Signifikanz vs. Relevanz: Denkfehler bei p-Werten

Autor: Wolfram Klatt
Veröffentlicht am: 10.08.2026

Dokument generiert über Gesundheitsmythen.de

ZUSAMMENFASSUNG (ABSTRACT)

Die Verwechslung von statistischer Signifikanz mit klinischer Relevanz verleitet dazu, minimale oder unbedeutende Effekte als wichtigen medizinischen Nutzen zu interpretieren. Der p-Wert sagt nur aus, ob ein Effekt unwahrscheinlich Zufall ist, nicht aber, wie spürbar er für den Patienten ist.

Signifikanz mit Relevanz verwechseln: Der Unterschied zwischen Statistik und Alltag in der Medizin

Statistische Signifikanz ist ein zentraler Begriff in der medizinischen Forschung, wird jedoch im Alltag – und selbst in der Gesundheitskommunikation – immer wieder missverstanden. Wenn in einem Bericht, in einem Ratgeberbuch oder in der Information zu einem neuen Behandlungsverfahren erläutert wird, eine Maßnahme habe in Studien eine „signifikante Wirkung“ gezeigt, verbinden viele Menschen damit automatisch eine *wichtige*, *große* oder *lebensverändernde* Wirkung. In der Fachsprache der Statistik bedeutet der Begriff „signifikant“ jedoch etwas ganz anderes: Er drückt lediglich eine Wahrscheinlichkeit aus, dass ein beobachtetes Ergebnis kein Zufall ist.

Diese Verwechslung von statistischer Signifikanz mit klinischer Relevanz – also dem tatsächlichen, spürbaren Nutzen für die Patientinnen und Patienten – kann weitreichende Folgen haben. Sie kann dazu führen, dass Menschen Hoffnungen in Therapien setzen, die zwar auf dem Papier einen Effekt zeigen, im echten Leben aber kaum einen spürbaren Unterschied machen. Dieser Artikel beleuchtet die Zusammenhänge objektiv, erklärt die dahinterstehenden Mechanismen und soll Ihnen helfen, medizinische Studienergebnisse in Zukunft noch fundierter einordnen zu können.

1. Die Definition: Was bedeutet der p-Wert wirklich?

Um dieses Konzept zu durchdringen, werfen wir einen kurzen Blick auf die wissenschaftliche Methode. Forschergruppen untersuchen beispielsweise, ob ein neues Medikament, ein pflanzliches Extrakt oder eine spezielle physiotherapeutische Übung wirksam ist. Sie möchten herausfinden, ob ein Unterschied zwischen der behandelten Gruppe und einer Kontrollgruppe (die beispielsweise ein Placebo erhält) existiert – und ob dieser Unterschied möglicherweise nur durch Zufall entstanden ist. Hier kommt der p-Wert (Wahrscheinlichkeitswert) ins Spiel.

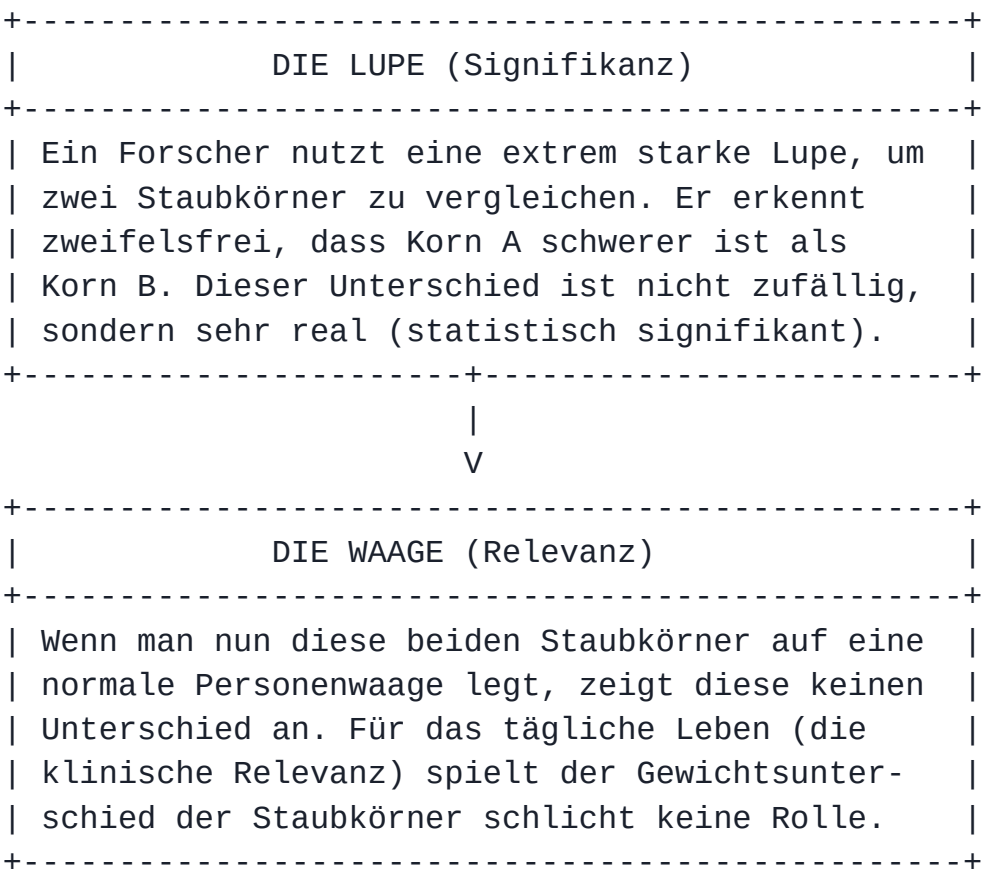
Der p-Wert beziffert, wie wahrscheinlich es ist, die in der Studie gesammelten Daten (oder noch extremere) zu erhalten, wenn man davon ausgeht, dass die untersuchte Maßnahme in Wirklichkeit *gar keinen* Effekt hat (die sogenannte Nullhypothese). In vielen Disziplinen wurde historisch eine Grenze festgelegt: Liegt der p-Wert unter 5 Prozent ($p < 0{,}05$), spricht man von einem „statistisch signifikanten“ Ergebnis.

Die sprachliche Hürde liegt in der Mehrdeutigkeit des Wortes „signifikant“. Im Alltag nutzen wir das Wort für Dinge, die von erheblicher Bedeutung oder enormem Ausmaß sind. In der Statistik heißt es jedoch nur: „Es ist unwahrscheinlich, dass dieser messbare Unterschied durch reinen Zufall entstanden ist.“

Der p-Wert signalisiert uns also, *dass* sehr wahrscheinlich ein Effekt vorliegt. Er sagt jedoch rein gar nichts darüber aus, *wie groß*, *wie spürbar* oder *wie bedeutsam* dieser Effekt für den Patienten ist. Genau hier entsteht häufig ein Missverständnis: Ein Effekt kann mathematisch hochsignifikant sein, ohne für die Patientin oder den Patienten im Alltag eine spürbare Relevanz zu besitzen.

Die Metapher der Lupe und der Waage

Um den Unterschied zwischen Signifikanz und Relevanz besser greifbar zu machen, kann die Analogie einer Lupe und einer Waage helfen:



Je größer eine Studie (je mehr Teilnehmende, was einer immer stärkeren Lupe entspricht), desto feinere Unterschiede lassen sich als signifikant nachweisen – auch wenn sie klinisch wie das Staubkorn auf der Personenwaage irrelevant bleiben.

2. Im klinischen Alltag: Wie Zahlen manchmal täuschen können

In der Praxis erleben Ärzte und Patienten immer wieder Situationen, in denen die statistische Signifikanz allein kein guter Ratgeber für medizinische Entscheidungen ist.

Ein klassisches, anschauliches Szenario betrifft beispielsweise die Blutdruckbehandlung. Nehmen wir an, in einer großen, internationalen Studie mit 40.000 Teilnehmenden wird untersucht, wie gut ein neues Präparat im Vergleich zum bewährten Standardmedikament wirkt. Das Ergebnis zeigt: Das neue Medikament senkt den systolischen Blutdruck im Durchschnitt um

1 mmHg (Millimeter Quecksilbersäule) stärker als das alte Präparat. Aufgrund der enormen Gruppengröße ist dieser Unterschied statistisch hochsignifikant ($p < 0{,}001$). Die Wahrscheinlichkeit, dass dieser Unterschied auf bloßem Zufall beruht, tendiert gegen null.

Der Hersteller darf nun vollkommen legitim mitteilen, dass sein Präparat eine signifikant bessere Wirkung zeigte. Für einen Menschen mit Bluthochdruck (der beispielsweise Werte von 160 mmHg aufweist) spielt es jedoch im Alltag und für seine gesundheitliche Prognose keine spürbare Rolle, ob sein Wert auf 140 mmHg oder auf 139 mmHg gesenkt wird. Der Unterschied von 1 mmHg ist klinisch gesehen unbedeutend. Wenn das neue Medikament nun mit höheren Kosten oder einem unbekannteren Nebenwirkungsprofil einhergeht, müssen Arzt und Patient gemeinsam abwägen, ob die statistische Überlegenheit wirklich eine Umstellung der Therapie rechtfertigt.

Ein weiteres Beispiel findet sich in Studien zu Therapien bei leichten depressiven Verstimmungen. Die Wirksamkeit wird hierbei oftmals anhand etablierter Fragebögen gemessen, die bestimmte Skalen (beispielsweise von 0 bis über 50 Punkten) nutzen. In großen Studien kann bereits eine Verbesserung um 1,5 bis 2 Punkte auf dieser Skala statistisch signifikant sein. Ein solcher Punktsprung ist zwar verlässlich messbar (Signifikanz), das bedeutet aber nicht zwingend, dass Betroffene in ihrem Lebensalltag eine echte, spürbare Entlastung von ihren Beschwerden erleben (Relevanz).

3. Ursachen des Missverständnisses: Unsere Sehnsucht nach Eindeutigkeit

Warum fällt es uns so schwer, diese statistischen Feinheiten intuitiv zu begreifen? Dies betrifft medizinische Laien ebenso wie erfahrene Fachleute im Gesundheitswesen. Die Antwort liegt zum großen Teil in unserer menschlichen Kognition und in unserem verständlichen Bedürfnis nach Klarheit.

Unser Gehirn strebt nach Gewissheit und einfache Antworten, besonders wenn es um komplexe oder angstbesetzte Themen wie Krankheit und Gesundheit geht. In Stresssituationen greifen wir bevorzugt auf mentale Abkürzungen zurück (Heuristiken). Kategorien wie „Ja oder Nein“, „Wirkung bewiesen oder nicht bewiesen“ sind wesentlich leichter zu verarbeiten als ein komplexes Kontinuum von Effektgrößen und Wahrscheinlichkeiten.

Der Nobelpreisträger Daniel Kahneman beschrieb in seinen Modellen sehr anschaulich, wie unser intuitives Denken („System 1“) statistische Unschärfen oft ausblendet und stattdessen lieber klare kausale Geschichten konstruiert. Der p-Wert ($p < 0,05$) fungiert hierbei als hochwillkommene Vereinfachung. Der unbewusste Verstand verwandelt eine trockene mathematische Grenze in eine binäre Garantie: Ein p-Wert unter 0,05 vermittelt das beruhigende Gefühl „Es wirkt zweifelsfrei“, ein Wert darüber vermittelt „Es wirkt nicht“.

Sobald der Begriff „signifikant“ fällt, signalisiert unser Sprachverständnis Wichtigkeit. Wenn uns mit wohlwollender Überzeugung berichtet wird: „Dieses Verfahren hat in einer Studie signifikante Ergebnisse geliefert“, werten wir dies als wissenschaftliches Siegel. Das ist keine Naivität, sondern ein zutiefst menschlicher Vorgang. Wenn Menschen erkrankt sind, suchen sie Halt in der Wissenschaft. Es ist eine der vornehmsten Aufgaben der medizinischen Kommunikation, diese Ergebnisse verständlich und transparent einzuordnen.

4. Wie das Missverständnis im Gesundheitswesen sichtbar wird

Die Verwechslung von Signifikanz und Relevanz ist nicht auf einen bestimmten Bereich beschränkt, sondern durchzieht viele Sektoren des Gesundheitsmarktes – und sie wird mitunter auch für Marketingzwecke genutzt.

Im Bereich der Nahrungsergänzungsmittel, neuartiger Superfoods oder bestimmter alternativmedizinischer Ansätze finden sich mitunter Studien, die kleine Stichprobengrößen aufweisen. Wenn bei einer kleinen Studiengruppe zahlreiche unterschiedliche Gesundheitsindikatoren (Blutdruck, Schlaf, subjektives Wohlbefinden etc.) gleichzeitig gemessen werden, steigt allein durch die Wahrscheinlichkeit die Chance, dass einer dieser Werte am Ende rein rechnerisch eine „Signifikanz“ aufweist (in der Forschung als multiples Testen bekannt). Es kann vorkommen, dass Hersteller dann genau dieses eine signifikant gewordene Ergebnis betonen, auch wenn der Nutzen im Lebensalltag der Anwender marginal ist.

Auch bei der Entwicklung und Vermarktung konventioneller Arzneimittel kann dieser Mechanismus eine Rolle spielen. Große Pharmaunternehmen führen gelegentlich Studien mit zehntausenden Probanden durch (Megatrials). Durch die enorme statistische Power solcher Studien können auch kleinste, klinisch kaum wahrnehmbare Verbesserungen gegenüber bestehenden Therapien als „signifikant“ nachgewiesen werden. In der Kommunikation

gegenüber Fachpersonal und Patienten kann diese statistische Signifikanz dann in den Vordergrund rücken und eine klinische Überlegenheit suggerieren, die in der Realität vielleicht sehr gering ausfällt.

In beiden Szenarien wird eine kommunikative Grauzone betreten. Der Patient vertraut darauf, dass der statistische Nachweis *irgendeines* Effekts gleichbedeutend mit einem *wichtigen* gesundheitlichen Nutzen ist.

5. Welche Auswirkungen hat dies auf Patientinnen und Patienten?

Wenn Patientinnen und Patienten aufgrund der Verwechslung von Signifikanz und Relevanz Behandlungsentscheidungen treffen, kann das vielschichtige Konsequenzen haben, die stets respektvoll betrachtet werden müssen.

1. Finanzielle Belastungen und Ressourcen: Patienten investieren oft erhebliches Vertrauen und auch finanzielle Mittel in Gesundheitsprodukte, neuartige Therapieansätze oder auch in teure patentierte Medikamente, die möglicherweise keinen klinisch relevanten Zusatznutzen gegenüber günstigeren und bewährten Standardtherapien besitzen. Das Attribut „signifikante Wirkung“ fungiert hier als Rechtfertigung für Investitionen, die nicht immer mit einer entsprechenden Steigerung der Lebensqualität einhergehen.

2. Fragen der Behandlungsrisiken: In der Medizin gilt, dass jede Intervention und jede Substanz auch mit unerwünschten Nebenwirkungen verbunden sein kann. Wenn ein Medikament zwar statistisch signifikant wirkt, der Nutzen für den Patienten aber im Alltag unmerklich klein ist, geht er dennoch das volle Risiko für mögliche Nebenwirkungen ein. Eine gut balancierte Therapieentscheidung muss stets das Verhältnis zwischen spürbarem Nutzen und potenziellen Risiken im Blick behalten.

3. Unerfüllte Hoffnungen bei schwerwiegenden Erkrankungen: Besonders empathisch muss auf Situationen geblickt werden, in denen Patienten an schweren, fortgeschrittenen Erkrankungen leiden. Die Hoffnung auf zusätzliche Lebenszeit oder Schmerzlinderung ist zutiefst verständlich. Wenn in einer solchen Phase eine neue Therapie vorgestellt wird, die die Lebenszeit statistisch signifikant um beispielsweise zwei Wochen verlängert, diese Zeit aber durch starke Nebenwirkungen geprägt ist, stehen Patienten vor einer schwierigen ethischen und persönlichen Entscheidung. Eine rein auf Signifikanz basierende Kommunikation kann

Hoffnungen wecken, die durch die klinische Realität nur schwer erfüllt werden können.

4. Die Tücken der Laborwerte (Surrogatparameter): Manchmal werden in Studien nicht direkt die gesundheitlichen Ziele der Patienten gemessen, sondern lediglich Ersatzwerte im Labor (Surrogatparameter). Ein Präparat mag einen bestimmten Entzündungswert oder den Cholesterinspiegel signifikant verändern. Dies ist auf dem Papier ein Erfolg. Entscheidend bleibt jedoch, ob diese Veränderung im Blut Jahre später auch dazu führt, dass die Patientin oder der Patient seltener erkrankt, im Alltag mobiler bleibt oder länger lebt. Die Relevanz findet stets im Leben des Patienten statt, nicht nur auf dem Laborbefund.

7. Wie Sie echte Relevanz im Alltag erkennen können

Um medizinische Ergebnisse besser bewerten zu können und sich nicht allein von statistischen Schwellenwerten leiten zu lassen, helfen einige wenige, aber sehr effektive Leitfragen und mentale Werkzeuge:

1. Die klare Trennung der Begriffe verinnerlichen: Der wohl wichtigste Grundsatz der medizinischen Studienbewertung lautet: **Statistisch signifikant bedeutet nicht zwingend relevant.** Es bedeutet lediglich, dass ein beobachteter Effekt sehr wahrscheinlich nicht rein zufällig aufgetreten ist.

2. Nach der absoluten Effektgröße fragen: Wenn Ihnen eine Therapie als „signifikant besser“ empfohlen wird, ist es stets ratsam, eine ganz direkte Frage zu stellen: *Um wie viel besser genau?* Fragen Sie nach konkreten Werten, die Ihr Leben betreffen. Wie viele Millimeter sinkt der Blutdruck? Wie viele Tage weniger haben Sie Schmerzen im Monat? Wie viel mehr Bewegungsumfang erzielen Sie nach der Physiotherapie? Erst die Nennung absoluter Größen ermöglicht es Ihnen, den klinischen Nutzen für Ihren Alltag einzuschätzen.

3. Die „Number Needed to Treat“ (NNT) heranziehen: Ein bewährtes Konzept in der evidenzbasierten Medizin ist die NNT (Anzahl der zu behandelnden Patienten). Diese Kennzahl drückt aus, wie viele Menschen eine bestimmte Therapie erhalten müssen, damit *eine einzige* Person einen klaren, messbaren Vorteil gegenüber der Kontrollgruppe hat. Eine NNT von 5 bedeutet: Von fünf behandelten Personen profitiert einer zusätzlich, vier hätten auch ohne das Medikament (oder mit Standardtherapie) denselben Verlauf gehabt. Manchmal verbergen sich hinter hochsignifikanten Ergebnissen NNTs von 50 oder mehr. Fragen Sie Ihren behandelnden Arzt nach solchen Einschätzungen – sie helfen ungemein bei der Entscheidung, ob eine

Maßnahme für Sie individuell sinnvoll ist.

4. Fokussierung auf patientenrelevante Endpunkte: Letztlich sollte bei jeder medizinischen Maßnahme eine Frage im Zentrum stehen: Führt diese Maßnahme dazu, dass sich Ihre ganz persönliche Lebensqualität verbessert, dass sich Beschwerden messbar lindern lassen oder dass ernsthafte gesundheitliche Ereignisse verhindert werden? Ein messbar veränderter Laborwert ist in erster Linie für das Diagnoseprotokoll interessant – für Sie als Patientin oder Patient zählt das spürbare Ergebnis.

Indem wir lernen, über den Glanz der statistischen Signifikanz hinauszuschauen und nach der klinischen Relevanz zu fragen, stärken wir unsere Gesundheitskompetenz. Wir begegnen medizinischen Heilsversprechen kritischer und können gemeinsam mit Ärztinnen und Ärzten fundiertere, an unseren wahren Bedürfnissen orientierte Entscheidungen treffen.

Dieses Dokument ist ein offizieller Export von Gesundheitsmythen.de (COMUN e.V.). Alle Fakten und Quellenangaben sind auf der Website einsehbar.