

Bekannte Narrative in der Statistik

Christel Weiß

Medizinische Fakultät Mannheim der Universität Heidelberg

Zusammenfassung Narrative sind Erzählungen oder Mythen, Legenden, Überlieferungen oder Zitate, Witze oder Anekdoten. Kennzeichnend für Narrative ist, dass weniger deren Inhalt oder Wahrheitsgehalt von Bedeutung ist als vielmehr die Art, wie sie dargeboten werden und das, was sie bewirken sollen. Sie sind nicht dazu angetan, Zuhörer sachlich zu informieren oder zu belehren; vielmehr sollen sie Emotionen auslösen. Diese reichen von unterhaltsamer Belustigung über kritiklose Zustimmung und ungläubiges Staunen bis hin zu spontanem Ärger. Dadurch erhöhen Narrative die Aufmerksamkeit bei den Zuhörern und tragen im Idealfall dazu bei, dass sie sich aktiv mit dem dargebotenen Stoff auseinandersetzen und ihn längere Zeit in Erinnerung behalten.

Das Erscheinungsbild der Statistik als wissenschaftliche Disziplin ist geprägt von zahlreichen Narrativen auf unterschiedlichen Ebenen, von unzähligen Witzen und Vorurteilen, die dem Ruf dieses Fachs vorausseilen. Über Statistiker und Mathematiker kursieren diverse Anekdoten, die einerseits deren Beiträge für die Statistik würdigen und andererseits auch ihre menschliche Seite beleuchten. Dieser Beitrag setzt sich damit auseinander, wie Narrative unterschiedlichster Art die Statistik als Wissenschaft und deren Ansehen in der Öffentlichkeit beeinflussen. Anhand von zehn Beispielen werden klassische Narrative dargelegt und diskutiert.

1 Einleitung

„Traue keiner Statistik, die du nicht selbst gefälscht hast.“ Dieses Zitat, das dem britischen Staatsmann Winston Churchill (1874–1965) zugeschrieben wird, dürfte das populärste Bonmot sein, das wie kein anderes die gängige Meinung über Statistik widerspiegelt. Ob Churchill diese Worte jemals von sich gegeben hat, ist keineswegs belegt. Dennoch halten sie sich hartnäckig.

Dieser Satz versinnbildlicht die charakteristischen Eigenschaften eines Narrativen: Er ist allgemein bekannt, in weiten Kreisen akzeptiert und bedarf keiner weiteren Erläuterungen. Er bestätigt, was viele Menschen zu wissen meinen: Auf Statistiken kann man sich nicht verlassen! Nach allgemeinem Konsens sind bei der Betrachtung



Abbildung 1: Winston Churchill. Angeblicher Urheber des Narrativs: „Traue keiner Statistik, die du nicht selbst gefälscht hast“. Bildquelle: Wikipedia

von Zahlenwerken oder Prognosen, die mit statistischen Methoden hergeleitet wurden, Misstrauen und Zweifel angesagt. Der Wahrheitsgehalt eines Narrativs wird als evident angesehen, ohne hinterfragt zu werden. Dessen Zweck besteht ja gerade *nicht* darin, althergebrachte Weisheiten in Frage zu stellen oder kritische Diskussionen anzuregen. Vielmehr soll es ein Gefühl des sozialen Zusammenhalts vermitteln, weil der Erzähler eines Narrativs selbstverständlich voraussetzt, dass all seine Zuhörer in stiller Übereinstimmung seine Meinung teilen.

Der Begriff „Narrativ“ ist unscharf definiert. Seine etymologischen Wurzeln findet man im lateinischen Verb „narrare“ (erzählen). Das Oxford English Dictionary definiert ein „narrative“ als „eine Erzählung oder Darstellung, die benutzt wird, um eine Gesellschaft oder historische Periode zu erklären oder zu rechtfertigen“. Diese Definition geht zurück auf

den französischen Philosophen Jean-François Lyotard (1924–1998). Er unterscheidet szientifisches Wissen, das auf wissenschaftlicher Evidenz basiert, von narrativem Wissen, das auf traditionellen Überlieferungen in Form von Erzählungen und Anekdoten gründet.

Obwohl die Begriffe „Geschichte“, „Erzählung“ und „Narrativ“ oft synonym verwendet werden, gibt es dennoch feine Unterschiede. Eine Geschichte soll Wissen vermitteln. Dabei kommt es in erster Linie auf den Inhalt an, weniger auf stilistische Elemente. Die Erzählung ist eine literarische Gattung, die die Zuhörer oder Leser unterhalten soll. Hier sind vor allem der Stil und die Technik, mit der der Erzähler vorgetragen wird, wichtig. Bei einem Narrativ ist der zu erreichende Zweck relevant (vgl. El Quassil & Karig, 2021).

Wie können in einer spröden Disziplin, als die die Statistik im Allgemeinen angesehen wird, Narrative entstehen? Einerseits ist die Statistik ein Teilgebiet der Mathematik – also einer Wissenschaft, deren Strukturen klar definiert sind und deren Aussagen nach festgelegten Regeln durch logisches Denken hergeleitet werden. Mathematische Sätze sind wissenschaftlich evident im Sinne von „erwiesen“; auch statistische Algorithmen sind klar nachvollziehbar. Vertreter von empirischen Wissenschaften wie der Medizin wenden diese Methoden an, um Daten zu analysieren und so zu neuen Erkenntnissen zu gelangen. Auch in anderen Disziplinen wie der Volkszählung, Meinungsforschung, Spieltheorie, im Versicherungswesen, in der Meteorologie und Politologie haben sich statistische Methoden als unentbehrliches Hilfsmittel etabliert, um Strukturen zu beschreiben, Prozesse zu erklären oder Prognosen zu erstellen.

Andererseits sind die speziellen Analysemethoden zur Aufbereitung eines Datensatzes nicht eindeutig vorgegeben. Die Ergebnisse lassen zudem unterschiedliche Interpretationen zu; die daraus zu ziehenden Konsequenzen liegen keineswegs klar auf der Hand. Dies überfordert zuweilen Anwender, denen es obliegt, adäquate Methoden zur Datenanalyse auszuwählen, als auch Konsumenten, die quasi rund um die Uhr über diverse Kommunikationsmedien mit Durchschnittswerten, grafischen Darstellungen, Wahrscheinlichkeiten oder Umfrageergebnissen konfrontiert werden. In aller Regel werden diese Informationen von einem breiten Publikum ohne tiefgründiges Nachfragen zur Kenntnis genommen. Auch Wissenschaftler, die statistische Methoden anwenden, unterliegen hin und wieder allzu menschlichen Schwächen und interpretieren ein Ergebnis nach ihren Wunschvorstellungen – ohne sich darüber Gedanken zu machen, ob ihre Schlussfolgerungen stimmig sind und inwieweit das Ergebnis praktisch relevant ist. Derlei Fehlinterpretationen können dazu führen, dass Missverständnisse entstehen, die als falsche Thesen oder unzulässige Interpretationen – häufig mit Narrativen pseudowissenschaftlich untermauert – verbreitet werden.

2 Eine Auswahl von Narrativen aus der Statistik

2.1 Narrativ der Skeptiker: Mit Statistik lässt sich alles beweisen

Dieses bekannte Narrativ erinnert an den bereits zitierten Rat „Traue keiner Statistik, die du nicht selbst gefälscht hast.“ Zweifelsohne ist dieser Satz ironisch gemeint. Er suggeriert, dass Ergebnisse, die mit statistischen Methoden gewonnen wurden, beliebig interpretierbar seien. Außerdem unterstellt er Statistikern, dass sie unseriös arbeiten, oder dass sie unlautere Tricks anwenden, um die von ihnen gewünschten Ergebnisse zu erhalten. Die Herkunft der Daten, deren Korrektheit und Vollständigkeit lassen sich tatsächlich kaum überprüfen.

Fakt ist: Mit Statistik lässt sich *nichts* beweisen. Man kann zwar durch effiziente Analysen Effekte nachweisen (vorbehaltlich einer Irrtumswahrscheinlichkeit) oder Prognosemodelle erstellen. Allerdings besagt das Ergebnis einer statistischen Analyse nichts über dessen Ursachen und die daraus zu ziehenden Schlussfolgerungen, und ein statistisches Modell ist nur so realistisch, wie die Daten, auf denen es basiert, und die Annahmen, die ihm zugrunde liegen.

Worauf ist dieses oft zitierte, harsche Vorurteil zurückzuführen? Möglicherweise hat dies damit zu tun, dass sich manche Prognosen oder Umfrageergebnisse im Nachhinein als falsch herausstellten (was dann meist der Statistik angelastet wird), oder dass hin und wieder Ergebnisse statistischer Analysen verbreitet werden, deren Sinn sich nicht erschließt. Riesige Datenmengen mit einer Vielzahl von Variablen verleiten

geradezu dazu, eine Unmenge von Vergleichen durchzuführen – irgendwann wird man mit an Sicherheit grenzender Wahrscheinlichkeit einen Effekt aufspüren, der sich statistisch absichern lässt (obwohl er vielleicht in Wirklichkeit gar nicht existiert). So werden mitunter Aufsehen erregende Ergebnisse produziert, die zufällig zu stehen kommen, aber jeglicher Sinnhaftigkeit entbehren.

Eine weitere Ursache für dieses Narrativ ist dadurch gegeben, dass Statistiken sich nicht selbst erklären, wodurch Raum für diverse Interpretationsmöglichkeiten gegeben ist. Ein Beispiel möge dies veranschaulichen: In der Entbindungsstation eines Universitätsklinikums werden 45 % aller schwangeren Frauen durch einen Kaiserschnitt entbunden; in einer Klinik, die „sanfte Geburten“ propagiert, beträgt dieser Anteil weniger als 20 %. Voreingenommene Menschen neigen beim Anblick dieser Zahlen eventuell spontan dazu, lang gehegte Vorurteile bestätigt zu sehen: Damit sei der „Beweis“ erbracht, dass Ärzte in Unikliniken nur auf Apparatemedizin vertrauen, auf die Bedürfnisse der Frauen nicht eingehen oder gar aus monetärem Interesse handeln. Nichts davon belegen diese Zahlen! Man muss bedenken: In Universitätskliniken ist der Anteil der Risikoschwangerschaften weitaus höher als in anderen Einrichtungen. Es ist naheliegend, dass ein sehr großer Anteil der Frauen mit Risikoschwangerschaften per Kaiserschnitt entbunden werden. Dann wäre das Vorurteil widerlegt! Dies sind zwei grundsätzlich verschiedene Interpretationen derselben Zahlen.

Es ist wichtig, die Ursache eines Ergebnisses kritisch zu hinterfragen (ehe man sich über die Statistik mokiert) und darüber nachzudenken, ob es wirklich so sensationell ist, wie es dargeboten wird. Weitere Beispiele für kuriose Fehlschlüsse samt Hinweisen zu deren Entlarvung werden in den Büchern von Gigerenzer (2015), Gritzmann (2024) und Krämer (2015) unterhaltsam präsentiert.

2.2 Narrativ der Durchschnittsmenschen: Der Mittelwert ist ein typischer Wert

Zweifelsohne ist der Mittelwert das bekannteste Lagemaß. Jeder kennt ihn, er ist leicht zu berechnen und ermöglicht Orientierung. Häufig dient der Mittelwert als grobe Richtschnur zur Einschätzung, ob ein einzelner Wert eher als hoch oder als niedrig einzustufen ist: sei es der Intelligenzquotient, das Gehalt eines Arbeitnehmers, der Benzinverbrauch eines Autos oder der Laborwert eines Patienten.

Dennoch stoßen der Mittelwert und dessen Interpretation aus verschiedenen Gründen auf Irritationen. Der erste Grund liegt darin, dass ein Mittelwert häufig kein Wert der Datenreihe ist, aus der er berechnet wurde. Der US-amerikanische Präsident Franklin D. Roosevelt (1882–1945) formulierte seine Vorbehalte mit dem Hinweis:

„Laut Statistik haben ein Millionär und ein armer Kerl je eine halbe Million“. Ein anderes Beispiel: „Im Durchschnitt hat jeder Mensch einen Hoden und einen Eierstock“. Die Fertilitätsrate 1,35 (bezogen auf das Jahr 2023 in Deutschland) besagt, dass eine „Durchschnittsfrau“ im Laufe ihres Lebens 1,35 Kinder zur Welt bringt – obwohl dies für keine einzige Frau zutrifft. Diese Beispiele widersprechen der These, dass der Mittelwert ein typischer Wert sei, vehement.

Ein weiterer Grund, weshalb der Mittelwert zuweilen mit Argwohn bedacht wird, ist die Tatsache, dass er sehr stark von Ausreißern beeinflusst werden kann, insbesondere bei einem geringen Stichprobenumfang. In diesen Fällen würde er ein verzerrtes Bild der tatsächlichen Verhältnisse wiedergeben. Wenn beispielsweise die Gehälter von vier Angestellten 5 000, 5 000, 6 000 und 9 000 Euro betragen und der Chef mit 20 000 Euro pro Monat entlohnt wird, ergibt sich rechnerisch ein Durchschnittsgehalt von 9 000 Euro, obwohl drei von fünf Gehältern deutlich geringer ausfallen. Der Median (das ist der Wert in der Mitte der nach der Größe sortierten Datenreihe) von 6 000 Euro oder der Modus (auch Modalwert genannt; das ist der Wert, der am häufigsten in der Stichprobe vertreten ist) in Höhe von 5 000 Euro geben die tatsächlichen Verhältnisse realistischer wieder.

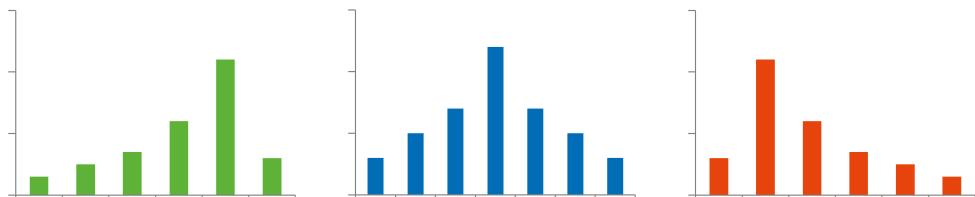


Abbildung 2: Eingipelige Verteilungen, von links nach rechts: rechtsgipflig (linksschief, Mittelwert < Median), symmetrisch (Mittelwert = Median), linksgipflig (rechtsschief, Mittelwert > Median). Selbst erstellte Diagramme

Ein dritter Grund: Bei manchen Verteilungen ist der Mittelwert als Lagemaß generell ungeeignet. Ein Beispiel stellt die Lebenserwartung im Mittelalter dar, die bei durchschnittlich etwa 40 Jahren lag. Jedoch starb damals kaum ein Mensch im Alter von 40 Jahren. Bis zu 40 % der Menschen überlebten das Säuglings- oder Kleinkindalter nicht; viele starben in jungen Jahren an Krankheiten oder in Kriegen. Sobald ein Individuum jedoch diese kritischen Jahre überlebt hatte, bestand die Chance, ein höheres Alter von 60 oder mehr Jahren zu erreichen. Tatsächlich handelt es sich bei der Lebenserwartung im Mittelalter um eine bimodale (U-förmige) Verteilung mit zwei „Gipfeln“ im jungen und im höheren Lebensalter.

Auch bei schiefen Verteilungen ist der Mittelwert nur bedingt geeignet. In der Medizin sind viele Merkmale (etwa physiologische Parameter oder Laborwerte) links-

gipflig verteilt mit einem Ausläufer am rechten Rand, der durch einige extrem hohe Werte verursacht wird. Diese hohen Werte führen dazu, dass der Mittelwert größer ist als der Median. Bei rechtsgipfligen Verteilungen ist der Mittelwert kleiner als der Median; ein Beispiel stellt die Schwangerschaftsdauer dar. Einige extrem hohe bzw. geringe Werte ziehen den Mittelwert nach oben bzw. nach unten, sodass er bei schießen Verteilungen nicht ohne Weiteres als „typisch“ bezeichnet werden kann.

Zuletzt sei auf die Problematik bei ordinal skalierten Merkmalen hingewiesen. Hier könnte der Mittelwert zu unsinnigen Schlussfolgerungen verleiten. Wenn beispielsweise der Therapieerfolg im Rahmen einer klinischen Studie mit Werten von 1 (vollständige Heilung) über 3 (Zustand unverändert) bis 5 (Patient verstorben) beurteilt wird, widerstrebt es dem gesunden Menschenverstand, den „mittleren Therapieerfolg“ zu berechnen und zu behaupten, 3 (unverändert) sei der Mittelwert zwischen 1 (vollständige Heilung) und 5 (verstorben). Woran liegt das? Der Unterschied zwischen „verstorben“ und „unverändert“ ist wesentlich gravierender als der Unterschied zwischen „unverändert“ und „geheilt“. Äquidistanz (gleichwertiger Abstand) ist demnach nicht gegeben – was für die Berechnung eines Mittelwerts eigentlich vorausgesetzt wird. Die Angabe des Medians ermöglicht dagegen eine sinnvolle Interpretation. Ein Median von 3 besagt: Bei der Hälfte der Patienten hat sich der Zustand verbessert, bei der anderen Hälfte verschlechtert.

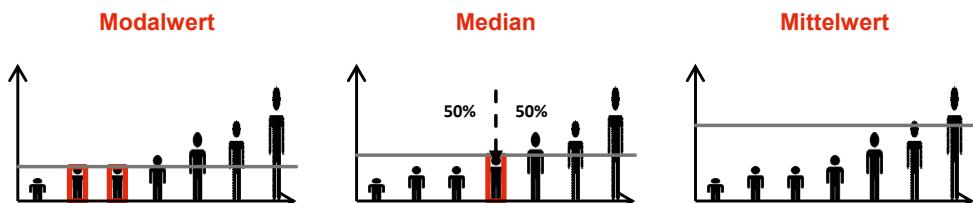


Abbildung 3: Vergleich Mittelwert, Median und Modus. Der Ausreißer am rechten Rand bewirkt, dass der Mittelwert größer ist als der Median. Selbst erstellte Diagramme

Ist der Mittelwert unbrauchbar? Keineswegs! Bei eingipfligen, symmetrischen Verteilungen (wie etwa der Normalverteilung) ist der Mittelwert ein geeignetes Lagemaß. Bei normalverteilten Daten liegen etwa zwei Drittel aller Werte in unmittelbarer Umgebung des Mittelwerts (höchstens eine Standardabweichung entfernt). Bei anderen Verteilungen sollte sich der Anwender fragen, ob andere oder zusätzliche Lagemaße sinnvoll sind. Bei schief verteilten und ordinal-skalierten Daten ist der Median besser als der Mittelwert geeignet; bei U-förmigen oder mehrgipfligen Verteilungen bieten sich Modalwerte an. Um die Form einer Verteilung zu beurteilen, eignet sich die Darstellung in Form eines Histogramms. Es liegt am Autor einer Publikation, geeignete Lagemaße zu präsentieren, und an den Lesern, diese adäquat zu interpretieren.

2.3 Narrativ der Nörgler: Seltene Ereignisse sind verdächtig

„Betrug! Schiebung!“ hörte man, als Rut Brandt auf einem Bundespresseball in den 1960er Jahren in einem Lotteriespiel ein Auto gewann. Bei mehreren Hundert Teilnehmern, die ein Los erstanden hatten, schien es extrem unwahrscheinlich zu sein, dass ausgerechnet die Gattin des Regierenden Bürgermeisters von Berlin den Hauptgewinn einfuhr (vgl. Brandt, 1992). Einige witterten ein abgekartetes Spiel und forderten Frau Brandt auf, das Auto zurückzugeben. Waren die Vorwürfe gerechtfertigt? Nichts spricht dafür, wenn man bedenkt, dass jeder Lotterteilnehmer exakt die gleiche Chance auf den Hauptgewinn hatte. Wenn eine weniger prominente Person gewonnen hätte, wäre diese ebenso überrascht gewesen wie Rut Brandt – auch wenn es keinen lautstarken Protest gegeben hätte. Frau Brandt behielt übrigens ihren Gewinn und schenkte ihr eigenes Auto einer sozialen Einrichtung.

Ob ein Ereignis als extrem unwahrscheinlich eingestuft wird, hängt nicht zuletzt vom Betrachter ab. Wenn bei der wöchentlichen Ziehung der Lottozahlen die Gewinnzahlen 1 bis 6 resultierten, würden viele sofort Zweifel äußern: Ging das mit rechten Dingen zu? War etwa die Lostrommel manipuliert? Die Zahlen 9, 11, 19, 23, 34 und 40 scheinen dagegen unverdächtig zu sein – außer für einen Menschen, der am 9. November 1934 um 23:40 Uhr geboren wurde. Dieser würde vielleicht glauben, dass eine höhere Macht im Spiel war. Schließlich ist die Wahrscheinlichkeit bei beiden Konstellationen mit etwa 1 zu 14 Millionen extrem gering.

Man muss sich jedoch klarmachen: Die Wahrscheinlichkeit ist für jede andere Zahlenkombination ebenso gering. In Deutschland spielen etwa 7 Millionen Menschen regelmäßig Lotto. Demnach ist zu erwarten (unter der Voraussetzung, dass alle Zahlenkombinationen etwa gleich häufig getippt werden), dass alle zwei Wochen ein Spieler zu einem glücklichen Gewinner wird. Keine 6er-Kombination hat höhere oder niedrigere Chancen als jede beliebige andere Kombination.

Ist es demnach egal, welche Zahlen auf dem Schein angekreuzt werden? Nicht unbedingt! Wenn zahlreiche Spieler dieselbe Zahlenkombination wählen, müssen diese sich den Gewinn teilen. Deshalb ein kleiner Tipp: Vermeiden Sie Muster auf Ihrem Lottoschein. Verzichten Sie auf die Zahlen 19 und 20, die als Teile von Geburtsjahren besonders häufig angekreuzt werden. Wenn Sie großes Glück haben, können Sie dann den Gewinn für sich allein behalten.

2.4 Narrativ für Politiker: Prognosen sind unzuverlässig

„Wahlen werden durch die Wähler, nicht durch Umfragen entschieden.“ Wie oft hört man diesen Satz vor politischen Wahlen, insbesondere von Vertretern der Parteien, denen ein schlechtes Wahlergebnis prognostiziert wird. Sie verweisen gerne auf Beispiele, bei denen die Wahlen anders ausgingen als die vorangegangenen Umfragen hatten erwarten ließen. Man denke etwa an die Bundestagswahl 2002. Lange Zeit sah die CDU/CSU wie der sichere Gewinner aus; dennoch kam es damals zur Fortsetzung der rot-grünen Koalition. Der Grund war eine Flutkatastrophe wenige Wochen vor der Wahl, die dazu geführt hatte, dass viele Wähler kurzfristig ihre Meinung änderten und – anders als ursprünglich geplant – einer der an der Regierung beteiligten Parteien ihre Stimme gaben. Auch der Ausgang der Wahl Donald Trumps im Jahr 2016 zum US-amerikanischen Präsidenten war überraschend. Bis kurz vor dem Wahltag galt es als gesichert, dass seine Gegenkandidatin Hillary Clinton gewinnen würde. Dies lag daran, dass die Ergebnisse der Umfragen teilweise falsch interpretiert wurden, und nicht zuletzt am komplizierten US-amerikanischen Wahlsystem.

Diese Geschehnisse mögen bei manchem Kandidaten die Hoffnung nähren, dass sich trotz schlechter Umfrageergebnisse die Dinge zum Besseren wenden. Doch diese Hoffnung ist trügerisch. Um die Verlässlichkeit von Wahlprognosen zu beurteilen, müsste man alle relevanten Umfragen in Augenschein nehmen und prüfen, wie viele Prognosen aus welchen Gründen sich im Nachhinein als falsch erwiesen. Es ist offenkundig, dass bei Umfragen vor einer Wahl nicht alle Befragten wahrheitsgemäß antworten, ihre Teilnahme verweigern oder dass sie ihre Meinung aufgrund nicht vorhersehbarer Ereignisse oder gar aufgrund vorangegangener Umfrageergebnisse kurzfristig ändern. Je nach Blickwinkel kann das dazu führen, dass das endgültige Wahlergebnis besser oder schlechter ist als erhofft. Es wäre aber naiv zu glauben, dass die Stimmung kurz vor einer Wahl kippt und der Wahlausgang alle Umfragen Lügen strafft – nur weil irgendwann einmal in der Vergangenheit eine Wahl anders ausgegangen war als Umfrageergebnisse hatten vermuten lassen. Das ist eher die Ausnahme als die Regel.

2.5 Narrativ der Diagnostiker: Befunde sind verlässlich

„Sie sind HIV-positiv“ oder „Bei Ihnen besteht der Verdacht auf ein Karzinom“. Solche Befunde werden die getestete Person stark beunruhigen. Bei negativen Befunden wird sie dagegen erleichtert aufatmen. Doch inwieweit sind diese Ergebnisse verlässlich?

Das Ziel eines diagnostischen Tests besteht darin, valide Informationen bezüglich des Gesundheitsstatus eines Patienten zu gewinnen. Ein Arzt führt einen Test

durch, um einen Verdacht zu bestätigen oder um eine bestimmte Krankheit auszuschließen. Im ersten Fall erwartet er einen positiven, im zweiten einen negativen Befund. Die Güte eines diagnostischen Tests wird durch zwei Kriterien bestimmt: der Sensitivität und der Spezifität. Die Sensitivität ist die Wahrscheinlichkeit, dass der Test eine vorhandene Krankheit nachweist; unter der Spezifität versteht man die Wahrscheinlichkeit, dass eine nicht erkrankte Person einen negativen Befund erhält. Falsche Ergebnisse sind freilich nicht ausgeschlossen. Wenn der Test eine vorhandene Krankheit nicht erkennt, ergibt sich ein falsch negativer Befund. Wenn er auf etwas anderes als die Krankheit reagiert, entsteht ein falsch positiver Befund.

Falsche Befunde können fatale Auswirkungen haben. Bei einem HIV-Test kann ein falsch negativer Befund dazu verleiten, dass die getestete Person andere unwissentlich ansteckt. Ein falsch positiver Befund würde zu einer immensen psychischen Belastung führen. HIV-Tests gelten jedoch als wenig fehleranfällig, wenn sie sachgerecht durchgeführt werden: Die Sensitivität beträgt 99,9 %, die Spezifität sogar 99,99 %. Demnach erhält nur eine von 1 000 infizierten Personen einen falsch-negativen und nur eine von 10 000 nicht infizierten Personen einen falsch-positiven Befund. Diese Zahlen suggerieren, dass falsche Befunde zwar theoretisch möglich, praktisch aber quasi ausgeschlossen sind. Für einen Arzt, der einen solchen Test durchführt, und die getestete Person steht jedoch eine andere Frage im Vordergrund: Wie hoch ist die Wahrscheinlichkeit, dass der erhaltene Befund korrekt ist? Mit welcher Wahrscheinlichkeit liegt bei einem positiven Befund wirklich eine Infektion vor, und mit welcher Wahrscheinlichkeit lässt sich bei einem negativen Befund eine Infektion ausschließen? Dabei handelt es sich um den positiven bzw. den negativen Vorhersagewert. Diese ergeben sich nicht direkt aus der Sensitivität und der Spezifität, sondern sind von der Prävalenz abhängig.

Wir betrachten nun folgendes Szenario: In einer Gruppe von Menschen, die keinem Risiko ausgesetzt sind, beträgt die Prävalenz 0,01 %. Das heißt: Nur einer von 10 000 ist infiziert. Wenn der Test an 10 000 Personen angewandt wird, sind die Häufigkeiten in Tabelle 1 (links) zu erwarten. Die infizierte Person wird einen positiven Befund erhalten. Bei den 9 999 nicht infizierten Personen wird sich jedoch wegen der Falsch-positiv-Rate von 0,01% ebenfalls ein positiver Befund ergeben (auf ganze Zahlen gerundet). Dies ist der einzige falsche Befund! Bezogen auf 10 000 Testungen ist dieser Anteil minimal; bezogen auf die beiden positiven Befunde sind dies jedoch 50%! Weil nur einer von zwei positiven Befunden korrekt ist, liegt der positive Vorhersagewert bei bescheidenen 50% – trotz der hohen Werte für Sensitivität und Spezifität. Der negative Vorhersagewert liegt dagegen bei 100 %. Freilich ist ein positiver Befund ein Alarmsignal: Dadurch hat sich die Prävalenz von 0,01% (vor dem Test) auf 50% (nach dem Test) erhöht. Aber für eine sichere Diagnose ist er unzureichend.

Anders stellt sich die Situation in einer Hochrisikogruppe mit einer Prävalenz von 20 % dar. Hier sind beide Vorhersagewerte mit 99,95 % (positiver) bzw. 99,98 % (negativer) sehr hoch (Tabelle 1, rechts).

Tabelle 1: Zu erwartende Häufigkeiten bei einem HIV-Test mit dem Prävalenzen 0,1% (links) und 10% (rechts) bei einer Sensitivität von 99,9% und einer Spezifität von 99,99%

	Prävalenz 0,01%			Prävalenz 20%		
	infiziert	nicht infiziert	Summe	infiziert	nicht infiziert	Summe
Befund positiv	1	1	2	1.998	1	1.999
Befund negativ	0	9.998	9.998	2	7.999	8.001
Summe	1	9.999	10.000	2.000	8.000	10.000

Aus diesen Überlegungen geht hervor: Ein Arzt muss bei der Bewertung eines positiven Befundes wissen, ob und welcher Risikogruppe die getestete Person angehört und wie hoch die Prävalenz einzustufen ist. Generell gilt: Bei geringer Prävalenz (wie sie etwa bei Screenings üblich ist) ist der positive Vorhersagewert in aller Regel sehr gering. Auch wenn die Prävalenz im Einzelfall nicht ohne Weiteres quantifizierbar ist, sollte der Arzt weitere Beobachtungen (die sich etwa aus der Anamnese ergeben) heranziehen und auch auf seine Erfahrung vertrauen, ehe er voreilig einen Verdacht ausspricht.

Der Psychologe Gerd Gigerenzer führte zu dieser Thematik umfangreiche Untersuchungen durch (vgl. Gigerenzer, 2013). Demnach informieren viele Broschüren für Krebsfrüherkennung nur unzureichend oder intransparent. Das erforderliche Wissen zur adäquaten Interpretation eines Testbefundes sei selbst bei manchen Experten, die derlei Tests anwenden und Patienten beraten, mangelhaft. Um kluge Entscheidungen treffen zu können, sollten sich deshalb sowohl Ärzte als zu testende Personen vorab umfassend mit dieser Thematik befassen. Solide Kenntnisse in Statistik können dabei nicht schaden.

2.6 Narrativ für Pessimisten: Ein hohes relatives Risiko birgt eine große Gefahr

Im Herbst 1995 meldete das „Committee of Safety of Medicines“ (britische Aufsichtsbehörde für Arzneimittelsicherheit), dass bestimmte Präparate der Antibabypille das Risiko für eine Thromboembolie um 100 % erhöhten (vgl. Furedi, 1999). Diese Nachricht versetzte viele betroffene Frauen in Panik und veranlasste sie, ihre Pille sofort abzusetzen. Die Folge waren zahlreiche unerwünschte Schwangerschaften und 13 000 zusätzliche Schwangerschaftsabbrüche. Die Ursachen für dieses Missverständnis lagen darin, dass manche Frauen glaubten, mit Sicherheit (100 %) eine Thromboembolie zu bekommen, wenn sie weiterhin die Pille nehmen würden. In Wirklichkeit erhöhte sich das Risiko von 1/7 000 auf 2/7 000. Dies ist tatsächlich eine Verdopplung oder eben eine Steigerung um 100 %. Diese Prozentzahl bezieht sich jedoch nur auf den minimalen Anteil aller Frauen, die eine Thromboembolie (auch ohne Pille) erleiden.

Relative Risiken werden gerne benutzt, wenn die absoluten Risiken gering sind. Schließlich klingt eine „Steigerung um 100 %“ eindrucksvoller als eine „Risikoerhöhung von 1/7 000 auf 2/7 000“. Relative Risiken bieten sich außerdem an, um die Vorteile eines Screenings zu veranschaulichen. Beispiel Mammographie-Screening: Eine Metaanalyse ergab, dass bei Frauen im Alter zwischen 40 und 49 Jahren, die regelmäßig gescreent werden, die brustkrebsbedingte Mortalität bezogen auf einen Zeitraum von 10 Jahren 2,9 Promille betrug, während sie in der Kontrollgruppe mit 3,6 Promille etwas höher war (vgl. Kerlikowske, 1997). Diese Zahlen belegen, dass in der Gruppe der gescreenten Frauen die Mortalität geringer ist. Sie zeigen aber auch, dass Screenen nicht jedes Karzinom und jeden dadurch bedingten Todesfall verhindern kann. Man kann nun leicht die Differenz der beiden Risiken berechnen: Sie besagt, dass 0,7 Promille aller Frauen vom Screening profitieren würden. Dies entspricht einer von 1 429 Frauen. Diese „Number Needed to Screen“ ist die Anzahl der Frauen, die regelmäßig gescreent werden müssen, damit eine profitiert. Das ist leicht nachvollziehbar: Von 1 429 Frauen, die regelmäßig gescreent werden, sterben 2,9 Promille innerhalb von 10 Jahren an einem Mammakarzinom – das sind 4 Frauen. In einer gleich großen Kontrollgruppe sterben 3,7 Promille, also 5 Frauen.

Diese Zahlen sind nicht sonderlich beeindruckend, insbesondere für jemanden, der das Screening kritisch betrachtet. Der Nutzen scheint gering zu sein (zumal zu bedenken ist, dass der positive Vorhersagewert bei einem Mammographie-Screening nur 10 % beträgt und dass deshalb viele Frauen unnötigerweise belastet werden). Man kann jedoch auch anders argumentieren: Das relative Risiko beträgt $3,6 \div 2,9 = 1,24$. Demnach ist das Risiko, an einem Mammakarzinom zu versterben, für eine Frau, die sich nicht screenen lässt, um 24 % erhöht. Ebenso irreführend ist die These: Durch

regelmäßiges Screenen lässt sich das Sterberisiko um nahezu 20 % senken – obwohl die Rechnung per se korrekt ist. Sie ergibt sich nach $(3,6 - 2,9) \div 2,9$.

Diese Zahlen klingen gefährlich! Sie verheimlichen indessen, dass die Basisrisiken minimal sind. Es ist deshalb wichtig, bei relativen Risiken einen kühlen Kopf zu bewahren und die absoluten Risiken in Augenschein zu nehmen.

2.7 Narrativ für Optimisten: Eine Assoziation besagt nichts über einen Einzelfall

Kurz nach dem zweiten Weltkrieg führten die britischen Epidemiologen Richard Doll (1912–2005) und Austin Bradford Hill (1897–1991) eine Fall-Kontroll-Studie durch, mit denen sie eine statistische Assoziation zwischen Rauchverhalten und Lungenkrebs nachwiesen. Die Ergebnisse wurden 1950 publiziert (vgl. Doll & Hill, 1950). Es folgte eine groß angelegte Kohortenstudie, an der etwa 35 000 britische Ärzte teilnahmen, um die Auswirkungen des Rauchverhaltens auf die Gesundheit zu untersuchen. Diese ergab, dass das Risiko eines starken Rauchers (der mindestens 25 Zigaretten täglich raucht), innerhalb eines Jahres an Lungenkrebs zu versterben, 14 Mal so hoch ist wie für einen Nichtraucher. Die Studie zeigte außerdem, dass Raucher auch für andere onkologische Erkrankungen anfälliger sind, und dass sich die Risiken mit steigendem Nikotinkonsum erhöhen (vgl. Doll & Hill, 2004).

Mit der Framingham-Studie – eine groß angelegte prospektive Kohortenstudie, die im Jahre 1948 in Framingham (Massachusetts) mit mehr als 5 000 Probanden startete – war der Nachweis erbracht, dass Rauchen auch ein Risikofaktor für kardiovaskuläre Erkrankungen darstellt (vgl. Doyle et al., 1962).

Viele Raucher wollen das nicht wahrhaben. Sie verweisen auf bekannte Persönlichkeiten wie Winston Churchill oder den ehemaligen Bundeskanzler Helmut Schmidt (1918–2015), die Jahrzehnte lang rauchten und dennoch ein sehr hohes Alter von über 90 Jahren erreichten. Andererseits gäbe es durchaus auch Patienten, die an Lungenkrebs versterben, obwohl sie niemals geraucht haben. Ein weiteres Argument: Die in den relevanten Publikationen angegebenen Inzidenzen der Raucher seien nicht allzu hoch. Hin und wieder kritisieren sie das Design der Fall-Kontroll-Studie von Hill and Doll: Mit einer retrospektiven Studie ließen sich keine kausalen Zusammenhänge nachweisen. Möglicherweise könnte die beobachtete Assoziation auf einen Confounder zurückzuführen sein (eine Störgröße, die einen Zusammenhang nur vortäuscht).

Diese Argumente lassen sich entkräften. Die angegebenen Inzidenzen beziehen sich lediglich auf ein Jahr. Für eine Person, die Jahrzehnte lang raucht, erhöhen sie sich entsprechend. Auch der Hinweis auf das retrospektive Design der Fall-Kontroll-

Studie von Doll und Hill ist nicht angemessen: Es war damals durchaus sinnvoll, diese Studie der sehr aufwendigen Kohortenstudie vorzuschenken. Aufgrund der hohen Fallzahlen und effizienten Analysemethoden besteht kein Zweifel, dass die erwähnten Studien die Gefährlichkeit des Rauchens eindrucksvoll belegen. In Deutschland sterben jedes Jahr schätzungsweise mehr als 100 000 Menschen an den Folgen des Rauchens. Die durchschnittliche Lebenserwartung von starken Rauchern ist zehn Jahre geringer als die von Nichtrauchern.

Freilich beinhalten diese Angaben keine Information bezüglich der Lebenserwartung und der Todesursache eines Individuums. Doch jeder Raucher sollte sich fragen, aus welchem Grund ausgerechnet er von onkologischen oder kardiovaskulären Erkrankungen verschont bleibt oder aus welchen Gründen er ein überdurchschnittlich hohes Alter erreichen sollte. Diese Einsichten kommen zu spät, wenn eine Krankheit ausgebrochen ist. Zwischen 80 % und 90 % aller Lungenkarzinome sind auf das Rauchen zurückzuführen. Der Gedanke „Auch Nichtraucher können betroffen sein“ ist dann nur ein schwacher Trost.

2.8 Narrativ für Ehrgeizige: Statistisch signifikant ist praktisch relevant

Wissenschaftliche Publikationen enthalten in aller Regel mindestens ein statistisch signifikantes Ergebnis. Dieses wird durch das Resultat eines statistischen Tests, den sogenannten p-Wert, quantifiziert. Dabei handelt es sich um die sogenannte Irrtumswahrscheinlichkeit. Diese gibt an, wie hoch die Wahrscheinlichkeit ist, dass das Testergebnis unter der Nullhypothese („Es gibt keinen Effekt“), also rein zufällig, zustande gekommen ist. Wenn der p-Wert unter 0,05 liegt, gilt nach allgemeinem Konsens das Testergebnis als „statistisch signifikant“. Dann bestehen guten Chancen, dass dieses Ergebnis Eingang in eine Publikation findet, die in einem wissenschaftlichen Journal erscheint.

Die Autoren und die Leser einer Publikation sollten sich jedoch von einem kleinen p-Wert nicht blenden lassen! Ein p-Wert besagt per se nichts über die Größe des nachgewiesenen Effekts, nichts über die Präzision der Schätzung, nichts über die Ursachen des Ergebnisses und ebenfalls nichts über dessen praktische Relevanz oder wissenschaftliche Brisanz. Man muss wissen: Der p-Wert hängt wesentlich von der Fallzahl ab. Mit einer sehr hohen Fallzahl lassen sich kleinste Unterschiede und schwächste Zusammenhänge statistisch absichern, während eine geringe Fallzahl ein signifikantes Ergebnis quasi unmöglich macht. Tricks, ein signifikantes Ergebnis zu erzielen, sind hinlänglich bekannt: Wenn genügend Merkmale erfasst sind, lassen sich mehrere potenzielle Zielgrößen analysieren, die wiederum von diversen Ein-

flussgrößen abhängen. Je mehr Tests durchgeführt werden, desto höher ist die Chance auf ein signifikantes Ergebnis.

Wenn ein Effekt nachgewiesen wird, obwohl keiner existiert, spricht man von einem α -Fehler. Bei dem üblichen Signifikanzniveau $\alpha = 0,05$ ist theoretisch bei einem von 20 durchgeführten Tests ein signifikantes Ergebnis zu erwarten. Da aber überwiegend signifikante Ergebnisse publiziert werden (und nicht-signifikante Ergebnisse kaum Einzug in die Fachliteratur finden), führt dies zu einem verzerrten Bild in der akademischen Öffentlichkeit, dem sogenannten Publikationsbias.

Wie kann sich ein Wissenschaftler vor einem α -Fehler schützen? Er sollte nicht wahllos diverse Ziel- und Einflussgrößen analysieren, sondern vor Studienbeginn eine konkrete Fragestellung aufstellen, eine präzise Hypothese formulieren und diese zunächst mit fachlichen Argumenten versuchen zu untermauern. Eine gut durchdachte Planung, sorgfältig durchgeführte Untersuchungen und eine effiziente Datenanalyse tragen das Ihre dazu bei, einen wirklichen Effekt nachzuweisen. Wie erkennt der Leser einer Publikation, ob ein statistisch signifikantes Testergebnis praktisch relevant ist? In keinem Fall genügt es, nur auf den p-Wert zu schielen. Zusätzlich sollte man die Größe eines nachgewiesenen Effekts in Augenschein nehmen, die Voraussetzungen und die Limitationen der Studie beachten und dann erst urteilen, wem die Ergebnisse nutzen und welche Konsequenzen sich daraus ergeben.

2.9 Narrativ für Tiefgründige: Korrelation bedeutet Kausalität

Menschen lieben es, Dingen auf den Grund zu gehen. Forscher sind bestrebt, eine plausible Ursache für einen nachgewiesenen Effekt zu finden, um eine Hypothese glaubwürdig zu belegen. Kann ihnen die Statistik dabei helfen? Nur bedingt und nicht auf triviale Weise.

Diverse statische Kenngrößen wurden entwickelt, die geeignet sind, die Stärke eines Zusammenhangs zu quantifizieren. Bei zwei quantitativen Merkmalen (z.B. Alter und systolischer Blutdruck) eignet sich ein Korrelationskoeffizient, bei zwei binären Merkmalen (z.B. Risikofaktor und Krankheit) bietet sich ein Assoziationsmaß wie die Odds Ratio an. Allerdings informieren diese Kenngrößen nicht darüber, ob ein nachgewiesener Zusammenhang kausal bedingt ist. In der Literatur kursieren zahlreiche Beispiele für sogenannte Schein- oder Nonsense-Korrelationen. So lässt sich beispielsweise zeigen, dass ein Zusammenhang zwischen der Geburtenrate und dem Storchenaufkommen in Deutschland über einen Zeitraum von mehreren Jahrzehnten besteht. Wenngleich sich der Hinweis erübrigt, dass hier *keine* Kausalität zugrunde liegt, stellt sich die Frage: Wie kommt diese Korrelation zustande? Ist sie etwa als Bestätigung des altbekannten Narrativ „Mit Statistik kann man alles beweisen“ zu

werten? Nein! Der statistische Zusammenhang ergibt sich wegen der Zeitläufte. Zu Beginn des 20. Jahrhunderts und in den Jahren davor war Deutschland ein Agrarland. Die Bedingungen für Klapperstörche waren optimal, die jährlichen Geburtenraten hoch. Beide Parameter verringerten sich in den Jahren der beiden Weltkriege und stiegen danach an. Im Zuge der Industrialisierung ab den 1960er Jahren gab es wieder weniger Störche, die Familienstrukturen änderten sich, es wurden weniger Kinder geboren. Der Lauf der Jahre bringt eine Vielzahl von gravierenden Änderungen mit sich; dies führt unweigerlich zu statistischen Zusammenhängen zwischen den sich ändernden Parametern.

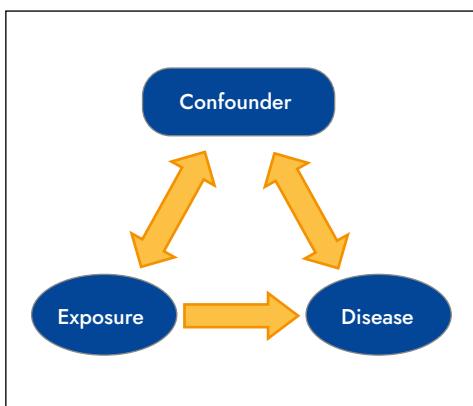


Abbildung 4: Zusammenwirken von Exposition, Krankheit und Confounder. Selbst erstellte Abbildung.

Jedoch sind Scheinkorrelationen nicht in jedem Fall so offensichtlich wie in diesem Beispiel. Insbesondere Fall-Kontroll-Studien sind anfällig für Confounder – das ist ein Merkmal, das sowohl mit der Erkrankung als auch mit der Exposition assoziiert ist. Ein Confounder kann dazu führen, dass zwischen zwei Faktoren ein Zusammenhang nachweisbar ist, der zwar statistisch signifikant, aber nicht kausal ist. Ein Beispiel: Eine Fall-Kontroll-Studie zeigt, dass Patienten mit einem Harnblasenkarzinom (Fälle) deutlich mehr Alkohol konsumieren als die gesunden Kontrollen. Lässt sich daraus schließen, dass Alkohol ein Risikofaktor für diese Erkrankung darstellt? Nicht unbedingt! Der eigentliche Risikofaktor ist das Rauchen. Der statistische Zusammenhang kommt zustande, weil der Confounder „Rauchen“ einerseits mit der Zielgröße (Erkrankung) und andererseits mit der Exposition (Raucher konsumieren häufig Alkohol) assoziiert ist. Der Verzicht auf Alkohol bei unverändertem Rauchverhalten würde das Erkrankungsrisiko nicht reduzieren.

Wie lässt sich der Einfluss eines Confounders verhindern? Theoretisch ideal sind experimentelle, randomisierte Studien, bei denen die Teilnehmer zufällig einer der zu vergleichenden Subgruppen zugeteilt werden. Dieses Design, bei dem mögliche Confounder auf beide Gruppen gleich verteilt werden, eignet sich bei Therapiestudien:

Die Patienten werden zufällig einer Therapiegruppe zugewiesen, die bezüglich ihrer Wirkung miteinander verglichen werden. Dies gewährleistet im Idealfall, dass die beiden Gruppen zu Beginn der Studie strukturgeleich sind und verhindert, dass Äpfel mit Birnen verglichen werden. Ein nachgewiesener statistisch signifikanter Unterschied zwischen den beiden Therapiegruppen impliziert, dass dieser kausal ist, und dass eine Therapie der anderen überlegen ist.

Bei Risikostudien ist die Randomisierung jedoch aus ethischen Gründen nicht anwendbar; hier ist man auf Beobachtungsstudien angewiesen. Wie lässt sich bei diesen Studientypen der Einfluss von Confoundern kontrollieren? Es ist hilfreich, vorab zu überlegen, welche potenziellen Confounder das Ergebnis einer Studie verzerrn könnten (in der medizinischen Forschung sind das häufig das Alter und das Geschlecht der Patienten sowie bekannte Risikofaktoren wie Alkohol- oder Nikotinabusus) und diese bei der Planung und bei der statistischen Analyse zu berücksichtigen. In jedem Fall sollte bei der Interpretation des Ergebnisses einer Risikostudie eine gewisse Vorsicht an den Tag gelegt werden. Der bereits erwähnte Epidemiologe Hill hat die nach ihm benannten Brad-Hill-Kriterien aufgestellt, die es ermöglichen, einen Kausalzusammenhang zwischen einer Exposition und einer Erkrankung zu beurteilen (vgl. Weiß, 2019).

2.10 Narrativ für Voreingenommene: Statistiker sind Nerds

Dieses Vorurteil scheint unausrottbar zu sein! Mathematikern und Statistikern werden zwar im Allgemeinen eine spezifische Intelligenz und fundiertes Fachwissen attestiert; ansonsten gelten sie als Eigenbrötler, weltfremd und sozial inkompotent. Sie befassen sich gerne mit mathematischen Problemen, sitzen stundenlang vor ihrem Computer und vergessen dabei die Welt um sich herum. Die Autorin dieses Beitrags würde diesem Narrativ freilich heftig widersprechen – wenngleich sie gestehen muss, dass es ein Fünkchen Wahrheit beinhalten könnte. Der Versuch, in einem unübersichtlichen Datenchaos Strukturen zu erkennen, Zusammenhänge zu entdecken oder ein komplexes statistisches Modell zu bilden, erfordert hochkonzentriertes Arbeiten, bei dem man ungerne gestört wird. Für Außenstehende mag das befremdlich wirken. Diese sollten jedoch – ehe sie sich der üblichen Klischees bedienen – bedenken, dass auch sie die Ergebnisse statistischer Analysen konsumieren und von ihnen profitieren – sei es direkt (indem sie etwa Umfrageergebnisse zur Kenntnis nehmen) oder indirekt (indem sie beispielsweise mit einer Therapie behandelt werden, deren Nutzen in einer klinischen Studie nachgewiesen wurde).

Schließlich sei angemerkt, dass „Nerd“ bei keinem Menschen ein Dauerzustand ist. Auch Statistiker haben soziale Kontakte, individuelle Vorlieben, Stärken, Schwä-

chen und Schrullen, viele haben sogar Humor! Deshalb besteht kein Anlass, sie mit negativ assoziierten Klischees zu versehen. Menschen sind verschieden – weil jeder einzigartig ist.

3 Schlussfolgerungen

Narrative sind verführerisch; sie vermitteln unterschwellige Botschaften. In einer Umgebung, in der nahezu permanent über diverse Kommunikationsmedien eine unübersehbare Fülle von Informationen verfügbar ist, fällt es schwer, diese zu sortieren und wichtige von weniger wichtigen Botschaften zu unterscheiden. Narrative sind dabei scheinbar hilfreich, da sie es vermögen, komplexe Sachverhalte verblüffend einfach zu erklären. Auf diese Weise befriedigen sie die Sehnsucht der Menschen nach Ordnung und einfachen Strukturen. Narrative können weder falsifiziert noch verifiziert werden. Sie werden selten kritisch reflektiert und noch seltener entlarvt. Der Erzähler eines Narrativs kann sich darauf verlassen, dass er allenthalben Zustimmung und selten Kritik erfahren wird. In diesem Sinne haben Narrative eine identitätsstiftende Funktion und erzeugen emotionale Resonanz. Die im vorigen Abschnitt vorgestellten Narrative werden ja nicht nur von einzelnen Laien oder Experten falsch verstanden oder wiedergegeben. Insbesondere Narrative, die mit einem bekannten Namen wie Winston Churchill assoziiert sind, sind quasi omnipräsent und werden besonders gerne zitiert. Die Gründe sind vielfältig: Einige Zeitgenossen mokieren sich liebend gerne über Mathematik und Statistik oder über Vertreter dieser Wissenschaften. Viele geben sich mit einfachen Lösungen zufrieden; anderen fehlt jegliches mathematische Verständnis. Es ist freilich weniger mühsam, ein Narrativ zu rezitieren (und sich dabei von vielen Mitmenschen verstanden zu wissen) als den Dingen auf den Grund zu gehen und sich danach vielleicht eingestehen zu müssen, dass man sich geirrt hatte.

Die im vorigen Abschnitt genannten Beispiele zeigen, dass der unbeirrbare Glaube an Narrative weitreichende Konsequenzen haben kann – nicht nur für diejenigen, die sie erzählen oder kritiklos zur Kenntnis nehmen. Nicht immer wird ein Narrativ zu einem Gefühl des sozialen Miteinanders führen, sondern manchmal auch spalten. Ein Statistiker wird allenfalls müde lächeln, wenn er mit Churchills Zitat konfrontiert wird (Narrativ 1), oder wenn er mit humorvollem Unterton als Nerd bezeichnet wird (Narrativ 10). Es ist beileibe kein Zeichen geistiger Genialität, wenn man nicht in der Lage ist, einfache Kenngrößen wie einen Mittelwert (Narrativ 2) oder einen statistisch nachgewiesenen Zusammenhang (Narrativ 9) sinnvoll zu interpretieren und mit diesem Unwissen kokettiert. Vielmehr ist anzunehmen, dass solche Menschen auch anderweitig oberflächlich urteilen. Zeitgenossen, die bei jedem unerwarteten Ereignis eine Verschwörung wittern, machen sich selbst und anderen das Leben schwer

(Narrativ 3). Politiker sind unglaublich, wenn sie ständig Narrative wiederholen (Narrativ 4). Wenn ein Mediziner bedingungslos dem Befund eines diagnostischen Tests vertraut, kann dies dazu führen, dass er eine falsche Diagnose stellt oder sogar unnötigerweise eine Therapie anordnet und damit seinen Patienten schadet (Narrativ 5). Menschen, die nicht in der Lage sind, ein relatives Risiko zu interpretieren, geraten leicht in Panik und lassen sich zu sinnlosen Handlungen verleiten (Narrativ 6). Optimisten, die ein Risiko nonchalant ignorieren, gefährden mitunter sich selbst und andere (Narrativ 7). Wissenschaftler, die sich von einem kleinen p-Wert blenden lassen, ohne dessen praktische Relevanz zu hinterfragen, tragen dazu bei, dass ein verzerrtes Bild in der akademischen Öffentlichkeit entsteht und eventuell sogar ihre eigene Forschungsarbeit in Verruf gerät (Narrativ 8).

Erfolgreich ist ein Narrativ jedoch nur dann, wenn es eine Resonanzgemeinde gibt, in der niemand widerspricht. Nicht-Widersprechen bedeutet nicht unbedingt, dass jedermann einem Narrativ Glauben schenkt. Ein anderer Grund könnte darin liegen, dass Zuhörer es wider besseres Wissen nicht wagen, in die Diskussion einzutreten und vehement zu widersprechen. Es ist in der Tat nervenaufreibend, eine andere Meinung zu vertreten, wenn a priori klar zu sein scheint, dass die Diskussionskontrahenten nur ungern bereit sind, ihre Argumente zu überdenken.

Narrative führen niemals zu neuen Erkenntnissen und sind deshalb auf Dauer langweilig. Was lässt sich dagegen unternehmen? Zunächst sollte jeder bereit sein, scheinbar festgefahrenen Meinungen kritisch zu hinterfragen, logisch zu überdenken, zu reflektieren und sich Irrtümer einzustehen. Dann sollte man sachlich argumentieren! Gegenargumente lassen sich nämlich – im Gegensatz zu Narrativen – belegen. Neue Aspekte (auch wenn sie für manche unbequem sein mögen) können eine Diskussion beleben und zu neuen Einsichten führen. Dafür erfordert es nur den Mut von Einzelnen zu widersprechen. Menschen würden noch auf Bäumen sitzen und in Höhlen hausen, es hätte keinen Fortschritt gegeben, wenn nicht zu allen Zeiten ein paar Mutige Widerspruch geleistet und Althergebrachtes in Frage gestellt hätten. „Sapere aude“: So formulierte dies der römische Dichter Horaz (65–8 v. Chr.). „Habe Mut, dich deines eigenen Verstandes zu bedienen“: So übersetzte diese Worte der Philosoph Immanuel Kant (1724–1804).

Literatur

- Brandt R:** *Freundesland. Erinnerungen.* Verlag Hoffmann und Campe, Seite 156; 1992
- Doll R, Hill AB:** Smoking and carcinoma of the lung; preliminary report. *British Medical Journal* 2, 739–748; 1950
- Doll R, Hill AB:** The mortality of doctors in relation to their smoking habits: a preliminary report. (Reprinted from Br Med J 1954). *British Medical Journal* 328, 1529–1533; 2004
- Doyle JT, Dawber TR, Kannel WB, Heslin AS, Kahn HA:** Cigarette Smoking and Coronary Heart Disease – Combined Experience of the Albany and Framingham Studies. *New England Journal* 266, 796–801; 1962
- El Ouassil S, Karig F:** *Erzählende Affen. Mythen, Lügen, Utopien.* Seite 153, Ullstein-Verlag; 2021
- Furedi A:** The public health implications of the 1995 ‚pill scare‘. *Hum Reprod Update* 5: 621–626; 1999
- Gigerenzer G:** *Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken.* Piper Taschenbuch Verlag. 4. Auflage; 2015
- Gigerenzer G:** HIV screening: helping clinicians make sense of test results to patients. *British Medical Journal* 347; 2013
- Gritzmann P:** *Plausibel, logisch, falsch: Auf den Holzwegen des gesunden Menschenverstandes.* Verlag C.H. Beck, 2. Auflage; 2024
- Kerlikowske K:** Efficacy of screening mammography among women aged 40 to 49 years and 50 to 50 years: Comparison of relative and absolute benefit. *J Natl Cancer Inst Mono* 22: 79–86; 1997
- Krämer W:** *So lügt man mit Statistik.* Campus-Verlag; 2015
- Weiβ C:** Kapitel 15.5: Nachweis einer Kausalität. In: *Basiswissen Medizinische Statistik*, Springer Verlag, 7. Auflage; 2019

Über die Autorin

Christel Weiβ ist Professorin für Biomathematik und Epidemiologie an der Medizinischen Fakultät Mannheim der Universität Heidelberg. In ihren Verantwortungsbereich fallen Lehrveranstaltungen für Studierende der Medizin und Masterkurs-Ab solventen, Seminare sowie die Beratung von Ärzten, wissenschaftlichen Mitarbeitern und Doktoranden bei der Planung und Durchführung von klinischen und epidemiologischen Studien. Frau Weiβ ist Autorin des Lehrbuchs „Basiswissen Medizinische Statistik“ (erschienen im Springer-Verlag, demnächst 8. Auflage), des Ratgebers „Promotion. Die medizinische Doktorarbeit – von der Themensuche bis zur Dissertation“ (zusammen mit Prof. Dr. Axel Bauer, erschienen im Thieme-Verlag, 4. Auflage) sowie Autorin oder Koautorin zahlreicher Papers und Buchbeiträge.

Korrespondenzadresse

Prof. Dr. Christel Weiß

Medizinische Fakultät Mannheim der Universität Heidelberg

Abteilung für Medizinische Statistik und Biomathematik,

Theodor-Kutzer-Ufer 1

68167 Mannheim, Germany

christel.weiss@medma.uni-heidelberg.de

<https://www.umm.uni-heidelberg.de/miism/biomedizinische-informatik/>

medizinische-statistik-biomathematik-und-informationsverarbeitung