

Über die Aussagekraft von Untersuchungsbefunden

Evidenzbasierte klinische Untersuchung

Adrian O. Goeldlin^a, Martin Perrig^b

^a Universität Bern, Berner Institut für Hausarztmedizin BIHAM

^b Universitätsklinik für Allgemeine Innere Medizin, Inselspital Bern

Auf grosses Interesse an der SwissFamilyDocs Conference 2015 stiess der interaktive Workshop über evidenzbasierte klinische Untersuchung. Die Autoren diskutierten mit dem erfahrenen und engagierten Publikum die Wertigkeit verschiedener klinischer Untersuchungen anhand von Fallvignetten.

Ist die Erhebung des Status im 21. Jahrhundert nur noch ein Relikt aus den Anfängen der Medizin, und beschränkt sich seine Bedeutung heute auf ein reines «Handauflegen», wie ein britischer Kollege bemerkt [1]?

Likelihood-Ratio: eine Zahl genügt

Klinische Untersuchungen sind, wie Laboranalysen oder bildgebende Verfahren, diagnostische Tests, deren Resultat eine Verdachtsdiagnose erhärtet oder abschwächt. Die Interpretation eines Testresultats erfolgt jeweils unter Berücksichtigung der Vortestwahrscheinlichkeit einer Krankheit in einem bestimmten Setting und der Qualität des Tests. Jeder Test zeichnet sich durch seine diagnostische Wertigkeit aus, die mittels Sensitivität und Spezifität angegeben wird. Zur Erinnerung: Tests mit hoher Sensitivität haben wenig falsch negative Resultate, Tests mit hoher Spezifität wenig falsch positive. Anstelle von Sensitivität und Spezifität kann jedoch die Aussagekraft eines Tests auch nur mit einer einzigen Zahl, der Likelihood-Ratio (LR), dargestellt werden:

$$LR+ = \frac{\text{Wahrscheinlichkeit eines positiven Befundes bei Patienten mit der gesuchten Krankheit}}{\text{Wahrscheinlichkeit desselben positiven Befundes bei Patienten ohne die gesuchte Krankheit}}$$

LR+ steht für ein positives Testresultat, welches die Posttestwahrscheinlichkeit erhöht, LR- für ein negatives Testresultat, welches die Posttestwahrscheinlichkeit verringert.

Für epidemiologisch Interessierte ergibt sich folgende Beziehung zwischen Sensitivität und Spezifität und Likelihood-Ratio:

$$\text{Positive LR} = \frac{\text{Sensitivität}}{1 - \text{Spezifität}}$$

$$\text{Negative LR} = \frac{1 - \text{Sensitivität}}{\text{Spezifität}}$$

Als Praktiker wird man kaum je eine Likelihood-Ratio berechnen, sondern auf entsprechende Angaben aus Standardwerken zurückgreifen [2, 3]. Um die Aussagekraft eines Tests abzuschätzen, reichen für den Alltagsgebrauch Näherungswerte. Annäherungsweise gilt für den Einfluss der Likelihood-Ratio auf die Nachtestwahrscheinlichkeit beispielsweise (Tab. 1):

- Likelihood-Ratio von 2 erhöht die Nachtestwahrscheinlichkeit um 15%.
- Likelihood-Ratio von 5 erhöht die Nachtestwahrscheinlichkeit um 30%.
- Likelihood-Ratio von 10 erhöht die Nachtestwahrscheinlichkeit um 45%.

Tabelle 1: Tabellarische Darstellung des approximativen Abschätzens der Nachtestwahrscheinlichkeit bei bekannter Vortestwahrscheinlichkeit und bekannter Likelihood-Ratio.

Likelihood-Ratio	...	0,1	0,2	0,5	1	2	5	10	...
Δ Nachtestwahrscheinlichkeit		-45%	-30%	-15%	±0%	+15%	+30%	+45%	

- Analog reduzieren Likelihood-Ratios von 0,5, 0,2 und 0,1 die Nachtstestwahrscheinlichkeit um 15%, 30% bzw. 45%.

Eine genaue Bestimmung der Nachtstestwahrscheinlichkeit ermöglicht die Anwendung des Fagan-Nomogramms [4], bei dem die Vortestwahrscheinlichkeit und Likelihood-Ratio eingetragen, mit einer Linie verbunden und danach die Nachtstestwahrscheinlichkeit abgelesen wird (vgl. auch Abb. 1). Da wir in der klinischen Medizin keine Standardbedingungen vorfinden, sind für die Anwendung der Likelihood-Ratios folgende zwei Voraussetzungen zu berücksichtigen:

- Likelihood-Ratios für klinische Tests sind nur gültig in dem Setting, in dem sie erhoben wurden.
- Damit die Aussagekraft eines klinischen Tests abgeschätzt werden kann, muss die Untersuchung richtig durchgeführt werden.

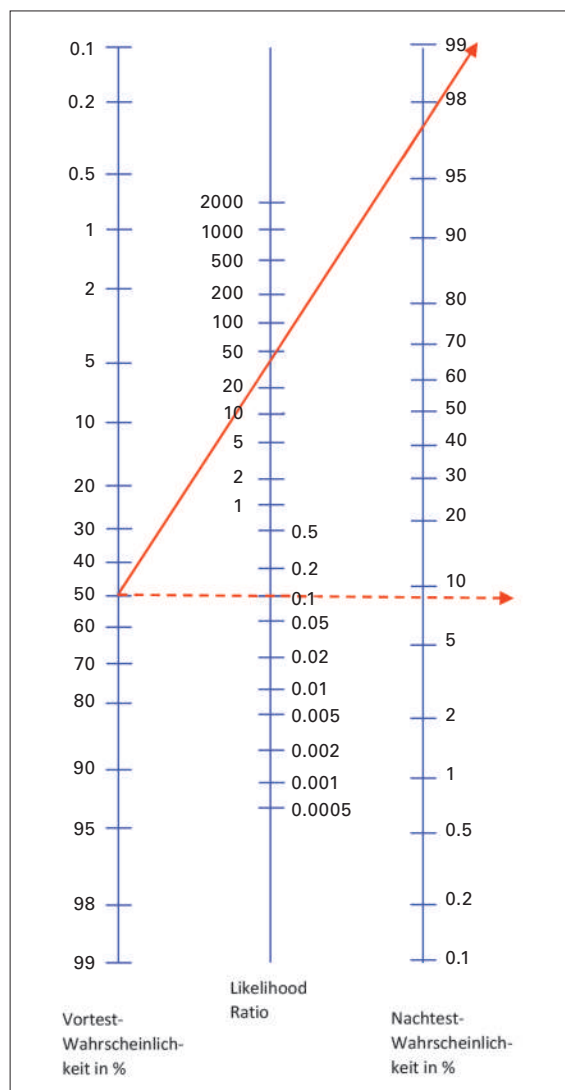


Abbildung 1: Fagan-Nomogramm.

Untersuchung nach einer Knieverletzung beim Fussballspiel

Ein Fussballspieler erleidet eine Knieverletzung und sucht Sie in Ihrer Sprechstunde auf. Aufgrund Ihrer sorgfältigen Anamnese über den Unfallhergang und Ihrer klinischen Erfahrung schätzen Sie die Wahrscheinlichkeit, dass er eine vordere Kreuzbandruptur erlitten hat, auf 50%. Welche klinische Untersuchung hilft Ihnen nun am raschesten weiter mit dieser Fragestellung? Als erfahrene Klinikern oder erfahrener Kliniker werden Sie wahrscheinlich die «vordere Schublade» prüfen, d.h. den Lachman-Test durchführen. Wie aussagekräftig ist diese Untersuchung betreffend die Wahrscheinlichkeit einer vorderen Kreuzbandruptur?

Bei richtiger Durchführung beträgt die Likelihood-Ratio eines positiven Lachman-Tests 42, diejenige eines negativen 0,1 [5, zitiert nach 6]. Tragen wir jetzt die Vortestwahrscheinlichkeit von 50% und die positive Likelihood-Ratio von 42 in ein Fagan-Nomogramm ein (Abb. 1), können wir rasch ablesen, dass in dieser Situation bei positivem Lachman-Test die Wahrscheinlichkeit einer vorderen Kreuzbandruptur ca. 97% beträgt. Bei negativem Test liegt die Wahrscheinlichkeit einer vorderen Kreuzbandruptur unter 10%.

Kombination verschiedener klinischer Befunde bei Pneumonieverdacht

Schon holen Sie die nächste Patientin aus dem Wartezimmer: eine 53-jährige Frau mit einer starken Erkältung, Heiserkeit und Myalgien. Begonnen hat die Erkrankung vor über einer Woche mit trockenem Husten. Seit 2 Tagen wird der Husten zunehmend produktiv, und der Allgemeinzustand der Patientin verschlechtert sich. Sie hat Fieber 38,3°C axillär, der Puls beträgt 110/min, im Status hören Sie rechts basal Rasselgeräusche.

Aufgrund dieser Anamnese wird klar, dass differentialdiagnostisch vor allem eine Bronchitis, ein Infekt der oberen Luftwege, ein Asthma, eine Sinusitis oder eine Pneumonie in Frage kommt. Natürlich wollen Sie vor allem die Lungenentzündung nicht verpassen, da die Mortalität und Morbidität einer unbehandelten bakteriellen Pneumonie auch im Jahr 2016 recht hoch ist. Helfen Ihnen die erhobenen Untersuchungsbefunde weiter bei der Frage, ob eine Pneumonie vorliegt?

Tabelle 2 zeigt, dass Fieber, Tachykardie und Rasselgeräusche als Einzelbefunde keinen sehr grossen Einfluss auf die Nachtstestwahrscheinlichkeit haben. Weder das Vorliegen noch das Fehlen eines einzelnen dieser Befunde erhöht oder erniedrigt die Wahrscheinlichkeit einer Pneumonie sehr relevant. Es stellt sich die Frage,

Tabelle 2: Sensitivität, Spezifität und Likelihood-Ratios ausgewählter Befunde bezüglich Vorliegen einer Pneumonie. Nach McGee [3].

Zeichen	Sensitivität in %	Spezifität in %	LR+ (vorhanden)	LR- (fehlend)
Fieber >37,8°C	27–69	49–94	2,2	0,7
Puls >100/min	17–65	60–92	1,7	0,8
Rasselgeräusche	19–67	36–94	1,8	0,8

ob und in welchem Ausmass eine Kombination dieser Befunde die Wahrscheinlichkeit verändert?

Gemäss Literatur dürfen Einzelbefunde kombiniert werden, allerdings maximal drei Befunde, Tests oder Zeichen. Diese dürfen pathophysiologisch nicht voneinander abhängig sein [3]. Fieber und Tachykardie beispielsweise sind nicht per se voneinander unabhängig. Rasselgeräusche (LR+ 1,8 für Pneumonie) und Fieber (LR+ 2,2 für Pneumonie) haben jedoch keine gemeinsame pathophysiologische Grundlage. Also dürfen wir die beiden kombinieren. Legen wir eine Vortestwahrscheinlichkeit von 20% für das Vorliegen einer Pneumonie im hausärztlichen Setting zugrunde, wird diese durch das Vorliegen von Fieber auf ca. 30% erhöht (Eintrag im Fagan-Nomogramm, exakte Berechnung: 35,5%). Diese neue Vortestwahrscheinlichkeit wird durch das Vorliegen von Rasselgeräuschen auf 48% erhöht (Eintrag im Nomogramm, exakte Berechnung: 49,8%). Nehmen wir die im ersten Abschnitt besprochenen Näherungswerte zu Hilfe (LR+ beider Zeichen je ca. 2), dann ergibt sich eine Erhöhung der Vortestwahrscheinlichkeit von 20% auf $20 + 15 + 15 = 50\%$, was sehr nah beim errechneten Wert liegt.

Will man mehr als drei Befunde oder Befunde mit gemeinsamer pathophysiologischer Grundlage kombinieren, darf dieses Vorgehen nicht angewendet werden. In diesem Fall werden sie als Score in klinischen

Studien validiert und erhalten als Ganzes eine eigene Likelihood-Ratio. Für den Fall der Pneumonie haben dies Heckerling und Kollegen [7] gemacht. Die sogenannte Heckerling-Regel finden Sie in der Tabelle 3.

Übereinstimmung zwischen Untersuchern: ein wichtiger Faktor

Schon länger betreuen Sie einen 65-jährigen Raucher (30 py) mit koronarer Herzkrankheit. Nun konsultiert er Sie wegen seit drei Wochen zunehmender Anstrengungsdyspnoe und produktiven Hustens. Sie denken an Herzinsuffizienz oder COPD. Klinisch finden Sie einen erhöhten Jugularvenendruck, einen dritten Herzton, bilaterale Rasselgeräusche und wenig expiratorisches Giemen.

Unterstützen diese Befunde die Hypothese Herzinsuffizienz? Tatsächlich bestätigen Studien bei diesen Befunden folgende hohe Likelihood-Ratios für das Vorliegen einer Herzinsuffizienz [2]: 3. Herzton (LR+ 11), Hepatojugulärer Reflux (LR+ 6,4), gestaute Halsvenen (LR+ 5,1). Der 3. Herzton scheint ein überaus zuverlässiger Untersuchungsbefund zu sein. Leider zeigt jedoch ausgerechnet dieser Befund eine ausgesprochene Untersucherabhängigkeit (Tab. 4). Je nach Studie kann die Übereinstimmung zwischen verschiedenen Untersuchern über das Vorliegen eines 3. Herztons zwischen «gut» und «schlechter als der Zufall» liegen.

Tabelle 3: Heckerling-Regel [7]. Bewertete Befunde: Temperatur >37,8°C, Herzfrequenz >100/min, Rasselgeräusche, abgeschwächtes Atemgeräusch, keine Anamnese von Asthma.

Anzahl Befunde	Wahrscheinlichkeit für Pneumonie
0	<1
1	ca. 1%
2	ca. 3%
3	ca. 10%
4	ca. 25%
5	ca. 50%

Tabelle 4: Übereinstimmung des Erkennens von Untersuchungsbefunden bei verschiedenen Untersuchern. Nach McGee [3].

Klinisches Zeichen	Kappa-Wert
Positiver hepato-jugulärer Reflux	0,92
Giemen	0,43–0,93
Gestaute Halsvenen	0,38–0,69
Rasselgeräusche	0,21–0,63
3. Herzton	–0,17–0,75
Kappa-Wert:	Übereinstimmung:
0–0,2	sehr gering
0,2–0,4	gering
0,4–0,6	mittelgradig
0,6–0,8	gut
0,8–1,0	fast perfekt
Werte <0	schlechter als Zufall

Take-home messages für die Praxis

1. Die Aussagekraft eines klinischen Tests kann anhand der Likelihood-Ratio einfach abgeschätzt werden.
2. Die Aussagekraft eines klinischen Tests ist abhängig von einer korrekten Untersuchungstechnik.
3. Das Wissen über die Aussagekraft von Untersuchungsbefunden ermöglicht es, die Wahrscheinlichkeit von Differentialdiagnosen klinisch einzuschätzen und auf dieser Basis weiterführende kosteneffiziente Abklärungen zu veranlassen.

Korrespondenz:
Dr. med. Martin Perrig, MME
Universitätsklinik für
Allgemeine Innere Medizin
Murtenstrasse
CH-3010 Bern
Martin.Perrig[at]insel.ch

Dank

Wir danken Frau Dr. med. et phil. Maria Wertli, Universitätsklinik für Allgemeine Innere Medizin, Inselspital Bern, für die kritische Durchsicht des Artikels.

Literatur

- 1 Spence D. Bad medicine: chest examination. 2012;345:e4569.
- 2 Simel DL, Rennie D, Keitz SA. The rational clinical examination: Evidence-based clinical diagnosis. New York: McGraw-Hill Medical; 2009. McGraw Hill professional.
- 3 McGee SR. Evidence-based physical diagnosis. 3rd ed. Philadelphia: Elsevier/Saunders. 2012.
- 4 Fagan TJ. Letter: Nomogram for Bayes theorem. N Engl J Med. 1975;293(5):257. doi:10.1056/NEJM197507312930513.
- 5 Lee JK, Yao L, Phelps CT, Wirth CR, Czajka J, Lozman J. Anterior cruciate ligament tears: MR imaging compared with arthroscopy and clinical tests. Radiology. 1988;166(3):861–4. doi:10.1148/radiology.166.3.3340785.
- 6 Solomon DH, Simel DL, Bates DW, Katz JN, Schaffer JL. Does This Patient Have a Torn Meniscus or Ligament of the Knee? Value of the Physical Examination. JAMA. 2001;286(13):1610–20. doi:10.1001/jama.286.13.1610.
- 7 Heckerling PS, Tape TG, Wigton RS, et al. Clinical Prediction Rule for Pulmonary Infiltrates. Ann Intern Med. 1990;113(9):6–670.