Evidence-based Medicine: Wie beurteile ich eine Studie zu einem diagnostischen Test?

Evidence-based Medicine: How to read an article about a diagnostic test?

U. Glenck[a], D. Pewsner[b], H. C. Bucher[c]

Einführung

Der technologische Fortschritt hat in den letzten drei Jahrzehnten eine Vielzahl neuer und verbesserter diagnostischer Möglichkeiten geschaffen. Die Testmöglichkeiten sind heute so zahlreich, dass Mediziner Gefahr laufen, mit dem ungezielten Einsatz von Tests Gesunde zu Kranken zu machen. Ganz besonders gilt dies, wenn Tests zum Screening einer nach klinischen Kriterien gesunden Population eingesetzt werden. Neben der kompetenten ärztlichen Arbeit mit der Gesundheit des einzelnen Patienten und ganzer Bevölkerungsgruppen vor Augen, verlangt auch das wachsende Bewusstsein über die beschränkten ökonomischen Mittel den rationalen und gezielten Einsatz der diagnostischen Möglichkeiten.
Angesichts dieser Entwicklung kommt der Information über die Leistungsfähigkeit diagnostischer Tests für klinisch tätige Ärzte eine zentrale Bedeutung zu. Viele Kliniker sind jedoch über die relevanten Kriterien und Fragen zur Überprüfung der Tauglichkeit von diagnostischen Tests zu wenig informiert oder haben zu wenig Zeit, sich diese Information zu beschaffen und kritisch zu evaluieren. Da die Resultate klinischer Forschung heute auch Kolleginnen und Kollegen in der peripheren Praxis zugänglich sind, soll in dieser Arbeit gezeigt werden, wie klinische Studien über den Wert diagnostischer Tests beurteilt und bei ihrer Anwendung in der täglichen Praxis die richtigen Schlussfolgerungen gezogen werden können.

Das Szenarium aus dem praktischen Alltag

Sie sehen in Ihrer Sprechstunde eine 27jährige Frau, die über ihr einseitig geschwollenes Bein beunruhigt ist. Dieses Symptom sei in den letzten 36 Stunden erstmals aufgetreten. Ihre Patientin ist Nichtraucherin, Mutter eines zweijährigen Knaben und nimmt einen Ovulationshemmer ein. Eine positive Familienanamnese bezüglich thromboembolischer Krankheiten liegt nicht vor. Die Untersuchung ergibt ein leichtes nicht eindellbares Ödem im Malleolarbereich, keine wesentliche Druckdolenz der Wade und eine Umfangdifferenz gegenüber der gesunden Seite von 1,5 cm. Mit Ausnahme von Besenreisern seit der Schwangerschaft bestehen keine Varizen.

Das diagnostische Problem

Das Risiko, dass die Patientin eine tiefe Beinvenen-Thrombose (TVT) hat, ist ziemlich klein. Sie erinnern sich jedoch an einen ähnlichen Fall, bei welchem Sie eine Thrombose verpasst hatten. Diesmal wollen Sie mehr Sicherheit, bevor Sie die Patientin ohne Antikoagulation entlassen. Eine notfallmässige Phlebographie scheint Ihnen übertrieben. Sollen Sie eine Duplex-Sonographie veranlassen? Sie haben kürzlich ein Testkit zur Bestimmung der D-Dimere (SimpliRed®) angeschafft. Da der Test bei der Patientin negativ ausfällt, entlassen Sie die Patientin ohne Diagnose, aber mit Ausschluss der tiefen Venenthrombose. Nach Feierabend blättern sie noch ein wenig durch die Literatur, die Sie sich zur Lektüre bereit gelegt haben und stossen in «Ars Medici» [1] auf einen Artikel, in welchem die Autoren Ihren D-Dimer-Test zum Ausschluss der Thrombose als ungeeignet darstellen. Bei einer Sensitivität von 80% werde eine von fünf tiefen Thrombosen mit einem negativen Testresultat verpasst!

[a] Allgemeinpraxis, Ottenbach
[b] Internistische Praxis, Bern
[c] Medizinische Universitäts-Poliklinik, Kantonsspital Basel

Korrespondenz:
PD Dr. Heiner C. Bucher
Medizinische Universitäts-Poliklinik
Kantonsspital Basel
CH-4031 Basel

E-mail: hbucher@uhbs.ch

Die klinische Fragestellung

Sie sind verunsichert und stellen sich die folgende Frage: Wie verlässlich ist der D-Dimere-Test für den Ausschluss einer tiefen Venenthrombose (TVT) bei niedriger oder mittelhoher Erkrankungswahrscheinlichkeit?

Die Literatursuche

Nach dem Nachtessen setzen Sie sich zur Literatursuche an Ihren Computer. Im Internet finden Sie Zugang zu Medline (http://www.hin.ch). Da Sie den von Ihnen benützten Test evaluieren wollen, geben Sie «D-dimers», «Simplired» und «Thrombosis» als Suchwörter ein und finden 25 Arbeiten zum Thema.
Eine kurze Durchsicht der Titel und Zusammenfassungen dieser Arbeiten zeigt Ihnen, dass Studien zu widersprüchlichen Schlussfolgerungen bezüglich der Nützlichkeit des SimpliRed-D-Dimer-Tests zum Ausschluss von tiefen Venenthrombosen kommen [2, 3]. Welchen Arbeiten können Sie trauen? Sie erinnern sich, in der Zeitschrift «Evidence-Based Medicine» [4] eine Besprechung von zwei Arbeiten gelesen zu haben, welche die Brauchbarkeit des D-Dimer-Schnelltests untersuchen [5, 6]. «Evidence-Based Medicine» bespricht und kommentiert Originalarbeiten, welche bestimmten Qualitätsmerkmalen genügen. Wir wollen in diesem Artikel eine dieser Arbeiten [5], welche aus dem Kantonsspital Basel stammt, genauer ansehen. Sie können sich die Originalarbeit über die Website des Journals of Vascular Surgery (http://www.harcourthealth.com), welche Sie mit einem Webbrowser gefunden haben, herunterladen.

Die Evidenz überprüfen (das «critical appraisal»)

Welchen Studienergebnissen können wir trauen? Sackett et al. haben Kriterien zum «critical appraisal» von Studien zu diagnostischen Tests entwickelt [7]. Hierbei sollten drei Hauptfragen beantwortet werden:
- Ist die Evidenz zur Treffsicherheit des diagnostischen Tests valide und methodisch korrekt überprüft worden?
- Kann der Test Patienten, welche die gesuchte Krankheit haben, von solchen unterscheiden, die sie nicht haben?
- Kann ich diesen geprüften Test an meinem Patienten anwenden? Ist er beim Entscheid über weitere Diagnostik oder Therapie hilfreich?

Ist die Evidenz zur Treffsicherheit des diagnostischen Tests valide und methodisch korrekt überprüft worden?

Die erste Frage richtet sich nach der methodischen Qualität der Studie. Wurde ein Test mit unzureichender Methodik evaluiert, erübrigt sich eine weitere Lektüre. Die Qualität der Methodik lässt sich anhand von 4 Punkten überprüfen (Tabelle 1).

- *Gibt es einen unabhängigen, verblindeten Vergleich mit einem Standardtest («Gold-Standard»)?*

Ob die Ergebnisse einer Studie vertrauenswürdig sind, hängt von der Studienmethodik ab. Bei der Lektüre müssen wir prüfen, inwiefern der Test und die Testergebnisse die «wahren» Verhältnisse widerspiegeln. Wir müssen uns versichern, dass ein zu evaluierender Test mit einem angemessenen Referenz- oder Standardtest verglichen wird. Der Referenztest muss der Test mit der höchsten ausgewiesenen Zuverlässigkeit («Gold-Standard») sein, z.B. eine Biopsie, Autopsie oder ein Langzeit-Follow-up. Falls der Referenztest sinnvoll erscheint, muss weiter überprüft werden, ob die Ergebnisse des Tests und der Referenztest **unabhängig** evaluiert wurden. Das heisst, Personen, welche die Resultate des Tests ab-

Tabelle 1.
Kriterien zur Beurteilung der Methodik einer Studie zu einem diagnostischen Test [7].

Ist die Evidenz zur Treffsicherheit des diagnostischen Tests valide und methodisch korrekt überprüft worden?
– Gibt es einen unabhängigen, verblindeten Vergleich mit einem Standardtest («gold standard»)?
– Schloss die Studie ein genügend breites Spektrum von Patienten ein, an welchen der Test in der klinischen Praxis angewendet wird?
– Beeinflussten die Ergebnisse des zu evaluierenden Tests die Entscheidung, ob der Standardtest angewendet wurde?
– Wurde die Testmethodik genügend detailliert beschrieben, um die Wiederholung des Tests zu ermöglichen? Wurde der Test unabhängig von der vorliegenden Studie an einem zweiten Kollektiv untersucht?
Hilft der Test Patienten, welche die gesuchte Zielkrankheit haben, von solchen unterscheiden, die sie nicht haben?
Kann ich diesen geprüften Test an meinen Patienten anwenden? Ist er im Entscheid über weitere Diagnostik oder Therapie hilfreich?
– Ist die Reproduzierbarkeit und die Interpretation der Testergebnisse in meinem klinischen Umfeld gegeben?
– Kann eine klinisch sinnvolle Vortest-Wahrscheinlichkeit für die Zielkrankheit meines Patienten geschätzt werden?
– Liegt die Nachtest-Wahrscheinlichkeit in einem Bereich, welche das Management meines Patienten ändern wird?

lesen, müssen gegenüber den Ergebnissen des Gold-Standards (und umgekehrt) **verblindet** sein (Blinded Assessment). Je grösser die Wahrscheinlichkeit, dass die Interpretation des Tests durch Kenntnis des Ergebnisses der Referenzuntersuchung beeinflusst werden kann, was vor allem bei bildgebenden Verfahren der Fall ist, desto wichtiger ist die Verblindung.

In unserem Beispiel ist diese Forderung erfüllt. Die Interpreten des SimpliRed-Tests wussten nichts von den Resultaten der Duplex-Sonographie und umgekehrt. Als Gold-Standard wurde die Duplex-Sonographie eingesetzt, obwohl die Phlebographie als Gold-Standard gilt. Die Autoren rechtfertigen ihr Vorgehen mit der hohen hausinternen Sensitivität (94–97%) und Spezifität (91–96%) der Duplex-Sonographie gegenüber der Phlebographie. Da der Gold-Standard bei allen Patienten angewendet werden muss, ist die weniger invasive Methode der Duplex-Sonographie vertretbar. Methodisch ist die Verlässlichkeit der Duplex-Sonographie jedoch im Vergleich mit der Phlebographie zur Diagnostik der tiefen Venenthrombose (TVT) geringer. Dies kann im konkreten Beispiel zu einer Beschönigung der tatsächlichen Leistungsfähigkeit des D-Dimer-Tests geführt haben. Durch die Nichtverwendung des Gold-Standards können leichtgradige Thrombosen verpasst und somit die tatsächliche Sensitivität und Spezifität des Tests überschätzt worden sein.

– *Schloss die Studie ein Spektrum von Patienten ein, das demjenigen, an welchen der Test in der klinischen Praxis angewendet wird, entspricht?*

Ein diagnostischer Test ist nur in dem Ausmass nützlich, als er zwischen gesuchten Zielcharakteristiken zu unterscheiden hilft, welche sonst nicht identifizierbar sind. Praktisch jeder diagnostische Test kann zwischen Schwerkranken und Gesunden unterscheiden. Der wahre diagnostische Wert einer Untersuchung wird deshalb nur in einer Studie erhellt, die Patienten einschliesst, welche die gesuchte Zielkrankheit in unterschiedlicher Ausprägung haben. Dies entspricht auch unserer Erfahrung, denn im klinischen Alltag präsentieren sich Krankheitsbilder selten in der typischen Lehrbuchform. Beispielsweise fanden sich bei Patienten mit kolorektalem Karzinom wesentlich höhere Serumspiegel des carcinoembryogenen Antigens (CEA) als bei Normalpersonen [8]. Diese Ergebnisse liessen vermuten, der Test eigne sich als Screeningtest für das kolorektale Karzinom. Folgeuntersuchungen an einem breiteren Kollektiv bestehend aus Patienten mit weniger fortgeschrittenem kolorektalem Karzinom und solchen mit anderen Tumoren oder Grundkrankheiten zeigten hingegen, dass der CEA-Test häufig zu falsch positiven Ergebnissen führte. Der CEA-Test wurde deshalb als Screeningtest verlassen.

In unserem Beispiel wurden über einen bestimmten Zeitraum **alle** Patienten mit Verdacht auf tiefe Venenthrombose (TVT), die der angiologischen Abteilung zugewiesen wurden**,** untersucht. Die Ausschlusskriterien sind klar definiert, und die Patienten entstammten zur Hälfte aus dem ambulanten Bereich. Die aus dem ambulanten und stationären Bereich rekrutierten Patienten wiesen bezüglich proximaler und distaler TVT ähnliche Prävalenzen von rund 5% bzw. 15% auf. Das Patientenspektrum ist somit breit und für die ambulante Praxis repräsentativ. Jedoch wissen wir nicht mit Sicherheit, ob nicht bei einem Teil der untersuchten Patienten bei der Überweisung bereits ein D-Dimer-Test vorlag und somit bereits eine Vorselektion erfolgte, die zu einer Beschönigung der Sensitivität führen würde.

– *Beeinflussten die Ergebnisse des zu evaluierenden Tests die Entscheidung, ob der Standardtest angewendet wurde?*

Die Eigenschaften eines zu evaluierenden Tests werden verzerrt, falls dessen Ergebnisse die Entscheidung beeinflussen, ob Patienten mit dem Referenztest untersucht werden oder nicht. Diese Fehlermöglichkeit wird in der Literatur als **«verification bias»** bezeichnet. Ein Beispiel soll dies verdeutlichen. In einer Studie zur Evaluation der Ventilations-Perfusions-Szintigraphie zur Diagnose der akuten Lungenembolie erhielten Patienten mit niedriger Wahrscheinlichkeit einer Lungenembolie im Szintigraphiebefund weniger häufig eine Angiographie (69%), als Patienten mit hochverdächtigem Szintigraphiebefund (92%) [9]. Es ist verständlich, dass Kliniker wenig geneigt sind, Patienten weiteren invasiven Untersuchungen auszusetzen, wenn die Chance eines zusätzlichen Informationsgewinns gering ist. Die Autoren dieser Studie versuchten jedoch, diesen Bias zu korrigieren, indem sie Patienten mit normalen Szintigraphiebefunden über ein Jahr bezüglich des Auftretens von Lungenembolien nachkontrollierten.

In der vorliegenden Studie wurde bei allen Patienten als «Gold-Standard-Untersuchung» die Duplex-Sonographie durchgeführt (der eigentliche Gold-Standard ist die Phlebographie). Leider haben die Autoren es unterlassen, durch eine Verlaufsbeobachtung und eine spätere Zweituntersuchung bei Testnegativen oder Untersuchten mit niedriger Erkrankungswahrscheinlichkeit sicherzustellen, dass sie mit der Duplex-Sonographie keine relevante Thrombose verpasst haben.

– *Wurde die Testmethodik genügend detailliert beschrieben, um die Wiederholung des Tests zu ermöglichen?*
– *Wurde der Test unabhängig von der vorliegenden Studie an einem zweiten Kollektiv untersucht?*

Studien zu diagnostischen Tests sollten die Testmethoden exakt beschreiben, um deren Replizierbarkeit zu ermöglichen und um überprüfen zu können, ob der Test unter klinischen Alltagsbedingungen anwendbar ist. Wichtige Angaben zu Technik, Art der verwendeten Geräte oder Labormethoden sollten erwähnt sein. Die Evaluation eines Tests in einer weiteren Patientenpopulation ermöglicht wichtige Aussagen zu dessen Replizierbarkeit.
Die Autoren haben die Testanordnungen im Detail und nachvollziehbar beschrieben, auf eine Replikation in einer weiteren Patientenpopulation wurde aber verzichtet. Eine zweite Studie aus London [6] zur gleichen Fragestellung, welche in der gleichen Nummer von «Evidence-Based Medicine» besprochen wird, kommt jedoch zu vergleichbaren Ergebnissen. Abschliessend kommen Sie zum Schluss, dass die Studie trotz gewisser methodischer Mängel, in bezug auf die Verlässlichkeit des D-Dimer-Testes zum Ausschluss einer tiefen Thrombose aussagekräftig ist.

Hilft der Test Patienten, welche die gesuchte Zielkrankheit haben, von solchen zu unterscheiden, bei denen dies nicht der Fall ist?

Wenn wir einen diagnostischen Test anwenden, gehen wir von einem klinischen Verdachtsmoment aus, d.h. von einer bestimmten Wahrscheinlichkeit (oder Vortest-Wahrscheinlichkeit), dass die gesuchte Zielkrankheit vorliegt. Der Test, den wir verwenden, soll uns mit möglichst grosser Sicherheit eine Zielkrankheit bestätigen oder ausschliessen. Ein positives Testresultat soll eine möglichst hohe, ein negatives eine möglichst niedrige Nachtest-Wahrscheinlichkeit für das Vorliegen der vermuteten Krankheit erzeugen. Die wichtigste Testqualität wird mit Sensitivität und Spezifität ausgedrückt. Ein modernes, im Vergleich zu Sensitivität und Spezifität aussagekräftigeres und anwenderfreundlicheres Konzept zur Beurteilung von Testqualitäten sind «Likelihood Ratios». Die Likelihood-Ratio [7] beschreibt, von der Vortest-Wahrscheinlichkeit unabhängig, die Leistung eines Tests. Sie fasst die Testqualitäten der Sensitivität und Spezifität in einer einzigen Zahl zusammen und ist ein objektiver Parameter der Testleistung. Sie ist definiert als

Tabelle 2. Vergleich des D-Dimer-Tests mit dem Standardtest (Duplex-Sonographie) zur Diagnose einer tiefen Venenthrombose (TVT) nach [7].

Ergebnis des Tests unter Evaluation	Referenztest (Duplex-Sonographie)		
	TVT vorhanden	TVT nicht vorhanden	
D-Dimer positiv	**a** Richtig positive: 72	**b** Falsch positive: 135	Alle Testpositive a + b: **207**
D-Dimer negativ	**c** Falsch negative: **11**	**d** Richtig negative: **180**	Alle Testnegative c + d: **191**
	Alle Kranken a + c: **83**	Alle Gesunden b + d: **315**	Alle Testpersonen a + b + c + d: **398**
Sensitivität Rate der Erkrankten mit positivem Test	Richtig positive Rate	a / (a + c)	(72 / 83) × 100% = 86,7%
Spezifität Rate der Gesunden mit negativem Test	Richtig negative Rate	d / (b + d)	(180 / 315) × 100% = 57,1%
Likelihood Ratio für einen positiven Test	Richtig positive Rate / Falsch positive Rate	Sensitivität / (1-Spezifität)	86,7% / (100% – 57,1%) = 2,02
Likelihood Ratio für einen negativen Test	Falsch Negative Rate / Richtig Negative Rate	(1-Sensitivität) / Spezifität	(100% – 86,7%) / 57.1% = 0,23
Positive prädiktiver Wert (PPV)	Rate der richtig positiven Tests von allen Testpositiven	a / (a + b)	(72 / 207) × 100% = 34,8%
Negativer prädiktiver Wert (NPV)	Rate der richtig negativen Tests von allen Testnegativen	d / (c + d)	(180 / 191) × 100% = 94,2%
Prävalenz	Vortest-Wahrscheinlichkeit	a + c/ (a + b + c + d)	(83 / 398) × 100% = 20,8%

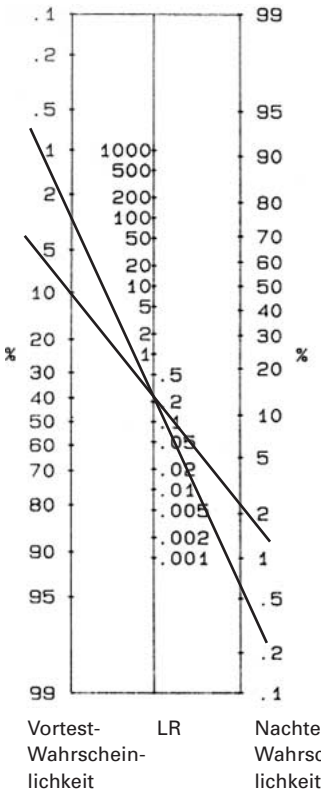

Abbildung 1.
Nomogramm nach Fagan [10]. Bestimmung der Nachtest-Wahrscheinlichkeit aus Vortest-Wahrscheinlichkeit und Likelihood Ratio (LR).

Verhältnis der Wahrscheinlichkeiten eines bestimmten Testresultats bei Kranken und Gesunden. Die Likelihood Ratio für einen positiven Test (LR+) besagt, wieviel mal wahrscheinlicher sich ein bestimmtes positives Testresultat bei einem Kranken als bei einem Gesunden findet. Die Likelihood Ratio für ein negatives Resultat (LR–) sagt aus, wieviel mal wahrscheinlicher sich ein negatives Testresultat bei einem tatsächlich Erkrankten im Vergleich zu einem Gesunden findet.

Die Likelihood Ratio für ein **positives** Testergebnis ist das Verhältnis der Rate der richtig positiven Testergebnisse zur Rate der falsch positiven Testergebnisse (Tabelle 2). Die Rate der richtig positiven Testergebnisse gemessen an allen Patienten mit der gesuchten Zielkrankheit entspricht der Sensitivität eines Test (a / a + c). Die Rate der falsch positiven Testergebnisse (b / b + d) gemessen an allen Gesunden entspricht dem Wert (1-Spezifität). Somit ist die LR+ gleich Sensitivität / 1-Spezifität.

Das Wahrscheinlichkeitsverhältnis für ein **negatives** Testergebnis (LR–) ist definiert als das Verhältnis der Rate der falsch negativen Testergebnisse zu derjenigen der richtig negativen Testergebnisse (Tabelle 2). Die Rate der falsch negativen Testergebnisse (c / a + c) entspricht dem Wert (1-Sensitivität), diejenige der richtig negativen Testergebnisse der Spezifität eines Tests (d / b + d).

Einer der grossen Vorteile der Likelihood Ratio ist, dass mittels der Vortest-Wahrscheinlichkeit unmittelbar die Nachtest-Wahrscheinlichkeit eines Testergebnisses berechnet werden kann. Die Vortest-Wahrscheinlichkeit wird auch als die Prävalenz einer Zielkrankheit in einer Bevölkerungs- oder Patientengruppe bezeichnet. Die Umrechnung von Vortest- auf Nachtest-Wahrscheinlichkeit geschieht am einfachsten mit dem Nomogramm von Fagan [10] (Abbildung 1). Die genauen Rechenoperationen dazu sind aus anderer Quelle zu entnehmen [7, 11]. Likelihood Ratios haben gegenüber den herkömmlichen Konzepten von Sensitivität und Spezifität wichtige Vorteile, indem für einen gegebenen Test die Nachtest-Wahrscheinlichkeiten anhand verschiedener Grenzwerte berechnet werden können. Testergebnisse können mit dem Einsatz von Likelihood Ratios somit differenzierter anhand unterschiedlicher Grenzwerte interpretiert werden (Tabelle 3). Wenn sich beispielsweise bei einem Patienten mit einer geschätzten Vortestwahrscheinlichkeit von 40% im Belastungs-EKG die ST-Strecke um weniger als 0,5 Millimeter senkt, verringert dies die Wahrscheinlichkeit einer koronaren Herzkrankheit um den Faktor 0,23 auf 13%, wohl keine Indikation für weitere Abklärungen. Im gleichen Fall erhöht eine ST-Senkung von 1,5 Millimeter die Wahrscheinlichkeit von 40% auf 88%, eine sichere Indikation für ein Koronarangiogramm.

Likelihood Ratios in der Grössenordnung von >10 bei positivem Test oder <0,1 bei negativem führen in der Regel zu ausgeprägten Veränderungen der Vortest-Wahrscheinlichkeit, die häufig, je nach Vortest-Wahrscheinlichkeit, eine gesuchte Zielkrankheit verlässlich ein- bzw. ausschliessen lassen. Likelihood Ratios zwischen 5 und 10 bzw. 0,1 und 0,2 ergeben mässige Veränderungen der Vortest-Wahrscheinlichkeit. Wahrscheinlichkeitsverhältnisse zwischen 2 und 5 bzw. 0,5 und 0,2 bewirken geringe, jedoch manchmal wichtige Veränderungen der Vortest-Wahrscheinlichkeit. Likelihood

Tabelle 3.
Abhängigkeit von Sensitivität, Spezifität und Likelihood Ratio von Grenzwerten (cut-offs) am Beispiel der ST-Senkung im Belastungs-EKG zur Diagnose der koronaren Herzkrankheit [15].

ST-Senkung mm	Sensitivität (%)	Spezifität (%)	Likelihood Ratio
<0,5			0,23
0,5-0,99	86	77	0,92
1,0-1,49	65	89	2,1
1,5-1,99	42	98	4,2
2,0-2,49	33	99	11
>2,5	20	>99	39

Ratios zwischen 1 und 2 und 0,5 und 1 verändern die Vortest-Wahrscheinlichkeit in einem klinisch kaum relevanten Ausmass.
Unser Beispiel zeigt, dass die Likelihood Ratio für ein positives Testresultat (LR+) für D-Dimere 2,0 ist (Tabelle 2). Das Ergebnis besagt also, dass eine derartige Konstellation 2mal häufiger bei einem Patienten vorkommt, der tatsächlich eine tiefe Venenthrombose (TVT) aufweist, als bei einem ohne diese Krankheit. Die Likelihood Ratio ist damit derart nahe bei 1 (eine Likelihood Ratio von 1 besagt Vortest-Wahrscheinlichkeit = Nachtest-Wahrscheinlichkeit), dass damit der Bestimmung des D-Dimer zur **Bestätigung** der Diagnose einer TVT keine Bedeutung zukommt. Hingegen ist die Likelihood Ratio für einen negativen D-Dimer-Test mit 0,23 in einem Bereich, in welchem sich insbesondere bei niedriger Vortest-Wahrscheinlichkeit eine TVT mit genügender Sicherheit ausschliessen lässt. Dies entspricht der hohen Sensitivität von 87% des Testes. (Bei einem Test mit einer **hohen** Sensitivität lässt sich bei einem **negativen** Testergebnis eine gesuchte Zielkrankheit mit grosser Wahrscheinlichkeit **aus**schliessen.)

Kann ich diesen geprüften Test an meinen Patienten anwenden? Ist er beim Entscheid über weitere Diagnostik oder Therapie hilfreich?

Falls die methodischen Aspekte einer Studie zur Evaluation eines Tests valide sind (interne Validität), stellt sich die Frage zur Anwendbarkeit des Test in der eigenen konkreten Praxis (externe Validität). Hierbei sollten die folgenden Fragen evaluiert werden (Tabelle 1).

– *Ist die Reproduzierbarkeit und die Interpretation der Testergebnisse in meinem klinischen Umfeld gegeben?*

Eine wichtige Eigenschaft eines diagnostischen Tests, die es zu kennen gilt, ist dessen Reproduzierbarkeit. Der Test soll bei stabilen Patienten oder Testkonditionen bei wiederholter Anwendung die gleichen Ergebnisse liefern. Kann ein Testresultat schlecht reproduziert werden, mag dies mit technischen oder methodischen Problemen oder mit der Subjektivität der Testinterpretation zusammenhängen. Deshalb sollten Angaben zur Reproduzierbarkeit vorhanden sein. Dies spielt vor allem eine Rolle, wenn die Testinterpretation spezielle Fähigkeiten der Untersucher erfordert wie die Interpretation von Computertomographien, Ultraschalluntersuchungen. Wenn die Reproduzierbarkeit eines Tests nur mittelmässig ist, der Test aber dennoch zwischen pathologischen und normalen Befunden unterscheiden hilft, besteht weniger Anlass zur Sorge. Ist die Reproduzierbarkeit eines Test sehr hoch und die Beurteilung durch die Untersucher variert nur geringgradig, dann ist der Test in seiner Anwendung entweder einfach oder die Testinterpreten sind hoch qualifiziert. Falls letzteres zutrifft, kann der Test in einem anderen klinischen Umfeld mit weniger qualifizierten Interpreten schlechter reproduzierbar sein.
Angaben zur Reproduzierbarkeit fehlen in unserem Beispiel. Für die technischen Details wird auf die Literatur verwiesen. Die Interpretation des semiquantitativen Tests wird nach definierten Kriterien auf positiv oder negativ vereinfacht. Dies entspricht nach unserer Erfahrung der Wirklichkeit des praktischen Alltags.

– *Kann eine klinisch vernünftige Vortest-Wahrscheinlichkeit für die Zielkrankheit meines Patienten geschätzt werden?*

Die Ermittlung der Vortest-Wahrscheinlichkeit ist ein Schlüsselpunkt im Alltag eines jeden Klinikers. Wir können uns hierzu auf unsere Erfahrung abstützen, doch ist unser Erinnerungsvermögen an ähnliche Patienten vielfach einseitig, da wir uns an verpasste Diagnosen eher erinnern oder von der Furcht, seltene behandelbare Krankheiten nicht zu verpassen, geleitet werden. Falls die Merkmale der Studienpopulation im Vergleich mit unseren Patienten dies zulassen, können wir auch von den Vortest-Wahrscheinlichkeiten der Studienpopulation ausgehen.
Die Studie in unserem Beispiel zeigt eine weitere Möglichkeit auf. Die Autoren haben sich in der Abschätzung der Vortest-Wahrscheinlichkeit einen validierten Score von Wells et al. zu Nutzen gemacht (Tabelle 4) [12]. Darin werden wichtige anamnestische und klinische Kriterien zur Wahrscheinlichkeitsschätzung einer TVT in ein Punktesystem zusammengefasst. Die Anwendung dieses klinischen Scores muss natürlich ebenfalls als Test mit Sensitivität, Spezifität und Likelihood Ratios angesehen werden. Sind viele Kriterien erfüllt, erreicht der «Wells-Test» eine hohe Punktezahl. Die Vortest-Wahrscheinlichkeit steigt von 25,5% (Prävalenz in der untersuchten Patientengruppe von Wells) auf eine Nachtest-Wahrscheinlichkeit von 80%. Umgekehrt sinkt die Vortest-Wahrscheinlichkeit bei kleiner Punktezahl, wie im Beispiel unseres klinischen Szenarios, auf eine Nachtest-Wahrscheinlichkeit von 3–10%. Dies ist dann die Vortest-Wahrscheinlichkeit für den nächsten Test, die Bestimmung der D-Dimere.
Eine kritische Evaluation der Kriterien von Wells übersteigt den Rahmen dieser Arbeit,

aber Kriterien zum «Critical Appraisal» einer Entscheidungsregel (Decision Rule) bestehen [7, 13]. Die Entscheidungsregel von Wells und die angegebene Prävalenz in der Studie sind brauchbare Richtlinien zum Abschätzen der Vortest-Wahrscheinlichkeit einer TVT.

– *Liegt die Nachtest-Wahrscheinlichkeit in einem Bereich, welche das Management meines Patienten ändern wird?*

Falls der diagnostische Test negativ ist oder eine Likelihood Ratio von 0,1 vorliegt, kann die Nachtest-Wahrscheinlichkeit so niedrig werden, dass wir eine Krankheit ausschliessen können. Das negative Testergebnis hat uns unter die Testschwelle geführt, bei welcher wir weitere diagnostische Massnahmen erwägen würden (Abbildung 2). Umgekehrt, würden wir bei einer mittleren Wahrscheinlichkeit auf der Basis unserer diagnostischen Hypothese falls notwendig eine Behandlung einleiten und die Diagnose in der Zwischenzeit erzwingen. Bei einer hohen Nachtest-Wahrscheinlichkeit für die Zielkrankheit kann je nach Klinik und Krankheit keine weitere Diagnostik notwendig sein und die Behandlung unmittelbar einsetzen oder fortgeführt werden.

Tabelle 4.
Klinisches Modell zur Schätzung der Vortest-Wahrscheinlichkeiten einer tiefen Venenthrombose (TVT) nach Wells [12].

Prävalenz der TVT in dem von Wells untersuchten Patientengut: 25,5%

Klinisches Symptom	Score
Aktives Tumorleiden (unter aktueller Behandlung oder palliativer Behandlung oder Behandlung innerhalb der letzten 6 Monate)	1
Paralyse, Parese oder kürzliche Immobilisation durch Schienung der unteren Extremitäten	1
Kürzlich während mehr als 3 Tage bettlägerig, oder grösserer chirurgischer Eingriff innerhalb der letzten 4 Wochen	1
Lokalisierter Hartspann entlang dem tiefen Venensystem	1
Ganzes Bein geschwollen	1
Wadenschwellung > 3 cm im Vergleich zum asymptomatischen Bein (Messpunkt 10 cm unterhalb der Tuberositas tibiae)	1
Eindellbares Ödem (ausgeprägter als im asymptomatischen Bein)	1
Vorhandensein von oberflächlichen Kollateralvenen (keine Varikosis)	1
Alternative Diagnose gleich wahrscheinlich oder wahrscheinlicher als TVT	–2

Auswertung:

Punktezahl	Wahrscheinlichkeit einer TVT	Wahrscheinlichkeit (%)
0 oder <0	Gering	3–10
1 oder 2	Mittel	17–33
3 oder mehr	Hoch	75–85

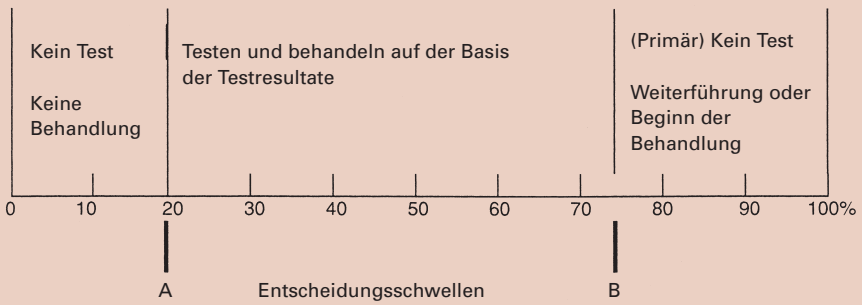

Abbildung 2.
Vortest-Wahrscheinlichkeit und Entscheidungsschwellen nach Sackett et al [7].

Auflösung des Szenarios

Gemäss den Wells-Kriterien (Tabelle 4) schätzen Sie eine niedrige Vortest-Wahrscheinlichkeit für eine DVT von 3% bis 10% [12, 14]. In der ambulanten Population der Studie am Kantonsspital Basel betrug die Prävalenz einer TVT 20,8%. Sie finden die Vortest-Wahrscheinlichkeit von 3% bzw. 10% in der linken Skala des Fagan-Nomogramms. In der mittleren Skala suchen Sie die Likelihood Ratio von 0,23 und lesen auf der rechten Skala eine Nachtest-Wahrscheinlichkeit von 1% bzw. 2,5% ab. Somit beträgt die geschätzte Wahrscheinlichkeit, dass unsere Patientin aufgrund des negativen D-Dimer-Tests eine TVT aufweist, rund 1% bis im ungünstigsten Fall 2,5% Das Beispiel illustriert, dass die Testausbeute bei gegebener Likelihood Ratio von der Vortest-Wahrscheinlichkeit abhängt. Je höher die Vortest-Wahrscheinlichkeit, desto eher lässt sich eine Diagnose bei gegebener Likelihood Ratio bestätigen. Umgekehrt, je niedriger die Vortest-Wahrscheinlichkeit ist, desto eher kann eine Diagnose bei gegebener Likelihood Ratio ausgeschlossen werden.

Inzwischen ist es Mitternacht. Sie sind nun überzeugt, dass Sie bei Ihrer 27jährigen Patientin mit einer niedrigen Vortest-Wahrscheinlichkeit und negativem D-Dimer-Test keine Thrombose verpasst haben und können mit dem «Restrisiko» von ungefähr 2% beruhigt einschlafen. Sie haben Ihre Patientin auf den folgenden Tag nochmals einbestellt und werden den Verlauf beurteilen («*Test of Time*»). Eine absolute Sicherheit in der Medizin gibt es schliesslich nur in den Köpfen gewisser Juristen. Der nächste Feierabend ist für einen Konzertbesuch mit dem Partner, der Partnerin reserviert.

Literatur

1 Aus der Au C, Wuillemin WA. D-Dimer-Test bei Verdacht auf venöse Thromboembolie. Ars Medici 1999; 11:751.
2. Farrell S, Hayes T, Shaw M. A negative SimpliRED D-dimer assay result does not exclude the diagnosis of deep vein thrombosis or pulmonary embolus in emergency department patients. Ann Emerg Med 2000;35:121-5.
3. Anderson DR, Wells PS, Stiell I, MacLeod B, Simms M, Gray L, et al. Management of patients with suspected deep vein thrombosis in the Emergency Department: Combining use of a clinical diagnosis model with D-dimer testing. J Emerg Med 2000;19:225-30.
4. Evidence-Based Medicine. BMJ Publishing Group; 2000. Vol. 5; 3; 92-93
5. Aschwanden M, Labs KH, Jeanneret C, Gehrig A, Jaeger KA. The Value of rapid D-dimer testing combined with structured clinical evaluation for the diagnosis of deep vein thrombosis. J Vasc Surg 1999;30: 929-35.
6. Lennox AF, Konstantinos TD, Serunkuma S,Daskalopoulou SE, Nicolaides AN. Combination of a clinical risk assessment score and rapid whole blood D-dimer testing in diagnosis of deep vein thrombosis in symptomatic patients. J Vasc Surg 199930:794-803.
7. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Hayes RB. Evidence-based Medicine. How to practice and teach EBM. Second Edition New York, Edinburgh, London, Toronto, Philadelphia, St. Louis, Sidney: Churchill Livingstone; 2000.
8. Thomson DMP, Krupey J, Freedman SO, Gold P. The radioimmunoassay of circulating carcinoembryonic antigen of the human digestive system. Proc Natl Acad Sci USA. 1969;64:161-7.
9. The PIOPED Investigators Value of ventilation/perfusion scan in acute pulmonary embolism: results of the Prospective Investigation of Pumonary Embolism Diagnosis (PIOPED). JAMA 1990;263:2751-9.
10. Fagan TJ. Nomogram for Bayes theorem. N Engl J Med 1975;293: 257.
11. Pewsner D, Bleuer J, Bucher HC, Battaglia M, Jüni P, Egger M. Der Intuition auf der Spur? Das Bayes'sche Theorem und die Diagnostik in der Grundversorgung. Swiss Medical Forum 2001;1:41-5 und 46-52.
12. Wells PS, Anderson DR, Bormanis J, Guy F, Mitchel M, Gray L, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. Lancet 1997; 350:1795-8.
13. Richardson WS, Wilson M, Guyatt GH, Cook DJ, Nishikawa J for the Evidence-Based Medicine Working Group. Users' Guide to the medical literature: XV. How to use an article about disease probability for differential diagnosis. JAMA 1999; 281: 2114-9.
14. Wells PS, Hirsh J, Anderson D, Lensing A, Fortster G, Kearon C. Accuracy of clinical assessment of deep-vein thrombosis. Lancet 1995;345:1326-30.
15. Bennett NM, Paris MC. Cadiovascular Problems. In: Black ER, Bordley DR; Tape TG; Panzer RJ (editors). Diagnostic strategies for common medical problems. Second Edition Philadelphia, Pennsylvenia: American College of Physicians; 1999. pp. 47-60.

Weiterführende Literatur zum Thema

Bucher HC, Schmidt JG, Steurer J. Kritische Beurteilung einer Arbeit zu einem diagnostischen Test. Schweiz Rundsch Med Prax 1998;87:1096-102.

Greenhalg T. How to read a paper. The basics of evidence based medicine. London: MJ Publishing Group; 1997.