

Wahrscheinlichkeitsverhältnis (Likelihood Ratio) - Alternative zu Sensitivität und Spezifität

In einem früheren EbM-Splitter [6] wurden die Maßzahlen Sensitivität und Spezifität vorgestellt, welche die Beurteilung der Güte eines diagnostischen Tests ermöglichen. Nur wenig benutzt werden zu diesem Zweck die sogenannten Wahrscheinlichkeitsverhältnisse (*Likelihood Ratios*), welche in dieser Arbeit anhand eines Beispiels zur Diagnostik von Kiefergelenkschmerzen [7] dargestellt werden.

Wahrscheinlichkeitsverhältnisse

Das Wahrscheinlichkeitsverhältnis gibt das Verhältnis der Wahrscheinlichkeiten wieder, dass bei einer *erkrankten* Person ein positives oder negatives Testergebnis beobachtet wird, im Vergleich zur entsprechenden Wahrscheinlichkeit bei einer gesunden Person.

Betrachtet man ein positives Testergebnis, so spricht man von einem *positiven Wahrscheinlichkeitsverhältnis (LR+)*, definiert als

$$LR+ = \frac{\text{Wahrscheinlichkeit eines positiven Testergebnisses bei erkrankten Personen}}{\text{Wahrscheinlichkeit eines positiven Testergebnisses bei gesunden Personen}}$$

Das positive Wahrscheinlichkeitsverhältnis kann mittels Sensitivität und Spezifität berechnet werden: $LR+ = \frac{\text{Sensitivität}}{(1 - \text{Spezifität})}$. Ein Wert von Eins bedeutet, dass die Wahrscheinlichkeit für ein positives Testergebnis bei Gesunden und Kranken gleich groß ist. Die diagnostische Maßnahme kann somit nicht in sinnvoller Weise verwendet werden, um erkrankte Personen zu erkennen. Je größer das positive Wahrscheinlichkeitsverhältnis ist, desto besser ist die diagnostische Maßnahme geeignet, eine erkrankte Person zu erkennen. Eine Schätzung des positiven Wahrscheinlichkeitsverhältnisses ist mittels der geschätzten Werte von Sensitivität und Spezifität möglich bzw. direkt durch $\frac{a}{a+c} / \frac{b}{b+d}$ gegeben (siehe Tab. 2).

Betrachtet man ein negatives Testergebnis, so spricht man von einem *negativen Wahrscheinlichkeitsverhältnis (LR-)*, definiert als

$$LR- = \frac{\text{Wahrscheinlichkeit eines negativen Testergebnisses bei erkrankten Personen}}{\text{Wahrscheinlichkeit eines negativen Testergebnisses bei gesunden Personen}}$$

Das negative Wahrscheinlichkeitsverhältnis kann ebenfalls mittels Sensitivität und Spezifität berechnet werden: $LR- = \frac{(1 - \text{Sensitivität})}{\text{Spezifität}}$. Ein Wert von Eins bedeutet, dass die Wahrscheinlichkeit für ein negatives Testergebnis bei Gesunden und Kranken gleich groß ist. Die diagnostische Maßnahme kann daher nicht in sinnvoller Weise verwendet werden, um gesunde Personen zu erkennen. Je kleiner das negative Wahrscheinlichkeitsverhältnis ist, desto besser ist die diagnostische Maßnahme geeignet, eine gesunde Person zu erkennen. Eine Schätzung des negativen Wahrscheinlichkeitsverhältnisses ist mittels der geschätzten Werte von Sensitivität und Spezifität möglich bzw. direkt durch $\frac{c}{a+c} / \frac{d}{b+d}$ gegeben. Für beide Wahrscheinlichkeitsverhältnisse kann man Konfidenzintervalle berechnen [8].

In Anlehnung an Jaeschke et al. [4] kann eine grobe Einteilung der Wahrscheinlichkeitsverhältnisse (LR) vorgenommen werden, anhand derer die Güte einer diagnostischen Maßnahme bewertet werden kann (Tab. 1).

Wahrscheinlichkeitsverhältnisse	Interpretation
LR+ > 10 bzw. LR- < 0,1	„überzeugende diagnostische Evidenz“
LR+ 5 - 10 bzw. LR- 0,1 - 0,2	„hohe diagnostische Evidenz“
LR+ 2 - 5 bzw. LR- 0,2 - 0,5	„schwache diagnostische Evidenz“
LR+ 1 - 2 bzw. LR- 0,5 - 1	„kaum relevante diagnostische Evidenz“

Tabelle 1 Wahrscheinlichkeitsverhältnisse und ihre Interpretation
(in Anlehnung an [4])

Zur umfassenden Bewertung der Güte einer diagnostischen Maßnahme muss man dementsprechend positives und negatives Wahrscheinlichkeitsverhältnis betrachten.

„Weder Plausibilität noch der Umstand, dass etwas an Universitäten gelehrt wird, noch die Tatsache, dass etwas gängige Meinung ist, sind Garantien für dessen Richtigkeit.“

Fredmund Malik,

Leiter des Management Zentrum St. Gallen und Bestsellerautor [5]

Beispiel: Effusionen im Kernspintogramm zur Diagnose von Kiefergelenkschmerzen

Shaefter et al. [7] gingen der Frage nach, ob mit Hilfe von (T2-gewichteten) Kernspintogrammen diagnostizierte Flüssigkeitsansammlungen (Effusionen) im Kiefergelenk als Gradmesser für vorhandene Kiefergelenkschmerzen angesehen werden können (vgl. Zeitschriftenreferat von Hugger [3]). In die Studie wurden 30 Frauen aufgenommen, bei denen unter Zugrundelegung standardisierter klinischer Diagnosekriterien [1] eine ein- oder beidseitige anteriore Diskusverlagerung mit Reposition diagnostiziert wurde. Insgesamt 14 der 30 Frauen litten seit wenigstens 6 Monaten an beidseitigen Kiefergelenkschmerzen; diese 14 Frauen wurden als „erkrankt“ eingestuft. Die restlichen 16 Frauen wurden als Kontrollgruppe (keine Kiefergelenkschmerzen) geführt.

Bei allen Frauen wurden jeweils beidseitig Kernspintographien der Kiefergelenke angefertigt, welche von einem Kliniker auf das Vorliegen von Effusionen geprüft wurden. Das Ergebnis dieser Studie ist in Tabelle 2 dargestellt. Man beachte, dass für jede Frau jeweils zwei Messwerte vorliegen (rechtes und linkes Kiefergelenk).

Für Sensitivität und Spezifität ergeben sich Werte von **86% (24/28)** und **28% (9/32)**. Aufgrund der sehr geringen Spezifität wird von den Autoren das Vorliegen von Effusionen

		Kiefergelenkschmerzen		Σ
		Ja	Nein	
Effusion im Kernspintomogramm	Ja	a 24	b 23	a+b 47
	Nein	c 4	d 9	c+d 13
Σ		a+c 28	b+d 32	N 60

Tabelle 2 Effusionen im Kernspintomogramm zur Diagnose von Kiefergelenkschmerzen [7]

nen im Kernspintomogramm als unzureichende Methode zur Diagnose von Kiefergelenkschmerzen eingestuft.

Für das positive Wahrscheinlichkeitsverhältnis $LR+$ ergibt sich ein Wert von **1,19** [$86\% / (100 - 28\%) = 0,86 / 0,72$]. Für das negative Wahrscheinlichkeitsverhältnis $LR-$ ergibt sich ein Wert von **0,51** ($0,14 / 0,28$).

Der Wert des positiven Wahrscheinlichkeitsverhältnisses liegt sehr nahe am Wert Eins, d.h. das Vorliegen von Effusionen im Kernspintomogramm ist keine geeignete Diagnosemethode, um Patienten mit Kiefergelenkschmerzen zu erkennen. Dies spiegelt sich auch in der Einteilung nach Jaeschke et al. [4] wider.

Der Wert des negativen Wahrscheinlichkeitsverhältnisses ist weiter von der Eins entfernt: Die Wahrscheinlichkeit für ein negatives Testergebnis ist bei Patienten mit Kiefergelenkschmerzen nur halb so groß wie bei schmerzfreien Personen. In der Einteilung nach Jaeschke et al. [4] fällt aber auch das negative Wahrscheinlichkeitsverhältnis in die Gruppe mit der geringsten Güte.

Schlussbemerkungen

In diesem EbM-Splitter wurden die positiven und negativen Wahrscheinlichkeitsverhältnisse als eine Alternative zu Sensitivität und Spezifität zur Beurteilung der Güte einer diagnostischen Maßnahme dargestellt. Da sich die Wahrscheinlichkeitsverhältnisse direkt mittels Sensitivität und Spezifität berechnen lassen (und umgekehrt), ist durch die Verwendung der Wahrscheinlichkeitsverhältnisse für die der Tabelle 2 zugrundeliegenden Daten kein zwingender Vorteil zu erkennen. Ein Vorteil der Wahrscheinlichkeitsverhältnisse im Vergleich zu Sensitivität und Spezifität besteht darin, dass Wahrscheinlichkeitsverhältnisse auch berechnet werden können, wenn mittels der diagnostischen Maßnahme eine Einteilung in mehr als zwei Gruppen vorgenommen wird. Wenn im verwendeten Beispiel bei der Bewertung des Vorliegens von Effusionen eine Einteilung in drei Gruppen durchgeführt worden wäre (ja/unklar/nein), könnte man weiterhin Wahrscheinlichkeitsverhältnisse berechnen. Sensitivität und Spezifität hingegen sind ausschließlich für Vierfeldertafeln definiert.

Nebenbemerkung: Vorsicht bei der Verwendung mehrerer Messungen bei einer Person!

In der Studie von Shaefer et al. [7] wurden bei 30 Frauen jeweils zwei Messungen (rechtes und linkes Kiefergelenk) in die Auswertung aufgenommen. Dies ist die gängige Vorgehensweise bei Diagnosestudien in der Zahnmedizin. Beispielsweise wurden in der Studie von Heinrich et al. [2], welche wir in einem früheren EbM-Splitter [6] als Beispiel verwendet haben, bei 186 Probanden insgesamt 5722 Approximallflächen untersucht. In Studien mit Mehrfachmessungen ist zur Beschreibung der Studiengröße primär die Anzahl

von Probanden und nicht die Anzahl von Messungen relevant.

Die verwendeten statistischen Methoden zur Auswertung von Diagnosestudien (z.B. Schätzung von Konfidenzintervallen für Sensitivität und Spezifität) basieren auf der Annahme der Unabhängigkeit der einzelnen diagnostischen Testergebnisse. Diese Annahme ist nicht erfüllt, falls für eine Person mehrere Messungen vorliegen. In der Studie von Shaefer et al. [7] beispielsweise ist nicht klar, ob der Kliniker die beiden Kernspintomogramme einer Patientin separat (d.h. ohne Kenntnis des zweiten Bildes) bewertet hat. Bei der gleichzeitigen Analyse der Kernspintomogramme wurden die beiden Ergebnisse einer Patientin möglicherweise nicht unabhängig voneinander interpretiert. Dazu kommt, dass bei systemisch-rheumatologischen Erkrankungen mit einer gewissen Wahrscheinlichkeit ähnliche Befunde im rechten und linken Kiefergelenk angetroffen werden. Neben diesen untersucher- und krankheitsbezogenen Faktoren kann die Verwendbarkeit und Interpretation eines Kernspintomogramms noch von technischen und patientenbezogenen (z.B. anatomischen) Faktoren abhängen.

Streng genommen ist somit die Berechnung von Wahrscheinlichkeitsverhältnissen bzw. von Sensitivität und Spezifität für die der Tabelle 2 zugrundeliegenden Daten nicht zulässig. Eine adäquate Auswertung würde in obigem Beispiel darin bestehen, eine Vierfeldertafel separat für die rechte und linke Seite zu erstellen und die Güte des diagnostischen Tests getrennt für die beiden Seiten zu beurteilen; diese detaillierte Information fehlt allerdings in der Publication von Shaefer et al. [7] wie auch in vielen anderen Darstellungen. Für die der Studie von Heinrich et al. [2] zugrundeliegenden Daten ist uns keine vergleichbare adäquate statistische Auswertungsmethode bekannt. Die Entwicklung von statistischen Verfahren für Diagnosestudien mit mehreren abhängigen Beobachtungen, wie sie in der Zahnmedizin typisch sind, stellt somit noch einen Bereich dar, in dem zukünftig statistisch-methodische Forschung notwendig ist.

Literatur

1. Dworkin S. F., LeResche, L.: Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. J Craniomandib Disord Facial Oral Pain 6, 301 (1992).
2. Heinrich, R., Künzel, W., Towfiq, H.: Approximate Kariesdiagnostik-Vergleich klinischer faseroptischer und röntgenographischer Diagnoseverfahren (Approximalkariesdiagnostik). Dtsch Zahn Mund Kieferheilkd Zentralbl 79, 535 (1991).
3. Hugger, A.: Shaefer, I. R., Jackson, D. L., Schiffman, E L., Anderson, Q. N.: Pressure-pain thresholds and MRI effusions in TMJ arthralgia. [Zeitschriftenreferat] Dtsch Zahnärztl Z 57, 5 (2002).
4. Jaeschke, R., Guyott, G., Sackett, D. L.: Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? J Am Med Assoc 271, 703 (1994). Mohk, F.: Führen, leiten, leben. 9. Aufl. Deutsche Verlags-Anstalt, Stuttgart ZOOS, n.
5. Schworzer, G., Türp, I. C., Antes, G.: EbM-Splitter: Die Vierfeldertafel (in Diagnosestudien): Sensitivität und Spezifität. Dtsch Zahnärztl Z 57, 333 (2002).
6. Shaefer, I. R., Jackson D. L., Schiffman, E L., Anderson, Q. N.: Pressure-pain thresholds and MRI effusions in TMJ arthralgia. J Dent Res 80, 1935 (2001).
7. Simel, D. L., Somo, G. F., Molchar, D. 8.: Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 44, 763 (1991).

Guido Schwarzer, Freiburg
Jens C. Türp, Freiburg / Basel
Gerd Antes, Freiburg