

Dr. med. Johannes G. Schmidt, Allgemeinpraxis und Institut für Klinische Epidemiologie,
Stiftung «Paracelsus Heute», Einsiedeln

Evidence-Based Medicine und der Wandel von «Qualität»

**Nicht die Kosten und die Qualität der Kanonen sind das Problem, sondern dass man aufhört,
damit auf Spatzen zu schiessen***

Zunehmend entpuppen sich als wirksam und nützlich empfundene medizinische Leistungen als zuwenig durchdacht, auch wenn der aufklärerische Akt einer zu Ende gedachten Medizin noch vermieden werden will. Noch vor einem Jahr wurde im Zusammenhang mit der Verordnung über präventive Leistungen ein Verzicht auf den Routine-Ultraschall in der Schwangerschaft von erbosten Gynäkologen in die Nähe von Mütter- und Kindsmord gerückt, was wenigstens vom «Blick» als publikationswürdig erachtet wurde. Im Rahmen einer vom Basler Frauenspital bewusst veranstalteten «Kontroverse» an der diesjährigen Gynäkologen-Weiterbildung geriet indessen der Versuch dieser «Blick»-Autoren, den angeblich unwiderlegbaren Nutzen des Routine-Ultraschalls darzulegen, zu einem Durcheinander basaler methodischer Konzepte. Evidence-based Medicine verbietet selektives und beliebiges Argumentieren, und ein wichtiges Prinzip ist die systematische Hinterfragung und Sichtung der wissenschaftlichen Evidenz.

Hierarchie der Evidenz

Klinische Erfahrungen und Spezialistenwissen haben zweifellos ihre «Richtigkeit», und niemand will auf das Wissen und entsprechende klinische Erfahrungen eines versierten Ultraschall-Spezialisten und Geburtshelfers verzichten. Die methodischen Kenntnisse der klinischen Epidemiologie führen aber zu zusätzlichen «Richtigkeiten». Zwischen diesen verschiedenen Richtigkeiten besteht eine gewisse Hierarchie, die Wissenschaft spricht von einer Hierarchie der Evidenz.

In der Medizin sind wir gewohnt und ausgebildet worden, uns auf eine untergeordnete Evidenz wie die zweifellos gutbelegte diagnostische Leistungsfähigkeit des Schwangerschafts-Ultraschalls zu beziehen, und noch immer werden daraus irrtümlich weitreichende Nutzen-Folgerungen abgeleitet [1]. Die übergeord-

nete Frage aber, ob ein Fortschritt in der diagnostischen Leistungsfähigkeit auch ohne Täuschung zu einem Nutzen-Ergebnis führt, ist in der Medizin bzw. für die meisten Mediziner immer noch neu. Die Berücksichtigung der Hierarchie der Evidenz ist immer brisanter geworden, weil die Zahl der Studien inzwischen nicht mehr zu übersehen ist, die gezeigt haben, dass auch selbstverständliche und plausible medizinische Massnahmen oft nicht nur keinen Nutzen, sondern auch Schaden gebracht haben. Ein diesbezüglicher «Klassiker» ist die CAST-Studie, welche drastisch gezeigt hat, wie eine «erfolgreiche» Behandlung pathologischer EKG-Erscheinungen am Schluss zu weit mehr Sterbefällen führte als eine Nichtbehandlung.

Evidence-based Medicine ist weniger eine Frage von Cochrane-Reviews und Meta-Analysen denn Ausdruck des methodischen Fortschritts in bezug auf die Aussagekraft verschiedener Arten von klinischen Studien. Es ist eine Revolution, die das herkömmliche Evidenz-Verständnis als ungültig entlarvt hat. Man hat gemerkt, dass nicht pathophysiologische Zusammenhänge und Plausibilitätsüberlegungen für einen anzunehmenden Nutzen gültig sind, sondern ob in geeigneten Studien in der Praxis gezeigt werden kann, dass es für die direkten Leidensparameter der Patienten insgesamt einen fassbaren Unterschied macht. Plausibilität und pathophysiologische Erklärung einer Wirkung sind sekundär geworden. Ebenso sekundär ist, ob eine Wirkung durch paraklinische Untersuchungsergebnisse – Labor, Röntgen, Ultraschall usw. – über das direkte Patientenleiden hinaus dokumentiert werden kann. Die geänderte Hierarchie der Evidenz wirft natürlich auch das medizinische Hierarchie-System durcheinander. Unter Opfern erreichte Facharztstitel entpuppen sich heute unvermittelt als Kataloge überholten Wissens, als des Kaisers neue Kleider, die keine Blöße mehr decken (nur ein naives Kind getraute sich in dieser Geschichte auszusprechen, was es wirklich sah, dass der Kaiser ja nackt sei; all die gebildeten, erwachsenen Leute und Hofschranzen mit mehr «Fachkompetenz» verliessen sich hingegen auf die Gelehrten-

* Nach einem Vortrag an der XXII. Fortbildungsveranstaltung der Universitäts-Frauenklinik Basel

meinung, dass die neuen Kleider etwas ganz Besonderes seien; David Sackett und seine Gruppe haben ihr Buch zur Klinischen Epidemiologie [2] dieser Geschichte gewidmet). Länger als etwa die Innere Medizin wollte die Gynäkologie die anstehende Geburt dieser neuen Hierarchie der Evidenz nicht sehen und hat ein überholtes Wissen über den Termin hinaus getragen, so dass sie nun selbst die komplikationsträchtigen Wehen einer Geburtseinleitung mitmachen muss. Dieses reale Leben kennt weder Anästhesie noch Kaiserschnitt, wenn es am Schluss die bestraft, die zu spät kommen.

Der Schein, der trügt: ein Fallbeispiel

Man kann die Klinische Epidemiologie als eine Meta-Methodik bezeichnen, die sich von pathophysiologischen Erklärungs-Vorstellungen und Überlegungen über die Wirkungsweise löst und von einer Betrachterebene von aussen unser Tun beurteilt. Diesen aufklärerischen Spagat zwischen therapeutisch notwendigem Glaube an unsere Interventionsmethoden und gleichzeitiger Infragestellung derselben sind wir nicht gewohnt, und diese Dehnung schmerzt zuerst einmal, solange wir ungeübt sind. Die intellektuelle Meistersung dieses Spagats erweist sich für die Ärzteschaft in der Tat als schwierig, etwa so schwierig, wie die Ablösung des ptolemäischen Weltbildes am Ende des Mittelalters. Auch damals war es für die alte, überholte Hierarchie so einfach, auf den Himmel zu zeigen und festzustellen, dass die Sonne doch jeden Tag um die Erde drehe. So einfach eben, wie uns heute ein klarer Nutzen des Ultraschall-Screenings glauben gemacht werden kann.

Eine Orchester-Mitspielerin erzählte mir nach einer Probe nicht ganz ohne Vorwurf, sie hätte Glück gehabt, dass bei ihr ein Routine-Ultraschall durchgeführt worden sei, denn man hätte so ihr Kind retten können. Im Ultraschall war nämlich eine Missbildung der Harnwege entdeckt worden, und durch eine intrauterine Punktion von Urin war offensichtlich das Leben des Kindes gerettet worden. Man sieht es also: Mit dem Ultraschall-Screening können Hydronephrosen intrauterin entdeckt werden (schon «Pyelondilatationen milder Ausprägung haben eine enorme präventivmedizinische Bedeutung» [1]), und bevor diese zu einer Nierenschädigung führen, kann dank Ultraschall-Früherkennung frühzeitig interveniert werden, die Stauung kann punktiert oder abgeleitet werden, und die Niere ist gerettet! So einfach ist das!

Zu einfach. Ein Jahr später ist die Mutter des inzwischen geborenen Kindes diesmal für ärztlichen Rat zu mir gekommen, weil jetzt das Kinderspital beschäftigt war, die extrauterinen routinemässigen Nachsorge-Untersuchungen durchzuführen, und die Mutter begann

an deren Sinnhaftigkeit zu zweifeln. Das Kind war von aussen gesehen vollständig gesund und entwickelte sich völlig normal, und die Mutter befand, die schon wieder geplante Szintigraphie sei für das Kind eine Zumutung. Die zugezogenen Arztberichte bestätigten die Befunde einer Nierenbeckenausgangs-Stenose mit Hydronephrosebildung. Die intrauterine Intervention hatte aus einer Punktion von 25 ml Urin bestanden, wobei die Punktion interessanterweise in die Blase erfolgte und nicht in den Ureter. Ob eine Punktion der Blase überhaupt eine nützliche Wirkung haben kann, ist hier aber nicht die entscheidende Frage. Die Fragen setzen viel grundsätzlicher an.

Die Bedeutung eines «abnormen» Befundes

Die klinische Bedeutung pathologischer Befunde – in diesem Fall der natürliche Verlauf («natural history») einer mit dem Ultraschall festgestellten Nierenabgangs-Stenose – lässt sich nur aus Kohortenstudien ablesen, in welchen ohne Selektion alle Fälle mit Ureterstenose eingeschlossen und vollständig (ohne drop-out) verfolgt worden sind. Solche Studien müssen «richtig» durchgeführt sein, indem Verlauf und Schwere der Krankheit «blind» von der Vorgeschichte beurteilt werden. Es gilt zu analysieren, wie die biologische Verteilung der Ultraschall-Messwerte bei Kindern aussieht, die im weiteren Verlauf ohne Intervention gesund geblieben sind, denn es ist ja längst bekannt, dass «abnorme» Untersuchungsergebnisse auch bei Gesunden vorkommen können und dass falsch pathologische Ergebnisse vor allem beim Routine-Screening relativ häufig sind. (Ein Haupthindernis für eine aufgeklärte Würdigung «pathologischer» Ergebnisse ist das unreflektierte Gleichsetzen von «normal» mit «gut» und von «abnorm» mit «böse», was dann immer wieder nach diagnostischem Exorzismus verlangt, als ob der vermeintlich gesichtete Satan sich durch wiederholte Ultraschall-Kontrollen in Schach halten liesse.) Es gilt weiter zu sehen, wie die Verteilung der Ultraschall-Messwerte bei Kindern ist, die tatsächlich klinisch fassbare Komplikationen entwickeln. Aus der Überlappung dieser Messwerte bei Gesunden und Kindern mit Komplikationen kann dann Spezifität und Sensitivität bzw. die Prädiktivität der Untersuchungsmethode bestimmt werden. Für eine nützliche Medizin entscheidend ist nicht die Spezifität in der Diagnose einer «Fehlbildung» an sich, sondern die Prädiktivität in der Vorhersage von echten klinischen Problemen (es ist ja auch nicht jeder Mensch ein pathologisches Problem, der über 1 Meter 90 gross gewachsen ist).

Es gibt keine zuverlässigen Studien zu dieser Frage. Wir wissen also nicht einmal, welche klinische Relevanz der als pathologisch bezeichnete Ultraschall-Befund bei unserem Kind hatte, es könnte sich um eine Norm-

variante gehandelt haben, um eine Laune der Natur, die uns Pathologien vorgaukelt, die bei genauerer Betrachtung selbstlimitierende Phänomene sind. Die Erfahrungen eines Ultraschall-Spezialisten könnten uns hier wichtige Hinweise geben, aber sie wären nicht mehr als das, und keinesfalls eine zuverlässige Evidenz.

Aber auch Studien, welche eine wichtige prognostische Bedeutung der Ultraschall-Befunde bei solchen Urogenital-Variationen tatsächlich festgestellt hätten, könnten erst nur einmal das Befürchtungs-Szenario untermauern; Evidenz für einen Interventionsnutzen würde noch immer fehlen.

Behandlung bei Anomalien: mehr Nutzen oder mehr Schaden?

Es ist heute kein Geheimnis mehr, dass ein präventiver Interventionsnutzen nur in kontrollierten Studien zuverlässig getestet werden kann. Wissen wir, ob es überhaupt einen Unterschied macht, ob Ureterstenosen ohne klinische Krankheitssymptome im Routine-Ultraschall gefunden werden oder erst beim Auftreten von Krankheitssymptomen? Wissen wir überhaupt, ob eine Punktion von 25 ml Urin eine nierenrettende Wirkung hat, wie es der Mutter dargelegt worden war?

Es fehlen richtige Studien, die sogar relativ einfach durchzuführen wären, wenn schon generell mit dem Ultraschall gescreent wird. Eine Studie hätte Föten einzuschliessen, bei welchen interveniert werden könnte oder sollte. Dann wird in der einen Zufallsgruppe interveniert, die anderen in Ruhe gelassen. Man könnte es auch so anstellen, dass man einen «pathologischen» Nierenbefund in der Studiengruppe mitteilt und in der Kontrollgruppe verschweigt. Aber auch wenn eine präventive Wirksamkeit von Ureteren-Manipulationen in kontrollierten Studien nachgewiesen wäre, fehlte immer noch die Antwort auf folgende Frage: Wie gross wäre der Unterschied, der durch eine Intervention erzielt werden kann? Ist eine vernünftige «Number needed to treat» zu erwarten, und sind erwünschte Wirkungen tatsächlich bedeutender als unerwünschte Wirkungen, d. h., besteht ein sicher positives Nutzen-Risiko-Verhältnis?

Die bisher übliche Gleichsetzung von Spezialisten-Wissen mit einer generelleren «Fachkompetenz» ist deshalb ein Trugschluss, weil es für eine zu Ende gedachte ärztliche Fragestellung nicht darum geht, wer in einer spezifischen Situation mit seinen Mitteln und Kompetenzen das Spektakulärste ausrichten kann, sondern darum, wo diese Mittel wirklich nützlich eingesetzt sind und wo sie eher mehr schaden als helfen. Die Medizin entwickelt sich offenbar anhand von Gerüchten. Wenn die unerhörten Fortschritte diagnostischer Möglichkeiten zu den Erfolgsgerüchten aus spezialisierten Kliniken geführt haben, so sind in der Periphe-

rie jetzt die neuen Geschichten eigentlich gesunder Patienten entstanden, bei denen zunehmend (falsch positive) Pathologien entdeckt werden, die gar keine sind. Während bisher Ultraschallspezialisten ihre in Einzelfällen nicht immer ganz unwahren Erfolgsgeschichten frühzeitiger Interventionsmöglichkeiten erzählt haben, dabei aber auch aus einem iatrogenen Verlust einer Säuglingsniere ein Erfolgsgerücht konstruieren [3], kommen nun auch andere Geschichten zunehmend an die Oberfläche: z.B. die des Göttikindes eines Urologen, welches an den Ureteren schwangerschafts-sonographisch so missgebildet war, dass ein dringender Abort empfohlen worden war; die «Missbildung» konnte tatsächlich postpartal verifiziert werden, führte aber in wenigen Wochen ohne jegliche Behandlung zu einer spontanen Remission. Das Kind ist heute kerngesund, und der glückliche Medizinergötti reibt sich die Augen – und bildet sich heute in Evidence-based Medicine [4] weiter. Geschichten dieser Art sind sehr zahlreich, wenn auch nicht immer so dramatisch. Die Aufgabe einer nutzbringenden Medizin ist nun ganz einfach, die Bedeutung dieser verschiedenen Gerüchte – verheissungsvolle und abschreckende – in adäquaten Studien näher zu quantifizieren und allenfalls in Utility-Analysen zu werten.

Die gültigste Evidenz, die wir heute ja bereits zur Verfügung haben, sind randomisiert kontrollierte Vergleichstudien an über 30 000 schwangeren Frauen, bei denen herauskam, dass die kindliche und mütterliche Gesundheit mit oder ohne Routine-Ultraschall identisch blieb [5]. Zwar wird immer wieder die Möglichkeit geltend gemacht, dass die inzwischen einige Jahre alten Studien in bezug auf den heutigen technischen Stand nicht mehr gültig seien; aber dies ist lediglich eine Spekulation, die nur durch geeignete neue Studien geklärt werden könnte. Und mit dem Fehlen genauer Daten über die prognostische Bedeutung und Prädiktivität von Ultraschall-«Pathologien» fehlen ja schon die Grundlagen für eine ernsthafte Nutzenhypothese. Zweifellos gibt es einzelne Fälle, wo der Routine-Ultraschall fast sicher von Nutzen war. Es gibt aber auch die einzelnen Fälle, wo jemand im Lotto das grosse Los gewinnt, doch wer verfällt dem Trugschluss, mit Lottospielen könnten ökonomische Probleme gelöst werden?

«Mittelalter» in der modernen Medizin

Wenn die Sonne scheinbar um die Erde geht und wenn der Routine-Ultraschall scheinbar Nieren rettet (oder dann halt angeblich die Lunge, wenn die Niere iatrogen verlorengeht [3]), so sind bei sorgfältiger Betrachtung auch andere Möglichkeiten vorhanden, die Zusammenhänge zu deuten. Die Vorgehensweise der Wissenschaft, Annahmen nicht nur vom Schein her zu akzeptieren, sondern zu versuchen, diese von

möglichst vielen Gesichtspunkten her zu falsifizieren, führt am Schluss zu gültigeren Ergebnissen. Der Nutzen des Routine-Ultraschalls in der Schwangerschaft entpuppt sich dann als Täuschung, die ohne grosse Mühe zu erkennen ist (ausser das intellektuelle Vermögen reicht nicht aus, die sich um die Erde drehende Sonne auch als Drehung der Erde um die Sonne vorstellen zu können).

Die Eltern dieses Kindes haben diese Überlegungen verstanden und sind heute erleichtert, dass sie sich nicht mehr dauernd über ihr Kind Sorgen zu machen brauchen. Sie sind auch nicht böse auf die erfolgten Eingriffe, sondern freuen sich, dass ihr Kind Glück gehabt hat und heute dank oder trotz Ultraschall-Screening gesund ist.

Für Fragen der «Qualitätssicherung» ist denn auch nicht die Durchführungs-Qualität allein das entscheidende Problem, sondern die fehlende intellektuelle und wissenschaftliche Qualität eines durchdachten Einsatzes von Leistungen. (Der überlegte Einsatz eines «veralteten» Gerätes ist immer noch besser als der unüberlegte Einsatz eines modernen Gerätes, welches den Qualitätsansprüchen einer sogenannten «Akademie für fetomaternalen Medizin» genügt, die sich anschickt, des Ultraschall-Kaisers neuen Kleider-Standard festzulegen.) Wichtig für die Qualität der Medizin ist deshalb, dass das zunehmende Interesse der Ärzteschaft an Evidence-based Medicine zu einem ausreichend profunden Verständnis des sich anbahnenden wissenschaftlich-fachlichen Wandels führt, welcher auch ganz entschiedene menschliche Schritte von uns Ärzten fordert. Absicherungs-Untersuchungen erzielen ihre sicher manchmal auch nützliche Wirkung meist nicht durch einen echten Informationsgewinn, sondern vielmehr durch die Magie der Absolutionsmacht, die modernen Geräten zugemessen wird. Wie schaffen wir es aber, diese nicht ganz harmlosen und zu teuren Absicherungs-Rituale etwa durch ein menschlich reifes Gespräch über die Bewältigung von Unsicherheit zu ersetzen? Noch besser ist es, den grassierenden Absicherungsbedarf gar nicht erst iatrogen zu induzieren, wenn Nutzenbeweise fehlen. Die Zukunft unseres Berufes wird vermutlich dadurch entschieden, dass wir nicht mehr nur simplistisch die rein technische Durchführung einer Leistung oder deren Preis, sondern in erster Linie deren gut durchdachten, situationsabhängigen Patientennutzen ins Zentrum unserer Diskussion stellen. Es geht heute darum, die Frage von Wirksamkeit mit der Frage der situationsabhängigen Gültigkeit, der «externen Validität» [6], zu vernetzen. Professionalität verlangt, dass wir wie ein Steuermann über der Sache stehen, statt uns so leichtgläubig wie andere auch in die Illusionen des medizintechnischen Fortschrittes zu verwickeln. Nur so kann auch der Segen moderner Technologie, gezielt und durchdacht eingesetzt, in eine medizinische Zukunft

gerettet werden. Die Geburt eines neuen, durchdachteren Evidenz-Verständnisses hat jedenfalls begonnen. Insbesondere Gynäkologen wissen ja, dass bei jeder Geburt die Phase der Geburtswehen unvermeidlich ist, aber auch vorübergeht. Gute Gynäkologen wissen auch, dass ein geglücktes Sich-Schicken in die Wehen am schnellsten und schmerzlosesten zum Erfolg führt. Auch wütendes Schreien sei während Momenten erlaubt, wie wollten wir es übelnehmen!

Als logische Folge dieses Wandels und der wachsenden Dynamik im medizinischen «Leistungstreit» scheint für die ökonomische Diskussion selbst eine Verlagerung von der Preiskontrolle zum bisher offenbar gar nicht in Betracht gezogenen Leistungswettbewerb angezeigt. Dabei geht es insbesondere darum, einen scheinbaren Leistungsabbau als Qualitätsverbesserung verstehen und «verkaufen» zu können. Denn nicht die Kosten und die Qualität der Kanonen sind das Problem, sondern dass wir aufhören, damit auf Spatzen zu schiessen. (Wenn wir Ärzte diesen Wandel gekonnt aufgreifen, können wir im übrigen auch auf «Evidence-based Economics» pochen.) □

Referenzen

1. Brühwiler H., Leserbrief, SÄZ Nr. 50/97
2. Sackett D.L., Haynes R.B., Tugwell P. Clinical Epidemiology; A basic Science for Clinical Medicine. Little, Brown and Company, Boston/Toronto 1991
3. Zimmermann R., Huch A. A large unilateral contained urinoma in a fetus: An indication for intrauterine urinary shunting. Fetal Diagn Ther 1993; 8: 268-272
4. Einsiedler Kurse: Zeitgemässe Beurteilung medizinischer Evidenz
5. Bucher H.C., Schmidt J.G. Routine Ultrasonography for Dating. In: Wildschut HJ, Weiner CP, Peters TJ, eds. When to Screen in Obstetrics and Gynecology. London: W.B. Saunders Co. Ltd., 1996: 100-107
6. Schmidt J.G. Die Vision einer pragmatischen klinischen Forschung oder das Ende der Diskussion über «Placebo» und «spezifische Wirkungen». Einführungsvortrag, 3. Wissenschaftliches Einsiedler Symposium 1997