich davon die Rede, was ein (positiver) Testbe-fund bedeutet und wie man die rele-vante statistische Information am besten kommunizieren kann, damit man die Bedeutung eines solchen Befunds besser versteht. Das The-ma „Risikokommunikation in der Medizin“ wird jedoch schon viel früher relevant, nämlich bei der Fra-ge, ob ein diagnostisches Verfahren überhaupt eingesetzt werden soll. Mit dieser Frage können wir gerade beim Beispiel der Mammographie bleiben, und zwar hier bei ihrer Ver-wendung im Rahmen der Brust-krebsfrüherkennung. Um zu ent-scheiden, ob sie an einem Mammo-graphie-Screening teilnehmen soll oder nicht, sollte eine Frau vernünftigerweise den Nutzen gegen die Nachteile abwägen. Vom psycholo-gischen Standpunkt aus spielt es aber eine große Rolle, wie der Nut-zen bzw. wie die Nachteile kommuni-ziert werden, und von daher sollte die Art dieser Kommunikation auch einen spürbaren Einfluss auf ihre Teilnahme-Entscheidung haben. Ein Beispiel: Der hauptsächliche Nutzen, eine Verminderung des Ri-sikos an Brustkrebs zu sterben, kann sowohl als relative wie auch als absolute Risikoreduktion darge-stellt werden. Im ersten Fall wird kommuniziert, dass die Brustkrebs-Sterblichkeit um etwa 25% redu-ziert wird, im zweiten Fall, dass sie um etwa 0.1% reduziert wird (22). Man beachte, dass beide Aussagen auf denselben Daten beruhen: Ohne Mammographie sterben innerhalb der nächsten zehn Jahre 4 von 1.000 Frauen an Brustkrebs, mit Mammo-graphie sind es 3 von 1.000. Mithin kann 1 der 4 Frauen, die bei Nicht-Teilnahme gestorben wären, geret-tet werden (relative Risiko-Reduk-tion = 25%). Genauso kann man aber auch sagen, dass 1 von 1.000 teilnehmenden Frauen diesen Nut-zen haben wird (absolute Risiko-Reduktion = 0.1%). Offensichtlich



stimmen beide Zahlen, aber nach welcher soll man sich richten? Für die Frau, die vor der Entscheidung steht, ob sie am Screening teilnehmen soll, ist die absolute Risiko-Reduktion (0.1%) die relevante Größe, denn dieser Wert bezieht sich auf genau jene 1.000 Frauen in dieser Entscheidungssituation. Die 25% hingegen beziehen sich nur auf die 4 Frauen, die ohne Screening innerhalb der nächsten zehn Jahre an Brustkrebs sterben würden, und niemand weiß zu jenem Zeitpunkt, welche 4 dies einmal sein werden. Da der Beitrag von Mühlhauser und Höldke im vorliegenden Heft genau dieses Thema aufgreift (siehe auch 22, 23), soll auf diesen Aspekt der Risikokommunikation an dieser Stelle nicht näher eingegangen werden.

Neben dem Format (Wahrscheinlichkeiten vs. natürliche Häufigkeiten) und der Bezugsgröße (absolute vs. relative Risiken), spielt es auch eine große Rolle, ob die statistischen Informationen in Bezug auf Gesundheit oder Krankheit kommuniziert werden. Diese Dimensionen lassen sich kombinieren. „Ihre Wahrscheinlichkeit, an Brustkrebs zu sterben, kann durch regelmäßige Teilnahme am Mammographie-Screening um 25% reduziert werden“ hat eine vollkommen andere psychologische Auswirkung als die Aussage „Ohne Teilnahme überleben 996, mit Teilnahme überleben 997 von 1.000 Frauen (die nächsten zehn Jahre, oder sie sterben an anderen Ursachen als Brustkrebs)“ (22). Während im letzten Fall Gesundheit als der Normalzustand erscheint, werden mit der ersten Aussage Ängste geweckt, denen dann die medizinischen Errungenschaften als Rettungsanker entgegen geworfen werden. Diese Gepflogenheit birgt jedoch die Gefahr in sich, dass sie eine Illusion der Absicherung nährt. Dazu Johannes Schmidt, Gründer der Stiftung *Paracelsus heute*: „Leichtgläubig haben wir uns in die Illusionen des medizintechnischen Fortschritts verwickeln lassen: Aller Verstandesbegabung

zum Trotz glauben wir daran, dass uns die moderne Technologie, die komplizierten Apparate und teuren Untersuchungen vor Krankheit und Leid zu bewahren vermöchten. Die Zahl der Behandlungen nimmt kontinuierlich zu, die Kosten steigen ungebremsst. Dabei ist die teuer mit Geld, Zeit und Ängsten erkaufte Sicherheit nur scheinbar.“ (24) Die Art und Weise, wie Risiken kommuniziert werden, kann je nach Intention dazu eingesetzt werden, diese Entwicklung voranzutreiben oder ihr entgegenzutreten.

Manchen Experten, und dies gilt sicher nicht nur für Mediziner, beschleicht nun vielleicht ein ungutes Gefühl bei der Vorstellung, sie sollten Klienten künftig explizit Unsicherheiten vermitteln. Erwarten die Patienten nicht, dass der Arzt Sicherheit verspricht und damit auch ausstrahlt? Unterminiert die ausführliche Darstellung von Wahrscheinlichkeiten, Fehlerquellen, diversen Einflussfaktoren und davon abhängigen Nutzenquantifizierungen nicht seine Autorität? Sicher ist diese Befürchtung nicht unbegründet. Akzeptiert man aber die Notwendigkeit einer adäquaten Patientenaufklärung, und antizipiert man, dass ein mehr oder weniger großer Anteil an Patienten dies vom Arzt sogar erwartet, so kommt man zu der Überlegung, dass gerade ein solcher offener Umgang mit Unsicherheit ein Vertrauensverhältnis begründen kann. In einigen Bereichen der Genetikberatung, gibt es bereits langjährige Erfahrungen mit den „challenges of uncertainty“ bei der Risikokommunikation (25, 26).

Darüber hinaus gilt es, ganz generell zu prüfen, wie denn der Wissensstand in der eigenen (aber auch in anderen) Disziplinen üblicherweise nach außen vertreten wird: Obwohl beispielsweise uneinheitliche Befunde und sich widersprechende Theorien unter Wissenschaftlern wohl bekannt sind, bekommt die Öffentlichkeit häufig „die Meinung des Experten“ zu hören, die von alle

dem nichts widerspiegelt. Dies fördert die irri-ge Annahme, dass Expertenaussagen automatisch mit Sicherheitsgarantien versehen sind. Ein nüchterner und unvoreingenommener Umgang der Öffentlichkeit mit Unsicherheit und Risiko ist nur zu erwarten, wenn auch Experten dieses Thema nicht scheuen.

Schlussfolgerungen

Der Schriftsteller H. G. Wells hat einmal gesagt, dass eine funktionierende Demokratie mündige Bürger braucht, die neben Lesen und Schreiben auch statistisches Denken gelernt haben. Was Letzteres betrifft, sind wir bestenfalls auf halbem Wege. Die hier erarbeiteten Erkenntnisse und Methoden können helfen, auf diesem Weg schneller voranzukommen. Das generelle Argument lautet zusammengefasst: Information braucht Repräsentation. Kognitive Strategien oder Algorithmen sind an bestimmte Repräsentationen adaptiert, beispielsweise an diejenigen, welche während der menschlichen Evolution relativ stabil waren. Wenn sich die Repräsentation in der heutigen Umwelt, hier durch Entwicklungen in der Mathematik, verändert hat, kann dies bestimmte Defizite in der Einschätzung von Risiken herbeiführen, bei Laien wie bei Experten. Diese Defizite sind künstlich geschaffen und auch wieder überwindbar. Die Repräsentation von Information in natürlichen Häufigkeiten ist eine einfache und praktikable Methode, von der Ärzte wie auch Laien bei der Interpretation und Kommunikation von Unsicherheiten und Risiken profitieren können. 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.



Abstract

How to improve the communication and understanding of medical test results

Interpreting medical test results demands statistical reasoning on the part of doctors: How great is the probability of falsely diagnosing an illness based on a positive test result? This probability can be determined with the assistance of Bayes's Rule. Several studies show, however, that doctors often experience problems with these kinds of Bayesian inferences. We demonstrate that doctors' judgments can be considerably improved when numerical information is presented in a form easily accessible to the ways humans process information. This is not the case with the utilization of applied probabilities that has become customary, but is the case when the problem is presented in terms of "natural frequencies" that result from the counting of observed isolated cases in a natural environment. In a series of studies, we varied the representation format of the relevant statistical information. If the information was not presented in the form of probabilities or percentages but simply in natural frequencies, medical experts, as well as laymen, were able to improve their judgments significantly. And dealing with probabilities and percentages can also be easily learned: Two training programs, which showed how probabilities can be translated into natural frequencies, placed the participants in the position of being able to obtain very good results when solving these tasks. Finally, we discuss the impact of a comprehensible risk and utility communication on the doctor-patient relationship.

Key words: Risk communication, Bayes theorem, Representation of information, statistical reasoning, medical training and postgraduate medical training

Danksagung: Wir danken der Deutschen Forschungsgemeinschaft (Ho 1847/1-2) für finanzielle Unterstützung.

Literatur

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

2. 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
3. Windeler J, Köbberling J (1986) Empirische Untersuchung zur Einschätzung diagnostischer Verfahren am Beispiel des Haemocult-Tests. *Klin Wschr* 64: 1106-1112
4. Casscells W, Schoenberger A, Grayboys T (1978) Interpretation by physicians of clinical laboratory results. *N Engl J Med* 299: 999-1001
5. Gigerenzer G, Hoffrage U (1995) How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol Rev* 102: 684-704
6. Daston LJ (1988) *Classical probability in the Enlightenment*. Princeton: Princeton University Press
7. 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: Cambridge University Press
8. Gallistel CR, Gelman R (1992) Preverbal and verbal counting and computation. *Cognition* 44: 43-74
9. Jonides J, Jones CM (1992) Direct coding for frequency of occurrence. *J Exp Psychol: Learn Mem & Cogn* 18: 368-378
10. Real LA (1991) Animal choice behavior and the evolution of cognitive architecture. *Science* 253: 980-986
11. Sedlmeier P, Hertwig R, Gigerenzer G (1998) Are judgments of the positional frequencies of letters systematically biased due to availability? *J Exp Psychol: Learn Mem & Cogn* 24: 754-770
12. Christensen-Szalanski JJ, Bushyhead JB (1981) Physicians' use of probabilistic information in a real clinical setting. *J Exp Psychol: Hum Percept & Perform* 7: 928-935
13. Kleiter GD (1994) Natural sampling: Rationality without base rates. In: Fischer GH, Laming D (eds) *Contributions to mathematical psychology, psychometrics, and methodology*. New York: Springer, pp. 375-388
14. Hoffrage U, Gigerenzer G (1998) Using natural frequencies to improve diagnostic inferences. *Acad Med* 73: 538-540
15. Stine GJ (1996) Acquired immune deficiency syndrome. Biological, medical, social and legal issues. Englewood Cliffs: Prentice Hall
16. Gigerenzer G, Hoffrage U, Ebert A (1998) AIDS counselling for low-risk clients. *AIDS Care* 10: 197-211
17. Sedlmeier P, Gigerenzer G (in press) Teaching Bayesian reasoning in less than two hours. *J Exp Psychol: General*
18. Sedlmeier P (1997) *BasicBayes: A tutor system for simple Bayesian inference*. *Behav Res Methods, Instr & Comput* 29: 328-336
19. Kurzenhäuser S, Hoffrage U (2000) Teaching Bayesian reasoning: Evaluation of a classroom tutorial. Manuscript in preparation
20. Elmore JG, Barton MB, Mocerri VM, Polk S, Arena PJ, Fletcher SW (1998) Ten-year risk of false positive screening mammograms and clinical breast examination. *N Engl J Med* 338: 1089-1096
21. Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF (1991) Psychological and behavioral implications of abnormal mammograms. *Ann Intern Med* 114: 657-661
22. Mühlhauser I, Hölldtke B (1999) Mammographie-Screening - Darstellung der wissenschaftlichen Evidenz-Grundlage zur Kommunikation mit der Frau. *Arznei-telegramm* 10/99: 101-108
23. Schmidt J (1999) Stell Dir vor es ist Mammografie *Ars Medici* 4/99: 240-244
24. Aus einer Informationsbroschüre der Stiftung Paracelsus heute (Ilgenweidstr 3, CH-8840 Einsiedeln. <http://www.paracelsus-heute.ch>)
25. Bortoff JL, Ratner PA, Johnson JL, Lovato CY, Joab SA (1998) Communicating cancer risk information: The challenges of uncertainty. *Patient Educ Couns* 33: 67-81
26. Hallowell N, Statham H, Murton F, Green J, Richards M (1997) "Talking about chance": The presentation of risk information during genetic counseling for breast and ovarian cancer. *J Genet Couns* 6: 269-286

Korrespondenzadresse:

Dr. Ulrich Hoffrage, Max-Planck-Institut für Bildungsforschung, Lentzeallee 94, 14195 Berlin
Tel.: 030-82406-273; Fax: 030-82406-394
e-mail: hoffrage@mpib-berlin.mpg.de

Ankündigung

12. Deutscher interdisziplinärer Schmerzkongress

Vom 15. bis 17. März 2001

Ort: Frankfurt, CCM

Ankunft: SCHMERZtherapeutisches Kolloquium e.V., Geschäftsstelle, Hainstr.2, 61476 Kronberg.

Tel.: 06173/9556 - 0, Fax: 06173/9556-14, E-mail: stk.zentrale@t-online.de



Urban & Fischer Verlag

Z. ärztl. Fortbild. Qual.sich. (ZaeFQ) (2000) 94: 713-719
<http://www.urbanfischer.de/journals/zaefq>

Editorial

Vereinzelt gibt es ihn noch: den Kranken, der zum Arzt geht und ohne zu hinterfragen befolgt, was die Autorität im weißen Kittel ihm sagt. Aber für eine zunehmende Anzahl von Menschen trifft dieses einfache Verhaltensmuster nicht mehr zu – sie wollen über diagnostische und therapeutische Maßnahmen aufgeklärt werden, die jeweiligen Implikationen verstehen und letztlich auch mitentscheiden. Getrieben wird diese Entwicklung durch mindestens zwei Faktoren. Zum einen macht sich in vielen gesellschaftlichen Bereichen eine Tendenz zur Individualisierung und Autonomie bemerkbar. Und so verwandelt sich auch der Patient vom passiven Leistungsempfänger zum mündigen und gleichberechtigten Partner im Gesundheitswesen, der das Gespräch mit dem Arzt sucht und diesem zuweilen sogar schon ausgedruckte Webseiten zu seiner Krankheit vorlegt – mit der Bitte um Kommentar. Zum anderen muss eine in weiten Bevölkerungskreisen verbreitete Unzufriedenheit mit dem derzeitigen Gesundheitswesen konstatiert werden. So wenden zum Beispiel immer mehr Menschen der Schulmedizin den Rücken zu und suchen ihr Heil bei der Alternativmedizin. Die Erfolge der Schulmedizin werden zwar nicht bestritten, aber es wird zunehmend hinterfragt, ob ihre diagnostischen und therapeutischen Maßnahmen immer in einem angemessenen Verhältnis zum Informationsgewinn bzw. zum Nutzen stehen.

Die evidenzbasierte Medizin könnte sowohl dem Bedürfnis nach Mündigkeit entsprechen, als auch eine Therapie für die Unzufriedenheit mit dem Gesundheitswesen sein. Dies kann sie allerdings nur, wenn sie nicht verkürzt als Meta-Analyse von randomisierten Studien verstanden wird, sondern wenn eines ihrer häufig übersehenen Mittel ebenfalls berücksichtigt wird: die Entscheidungsanalyse. Der Informationsgewinn eines medizinischen Tests oder der Nutzen einer therapeutischen Maßnahme mag statistisch signifikant sein, aber dies bedeutet noch lange nicht, dass der Einsatz im Einzelfall immer sinnvoll ist. Und spätestens hier wird offenbar, welche entscheidende Rolle der Kommunikation zukommt: Denn um über diesen Einsatz zu entscheiden, muss die Bedeutung des Testergebnisses bzw. das Ausmaß des Nutzens verstanden werden, und genau dies ist oftmals ein Kommunikationsproblem. Darüber hinaus müssen die persönlichen Werte und Sichtweisen des Patienten bei derartigen Entscheidungen berücksichtigt werden und auch hierzu ist das vertrauensvolle Gespräch mit dem Arzt unerlässlich.

So ist es zu begrüßen, dass die *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung* dem Thema „Qualität und Kommunikation“ ein Schwerpunktheft widmet. Die hier zusammengetragenen Arbeiten beleuchten verschiedene Aspekte. In dem Beitrag von Ortmann werden Umfrageergebnisse berichtet, die zeigen, wie wichtig das Gespräch und das individuelle Eingehen auf den Patienten sein kann – insbesondere wenn dieser mit funktionellen Störungen zum Arzt kommt, das heißt mit Krankheitsbildern, bei denen die klassische Schulmedizin außer oft unnützen Abklärungsritualen nur wenig zu bieten hat. Empathie ist eine Sache, Informationsvermittlung eine andere. Beides ist wichtig. In den beiden folgenden Beiträgen geht es um die verständliche Kommunikation von statistischer Information – bei Hoffrage et al. bezüglich des prädiktiven Werts eines diagnostischen Tests, bei Mühlhauser und Höldke bezüglich des Nutzens eines Mammographie-Screenings. Abschließend wenden sich Paschen und Bastek dem zentralen Medium der Kommunikation zu – der Sprache – und diskutieren die Vorteile, aber auch die Gefahren einer Fachsprache im Qualitätsmanagement.

Ich wünsche allen Lesern eine anregende Lektüre.

U. Hoffrage, Berlin