

# Diagnostische Forschung

## Alternative Konzepte für die Forschung und Kommunikation der Resultate

L. M. Bachmann<sup>a</sup>, J. Steurer<sup>a</sup>, G. ter Riet<sup>b</sup>

<sup>a</sup> Horten Zentrum für praxisorientierte Forschung und Wissenstransfer, Zürich

<sup>b</sup> Academic Medical Centre, Dept General Practice, Amsterdam, NL

In der Therapieforschung<sup>1</sup> wurden in den letzten Jahrzehnten robuste Methoden zur validen Durchführung von Studien und hilfreiche Darstellungen zur Kommunikation der Resultate entwickelt. Durch das wachsende Interesse einer breiten wissenschaftlichen Öffentlichkeit an Evidence-based Medicine (EBM) wurden diese Erkenntnisse effektiv verbreitet und sind heute aus der klinisch-medizinischen Forschung nicht mehr wegzudenken. Ein Beispiel dafür ist der NNT-Wert («number needed to treat»), der angibt, wie viele Patienten mit einem Medikament während einer bestimmten Zeit behandelt werden müssen, um eine Krankheit oder die Komplikation einer Erkrankung zu verhindern (Myokardinfarkt, Schlaganfall). Darstellungsformen von Resultaten wie diese erleichtern dem Arzt, die Wirksamkeit eines Medikaments einzuschätzen und in seine klinische Praxis zu integrieren.

Im Gegensatz dazu ist im Bereich der diagnostischen Forschung<sup>2</sup> relativ wenig passiert. Erstaunlicherweise, denn erst die richtige Interpretation und Verwendung von diagnostischer Information erlaubt es dem behandelnden Arzt, Entscheide bezüglich einer adäquaten Behandlung zu treffen. Die sogenannte «likelihood ratio» (LR) ist das einzige bisher ernsthaft portierte Konzept zur Interpretation und Integration von diagnostischer Forschung in die klinische Praxis. Die LR macht eine Aussage über die Wahrscheinlichkeit, ein bestimmtes Testresultat unter Erkrankten zu finden im Vergleich zu Nichterkrankten. In Kombination mit einem Nomogramm [1] lässt sich damit bei bekannter Vortestwahrscheinlichkeit die Nachtestwahrscheinlichkeit bestimmen.

<sup>1</sup> Studien, die sich mit der Wirksamkeit einer medizinischen Intervention beschäftigen

<sup>2</sup> Studien, die sich mit der Aussagekraft eines diagnostischen Tests beschäftigen

### Korrespondenz:

Dr. med. Lucas M. Bachmann

Horten Zentrum für praxisorientierte Forschung und Wissenstransfer

Universitätsspital Zürich

Bolleystrasse 40

CH-8091 Zürich

Die Ergebnisse von Diagnosestudien werden vorwiegend mit den Begriffen Sensitivität und Spezifität angegeben [2, 3]. Das vor über 40 Jahren beschriebene Konzept [4] erfreut sich fast ungebrochener Beliebtheit bei Autoren und Herausgebern medizinischer Fachzeitschriften. Seit Jahrzehnten wird den Studierenden und den Ärzten im Studium und in Fortbildungsveranstaltungen gelehrt, wie diese Begriffe definiert sind und wie diese Information in den klinischen Entscheidungsprozess integriert werden sollten. Trotz breiter Veröffentlichung hat das vorgeschlagene Konzept allerdings bis zum heutigen Tag nicht systematisch Einzug in die klinische Praxis gefunden. Es gibt Hinweise, die vermuten lassen, dass Entscheidungen auf der Basis von Testgenauigkeitswerten umständlich und fehlerhaft sind [3, 5–7].

In einer von Reid et al. [3] publizierten Studie wurden 300 Ärzte befragt, wie häufig sie die Sensitivität und Spezifität bei der Interpretation diagnostischer Tests anwenden. Gerade zwei Prozent verwenden diese Werte für die Berechnung der Nachtestwahrscheinlichkeit. Ähnliche Resultate wurden auch für die Verwendung der «likelihood ratio» in der Praxis gefunden. In der gleichen Untersuchung von Reid verwendeten weniger als ein Prozent der befragten Ärzte die LR zur Beurteilung der Erkrankungswahrscheinlichkeit.

Wenn man nach den Gründen fragt, warum diese Informationen nicht verwendet werden, so geben drei Viertel der Ärzte an, dass dieses Vorgehen im klinischen Alltag nicht praktikabel ist. Wenn die Sensitivität und Spezifität eines Testes bekannt sind, ist es für viele Ärzte schwierig, die Nachtestwahrscheinlichkeiten zu berechnen [5, 6]. In einer Studie konnten wir zeigen [7], dass Ärzte bei der Beurteilung der Krankheitswahrscheinlichkeit keinen Unterschied machen, ob man nun mitteilt, dass das Testresultat positiv ist oder zusätzlich zum Testresultat die Werte für die Sensitivität und Spezifität des betreffenden Tests angibt. In beiden Situationen haben die Ärzte die Wahrscheinlichkeit um das drei- bis fünffache überschätzt. Die Konsequenzen dieser Fehleinschätzung von Testresultaten für den einzelnen Patienten und/oder das Gesundheitssystem wurden noch kaum untersucht.

Die methodische Forschung der letzten Jahre hat ebenfalls gezeigt, dass ein diagnostisches Konzept, welches auf der Verwendung von LRs aufbaut, Mängel aufweist, sobald diagnostische Information verschiedener Tests integriert werden soll [8–11].

Dies rührt daher, dass Tests innerhalb einer Serie häufig abhängig sind, d.h., dass beispielsweise nach einem Bluttest, der Hinweise auf eine Myokardschädigung gibt, das EKG mit grösserer Wahrscheinlichkeit pathologisch ausfallen wird, als wenn die Information der Blutuntersuchung fehlt.

Studien, welche nicht nur einen Test untersuchen, sondern die ganze diagnostische Kette, wie sie in der klinischen Praxis angewendet wird, analysieren, sind aufwendig und selten. Am Beispiel der Abklärung von postmenopausalen Frauen mit Verdacht auf ein Endometriumkarzinom haben wir versucht, diese

diagnostische Kette in einer Studie abzubilden, und konnten so den Beitrag von Anamnesefragen und bildgebenden Verfahren bestimmen [12]. Es konnte gezeigt werden, dass eine Ultraschalluntersuchung nach der Beantwortung von einigen Anamnesefragen keinen grossen Beitrag liefert, ein Endometriumkarzinom ein- oder auszuschliessen, dies obschon die publizierte Sensitivität und Spezifität dieser Untersuchung durchwegs hoch ist [13]. In der klinischen Realität werden vor einer Ultraschalluntersuchung mehrere Tests (Anamnese und klinische Untersuchung) gemacht. Die Frage an die Forschung ist demnach, welche Aussagekraft die Informationen der Anamnese und der körperlichen Untersuchung haben und welche *zusätzlichen* Informationen der Ultraschall oder andere bildgebende Verfahren liefern. Wenn beispielsweise die Anamnese und klinische Untersuchungen bereits eine sehr gute Differenzierung zwischen Gesunden und Kranken erlauben, so verändert der Ultraschall die Klassifizierung von Gesunden und Kranken nicht wesentlich.

Die genannten Beispiele und Erläuterungen sollten aufzeigen, dass noch viele Probleme in der diagnostischen Forschung zu lösen sind, dass sich jedoch auch einige methodische Entwicklungsschritte der letzten Jahre in der Forschung abzuzeichnen beginnen. Die Bedeutung diagnostischer Forschung wächst stetig und vielerorts organisieren sich Forschergruppen, um Lösungsbeiträge zu liefern. Besonders erwähnenswert ist eine kürzlich lancierte Initiative, die «Bayes Library of Diagnostic Studies and Reviews», welche unter der Schirmherrschaft des neugegründeten Institutes für klinische Epidemiologie in Basel (BICE) und unterstützt von internationalen Experten gegründet wurde. Ziel dieses Projektes ist die Aufbereitung und Verbreitung diagnostischer Forschungsergebnisse (Sensitivitäten, Spezifitäten, Likelihood Ratios) für die klinische Praxis.

Des Weiteren werden zur Zeit von einem Expertengremium Richtlinien zur adäquaten Analyse und Publikation von diagnostischen Forschungsergebnissen entwickelt (STARD), die analog zur Therapieforschung (Consort Statement) [14] und zu Metaanalysen von Observationsstudien (MOOSE) [15] einen Standard für qualitativ hochstehende diagnostische Forschung setzen sollen.

Letztendlich muss es das Ziel der diagnostischen Forschung sein, valide und verständliche Resultate zu generieren. Die Resultate müssen derart dargestellt und kommuniziert werden, dass sie von den Ärzten genutzt und in den klinischen Entscheidungsprozess integriert werden können. Eine Voraussetzung dafür ist die Information, ob und inwiefern ein Testresultat

die Behandlung des Patienten beeinflusst, und eine Stufe höher, wie die diagnostische Information gekoppelt an eine effektive Behandlung den Outcome des Patienten verbessert.

Um diese Aspekte zu diskutieren, möchten wir im Rahmen eines Kolloquiums in Zürich am 14. Februar 2002 zum Thema: «Diagnostische Forschungsmethoden: Wege zur verständlichen Darstellung der Resultate diagnostischer Studien» ein Forum für die diagnostische Forschung schaffen.

## Literatur

- 1 Fagan TJ. Nomogram for Bayes theorem. [Letter]. *N Engl J Med* 1975;293(5):257.
- 2 Guggenmoos-Holzmann I, van Houwelingen HC. The (in)validity of sensitivity and specificity. *Stat Med* 2000;19(13):1783-92.
- 3 Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; 104(4):374-80.
- 4 Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis. *Science* 1959;130:9-21.
- 5 Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science* 2000; 290(5500):2261-2.
- 6 Gigerenzer G. The psychology of good judgment: frequency formats and simple algorithms. *Med Decis Making* 1996; 16(3):273-80.
- 7 Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating test accuracy terms to practicing physicians – a controlled study. *BMJ* in press 2001.
- 8 Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE. Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology* 1997;8(1):12-7.
- 9 de Korte PJ. Probability analysis in diagnosing Coronary Artery Disease. Maastricht: Universitaire Pers; 1993.
- 10 Miettinen OS. The modern scientific physician: 3. Scientific diagnosis. *CMAJ* 2001;165(6):781-2.
- 11 ter Riet G, Kessels AGH, Bachmann LM. Systematic reviews of evaluations of diagnostic and screening tests – Two issues were simplified [Letter]. *Br Med J* 2001;323:1188.
- 12 Bachmann LM, ter Riet G, Clark TC, Gupta JK, Khan KS. Probability analysis for diagnosis of endometrial disease in postmenopausal bleeding: An approach for a rational diagnostic workup. *Gynecol Oncol* submitted 2001.
- 13 Bronz L, Dreher E, Almendral A, Studer A, Haller U. Abklärung von postmenopausalen Blutungen (PMBP). *Schweiz Ärztezeitung* 2000;81:1635-46.
- 14 Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001;285(15): 1987-91.
- 15 Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283(15):2008-12.