

Reto Savoca<sup>1</sup>

# Interpretation von Laborresultaten: Bayes oder Bauchgefühl?

**Das Bayes'sche Theorem ist ein Versuch, mit dem Informationsgehalt eines binären (Labor-)Tests die anamnestisch geschätzte Krankheitswahrscheinlichkeit objektiv zu vergrössern oder zu verkleinern. Die Werkzeuge hierzu sind in Studien erhobene Daten zum jeweiligen Test wie Spezifität (= Negativitätsrate bei Gesunden) und Sensitivität (Positivitätsrate bei Kranken). Der positive bzw. negative prädiktive Wert macht eine Aussage darüber, wie wahrscheinlich es ist, ob ein einzelner Patient bei einem positiven Test krank bzw. bei einem negativen Test gesund ist. Die pos. und neg. prädiktiven Werte setzen jedoch voraus, dass man die Krankheitsprävalenz kennt.**

Alle, die sich intensiv mit Labor- oder anderen diagnostischen Resultaten befassen, müssen sich mit Begriffen wie «diagnostische Sensitivität und Spezifität», «positiver und negativer prädiktiver Wert» sowie «Likelihood-Ratio» beschäftigen. Viele tun sich schwer mit diesen Begriffen und erst recht mit deren Anwendung – und dies mit gutem Grund. Dieser Artikel versucht, etwas Licht in die Sache zu bringen.

**Sensitivität und Spezifität** sollen ein prävalenzunabhängiges, objektives Mass für die Leistungsfähigkeit eines diagnostischen Tests darstellen.

Die **Sensitivität** ist der prozentuale Anteil der Kranken, welche testpositiv werden oder:

**(Richtig Positive / Kranke)** bzw.  
**(Richtig Positive / (Richtig Positive + Falsch Negative)).**

Die **Spezifität** ist dagegen der prozentuale Anteil der Gesunden (= «Nichtkranken»), die testnegativ bleiben, also:

**(Richtig Negative / Gesunde)** bzw.  
**(Richtig Negative / (Richtig Negative + Falsch Positive)).**

Es ist einleuchtend, dass Sensitivität und Spezifität wesentlich vom Grenzwert oder Cut-off abhängen, also ab welchem Wert ein Test als «positiv» bezeichnet wird. Mehr noch: Die Werte für Sensitivität und Spezifität ändern sich gegenläufig, wenn der Grenzwert verschoben wird. Je höher also die Sensitivität, desto kleiner die Spezifität und umgekehrt (Abbildung 1).

Vereinfacht ausgedrückt, erkennen wir mit einem sehr tiefen Grenzwert zwar alle Kranken, markieren aber lei-

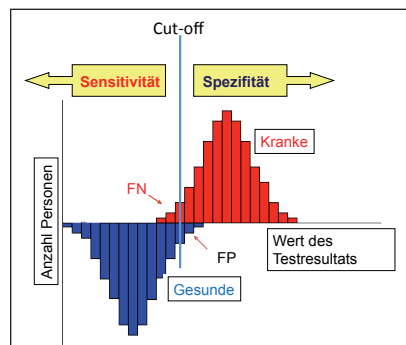


Abbildung 1

der sehr viele «Nichtkranke» ebenfalls fälschlicherweise mit «krank».

Oder wir verwenden einen sehr hohen Grenzwert: Jetzt sinkt der Anteil der Gesunden, die falsch positiv werden – dafür verpassen wir so einen Grossteil der Kranken.

## Aus der Praxis

Ein Beispiel dafür sind die Tests für das kardiale Troponin der vierten Generation: hier konnte der Grenzwert (Cut-off) dank der besseren analytischen Leistungsfähigkeit, sprich bessere Impräzision im tiefen Messbereich, nach unten gesetzt werden. Es gibt jetzt deutlich weniger akute Koronarsynndrome (ACS), die wegen negativer Troponinwerte verpasst werden (bessere Sensitivität!), hingegen nehmen jetzt «positive» Troponinwerte zu, die nicht auf ein ACS, sondern z.B. auf sonstige Myokardschäden zurückgeführt werden müssen («schlechtere» Spezifität bezüglich ACS!).

Leider immer noch beliebt sind «Krebs-tests» – häufig auch auf Wunsch von Patienten. Nehmen wir beispielsweise an, bei einem 35-jährigen offensichtlich gesunden Probanden ohne Informationen über Risiken sei ein Alpha-Feto-Protein-Test (AFP) positiv. AFP

hat bezüglich des hepatozellulären Karzinoms (HCC) eine Sensitivität von nur ca. 50–70% und eine Spezifität von ca. 80%.

Wie hoch ist also die Wahrscheinlichkeit, dass dieser Proband ein HCC hat? Natürlich ist sie nicht 50–70%, sondern hier kommt der **positive prädiktive Wert (PPV)** oder positiver Voraussagewert ins Spiel.

Der PPV berechnet sich aus:

**(Richtig Positive / Alle Positive) = (Richtig Positive / (Richtig Positive + FALSCH Positive)).**

Wie hoch der positive prädiktive Wert wird, hängt wesentlich von der Population ab, mit der man ihn berechnet (prävalenzabhängig). In diesem Beispiel sei die geschätzte Vortest-Wahrscheinlichkeit für ein HCC ca. 0,2% (Inzidenz: 5/100 000 p. Jahr).

Richtig Positive (RP):

70% (Sens.) von 0,2% (Wahrscheinlichkeit für Kranke) = 0,14%

Falsch Positive (FP):

20% (Unspezifität = 100% – Spezifität) von 99,8% (Wahrscheinlichkeit für Gesunde) = 19,96%

Positiv prädiktiver Wert:  $RP / (RP + FP) = 0,14 / (19,96 + 0,14) = 0,7\%$

Bei einem Individuum unbekannter Anamnese mit unbekanntem Risiko-status ist die Wahrscheinlichkeit für ein HCC ca. 0,2%, wenn kein Testresultat vorliegt und 0,7%, wenn ein positiver AFP-Test ermittelt wurde.

Die Wahrscheinlichkeit bzw. die Erhöhung der Wahrscheinlichkeit reicht sicherlich nicht aus, um weitere Abklärungen zu rechtfertigen, das positive Resultat ist aber häufig sehr beunruhigend für die Patienten.

<sup>1</sup> Dr. Reto Savoca, FAMH Klin.Chemie und Hämatologie, Laborleiter Zentrallabor, Spitäler Schaffhausen

## Interprétation de résultats de laboratoire: Bayes ou intuition?

Le théorème de Bayes est une tentative d'augmenter ou de diminuer objectivement la probabilité de maladie estimée par anamnèse en utilisant les informations fournies par un test (de laboratoire) binaire. Les outils utilisés à cet effet sont les données issues d'études concernant le test en question, comme la spécificité (= taux de négativité chez des sujets sains) et la sensibilité (= taux de positivité chez des sujets malades). La valeur prédictive positive ou négative renseigne sur la probabilité qu'un patient individuel soit malade en cas de test positif ou qu'il soit en bonne santé en cas de test négatif. Les valeurs prédictives positives et négatives présupposent néanmoins que la prévalence de la maladie soit connue.

Les faiblesses du théorème de Bayes incluent l'absence fréquente de données concernant la sensibilité et la spécificité en rapport avec différentes maladies et stades de maladies. Par ailleurs, cette approche ne tient pas compte des informations quantitatives (niveau de la valeur positive). Ces deux faiblesses altèrent la valeur du concept dans la pratique.

Selbstverständlich sieht die Sache wesentlich anders aus, wenn Risikofaktoren wie eine Zirrhose oder eine chronische Hepatitis-B oder C-Infektion bekannt sind. Dann steigt die Vortestwahrscheinlichkeit deutlich an und entspricht nicht mehr der «Gesamtbevölkerungswahrscheinlichkeit», sondern dem erhöhten persönlichen Risiko. Den Einfluss der Prävalenz auf den positiven (und negativen) prädiktiven Wert für diesen und einen zweiten Test mit besserer diagnostischer Leistung ist in Tabelle 1 zu sehen.

Das Gegenstück zum positiv prädiktiven Wert ist der **negativ prädiktive Wert** oder negativer Voraussagewert NPV. Der NPV berechnet sich aus: **(Richtig Negative / Alle Negative) = (Richtig Negative / (Richtig Negative + FALSCH Negative))**. Je höher der negativ prädiktive Wert,

Tabelle 1:	Test 1		Test 2	
	(Sens. 70%, Spez. 80%)		(Sens. 99%, Spez. 95%)	
Prävalenz (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)
0,2	0,7	99,9	3,8	100,0
0,4	1,4	99,8	7,4	100,0
1	3,4	99,6	16,7	100,0
2	6,7	99,2	28,8	100,0
4	12,7	98,5	45,2	100,0
8	23,3	96,8	63,3	99,9
16	40,0	93,3	79,0	99,8
32	62,2	85,0	90,3	99,5
50	77,8	72,7	95,2	99,0

desto höher ist die Wahrscheinlichkeit, dass ein testnegativer Patient die Krankheit NICHT hat. Genau wie beim PPV ist der Wert des negativ prädiktiven Wertes wesentlich von der untersuchten Population (Prävalenz!) abhängig.

### Likelihood-Ratio

Wollen wir ein Mass für den Informationsgewinn, den uns ein Test liefern kann, verwenden wir die Likelihood-Ratio. Die Likelihood-Ratio (LR+) für den POSITIVEN Test gibt darüber Auskunft, wie viel wahrscheinlicher ein positiver Test bei Kranken im Vergleich zu Gesunden ist, sie ist also ein Mass für die **Einschlusskraft** des Tests.

Die LR+ ist gleich **Sensitivität / (1-Spezifität)**.

Vereinfacht ausgedrückt: Je sensitiver UND je weniger unspezifisch ein Test ist, desto höher die Krankheitswahrscheinlichkeit bei einem positiven Test. Die Vortestwahrscheinlichkeit kann nun mit der LR+ multipliziert werden und man erhält eine höhere Nachtestwahrscheinlichkeit.

Ein Test mit einer LR+ von 1 hat keine Aussage – Vor- und Nachtestwahrscheinlichkeit sind gleich. Ein guter Test hat jedoch eine LR+ von 10 und mehr.

**ACHTUNG FALLSTRICK:** Wir benötigen, um mit Likelihood Ratios rechnen zu können, nicht die Wahrscheinlichkeit für eine Erkrankung in %, sondern in «odds» = «Chancen», oder anders ausgedrückt in **krank/gesund**. Beispiel: Die Vortestwahrscheinlichkeit sei 20%, die odds sind damit 2:8 = 1:4. Bei einer LR+ von 2 erhalten wir die Nachtestodds von 2:4 = 1:2 (oder 33%, wieder umgerechnet auf prozentuale Wahrscheinlichkeiten).

Das Gegenstück zur LR+ ist die LR-, also die Angabe, wie viel häufiger ein negatives Resultat bei einem Test bei Gesunden zu erwarten ist – die Ausschlusskraft eines Tests.

Die LR- ist gleich **(1-Sensitivität) / Spezifität**. Bei der LR- sind Werte <1 hilfreich, also möglichst 0,1 oder kleiner.

Bei der LR- sind Werte <1 hilfreich, also möglichst 0,1 oder kleiner.

### Caveats

Bei allem nötigen Respekt für die Statistik – die praktische Anwendung ist nicht unproblematisch, denn die Daten für Sensitivität und Spezifität sind nicht für alle Tests und alle Erkrankungen ermittelt worden, sie ändern je nach Krankheitsstadium markant – beim ACS beispielsweise innerhalb von Stunden!

Zudem wird die Sichtweise «Negativ – Positiv» einem quantitativen Test in keiner Weise gerecht: Hier geht viel Information verloren. Jeder Kliniker wird jedoch mit Blick auf extrem pathologische Werte diese Zusatzinformation intuitiv berücksichtigen.

Es schadet nicht, die beschriebenen Konzepte für einen sinnvollen Einsatz von (Labor)-Analytik im Hinterkopf zu haben; dennoch: Aufgrund der häufig fehlenden Information, mässiger Praktikabilität und hoher Komplexität: Bauchgefühl schlägt Bayes.

Korrespondenz:  
Reto.Savoca@spitaeler-sh.ch