

## **Sensitivität und Spezifität spielen in der medizinischen Praxis eine marginale Rolle**

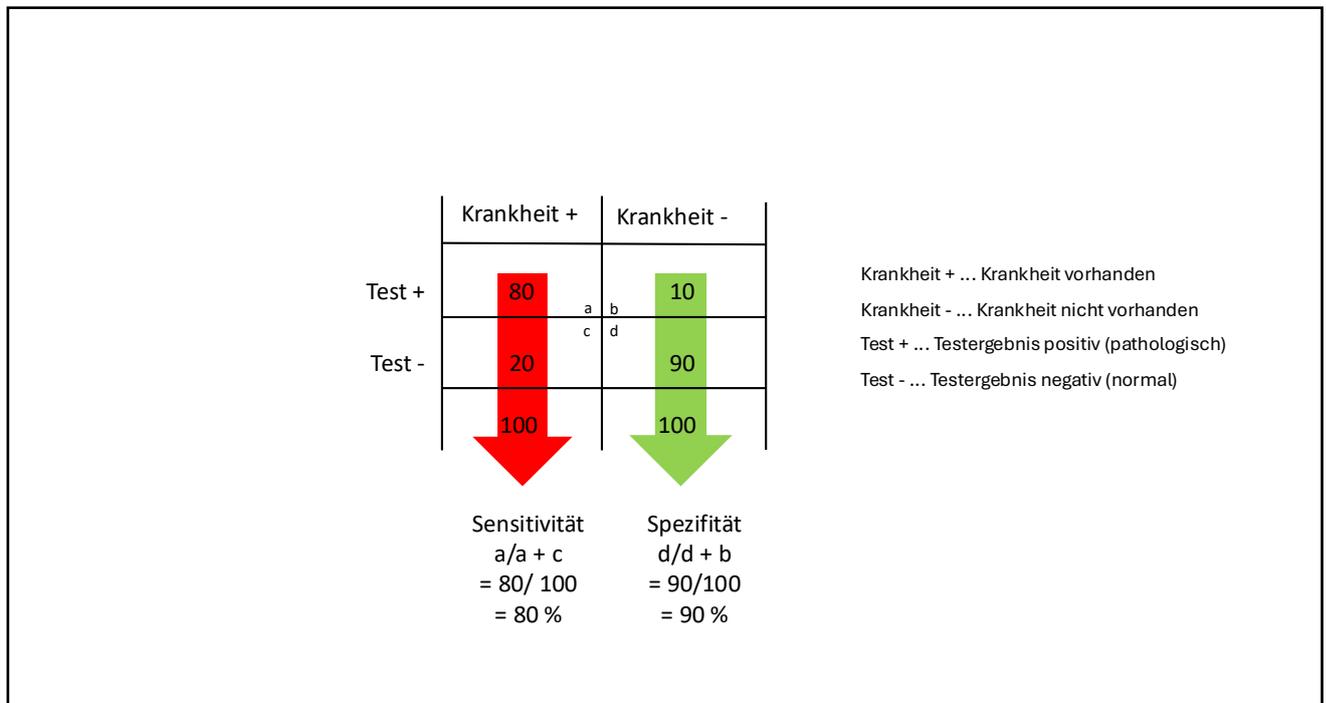
Prof. em. Dr. med. Johann Steurer

In dem kurzen Artikel wird das Konzept der Sensitivität und Spezifität eines medizinischen Tests beschrieben und an Beispielen erläutert. Es ist wichtig dieses Konzept zu kennen und zu verstehen. Es ist aber auch wichtig die Grenzen und die Gründe für die klinisch doch beschränkten Anwendungsmöglichkeiten der Sensitivität und Spezifität eines Tests zu kennen.

Um eine Krankheit nachzuweisen oder auszuschliessen werden nach der Anamnese und der körperlichen Untersuchung meist weitere Untersuchungen – Laboruntersuchungen; bildgebende Verfahren wie Ultraschall, Röntgen, CT oder MRT; Ergometrie; Lungenfunktionsprüfung; Endoskopien, Angiographien – durchgeführt. Abhängig von den Ergebnissen dieser Untersuchungen steigen oder sinken die Wahrscheinlichkeiten der vermuteten Krankheiten. Wenn zum Beispiel der CRP-Wert (C-reaktives Protein; ein im Blut messbarer Parameter, der bei Infektionen erhöht ist) bei einem Patienten mit verstopfter Nase und Husten deutlich erhöht ist, steigt die Wahrscheinlichkeit, dass der Patient eine Lungenentzündung und nicht nur einen banalen Schnupfen hat.

Der Effekt der Ergebnisse auf die Wahrscheinlichkeiten hängt von der Aussagekraft der durchgeführten Untersuchungen ab. Die Aussagekraft wird durch die Sensitivität und die Spezifität einer Untersuchung bestimmt. Die Sensitivität beschreibt wie viele der Patienten mit einer Erkrankung von dem Test korrekt als krank identifiziert werden. Die Spezifität hingegen gibt an wie viele Patienten, die nicht an der vermuteten Krankheit leiden, mit der Untersuchung korrekt identifiziert werden.

Abbildung 1: Sensitivität und Spezifität eines Tests; eine andere Darstellungsform im Addendum 1



Mit der Vierfelder-Tabelle (Abbildung 1) können noch weitere Werte berechnet werden – der positiv prädiktive und der negativ prädiktive Wert.

Der positiv prädiktive Wert sagt aus wie viele Patienten mit einem positiven Testergebnis an der vermuteten Krankheit leiden (nicht mit Sensitivität verwechseln!) – im gegebenen Beispiel  $a/a + b = 80/90 = 89\%$ . Der negativ prädiktive Wert sagt aus wie viele Patienten mit einem negativen Testergebnis nicht an der vermuteten Krankheit leiden (nicht mit Spezifität verwechseln!) – im gegebenen Beispiel  $d/c + d = 90/110 = 82\%$ .

Diese Werte von 89%, respektive 82 %, sind korrekt, wenn die Vortest-Wahrscheinlichkeit (Erklärung, siehe nächsten Paragraph), wie in Abbildung 1. 50 % beträgt. Wenn die Vortestwahrscheinlichkeit 10 % beträgt, sinkt der positiv prädiktive Wert von 89 % auf 44 % und die Spezifität steigt von 82 % auf 98 %.

Im Studium wird vermittelt wie im diagnostischen Prozess die Sensitivität und Spezifität angewandt werden sollen.

Folgendermassen. Nach der Anamnese und der körperlichen Untersuchung vermutet die Ärztin eine oder mehrere möglichen Krankheiten als Grund der Beschwerden. Beschränken wir uns der Einfachheit halber auf eine Krankheit. Die Ärztin schätzt aufgrund der Anamnese und der körperlichen Untersuchung die Wahrscheinlichkeit der vermuteten Krankheit. Diese Wahrscheinlichkeit wird als Vortestwahrscheinlichkeit bezeichnet. Unter Anwendung einer Formel (Bayessche Formel im Addendum 2) lässt sich aus der Vortestwahrscheinlichkeit und der Sensitivität und Spezifität eines diagnostischen Tests die Nachtestwahrscheinlichkeit berechnen. Auch wenn dieses Vorgehen theoretisch korrekt ist, bleibt die praktische Relevanz im medizinischen Alltag äusserst gering. Derartige Berechnungen werden in der Praxis praktisch nie durchgeführt (1).

### **Warum werden derartige Berechnungen nicht durchgeführt?**

Dafür gibt es mehrere Gründe.

1. Ärzte und Ärztinnen kennen nur für sehr wenige Untersuchungen die Werte für die Sensitivität und Spezifität und die Bedeutung der Begriffe werden häufig verwechselt – insbesondere die Sensitivität mit dem positiv prädiktiven Wert (2).
2. Die Schätzung der Vortestwahrscheinlichkeit ist schwierig. In einer Untersuchung (3) schätzten Ärztinnen und Ärzte mit langjähriger Praxiserfahrung auf der Basis von sehr sorgfältig erarbeiteten Fallbeschreibungen die Vortest-Wahrscheinlichkeiten eines akuten Koronarsyndroms (Herzinfarkt). Die geschätzten Werte für ein und dieselbe Fallbeschreibung schwankten teilweise zwischen 1 % und 99 %. Bei der Fallbeschreibung mit der geringsten Schwankungsbreite lag sie bei 40 % (die tiefste Schätzung war 0 % und die höchste 40 %). Das Ergebnis dieser Studie ist ein Hinweis, dass verlässliche Schätzungen von Wahrscheinlichkeiten einer Erkrankung praktisch unmöglich sind.

Andere Forscher kamen in anderen Studien zur gleichen Konklusion (4-6). An einer Studie, an der über 500 praktizierende Ärztinnen und Ärzte teilnahmen, überschätzen sie die Vortestwahrscheinlichkeiten sehr deutlich. Nach Vorliegen der Testergebnisse überschätzten die Teilnehmenden auch die Nachtestwahrscheinlichkeiten sehr deutlich (7).

Ein Studienteilnehmer kommentierte die Ergebnisse – das Ergebnis sei für ihn nicht überraschend denn, «So macht man nicht Medizin (isn't how you do medicine)». Der diagnostische Prozess basiert primär auf der Grundlage von Mustererkennung und dem Abgleich zwischen den Informationen aus Anamnese und weiteren Untersuchungen und den im Gehirn gespeicherten Krankheits-Skripten (8). Explizite Quantifizierungen von Wahrscheinlichkeiten sind der extreme Ausnahmefall.

3. Die Werte für die Sensitivität und Spezifität eines Tests sind keine «Naturkonstanten», sondern hängen von der Zusammensetzung der Population, in der der Test evaluiert wurde, ab. Das heisst, die Werte für die Sensitivität und Spezifität variieren von Studie zu Studie.

Beispiele: Zum Nachweis von SARS-CoV-2, dem Virus der COVID-19 Epidemie, wurde ein Antigentest (rapid antigen test) entwickelt. Die Sensitivität des Tests schwankte, je nach Studie zwischen 30 % und 100 % (9). Die Sensitivität war in Populationen mit Symptomen von COVID-19 höher als in asymptomatischen Populationen. Sie war auch höher in den ersten 7 Tagen nach Beginn von COVID-19 Symptomen als bei Patienten, bei denen die Symptome schon länger dauerten.

Ein weiteres Beispiel ist die Sensitivität der Ergometrie bei Verdacht auf eine klinisch relevante Koronarstenose. Bei Männern beträgt die Sensitivität 65 %, bei Frauen 30 %, bei Rauchern 61 % und bei Nichtrauchern 52 %, bei Patienten mit Erkrankung einer Koronararterie 39 % und bei Erkrankung aller drei Herzkranzgefässe 77 % (10).

Die Variabilität trifft auch für die Spezifität zu. Die Spezifität von D-Dimer (ein Test für die Diagnose einer tiefen Venenthrombose oder einer Lungenembolie) schwankt je nach Studie zwischen 29 % und 80 % (11).

Die Schätzung von Vortestwahrscheinlichkeiten ist schwierig, die Werte für die Sensitivität und Spezifität eines Testes variieren stark und die Berechnung der Nachtestwahrscheinlichkeit – wenn diese überhaupt gemacht werden – daher selten korrekt. Daher spielt es keine grosse Rolle, wenn die Kliniker die Werte für die Sensitivität und Spezifität eines Tests gar nicht kennen.

Nachdem die konzeptionellen Schwächen von Sensitivität und Spezifität erläutert sind, ergeben sich zwei weitere Fragen.

Warum werden nach wie vor zahlreiche Studien veröffentlicht, die ausschließlich Sensitivität und Spezifität diagnostischer Tests berichten?

Und welchen konkreten Beitrag leisten diese Parameter tatsächlich im Rahmen diagnostischer Prozesse?

Zur ersten Frage: Es existieren mittlerweile Hunderte von Studien zur diagnostischen Aussagekraft des D-Dimers – einem Laborparameter, dessen Erhöhung auf das mögliche Vorliegen einer venösen Thromboembolie (tiefe Venenthrombose/Lungenembolie) hinweist. Für bestimmte Subgruppen, etwa Patienten mit und ohne Niereninsuffizienz, ältere versus jüngere Personen oder andere spezifische Kollektive, kann die Untersuchung von Sensitivität und Spezifität durchaus klinisch relevant sein. Dennoch erscheint der überwiegende Teil – geschätzt über 90 % – dieser Publikationen redundant. Sie tragen kaum neue Erkenntnisse bei und scheinen primär der Erhöhung individueller Publikationszahlen zu dienen.

Zur zweiten Fragestellung: Diagnostische Tests mit ausgeprägt hoher Sensitivität oder Spezifität (solche Tests sind selten) können im diagnostischen Prozess von erheblichem Nutzen sein, da sie die Vortestwahrscheinlichkeit substanziell verändern – entweder durch weitgehenden Ausschluss oder durch relevante Erhöhung der Wahrscheinlichkeit für das Vorliegen der vermuteten Diagnose.

Ein Test mit sehr hoher Sensitivität eignet sich insbesondere zum Ausschluss einer Erkrankung: Fällt das Testergebnis negativ aus, ist das Vorliegen der gesuchten Erkrankung mit hoher Wahrscheinlichkeit ausgeschlossen – wenn auch nicht mit absoluter Sicherheit. Dies lässt sich mit der Eselsbrücke **SnNout** zusammenfassen (Sensitivität hoch, wenn Testergebnis negativ, vermutete Krankheit out).

*Beispiele:*

	<i>Krankheit</i>	<i>Sensitivität</i>
<i>proBNP<sup>1</sup></i>	<i>Herzinsuffizienz</i>	<i>~ 95 % (12)</i>
<i>D-dimer</i>	<i>Venenthrombose</i>	<i>~ 95 % (11)</i>
<i>Troponin</i>	<i>Herzinfarkt</i>	<i>~ 95 % (13)</i>

Wenn die Spezifität eines Testes sehr hoch ist und das Ergebnis des Testes positiv (also pathologisch) ist, dann ist die Wahrscheinlichkeit der vermuteten Krankheit sehr hoch (aber in der Regel nicht beweisend). Die Merkhilfe dafür: **SpPI**n (Spezifität hoch, Testergebnis positiv, vermutete Krankheit in).

*Beispiele:*

	<i>Krankheit</i>	<i>Spezifität</i>
<i>Hepato-jugulärer Reflux</i>	<i>Linksherzinsuffizienz</i>	<i>~ 94 % (14)</i>
<i>Anti-CCP-Antikörper</i>	<i>rheumatoide Arthritis</i>	<i>~ 95 % (15)</i>
<i>Legionella Antigen Test (Urin)</i>	<i>Legionellose</i>	<i>~ 99 % (16)</i>

Achtung! Auch wenn der Wert für die Sensitivität eines Tests sehr hoch ist, kann eine Krankheit – bei hoher Sensitivität und negativem Testergebnis – nicht ausgeschlossen werden. Der gleiche Vorbehalt gilt vice versa auch für einen Test mit hoher Spezifität. Im Falle vom D-Dimer ist die Sensitivität bei etwa 95 %. Erläuterung an zwei Beispielen.

*Beispiel 1:*

*Ein 40-jähriger Mann hatte vor einem Tag beim Ein- und Ausatmen ein paarmal einen Schmerz im Brustbereich. Er konnte mit einem Finger zeigen, wo der Schmerz war. Der Mann hatte noch nie eine Thrombose, er war nicht bettlägerig, ist nicht mehrere Stunden geflogen oder im Auto oder Zug gereist, er raucht nicht – er hat also keine Risikofaktoren für eine venöse Thrombose oder Lungenembolie. Diese kurze Schmerzepisode könnte aber eine Manifestation einer Lungenembolie sein. Die Wahrscheinlichkeit ist sehr gering, aber eine Lungenembolie darf man nicht verpassen, da sie potenziell*

---

<sup>1</sup> N-terminales pro-brain natriuretisches Peptid

*fatal ist. Die Bestimmung des D-Dimers ergibt einen Wert im Normbereich und damit ist eine Lungenembolie praktisch ausgeschlossen (korrekt wäre zu schreiben; die Wahrscheinlichkeit einer Lungenembolie ist so klein, dass keine weiteren Untersuchungen mehr indiziert sind). Siehe Addendum 2*

*Beispiel 2:*

*Eine 40-jährige Frau kommt in den Notfall wegen intermittierender Atemnot seit zwei Stunden. Sie kommt gerade retour von einer Reise nach Florida, sie nimmt Antikonzeptiva, sie raucht 20 Zigaretten/d und ist übergewichtig. Aufgrund der Beschwerden und den Risikofaktoren (lange Reise, Antikonzeptiva, Rauchen, Übergewicht) ist die Wahrscheinlichkeit für eine Lungenembolie hoch (wie hoch in konkreten Zahlen weiss ich nicht). Auch bei dieser Patientin ist das D-Dimer im Blut nicht erhöht. Es wäre falsch, obwohl die Sensitivität des D-Dimers sehr hoch ist, anzunehmen, dass eine Lungenembolie bei dieser Patientin so unwahrscheinlich ist, dass weitere Untersuchungen unnötig sind. Die Sensitivität ist etwa 95 % - es werden also etwa 5 % der Patienten mit dem Test nicht erfasst. Bei dieser Patientin ist eine CT-Untersuchung der Lunge zum Nachweis oder Ausschluss einer Lungenembolie indiziert. Siehe Addendum 2*

Was ist der Unterschied zwischen den beiden Patienten. Bei beiden besteht der Verdacht auf eine Lungenembolie und bei beiden ist das D-Dimer im Blut nicht erhöht.

Beim Patienten ist die Wahrscheinlichkeit einer Lungenembolie (keine Risikofaktoren, Beschwerden eher untypisch, da sie nur kurz andauerten) klein und der normale D-Dimer-Wert stärkt diese Einschätzung.

Bei der Patientin ist die Wahrscheinlichkeit einer Lungenembolie (mehrere Risikofaktoren; akut aufgetretene Atemnot, die für eine Lungenembolie spricht, zumindest mit einer Lungenembolie vereinbar sind) hoch bis sehr hoch. Der normale D-Dimer-Wert reduziert zwar die Wahrscheinlichkeit, aber nicht so stark, dass auf weitere Untersuchungen verzichtet werden könnte.

Ist die Sensitivität oder Spezifität eines Tests nicht so hoch wie in den zuvor genannten idealtypischen Beispielen, so ist sein Nutzen für den diagnostischen Prozess eingeschränkt.

Ein illustratives Beispiel hierfür ist das C-reaktive Protein (CRP): Fragt man klinisch tätige Ärzte nach der Sensitivität oder Spezifität dieses Markers, kann wahrscheinlich keiner eine präzise Angabe machen. Dennoch ist die Bestimmung des CRP-Wertes ein wertvoller Test und wird routinemässig genutzt, weil ein erhöhter CRP-Wert als unspezifischer, aber klinisch relevanter Hinweis auf eine entzündliche oder infektiöse Genese gilt.

*Beispiel 1 (vereinfacht, zur Erläuterung des Konzeptes):*

*Eine 30-jährige Frau kommt in die Sprechstunde wegen Husten seit 4 Tagen (Intensität des Hustens in den vergangenen Tagen zunehmend), verstopfter Nase und abendlich erhöhter Körpertemperatur (maximal 37.9 °C). Der behandelnde Arzt vermutet als Grund der Beschwerden: banale Erkältung oder eine Lungenentzündung. Er schätzt die Wahrscheinlichkeit für eine Erkältung aufgrund seiner Erfahrung höher ein als die einer Lungenentzündung. Er lässt den CRP-Wert bestimmen. Dieser ist nur minimal erhöht. Aufgrund dieser Information – ohne, dass er die Vortestwahrscheinlichkeit und die Werte für die Sensitivität und Spezifität des CRP-Tests kennt – wird er eine Lungenentzündung als sehr unwahrscheinlich einschätzen – sein «degree of belief». Bei einer Lungenentzündung sind die CRP-Werte in der Regel hoch. Der Arzt stellt sich nun die Frage, ob die Wahrscheinlichkeit einer Lungenentzündung so klein ist, dass er auf eine Antibiotikatherapie verzichten kann. Wenn er nicht sicher fühlt, wird er ein Röntgenbild des Thorax anfertigen. Ist das Thoraxbild normal fühlt er sich sicher genug, um auf eine Therapie mit Antibiotika zu verzichten.*

Die klinischen Beispiele – Lungenembolie, akute Bronchitis und Pneumonie – verdeutlichen zentrale Prinzipien des *diagnostischen Denkens*, insbesondere des *Bayesianischen Denkens*. Dieser Ansatz basiert auf der Anwendung von Wahrscheinlichkeiten, um diagnostische Entscheidungen unter Unsicherheit zu treffen.

Am Beginn jeder diagnostischen Überlegung steht die sogenannte Vortest-Wahrscheinlichkeit – also eine erste Einschätzung, wie wahrscheinlich eine bestimmte Erkrankung beim jeweiligen Patienten oder bei der Patientin ist. Diese Einschätzung basiert auf Informationen aus der Anamnese, gegebenenfalls der körperlichen Untersuchung sowie dem persönlichen Wissen. Im genannten Beispiel wird angenommen, dass eine Lungenembolie beim männlichen Patienten eher unwahrscheinlich, bei der weiblichen Patientin jedoch sehr wahrscheinlich ist.

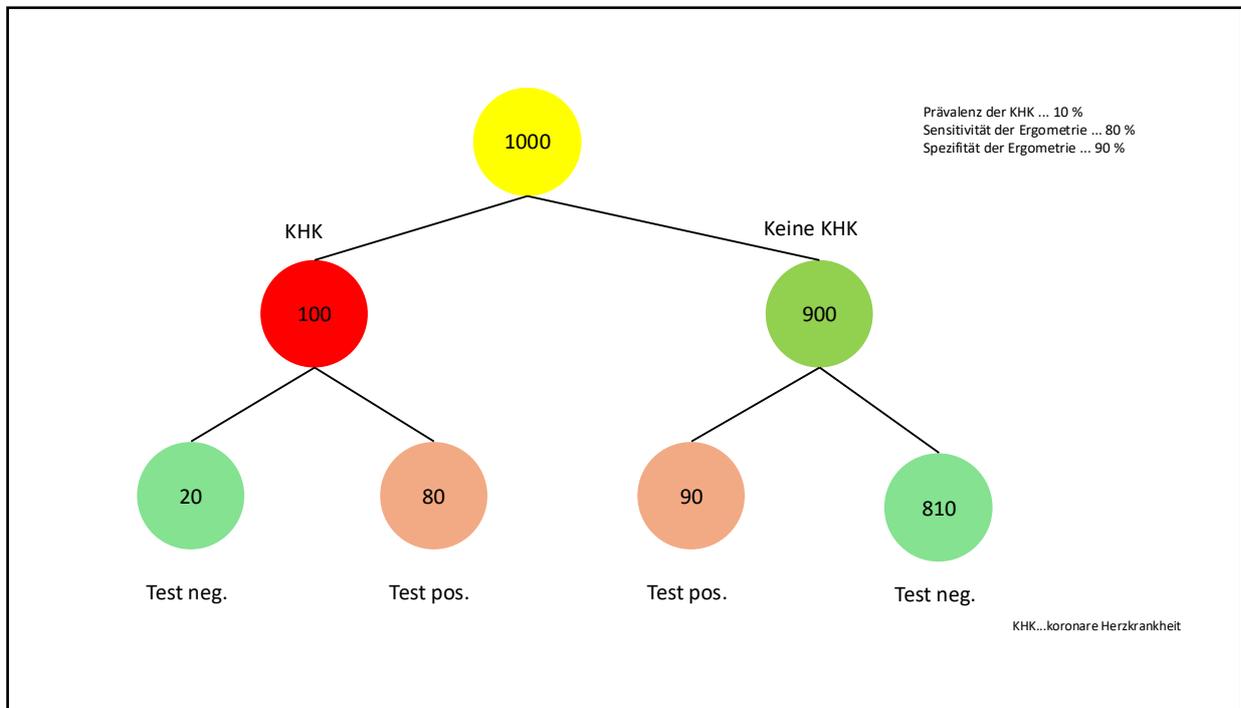
Ergeben sich im weiteren Verlauf neue Befunde/Informationen – beispielsweise ein erhöhter oder unauffälliger D-Dimer-Wert – wird diese ursprüngliche Einschätzung angepasst. Das Ergebnis dieser Neubewertung ist die Posttest-Wahrscheinlichkeit. Sie bildet die Grundlage für die Entscheidung, ob weitere diagnostische Maßnahmen (z. B. CT-Angiografie, Thoraxröntgenbild) erforderlich sind oder ob die Erkrankung mit ausreichender Sicherheit ausgeschlossen werden kann.

Zentral für dieses Vorgehen ist der Bayessche Wahrscheinlichkeitsbegriff. Anders als im frequentistischen Verständnis, bei dem Wahrscheinlichkeit als relative Häufigkeit in einer großen Zahl von Wiederholungen definiert wird (etwa die Chance, bei 1.000 Würfeln etwa  $1/6$  der Male eine Vier zu würfeln), versteht der Bayessche Ansatz Wahrscheinlichkeit als subjektiven Grad der Überzeugung – im Englischen als *degree of belief* bezeichnet.

Eine etwas ausführlichere Auseinandersetzung mit diesem Konzept folgt in einem separaten Artikel zum Thema Wahrscheinlichkeiten – der Essenz der medizinischen Entscheidungsfindung.

\*\*\*

Addendum 1: eine Alternative zur Vierfelder-Tafel für die Darstellung von Sensitivität und Spezifität.



Addendum 2: Bayes Theorem zur Berechnung der Posttestwahrscheinlichkeit  
Als Beispiel dient die Aussagekraft des D-Dimer Tests bei Verdacht auf eine Lungenembolie. Die Sensitivität des Dimers beträgt 95 % und die Spezifität liegt bei etwa 50 %. Die Vortestwahrscheinlichkeit (VTW) im ersten Beispiel bei 5 % und beim zweiten Beispiel bei 80 %.

Die (vereinfachte, aber korrekte) Formel von Bayes zur Berechnung der Posttestwahrscheinlichkeiten (PTW), wenn der D-Dimer-Wert im Normbereich ist (negatives Testergebnis)

$$\begin{aligned} \text{PTW} &= (1 - \text{Sensitivität}) \times \text{VTW} / (1 - \text{Sensitivität}) \times \text{VTW} + \text{Spezifität} \times (1 - \text{VTW}) \\ &= (1 - 0.95) \times 0.05 / (1 - 0.95) \times 0.05 + 0.5 \times (1 - 0.05) = 0.009 = 0.9 \% \end{aligned}$$

$$\begin{aligned} \text{PTW} &= (1 - \text{Sensitivität}) \times \text{VTW} / (1 - \text{Sensitivität}) \times \text{VTW} + \text{Spezifität} \times (1 - \text{VTW}) \\ &= (1 - 0.95) \times 0.80 / (1 - 0.95) \times 0.8 + 0.5 \times (1 - 0.8) = 0.28 = 28 \% \end{aligned}$$

## Referenzen

1. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *The American journal of medicine*. 1998;104(4):374-80.
2. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ*. 2002;324(7341):824-6.
3. Steurer J, Held U, Miettinen OS. Diagnostic probability function for acute coronary heart disease garnered from experts' tacit knowledge. *Journal of clinical epidemiology*. 2013;66(11):1289-95.
4. Whiting PF, Davenport C, Jameson C, Burke M, Sterne JA, Hyde C, et al. How well do health professionals interpret diagnostic information? A systematic review. *Bmj Open*. 2015;5(7):e008155.
5. Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *The American journal of medicine*. 1981;71(6):991-8.
6. Krouss M, Croft L, Morgan DJ. Physician Understanding and Ability to Communicate Harms and Benefits of Common Medical Treatments. *JAMA Intern Med*. 2016;176(10):1565-7.
7. Morgan DJ, Pineles L, Owczarzak J, Magder L, Scherer L, Brown JP, et al. Accuracy of Practitioner Estimates of Probability of Diagnosis Before and After Testing. *JAMA Intern Med*. 2021;181(6):747-55.
8. Heneghan C, Glasziou P, Thompson M, Rose P, Balla J, Lasserson D, et al. Diagnostic strategies used in primary care. *BMJ*. 2009;338:b946.
9. Parvu V, Gary DS, Mann J, Lin YC, Mills D, Cooper L, et al. Factors that Influence the Reported Sensitivity of Rapid Antigen Testing for SARS-CoV-2. *Front Microbiol*. 2021;12.
10. Moons KG, van Es G-A, Deckers JW, Habbema DJ, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and Bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology*. 1997;8(1):12-7.
11. Di Nisio M, Squizzato A, Rutjes AW, Buller HR, Zwinderman AH, Bossuyt PM. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. *J Thromb Haemost*. 2007;5(2):296-304.
12. Zaphiriou A, Robb S, Murray-Thomas T, Mendez G, Fox K, McDonagh T, et al. The diagnostic accuracy of plasma BNP and NTproBNP in patients referred from primary care with

suspected heart failure: results of the UK natriuretic peptide study. *Eur J Heart Fail.* 2005;7(4):537-41.

13. Byrne RA, Rossello X, Coughlan JJ, Barbato E, Berry C, Chieffo A, et al. 2023 ESC Guidelines for the management of acute coronary syndromes: Developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *European Heart Journal.* 2023;44(38):3720-826.

14. Marantz PR, Kaplan MC, Alderman MH. Clinical diagnosis of congestive heart failure in patients with acute dyspnea. *Chest.* 1990;97(4):776-81.

15. Niewold TB, Harrison MJ, Paget SA. Anti-CCP antibody testing as a diagnostic and prognostic tool in rheumatoid arthritis. *QJM.* 2007;100(4):193-201.

16. Shimada T, Noguchi Y, Jackson JL, Miyashita J, Hayashino Y, Kamiya T, et al. Systematic review and metaanalysis: urinary antigen tests for Legionellosis. *Chest.* 2009;136(6):1576-85.