

Das Praxislabor: die scholastischen Täuschungen der «Qualitätskontrolle»

Bei der externen Kontrolle des Einzellabors unterliegen Ringversuche einem Denkfehler: Abweichungen durch fehlerhaftes Arbeiten können von Zufalls-Abweichungen nicht unterschieden werden. Trotz korrekter Bestimmung kann ein Praxislabor bei Vorliegen einer grösseren Zufalls-Abweichung als ungenügend taxiert werden.

Johannes Schmidt

Allgemeinmedizin, Klinische Epidemiologie, Altchinesische Medizin

Ein Kommentar von Roman Fried, Verein für Medizinische Qualitätskontrolle, findet sich am Ende des Artikels.

Vor gut zwanzig Jahren sind mir nach kurzer Teilnahme an den Labor-Ringversuchen die Fehler bei dieser Art Qualitätskontrolle aufgefallen; zu meiner Ausbildung gehörte ein Studium der Klinischen Epidemiologie. Heute Sorge ich für Qualität in eigener Regie – *alterius non sit qui suus esse potest**.

Fehler durch Zufalls-Variabilität

Bei der externen Kontrolle des Einzellabors unterliegen Ringversuche einem Denkfehler: Abweichungen durch fehlerhaftes Arbeiten können von Zufalls-Abweichungen nicht unterschieden werden. Trotz korrekter Bestimmung wird ein Praxislabor deshalb als ungenügend taxiert, wenn eine grössere Zufalls-Abweichung das Ergebnis mitbestimmt. Trotz fehlerhafter Arbeit wird ein anderes Praxislabor als gut arbeitend taxiert, wenn eine fehlerbedingte Abweichung durch eine gegenläufige Zufalls-Abweichung wieder aufgehoben wird. Der Labor-Ringversuch verändert die Gesetze der Stochastik nicht, er übersieht oder ignoriert sie nur. Es wäre also eine Täuschung zu glauben, dank Qualitätskontrolle könne es Laborresultate ohne Fehler geben. Und es ist erst recht eine Täuschung zu meinen, das Praxislabor würde dank Überprüfung im Ringversuch fehlerfrei arbeiten. So verdunkeln Ringversuche die Qualität, auf die es ankommt: Nämlich, dass der Arzt bei der Interpretation eines Laborresultates stets weiss, dass Zufalls-Abweichungen unvermeidlich sind – und dass die Richtigkeit von Laborergebnissen meist weniger von der Labor-Präzision abhängt als von der Variabilität der Biologie (s. unten).

Die Zufalls-Einwirkung bildet sich ab in der Gauss'schen Kurve – einer mehr oder weniger breiten, glockenförmigen Kurve mit niedrigen, asymptotischen («unendlichen») Ausläufern an der Basis. Die Zufalls-Einwirkung verhindert, dass sich eine «fehlerfreie», dünne senkrechte Linie mit dem immer gleichen Wert (auf der Abszisse) bildet. Zufallsbedingte «Störungen» entstehen z. B. durch unvermeidbare Temperaturschwankungen, Schwankungen im Stromnetz, Schwankungen bei der Vermischung und anderem mehr. 68% der Ergebnisse liegen innerhalb einer Standardabweichung, 95% in-

nerhalb von 2 Standardabweichungen, und mit 1% Wahrscheinlichkeit kommen Ergebnisse mit über 2½ Standardabweichung vor. Durch grobe Zufalls-Ausreisser kann ein technisch korrektes Laborergebnis gelegentlich weit weg vom wahren Wert liegen.

Auch der Labor-Ringversuch zeigt diese Gesetzmässigkeiten. Zum einen weicht der Mittelwert aus allen Ringversuchs-Resultaten vom Zielwert des Zentrallabors immer etwas ab. Denn auch das Zentrallabor kann den wahren Wert der jeweiligen Probe nicht genau bestimmen, weil dieses ebenfalls durch Zufalls-Einflüsse gestört wird. Zum andern zeigt die Resultate-Verteilung eines Ringversuchs annäherungsweise die Form der Gauss'schen Glocke. Natürlich sind dabei Abweichungen, die durch falsches Arbeiten im Praxislabor zustande kommen, nicht ausgeschlossen, doch sie lassen sich von Zufalls-Abweichungen nicht auseinanderhalten. Wenn ein Praxislabor also die Note *ungenügend* erhält, weiss man nicht, ob ein Zufallsfehler oder falsches Arbeiten zu diesem Ergebnis geführt hat. Wenn ein Praxislabor eine gute Note erhält, weiss man wiederum nicht genau, ob nicht eine falsche Hantierung durch eine Zufalls-Abweichung ausgeglichen worden ist. Es lässt sich aufgrund der Verteilungs-Varianz der eingesandten Resultate lediglich eine Aussage darüber machen, ob im ganzen untersuchten Teilnehmer-

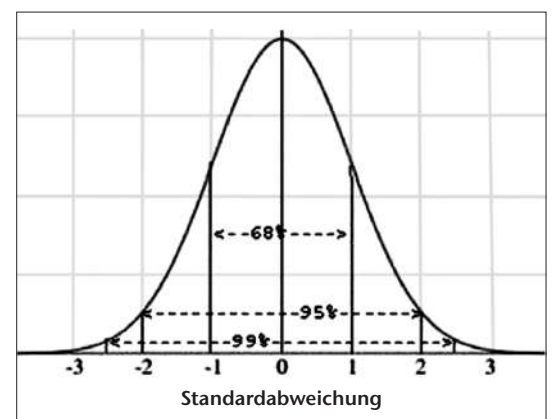


Abbildung 1:
Gauss'sche Normalverteilung.

* Wahlspruch von Paracelsus:
«Es sei kein bequemer
Scholastiker, wer eigenständig
urteilen kann.»

Korrespondenz:
Dr. med. Johannes Schmidt
Praxiszentrum Meinradsberg
Stiftung Paracelsus heute
CH-8840 Einsiedeln
schmidt[at]paracelsus-heute.ch

System mehr oder weniger Fehler gemacht worden sind, der Ringversuch ist jedoch nicht in der Lage, die Fehler zu lokalisieren. In der Industrie-Produktion, von wo Ringversuche herkommen, kann diese Art Kontrolle sinnvoll sein. Die Kontrolle des Einzella-bors kann auf diese Weise aber nicht geleistet werden.

Wir müssen uns bewusst bleiben, dass Laboreergebnisse eine Pathologie nie direkt anzeigen oder ausschliessen.

Effektive Qualitätssicherung heisst in der Praxis: Ist ein Laborresultat aufgrund einer unpassenden Klinik oder aus anderen Gründen nicht vertrauens-

würdig, kann und sollte es ignoriert oder wiederholt werden, evtl. mit Wiederholung der Blutentnahme. Bleibt der Zweifel bestehen, dann kann die Probe zusätzlich in ein Fremdlabor geschickt werden. So zeigt sich, ob das eigene Praxislabor einen Fehler gemacht hat – gezielt dann, wenn es darauf ankommt. Letztlich müssen wir Laborresultaten gegenüber immer misstrauisch sein, weil sie aus weiteren Gründen falsch sein können – falsch positiv oder falsch negativ.

Fehler durch biologische Variabilität

Dass ein pathologisches oder normales Laboreergebnis nicht für bare Münze genommen werden kann, liegt an weiteren Formen der Variabilität, welche falsch positive und falsch negative Ergebnisse produzieren: Variabilität in der biologischen Reaktion (z.B. Antikörperbildung, Kreuzreaktionen) und biologische Variabilität der Parameter bei Gesunden und Kranken, die sich stets mehr oder weniger stark überlappen. Entsprechend beinhaltet eine Laborbestimmung in der Arztpraxis grundsätzlich folgende Schritte:

- Klinische Einschätzung des Patientenproblems und der (Vortest-)Wahrscheinlichkeit einer vermuteten Diagnose bzw. Pathologie
- Identifikation von Labortests, die eine grössere diagnostische Gewissheit erlauben könnten
- Abschätzung der Vortest-Wahrscheinlichkeit und der zu erwartenden Nachtest-Wahrscheinlichkeiten und des daraus entstehenden Informationsgewinns
- Einschätzung des Informationsgewinns in Bezug auf dessen klinische Relevanz und therapeutische Konsequenzen
- Einschätzung der Nachteile möglicher Überdiagnose
- Durchführung oder Unterlassung der Laborbestimmung
- Wenn durchgeführt, Einschätzung des Resultates: eher richtig oder falsch positiv bzw. eher richtig oder falsch negativ?

Bei der Interpretation gilt es die Spezifität und Sensitivität eines Labortests zu berücksichtigen – wie uns allen bekannt ist. Eine Sensitivität von 95% bedeutet, dass mit 5% Wahrscheinlichkeit ein falsch negatives Resultat erfolgt und die Pathologie verpasst wird. Eine Spezifität von 95% bedeutet, dass mit 5% Wahrscheinlichkeit ein falsch positiver pathologischer Wert, also eine Pathologie-Fehlannonce, erfolgt. In der Praxis sind Pathologie-Fehlannoncen allerdings viel häufiger, als in der Spezifitäts-Zahl zum Ausdruck kommt. Ist der Pathologie-Verdacht vor dem Test z.B. nur 10% oder 1%, dann wird der Zuordnungs-Fehler verzehnfacht bzw. verhundertfacht, sodass ein pathologisches Ergebnis mit nur noch einer Wahrscheinlichkeit von 66% bzw. 16% richtig ist – also nicht mehr mit 95% (beim Beispiel einer Sensitivität und Spezifität von je 95%). In 34% bzw. 84% besteht dann eine Pathologie-Fehlannonce. Eine einfache Vierfelder-Tafel kann dies demonstrieren (s. Abb. 2).

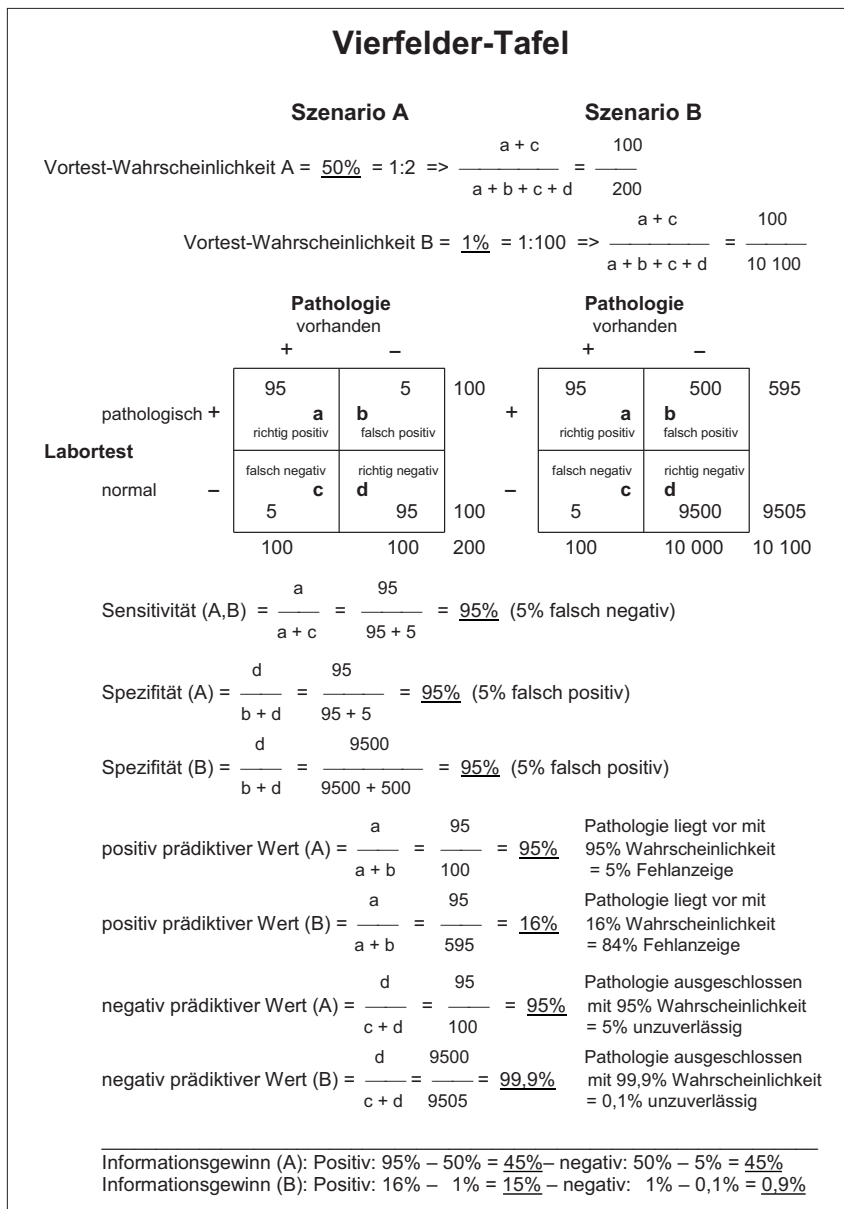


Abbildung 2: die Vierfelder-Tafel.

Der negativ prädiktive Wert als das Mass des Pathologie-Ausschlusses wird ebenfalls wesentlich von der Vortest-Wahrscheinlichkeit – von Anamnese und Klinik – mitbestimmt. Zwischen positiv prädiktivem Wert (bzw. Spezifität) und negativ prädiktivem Wert (bzw. Sensitivität) besteht ein Konkurrenz-Verhältnis, sodass eine Fehler-Minimierung beim einen den Fehler beim anderen erhöht. Man kann also nicht «zur Sicherheit» gerade mal alle möglichen Laborbestimmungen vornehmen, ohne das Risiko von Pathologie-Fehlanzeigen zu erhöhen. Ein Labortest bringt viel Informationsgewinn, wenn die Vortest-Wahrscheinlichkeit, die wir aufgrund der anamnestischen und klinischen Daten gewinnen, schon relativ hoch ist. Ein pathologisches Testergebnis in Szenario A bestätigt den Vortest-Krankheitsverdacht, indem es die Gewissheit von 50% auf 95% erhöht (wenn der Test 95% sensitiv und 95% spezifisch ist). Ein normales Testergebnis widerspricht dem Krankheitsverdacht und erhöht in Szenario A die Gewissheit, dass die entsprechende Pathologie doch nicht vorliegt, von 50% auf 95%. Ein Test in einem derartigen Szenario mit hoher Vortest-Wahrscheinlichkeit hat in der Regel therapeutische Konsequenzen. Ein pathologisches Testergebnis in Szenario B erhöht die Gewissheit hingegen lediglich von 1% auf 16%. Und ein normales Testergebnis in Szenario B erhöht die Gewissheit, dass die entsprechende Pathologie nicht vorliegt, von 99% auf 99,9%. Ein Test in einem derartigen klinischen Kontext hat in der Regel keine therapeutische Konsequenzen, birgt die Gefahr einer Pathologie-Fehlanzeige mit unerwünschten Auswirkungen und kostet Geld für nichts.

Diagnostik macht auch krank

Ein technisch noch so korrektes und wahres Laborresultat zeigt uns also noch nicht viel. Es kann aufgrund der biologischen Variabilität richtig oder falsch sein – richtig positiv, richtig negativ oder falsch positiv, falsch negativ. Wir müssen uns deshalb stets bewusst bleiben, dass Laboregebnisse eine Pathologie nie direkt anzeigen oder ausschliessen. Nicht wegen möglicher stochastischer und technischer Fehler in der Laborbestimmung selbst (die kaum mehr eine Rolle spielen), sondern hauptsächlich wegen der «störenden» biologischen Variabilität, die es verunmöglicht, scharf und zuverlässig zwischen normal und pathologisch zu unterscheiden. Absicherung und «Durch-Checken» ist deshalb schnell eine Täuschung und hat die Nebenwirkung falsch positiver und manchmal auch falsch negativer Fehlanzeigen. Wann immer kontrollierte Studien über den Nutzen des «Check-Ups» durchgeführt worden sind – bis zu 16 Jahre Follow-up –, zeigte sich nicht nur eine fehlende präventive Wirksamkeit, sondern eine unerwünschte Zunahme von Hospitalisationen – falsch positive Fehlanzeigen, die aufwendigere Abklärungen «benötigten».

So wird man in den Händen der Medizin nicht nur gesund, sondern auch krank: Durch falsch positive Befunde, die wiederholt nachkontrolliert werden «müssen», und durch Überdiagnose und Überbehandlung. Im Unterschied zum Universitätsspital ist in der Hausarzt-Praxis eine geringe Vortest-Wahrscheinlichkeit die Regel. Der Hausarzt muss deshalb anders denken. Er muss lernen, ein «pathologisches»

Durch eine externe Kontrolle lässt sich die Qualität einzelner Laborbestimmungen nicht prüfen.

Ergebnis oft als normal zu betrachten, denn entscheidend ist die Nachtest-Wahrscheinlichkeit. Sich nicht in kollektive Absicherungs-Rituale mit unerwünschten Wirkungen zu flüchten, sondern die natürliche Rest-Ungewissheit auf sich zu nehmen, ist eine wichtige ärztlich-menschliche Leistung. Genaue Anamnesen vermitteln Erfahrung und geben Sicherheit. Dort, wo die Prognose gut ist oder sich durch Interventionismus nicht eindeutig verbessern lässt, kann ein bewusster Diagnose-Verzicht am meisten Nutzen bringen. Zukunftsweisend ist die Diskussion, die über das PSA-Screening begonnen hat. Das Swiss Medical Board rät vom PSA-Screening ab, weil unerwünschte Überdiagnose und Überbehandlung den zu geringen, fraglichen präventiven Nutzen übertreffen. Denn beide, PSA und Histologie, haben eine zu fehlerhafte klinische Prädiktivität. Die Diskussion über das PSA-Screening zeigt allerdings auf, wie schwer es grossen Teilen der Ärzteschaft fällt, zu erkennen, dass Diagnostik – und im Speziellen Frühdiagnose – auch Nachteile hat. Doch man kann nicht übersehen, dass selbst Menschen mit gefürchteten Pathologien vielfach gesund bleiben, wenn wir diese nicht unnötig diagnostizieren.

Diagnostischer Informationsgewinn im klinischen Szenario

In der Kette von der ärztlichen Diagnostik bis zur Intervention braucht es ein tieferes Verständnis von fehlerhaften Zusammenhängen auf drei Stufen:

- zwischen Pathologie und effektivem Kranksein,
- zwischen pathologischem Untersuchungsergebnis und effektiver Pathologie und
- zwischen Messwert und wahren Wert.

Das führt dazu, dass in der Regel schon eine «halbpatzige», semi-quantitative Laborpräzision in der Praxis genügt, um die klinische Frage zu beantworten. Denn klinisch relevante Nachtest-Wahrscheinlichkeiten hängen nur wenig von der Laborpräzision ab. Wir brauchen eine optimale, aber nicht unbedingt eine

maximale Präzision, auf welche die dazu nicht einmal geeigneten Ringversuche allein abzielen. Scheinbare «Qualitätsverbesserung», die überdies die entscheidenden Fragen verdeckt, sollten wir als unerwünschte scholastische Betriebsamkeit erkennen. In unseren früheren Gesprächen hatte der damalige FMH-Präsident Hans-Heinrich Brunner diese Kritik akzeptiert, machte aber politische Gründe geltend. Es brauche dringend eine Labor-Qualitätskontrolle, um das Praxislabor zu erhalten. Zwanzig Jahre später bleibt das Praxislabor weiterhin unter Beschuss. Zu Recht, weil die falsche «Qualitätssicherung» keine echte Qualität verbessert und über die eigentlichen Probleme wie Überdiagnose und Labelling hinweggeschaut hat. Zu Unrecht, wo das gezielt eingesetzte Praxislabor das Management durch den Hausarzt verbessert und erleichtert. Diagnostik ist eine nicht immer einfache Gratwanderung zwischen Zuviel und Zuwenig. Es gehört zur ärztlichen Aufgabe, *diese* Qualität richtig zu verstehen und der Öffentlichkeit gegenüber zu vertreten. Hausärzte sind die Spezialisten, die den Wald vor lauter Bäumen noch sehen sollten – um den Wald und seine Bäume zu bewahren vor den Nebenwirkungen und unnötigen Kosten von zuviel spezialisierten Maximal-Interventionen. Auch bei der Labor-Qualitätssicherung wollen die Spezialisten der Klinischen Chemie nur das Beste – nur ist das schnell etwas anderes, als in der Praxis nötig und effektiv ist.

Durch eine externe Kontrolle lässt sich die Qualität einzelner Laborbestimmungen nicht prüfen. Und weil in der Praxis zu viel Labor gemacht wird, ist der jeweilige Informationsgewinn vielfach so gering, dass allfällige technische Fehler gar nicht mehr relevant sind. Auf den heutigen Ringversuchen lässt sich vielleicht aufbauen, wenn die entsprechenden Kontrollzentren sich darauf beschränken, uns mit ihren Auswertungen die Existenz der Zufalls-Variabilität zu demonstrieren und in Erinnerung zu rufen statt hinter irreführenden Benotungen zu verstecken. Ein Speziemen-Versand als Teil der Qualitätsbemühungen könnte dann seinen Platz behalten, wenn eine Ergänzung erfolgt, etwa durch ein klinisches Szenario mit entsprechenden Fragen zum diagnostischen Informationsgewinn. Ein kunstgerechtes «Medical Decision Making» und die Bewertung des Labors im operativen Kontext der ärztlichen Kunst würde dann Gegenstand «richtig positiver» Qualitätsbemühungen. Das wachsende Problem unreflektierter, iatrogenen Überdiagnose, das neuerdings unter dem Begriff der Quartär-Prävention thematisiert wird, könnte so die nötige Aufmerksamkeit bekommen.

Literatur zum Thema findet sich im Internet unter www.saez.ch → Aktuelle Nummer oder → Archiv → 2012 → 48.

Kommentar von Dr. Roman Fried, Verein für medizinische Qualitätskontrolle

Variabilität ist ein zentraler Aspekt in der Arbeit jedes medizinischen Laboratoriums. Deshalb bin ich Dr. Schmidt sehr dankbar für seine Ausführungen und für die Möglichkeit, diese durch meine Gedanken zu ergänzen.

Ringversuche haben nichts mit Schulnoten zu tun! Externe Qualitätskontrolle ist ein Werkzeug, um die Qualität des Labors zu verbessern. Nicht das Labor oder die zuständige Mitarbeiterin sind «ungenügend», sondern ein Messwert liegt ausserhalb der Toleranz. Niemand wird bestraft, sondern das Labor wird durch den ungenügenden Messwert auf ein mögliches Problem aufmerksam gemacht, analysiert die Situation und trifft die nötigen Verbesserungs-massnahmen.

Wenn ein Labor eine Glukose-Konzentration misst, die 20% höher als der Zielwert ist, kann das Zufall sein, wie Dr. Schmidt in seinem Artikel schreibt. Wenn aber dank der internen Qualitätskontrolle bekannt ist, dass das Gerät einen Variationskoeffizient von 5% besitzt, dann ist die Chance, dass trotz des ungenügenden Messwerts alles in Ordnung ist, kleiner als ein Promille. Das Beispiel zeigt, dass die entscheidende Massnahme im Umgang mit Variabilität die Quantifizierung ist. Sobald die Variabilität in Zahlen gefasst wird, kann man trotz Unsicherheit sichere Entscheide fällen.

Interne und externe Qualitätskontrollen sind die wichtigsten Werkzeuge, mit denen die Leistungsdaten eines Analysensystems ermittelt und überwacht werden. Ein Verzicht auf diese Massnahmen ist deshalb unverantwortlich.

Es ist leider ein Irrtum zu meinen, durch Qualitätskontrollen würde die Leistungsfähigkeit eines Analysensystems verbessert. Die Aufgabe der Kontrollen ist es, das Labor zu alarmieren, wenn sich die Leistung des Analysensystems verschlechtert. Analytische Probleme können zu Fehlern führen, die so gross sind, dass die Laborwerte auch aus klinischer Sicht unbrauchbar werden.

Wer nicht bereit ist, den Aufwand für eine korrekte, zeitgemässe Analytik aufzubringen, sollte seine Patientenproben lieber in ein Auftragslabor schicken, statt «halbhatzige» Analysen zu machen. Selbstverständlich kann die technische Qualitätskontrolle alleine nicht garantieren, dass ein Labor korrekt arbeitet. Der Verein für medizinische Qualitätskontrolle organisiert schon seit vier Jahren einen Ringversuch zur Prä- und Postanalytik in dessen Rahmen auch der diagnostische Informationsgewinn diskutiert wird. Wir engagieren uns auch stark im Bereich der Aus- und Weiterbildung der Medizinischen Praxisassistentinnen, deren sorgfältige und kompetente Arbeit die Grundlage für eine zuverlässige Laboranalytik bildet.

Literatur zum Thema

- Dubs L. Die arthroskopische Gelenksäuberung bei Kniearthrose. *Ars Medici*. 2010; www.paracelsus-heute.ch/cms/literatur/PDF/ArsMedici2011_Dubs.pdf
- Feinstein AR. The hard data creed in current clinical practice: Its spurious validity and the challenge to define meaningful clinical variables. In: Schmidt JG, Steele RE (Hrsg.). *Kritik der medizinischen Vernunft: Schritte zu einer zeitgemässen Medizin – Ein Lesebuch*. Mainz: Kirchheim-Verlag; 1994. www.paracelsus-heute.ch/cms/literatur/011_wiss_einsiedler_symposiumen/PDF_1_symposium/81Feinstein.pdf
- Schmidt JG. Früherkennung und Umgang mit Risikofaktoren. In: Kochen MM (Hrsg.). *Allgemeinmedizin und Familienmedizin (Lehrbuch Duale Reihe)*. Stuttgart: Thieme Verlag; 2006. S. 25–38. www.paracelsus-heute.ch/cms/literatur/PDF/Lehrbuch2006.pdf
- Schmidt JG. Evidence-based guidelines: should guide us in asking for the evidence and information we need. *Eur J Gen Practice*. 1999;5:3–7. www.paracelsus-heute.ch/cms/literatur/PDF/EJGP1999.pdf
- Schmidt JG. Ist Verzicht immer Nihilismus? Über das Handeln aufgrund rationaler Erkenntnisse (Vortrag am Deutschen Internistenkongress 1997). In: Köbberling J (Hrsg.). *Zeitfragen der Medizin*. Berlin, Heidelberg: Springer-Verlag; 1997. S. 115–125. www.paracelsus-heute.ch/cms/literatur/PDF/Springer1997.pdf
- Schmidt JG. Check-up des Check-ups – Qualitätssicherung oder Schildbürgerstreich? *Ars Medici*. 1995;85(11):735–7. www.paracelsus-heute.ch/cms/literatur/PDF/ArsMedici1995.pdf
- Schmidt JG, Bucher HC, Steurer J. Zeitgemässe Beurteilung medizinischen Wissens – «Critical Appraisal» und Methoden der klinischen Epidemiologie für die praxisbezogene Bewertung von Studien. *Kurs-Skriptum 1999*. www.paracelsus-heute.ch/cms/stiftung/methodik_studien/Kursskriptum.pdf
- Sox HC, Higgins MC, Blatt MA, Marton KI. *Medical Decision Making*. American College of Physicians. 2006. www.paracelsus-heute.ch/cms/aktuell/PDF/Sox.pdf
- Steurer J, Bucher HC, Schmidt JG. Kritische Beurteilung einer Arbeit zu einem diagnostischen Test. *Schweiz. Rundsch. Med. Praxis*. 1998;87:1096–102. www.paracelsus-heute.ch/cms/literatur/PDF/Praxis1998_Diagnostik.pdf
- Stiftung Paracelsus heute. *Medizinische Diagnostik: Information und Aberglaube. Patient sein heute – Medizin auf dem Prüfstand kritischer Wissenschaft 1999; Ausgabe 8/9*. www.paracelsus-heute.ch/cms/literatur/015_patienteninfo_stiftung/PDF/01418_Diagnostik.PDF
- Welch HG, Schwartz L, Woloshin Steve. *Overdiagnosed: Making People Sick in the Pursuit of Health*. Beacon Press; 2012.