

Der Intuition auf der Spur? Das Bayes'sche Theorem und die Diagnostik in der Grundversorgung

Teil II

D. Pewsner^a, J. P. Bleuer^b, H. C. Bucher^c, M. Battaglia^d, P. Jüni^{d,e}, M. Egger^{d,f}

Einleitung

Im ersten Teil dieses Artikels haben wir aufgezeigt, dass in der Diagnostik die Anwendung des Bayes'schen Theorems eine rationale Stütze der Intuition darstellen kann. Am Beispiel des Landwirts mit Verdacht auf akute Appendizitis erörterten wir die bei der Differentialdiagnose geltenden Regeln zur Informationsgewinnung. Auch wurden neben Sensitivität und Spezifität die Vorteile der Likelihood Ratios als objektive Parameter der Testleistung beleuchtet.

Im folgenden wird beschrieben, wie das Bayes'sche Theorem quantitativ die Diagnostik zu bereichern vermag. Dies im Hinblick auf das Potential, welches eine gut zugängliche diagnostische Datenbank für Grundversorger in sich birgt.

Ein Landwirt mit Bauchschmerzen: Wie wahrscheinlich ist eine Appendizitis?

Wie gelangen wir bei unserem Landwirt mit Hilfe des Bayes'schen Theorems zur Nachtestwahrscheinlichkeit einer Appendizitis? Als erster Schritt muss die Vortestwahrscheinlichkeit einer Appendizitis geschätzt werden. Wagner et al. [4] konnten zeigen, dass in der ambulanten Praxis die Prävalenz dieser Erkrankung bei Erwachsenen mit Bauchschmerzen 0,7–1,6% beträgt (Tab. 1). Die Appendizitisprävalenz bei Patienten mit akuten Bauchschmerzen ist in unserem Fall nicht relevant, da die Bezeichnung «akut» schon die Anwendung klinischer Tests vorwegnimmt. Die Prävalenz von 1% geben wir in den «Trichter» des ersten «Wahrscheinlichkeitsumwandler», dem ersten Test (Abb. 1). Nun benötigen wir zur Charakterisierung dieser «Wahrscheinlichkeitsumwandler» Angaben zu den Testqualitäten von Anamnese, klinischen Untersuchungsbefunden und Labor. Diese sind in Tabelle 2 zusammengefasst,

wobei nur die Likelihood Ratios angegeben sind und auf die Nennung von Sensitivität und Spezifität verzichtet wurde. Am Beispiel des Erbrechens wird deutlich, dass auch das Vorhandensein eines Symptoms eine LR+ von weniger als 1 aufweisen kann. Erbrechen in der Anamnese reduziert somit die Wahrscheinlichkeit einer Appendizitis. Es fällt auf, dass nur die Angabe von Schmerzen im Bereich des rechten unteren Abdomens mit einer LR+ von 7 als isolierter Test einen wesentlichen Informationsgewinn erbringt. Als Entzündungsparameter sind Leukozytenzahl und Fieber angeführt (Tab. 3). Das C-reaktive Protein und die Blutsenkungsreaktion erbringen schlechtere Testleistungen. Da diese Parameter nicht dichotom sind, entfällt die Unterscheidung in LR+ und LR-. Dies verdeutlicht einen weiteren Vorteil der Likelihood Ratio: Die Testausbeute kann durch die Berechnung von LRs für verschiedene Grenzwerte eines Tests optimiert werden. Als isolierter Test können einzig Leukozytenwerte über $15 \times 10^9/L$ als aussagekräftig angesehen werden.

Die folgenden Angaben sind somit bei unserem Landwirt mit Bauchschmerzen als Tests für eine Appendizitis verwertbar (zur Vereinfachung wurden die Werte gerundet): (I) *Schmerzen im rechten Unterbauch* (LR+ = 7,3); (II) *Migration der Schmerzen in den rechten Unterbauch* (LR+ = 3,2); (III) *noch nie verspürter, neuartiger Schmerz* (LR+ = 1,5) und (IV) die *Leukozytose von $13 \times 10^9/L$* (LR+ = 2,4). Die Vortestwahrscheinlichkeit von ungefähr 1% (Tab. 1) wird somit durch mehrere Tests, welche als Sequenz von «Wahrscheinlichkeitsumwandlern» (Abb. 1) angesehen werden können, modifiziert. Als Glied einer solchen «Testkette» kann auch der in Isolation wenig aussagekräftige Test der Neuartigkeit des Schmerzes mit einer LR+ von nur 1,5 sinnvoll sein. Die Nachtestwahrscheinlichkeit des vorangegangenen Tests entspricht dabei der Vortestwahrscheinlichkeit des nächsten Tests. Wichtig ist, dass nur voneinander unabhängige Testgrößen

^a Praxis für Innere Medizin FMH, Bern

^b Mediscope AG, PF, 3000 Bern 23

^c Medizinische Universitäts-Poliklinik, Kantonsspital, Basel

^d Institut für Sozial- und Präventivmedizin, Universität Bern

^e Rheumatologische Universitätsklinik, Inselspital, Bern

^f Division of Health Services Research, Department of Social Medicine, University of Bristol

Korrespondenz:
Dr. Daniel Pewsner
Postfach
CH-3000 Bern 26

daniel.pewsner@bluewin.ch

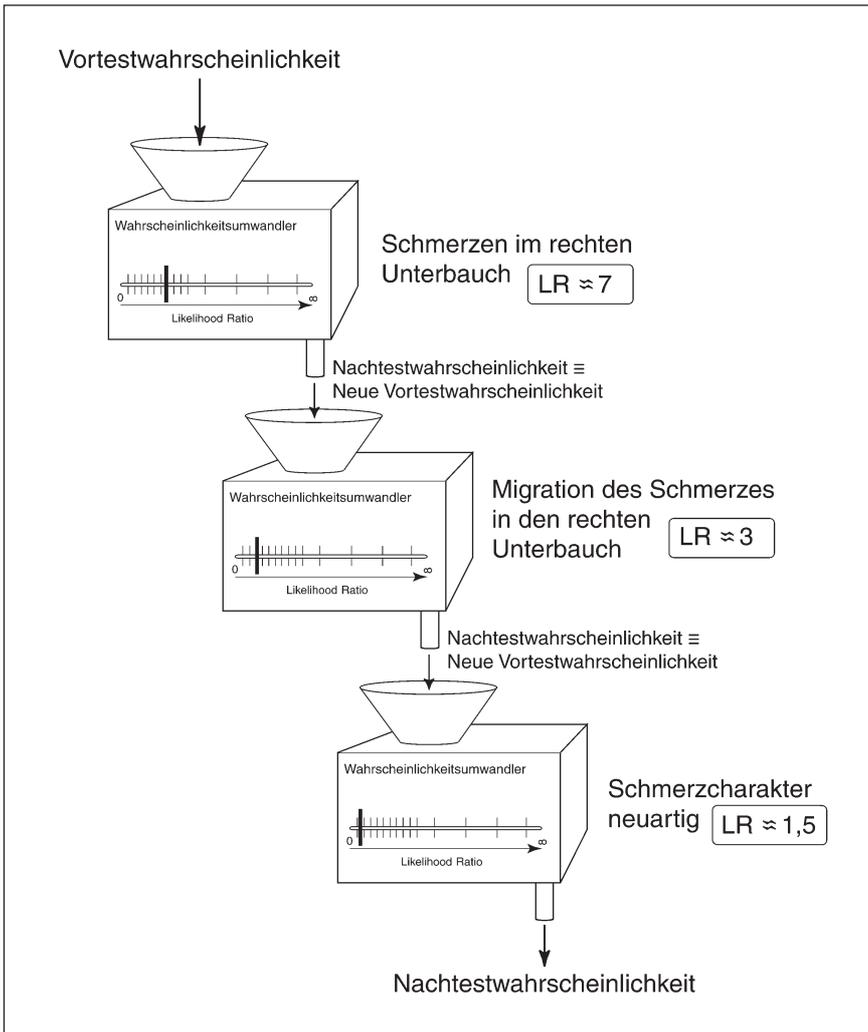


Abbildung 1. Eine Kaskade von Wahrscheinlichkeitsumwandlern modifiziert die Vortestwahrscheinlichkeit, wobei die Nachttestwahrscheinlichkeit des vorangegangenen Tests der Vortestwahrscheinlichkeit des nächsten Tests entspricht. Die Kaskade darf in ihrer Gesamtheit als ein Test angesehen werden, dessen Likelihood Ratio dem Produkt der Likelihood Ratios der einzelnen Tests entspricht. Die mathematische Anwendung dieser Kaskade ist auch gestattet, wenn Tests nicht konsekutiv durchgeführt werden.

Berücksichtigung finden. So wäre es falsch, die Entzündungsparameter Fieber und Leukozytose zu kombinieren. Viel mehr müssen wir uns auf diejenigen Entzündungsparameter mit der besseren Testleistung beschränken. Die Likelihood Ratio der ganzen Testserie ergibt sich aus dem Produkt der Ratios der einzelnen in dieser Serie angewandten Tests:

$$\text{Gesamt-Likelihood-Ratio} = 7,3 \times 3,2 \times 1,5 \times 2,4 = 84$$

Diese Konstellation von Testresultaten wird sich somit 84mal wahrscheinlicher bei Patienten mit akuter Appendizitis finden als bei solchen ohne entzündete Appendix. Bei Kenntnis der Vortestwahrscheinlichkeit von etwa 1% und einer Likelihood-Ratio von 84 kann die Nachttestwahrscheinlichkeit berechnet werden. Dies, indem wir von der «Währung» der Prozente der Vortestwahrscheinlichkeit in die

«Währung» der Odds wechseln, diese Vortest-Odds mit 84 multiplizieren und darauf die erhaltenen Nachttest-Odds wieder in Prozent umformen. Diese umständliche Arbeit nimmt uns das *Fagan-Nomogramm* [17] ab (Abb. 2): Mit dessen Hilfe gelangt man, ausgehend von der Vortestwahrscheinlichkeit, über die Likelihood-Ratio direkt zur Nachttestwahrscheinlichkeit: Die Vortestwahrscheinlichkeit von 1% wird auf der linken Skala aufgesucht und mit dem Lineal eine Linie zur mittleren Skala der Likelihood-Ratio von 84 gezogen. Auf der rechten Skala kann dann der Wert der Nachttestwahrscheinlichkeit abgelesen werden: Er beträgt etwa 45% (Abb. 2). Da die untersuchten Patientengruppen eine höhere Appendizitisprävalenz als in der Grundversorgung aufweisen, wurden die Likelihood-Ratios wahrscheinlich eher überschätzt (*Spektrumbias*, siehe Tab. 5). Die Nachttestwahrscheinlichkeit dürfte realistischere zwischen 35 und 45% liegen, einem Wert, der wahrscheinlich gut mit der vom Leser intuitiv geschätzten Wahrscheinlichkeit übereinstimmt. Intuition und berechnete Wahrscheinlichkeit stimmen jedoch nicht immer überein. So würden wohl viele Ärzte/-innen die Wahrscheinlichkeit einer Lungenembolie bei einem Patienten mit Status nach Hüftgelenkersatz und typischen Symptomen und Befunden überbewerten (Tab. 6).

Wie weiter? Die Bedeutung der Nachttestwahrscheinlichkeit

Welche Nachttestwahrscheinlichkeit gefordert werden muss, damit eine Diagnose gestellt bzw. ausgeschlossen werden darf, hängt von vielen Faktoren ab. In diesem Zusammenhang sei kurz auf das Konzept der Entscheidungsschwellen eingegangen [1, 3, 5]. Eine «Diagnose stellen» impliziert, dass wir den Patienten so behandeln, als ob die entsprechende Erkrankung vorliegen würde, auch wenn die Wahrscheinlichkeit, dass die Krankheit in der Tat vorliegt, keine 100% erreicht. Die «Test-Therapie-Schwelle» definiert die Krankheitswahrscheinlichkeit, bei der die beiden Optionen «weitere Abklärung» und «Therapie» gleichwertig sind. Beim Überschreiten dieses Grenzwerts wird die Diagnose gestellt und somit eine Behandlung eingeleitet, wogegen beim Unterschreiten dieses Werts weiter abgeklärt wird, bis die «Test-Therapie-Schwelle» doch noch überschritten oder die «Test-Keine-Therapie-Schwelle» unterschritten wird. Dieser zweite Grenzwert besagt, dass die beiden Möglichkeiten «weitere Abklärung» und «keine Therapie» gleichwertig sind. Beim Unterschreiten dieses Werts wird weder behandelt noch weiter abgeklärt, die Diagnose gilt als ausgeschlossen. Aus dem Gesagten folgt, dass die

Tabelle 1.
Vortestwahrscheinlichkeit (Prävalenz) der Appendizitis.

Patientenkollektiv	Vortestwahrscheinlichkeit
Ambulanter Bereich: Erwachsene mit Bauchschmerzen	0,7–1,6%
Notfallstationen: Patienten unter 60jährig mit akuten Bauchschmerzen	12–26%
Notfallstationen: Patienten über 60jährig mit akuten Bauchschmerzen	4%

Quelle: Wagner et al. [4]

Tabelle 2.
Anamnese und Klinik bei der akuten Appendizitis.

Anamnestische Angaben und Symptome	Likelihood-Ratio für das Vorliegen von Symptom oder Befund (LR+)
Schmerz rechtes unteres Abdomen	7,31
Neuartiger Schmerz	1,50
Schmerz vor Erbrechen	2,76
Migration Schmerz in rechtes unteres Abdomen	3,18
Inappetenz	1,27
Nausea	1,20
Erbrechen	0,92
Untersuchungsbefunde	
Défense	3,76
Psoaszeichen	2,38
Loslassschmerz	6,30
Douglasschmerz	5,34
Fieber (rektal)	Siehe Tabelle 3

Quelle: Wagner et al. [4]

Tabelle 3.
Testleistung der Entzündungsparameter bei Appendizitis.

Entzündungsparameter	Likelihood Ratio
Leukozytenzahl ($\times 10^9/L$)	
<8,0	0,16
8 bis ≤ 10	0,83
10 bis ≤ 12	1,12
12 bis ≤ 15	2,44
>15	7,03
Rektale Temperatur in °C	
<37,7	0,44
37,7–37,9	1,61
38,0–38,4	1,77
$\geq 38,5$	3,01

Quelle: Andersson et al. [16]

Anwendung eines Tests nur dann sinnvoll ist, wenn wir uns im Bereich des «Niemandlands» des Testbereichs befinden und zudem eine Chance besteht, dass dieser Test uns in den Bereich jenseits der «Test-Therapie-Schwelle» (Behandlung) oder jenseits der «Test-Keine-Therapie-Schwelle» (Nichtbehandlung) bringt. Im Fall des Landwirts ist bei einer Nachttestwahrscheinlichkeit von 35–45% die Indikation zur Operation nicht gegeben. Allerdings berechtigt uns diese Nachttestwahrscheinlichkeit nicht, den Patienten beruhigt nach Hause zu schicken. Vielmehr befinden wir uns noch im «Niemandland» zwischen den beiden Entscheidungsschwellen. Somit ist der Einsatz weiterer Tests, die das Potential aufweisen, uns über die eine oder andere Schwelle zu tragen, notwendig. Die gezielte Anwendung des «Tests of Time» ohne Hospitalisation (dessen Charakteristika schwer abzuschätzen sind) hängt neben der Patientenpräferenz und der Distanz vom Spital auch von der Verfügbarkeit weiterer diagnostischer Ressourcen ab. So können die abdominale Sonographie (Sensitivität = 76%, Spezifität = 91%, LR+ = 8, LR- = 0,25 [7]) und noch mehr die Computertomographie der Appendix (Sensitivität = 96%, Spezifität = 89%, LR+ = 9, LR- = 0,04 [7]) bzw. die Spiral-Computertomographie mit Kontrastmitteleinlauf (Sensitivität = 98% und Spezifität = 98%, LR+ = 48, LR- = 0,02 [8]) weitere zeit- und kostensparende Abklärungsschritte darstellen. Aufgrund ihrer hohen Sensitivität erfüllt die Computertomographie die Bedingungen eines *SnNout-Tests* (Tab. 4): ein negativer Befund schliesst eine Appendizitis mit grosser Zuverlässigkeit aus, wenn die Vortestwahrscheinlichkeit nicht sehr hoch ist. Die Spiral-Computertomographie mit Kontrastmitteleinlauf erfüllt sogar die Kriterien eines *SnNout-* und *SpPin-Tests* (Tab. 4). Im vorliegenden Fall wird der Patient nach einem ausführlichen Gespräch in Anwesenheit seiner Frau nach Hause entlassen, aber täglich nachkontrolliert. Am zweiten Tag haben die Bauchschmerzen deutlich abgenommen und der Patient ist eine Woche später wieder voll einsatzfähig. Der «Test of Time» hat zum Ausschluss einer (behandlungsbedürftigen) Appendizitis geführt.

Der Intuition auf der Spur: Brauchen wir eine «Bayes Collaboration»?

Die Kenntnis der Prävalenz von Erkrankungen wie auch der diagnostischen Leistungsfähigkeit von Anamnese, Klinik und Tests bildet ein wertvolles Werkzeug guter Diagnostik. Selbstverständlich sind damit nicht alle Geheimnisse der ärztlichen Kunst erklärt. Das Bayes'sche

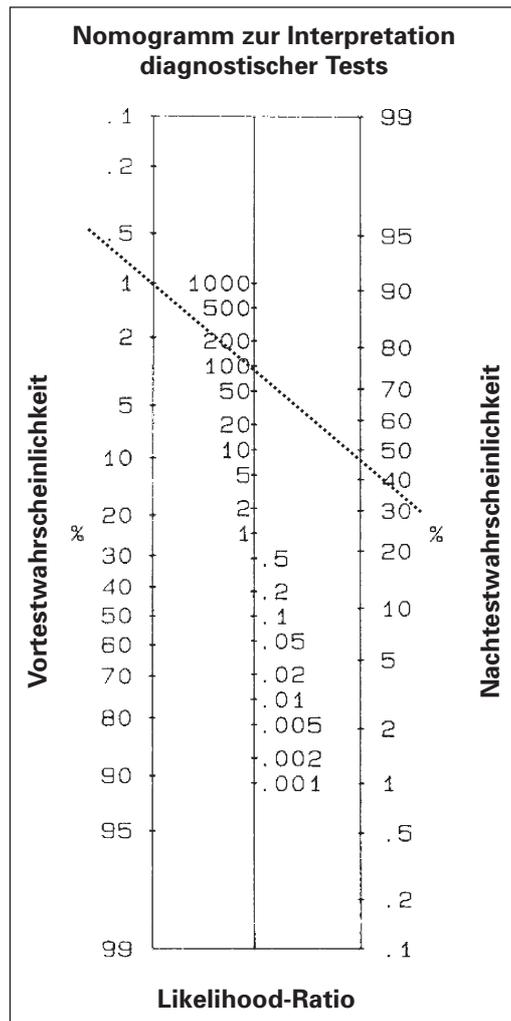


Abbildung 2

Mit Hilfe des Fagan-Nomogramms [17] lässt sich bei Kenntnis der Vortestwahrscheinlichkeit in Prozent und der Likelihood Ratio des oder der verwendeten Test(s) die Nachtestwahrscheinlichkeit in Prozent direkt ablesen. Damit kann die Umrechnung in Odds umgangen werden. Beim Landwirt mit Bauchschmerzen beträgt die Vortestwahrscheinlichkeit 1% und die Likelihood-Ratio der angewandten Tests 63. Es ergibt sich eine Nachtestwahrscheinlichkeit von 40%.

Theorem bringt uns auf der Spur der schwer fassbaren und störungsanfälligen Intuition einen Schritt weiter, indem es diese auf eine rationalere Grundlage stellt. Es kann und soll allerdings die Frucht langjähriger Erfahrung nicht ersetzen.

Zurzeit können Grundversorger und Spezialisten wenig von Thomas Bayes profitieren, da die notwendigen Informationen über Vortestwahrscheinlichkeiten und Testcharakteristika

nicht in übersichtlicher und rasch zugreifbarer Form vorliegen. Der Aufbau und Unterhalt einer derartigen Datenbank sollte idealerweise als «*Bayes Collaboration*» analog zur Cochrane Collaboration von einem internationalen Netzwerk interessierter Fachleute in Angriff genommen werden [9]. Die systematische, möglichst umfassende Suche und kritische Auswahl der Artikel zu Testqualität und Prävalenz bildet die Grundlage für den Aufbau der geplanten Datenbank. Wesentlich für die Auswahl der Studien sind deren methodische Qualität sowie Relevanz für die Grundversorger. In einer Übersicht der diagnostischen Literatur in qualitativ hochstehenden medizinischen Zeitschriften (NEJM, JAMA, BMJ und Lancet 1990–93) entsprachen nur 18% der untersuchten Studien grundlegenden methodologischen Anforderungen [6] (Tab. 5). Immerhin genügten schon 30% bei den von Lijmer et al. [10] 1996–97 untersuchten Arbeiten ähnlichen Kriterien. Es kann deshalb angenommen werden, dass die Methodik sich in den letzten Jahren verbessert hat.

Wir haben im Rahmen eines von der Schweizerischen Akademie der Medizinischen Wissenschaften unterstützten Pilotprojekts mit dem Aufbau einer diagnostischen Datenbank für Grundversorger begonnen: Es sollen häufige und im Alltag relevante Krankheiten berücksichtigt werden. Als Beispiele seien die tiefe Venenthrombose, der akute Myokardinfarkt, die chronische Angina pectoris, die akute Lungenembolie, die Pneumonie, die akute Sinusitis, die Meningitis, Kopfschmerzen und Rückenschmerzen genannt. Ziel ist es, Praktikern und Klinikern ein Werkzeug in die Hand zu geben, welches diagnostische Abläufe in der Patientenbetreuung verbessern kann. Die Datenbank soll auch in elektronischer Form zugänglich gemacht werden. Es ist vorgesehen, diese Sammlung durch ständigen Einbezug neu erscheinender Studien zu erweitern und auf dem neuesten Stand zu halten. Dies ist ein umfangreiches und langfristiges Unterfangen, und wir sind an Anregungen und der Mitarbeit weiterer PraktikerInnen und KlinikerInnen interessiert.

Verdankung

Wir bedanken uns bei Dres. A. Jaggi und Ch. Junker, Bern, für die kritische Durchsicht des Manuskriptes.

Tabelle 4. «SpPins» und «SnNouts».

Aus der Definition der Sensitivität kann die Faustregel abgeleitet werden, dass ein negativer Test mit hoher Sensitivität (tiefe falsch negative Rate) den Ausschluss einer gesuchten Erkrankung (z.B. D-Dimere [ELISA] bei der Frage nach Lungenembolie) erlaubt, bei positivem Resultat hingegen keinen «Einschluss». Anschaulicher ist die Vorstellung, dass ein Fischer ein sehr engmaschiges Netz (= hohe Sensitivität) verwendet, um auch kleine Fische nicht zu verpassen. Dafür muss er viel Unrat im Netz in Kauf nehmen (ein schweres Netz darf nicht zum Schluss verleiten, dass ein Fisch gefangen worden ist). Diese Regel wird in der angelsächsischen Literatur häufig als «SnNout» bezeichnet [2, 15].

«SnNout»: «Sensitivity Negative rule the target disorder out»: Ein negatives Testresultat bei einem hochsensitiven Test erlaubt Ausschluss einer Erkrankung.

Umgekehrt gilt die Faustregel, dass ein positiver Test mit hoher Spezifität (tiefe falsch positive Rate) den «Einschluss» einer gesuchten Erkrankung (z.B. positive Tbc-Sputumkultur) gestattet, nicht jedoch deren Ausschluss bei negativem Resultat. Derartige Befunde werden auch als «pathognomonisch» bezeichnet. Ein weitmaschiges Netz oder viel mehr noch eine Reuse lässt zwar viele Fische durchschlüpfen, die Wahrscheinlichkeit ist jedoch hoch, dass bei einem seltenen Fang tatsächlich ein gesuchter Fisch gefangen worden ist. Es wird vom «SpPin» gesprochen.

«SpPin»: «Specificity Positive rule the target disorder in»: Ein positives Testresultat bei einem hochspezifischen Test erlaubt Einschluss einer Erkrankung.

Einschränkend sei angefügt, dass die Aussagekraft eines Tests – sei sie bezüglich Ausschluss oder Einschluss – von Sensitivität und Spezifität gemeinsam abhängt. Die Faustregeln «SnNout» und «SpPin» gelten deshalb nur, wenn die dazugehörige Spezifität bzw. Sensitivität nicht weniger als 40–50% beträgt; eine Bedingung, welche die meisten gebräuchlichen Tests allerdings erfüllen.

Tabelle 5. Verzerrungen bei der Evaluierung diagnostischer Tests: Spektrumbias, Verifikationsbias und Reviewbias.

In Populationen mit Häufung einer bestimmten Pathologie finden sich vermehrt fortgeschrittene Stadien dieser Erkrankung, was potentiell zur Verbesserung der Testsensitivität (weniger falsch Negative) führt [10, 11]. Dies wird als *Spektrumbias* bezeichnet. Die Veränderung des Patientenspektrums kann auch das Vorkommen anderer differentialdiagnostisch relevanter Erkrankungen beeinflussen, was die Spezifität verändert (Anzahl falsch Positiver). Die im Haupttext vorgestellten Testqualitäten für Symptome und Befunde stammen von Patienten, die wegen Verdachts auf Appendizitis in Notfallstationen eingewiesen wurden. Die meisten haben deshalb die «Hürde» der Grundversorgung hinter sich und somit eine entsprechend höhere Vortestwahrscheinlichkeit bzw. Prävalenz einer Appendizitis.

Bei der Appendizitis, aber auch bei anderen Erkrankungen, stellt sich das Problem der Anwendung des Goldstandards, in diesem Fall der Operation mit histologischer Untersuchung. Bei idealem Studiendesign müsste bei allen untersuchten Patienten die Diagnose mit dem Goldstandard verifiziert werden. Da im Falle der Appendizitis die Operation aber nur bei dringendem Verdacht verantwortbar ist, wird eine negative Diagnose nicht entsprechend verifiziert. Stattdessen gelangt ein dem Goldstandard unterlegener Referenztest (z. B. «Test of Time») zum Einsatz, was zu einem sogenannten *Verifikationsbias* führen kann [13, 14].

Beim Vergleich der Leistung eines neuen Tests mit dem Referenztest müssen die Ergebnisse unabhängig evaluiert werden. Personen, welche die Ergebnisse eines neuen Tests beurteilen, sollten gegenüber dem Goldstandard «verblindet» sein [10]. Ansonsten besteht die Gefahr der «Beschönigung» der Sensitivität und Spezifität des neuen Tests, da wir erfahrungsgemäss Dinge besser erkennen, wenn wir wissen, wo und ob wir sie suchen müssen. Im Zusammenhang mit der Appendizitis spielt dieser sogenannte *Reviewbias* vor allem bei den bildgebenden Verfahren der Sonographie und der Computertomographie eine Rolle.

Tabelle 6.
Wie nützlich sind Anamnese und Klinik bei Verdacht auf Lungenembolie?

Ein 65-jähriger Mann mit Status nach Hüftgelenkersatz vor einem Monat kommt mit Dyspnoe, neu aufgetretenen gelegentlichen Palpitationen und Pleuraschmerz in die Praxis. Bei der Untersuchung findet sich eine Tachypnoe und Tachykardie. Lungenembolie? Nicht unbedingt: In einer methodisch guten Arbeit untersuchten Stein et al. [14] 365 Patienten mit Verdacht auf Lungenembolie. Bei 117 Patienten wurde die Lungenembolie angiographisch verifiziert (Vortestwahrscheinlichkeit = 32%). Keine der typischen anamnestischen oder klinischen Angaben weist bezüglich der Lungenembolie eine gute diagnostische Aussagekraft auf. Dies gilt sogar für eine konkomitierende tiefe Venenthrombose. Die Verknüpfung der Aussagen mehrerer Tests erlaubt höchstens, den Verdacht auf eine Lungenembolie auszusprechen oder diesen abzuschwächen (in diesem Beispiel verwandeln die beschriebenen Symptome und Befunde mit einer Gesamt-LR von 2,5 die Vortestwahrscheinlichkeit von etwa 32% in eine Nachtestwahrscheinlichkeit von 54%). Erst die apparative Diagnostik ist in der Lage, die Diagnose zu stellen oder auszuschliessen.

	Sensitivität (%)	Spezifität (%)	Likelihood (LR+)
Anamnestische Angaben und Symptome			
Immobilisation	56	67	1,7
Vorangegangene Operation	54	69	1,7
Dyspnoe	73	28	1,0
Pleuraschmerz	66	41	1,1
Husten	37	64	1,0
Hämoptyse	13	92	1,6
Palpitationen	10	82	0,6
Angor ohne Ausstrahlung	4	94	0,7
Untersuchungsbefunde			
Tachypnoe (>20/min)	70	32	1,0
Tachykardie (>100/min)	30	76	1,3
4. Herzton	24	86	1,7
Verstärkter Pulmonalklappenschluss	23	87	1,8
Tiefe Venenthrombose	11	89	1,0
RV-Heben	4	98	2,0
Pleurareiben	3	98	1,5
3. Herzton	3	96	0,7
Zyanose	1	98	0,5

Quelle: Stein et al. [14]

Literatur

- 1 Sox HC, Blatt MA, Higgins MC, Marton KI. Medical Decision-Making. Boston: Butterworth-Heinemann; 1987.
- 2 Sackett DL, Richardson WS, Rosenberg W, Haynes RB. Evidence-based Medicine. New York: Churchill Livingstone; 1997.
- 3 Dubs L. Diagnostik mit Hilfe der Entscheidungsanalyse. Schweiz Arztezeitung 1999;80:785-7.
- 4 Wagner JM, McKinney WP, Carpenter JL. Does this patient have appendicitis? JAMA 1996;276:1589-94.
- 5 Kassirer JP, Gorry GA. Clinical problem solving: a behavioral analysis. Ann Intern Med 1978;89:245-55.
- 6 Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA 1995; 274:645-51.
- 7 Balthazar EJ, Birnbaum BA, Yee J, Megibow AJ, Roshkow J, Gray C. Acute Appendicitis: CT and US Correlation in 100 Patients. Radiology 1991;1889:21-4.
- 8 Rao PM, Rhea JT, Novelline RA, Mostafavi AA, McCabe CJ. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. N Engl J Med 1998;338:141-6.
- 9 Pewsner D, Bleuer JP, Jüni P, Battaglia M, Bucher HC, Egger M. Do we need a Bayes Collaboration? Proposal for a Diagnostic Database. 8th Cochrane Colloquium, Cape Town, South Africa, 25-29 October 2000.
- 10 Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282: 1061-6.
- 11 Deeks JJ. Systematic Reviews of Evaluations of Diagnostic and Screening Tests. In Egger M, Smith GD, Altman DG, eds. Systematic Reviews in Health Care: Meta-Analysis in Context. London: Br Med J Books; 2000.
- 12 Panzer RJ, Suchman AL, Griner PF. Workup bias in prediction research. Med Decis Making 1987;7:115-9.
- 13 Magid D, Schwartz B, Craft J, Schwartz JS. Prevention of Lyme disease after tick bites. A cost-effectiveness analysis. N Engl J Med 1992; 327:534-41.
- 14 Stein PD, Terrin ML, Hales CA, Palevsky HI, Saltzman HA, Thompson BT, et al. Clinical, laboratory, roentgenographic, and electrocardiographic findings in patients with acute pulmonary embolism and no pre-existing cardiac or pulmonary disease. Chest 1991;100:598-603.
- 15 Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13:313-21.
- 16 Andersson RE, Hugander AP, Ghazi SH, Ravn H, Offenbartl SK, Nystrom PO, et al. Diagnostic Value of Disease History, Clinical Presentation and Inflammatory Parameters of Appendicitis. World J Surg 1999;23: 133-40.
- 17 Fagan TJ. Nomogram for Bayes theorem. N Engl J Med 1975;293:257.