

*Wenn ihr Mediziner nicht sagt,
wie Gesundheitsleistungen zu bewerten sind,
werden wir, die Ökonomen es tun.*
George Silberman, 1992

1 Einleitung

1.1. Diagnostik im klinischen Alltag

Moderne bildgebende, laborchemische und endoskopische Diagnosemethoden sind segensreich und haben fraglos zu einem besseren Verständnis der Pathogenese bestimmter Erkrankungen, ihrer Klassifikation und stadiengerechten Therapie beigetragen. Die Euphorie, die die Einführung neuer radiologischer Technologie wie der Magnetresonanztomografie oder molekularbiologischer Methoden hervorbringt, birgt jedoch die Gefahr, ihren tatsächlichen Wert für Anbieter, Konsumenten und Kostenträger von Gesundheitsleistungen nicht mehr kritisch zu hinterfragen.¹

Es ist derzeit unbekannt, ab welchem Grenzwert Ärzte die tägliche Flut an diagnostischen Informationen nicht mehr verarbeiten können. Automatisierte laborchemische Verfahren liefern z.B. (zum großen Teil technisch bedingt) Werte, die im klinischen Alltag selten oder *nie* zu einer Therapieentscheidung beitragen (klassisches Beispiel: die Mitbestimmung der Natrium- und Chlorid-Konzentration bei Anforderung der Kalium-Konzentration im Serum). Dieser Informationsüberstand wird gemeinhin als unvermeidbar akzeptiert.

Viele apparative diagnostische Verfahren verstärken die „gefühlte Sicherheit“ von Patient und Arzt^{2,3,4}, der im Zweifelsfall entsprochen wird (und sogar entsprochen werden muss).

Im Idealfall bestätigen die radiologisch und laborchemisch gewonnenen Daten die Ergebnisse der körperlichen Untersuchung. Dies führt auf der Seite des Arztes zur

¹ Contopoulos-Ioannidis DG, Ntzani E, Ioannidis JP. Translation of highly promising basic science research into clinical applications. *Am J Med* 2003;114:477-484.

² Shields G, King W, Fulks S, Fallon LF. Determinants of perceived safety among the elderly: an exploratory study. *J Gerontol Soc Work* 2002;38:73-83.

³ Ansari WE. Educational partnerships for public health: do stakeholders perceive similar outcomes? *J Public Health Management Practice* 2003;9:136-156.

⁴ Innvær S, Vist G, Trommald M, Oxman A. Health policy-makers' perceptions of their use of evidence: a systematic review. *J Health Serv Res Policy* 2002;7:239-244.

Erhöhung des Selbstwertgefühls, auf der Seite des Patienten zu einer Stärkung der Vertrauensbasis.

Im ungünstigsten Fall wird Diagnostik allerdings als Mittel zum Zweck betrieben, sei es aus reiner Verlegenheit (weil eine Erkrankung gesichert ist, spezifische Therapiemaßnahmen jedoch entweder nicht erforderlich oder verfügbar sind) oder aber finanzielle Interessen zur Entscheidung für einen diagnostischen Test beitragen.

Für therapeutische und diagnostische Interventionen gilt gleichermaßen, dass ihr Nutzen sowohl für das Individuum als auch die Gesellschaft auf drei Ebenen bewertet werden muss.⁵

Die entstehenden Kosten sollten nur dann solidarisch getragen werden, wenn auf allen der folgenden Bewertungsebenen zugunsten der interessierenden Maßnahme abgestimmt wird:

- 1) Wirksamkeit (Effektivität, *efficacy*, „does it work in clinical practice“). Im Fall diagnostischer Methoden beschreibt die Effektivität, wie der Test zwischen Menschen mit und ohne ein bestimmtes Merkmal bzw. einer bestimmten Erkrankung diskriminiert. Sie wird ausgedrückt durch die bekannten diagnostischen Kenngrößen wie Sensitivität, Spezifität, diagnostische Genauigkeit, Likelihood Ratios und Receiver Operating Characteristics (ROC) Kurven. Ein nicht effektiver Test sollte schon auf dieser Bewertungsstufe verworfen werden.
- 2) Wirkungsgrad (*effectiveness*, „does it work in real life“). Das Ergebnis eines diagnostischen Tests sollte einen messbaren Einfluss auf die Entscheidungen des Arztes nehmen. Dies kann z.B. die Vermeidung weiterführender invasiver diagnostischer Maßnahmen oder aber eine stadiengerechte Therapie sein.
- 3) Nutzwert (Effizienz, *efficiency*, „does it contribute to a more efficient use of resources“). Die auf der Basis des Testergebnisses getroffenen Entscheidungen sollten einen Vorteil für den individuellen Patienten (eine Verlängerung oder eine qualitative Verbesserung des Lebens) oder der Solidargemeinschaft (eine Reduktion der Prävalenz einer Erkrankung, ihrer Folgen oder ihrer direkten und indirekten Kosten) mit sich bringen.

Dieses hierarchische Modell wurde vielfach modifiziert; Fryback und Thornbury schlugen sogar sechs Bewertungsebenen vor⁶: die technische Machbarkeit, die diagnostische Genauigkeit, den Einfluss auf die klinische Entscheidungsfindung (die Differenz zwischen prior und posterior probability), den therapeutischen

⁵ Kielhorn A, Graf von der Schulenburg JM. *The health economics handbook*. Tattenhall: Adis, 2000.

⁶ Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88-94.

Wirkungsgrad (die Dimension des Einflusses auf Therapieentscheidungen), das Outcome (Änderungen in Quality-Adjusted Life Years [QALYs]) und den Gewinn für die Gesellschaft (Cost-benefit, Cost-effectiveness, Cost-utility).

Unabhängig davon, wie detailliert die verschiedenen Ebenen aufgeschlüsselt werden, ist der Wirkungsgrad einer medizinischen Maßnahme schwierig zu messen, da neben objektiven Einflüssen eines Testergebnisses auf das Management einer Erkrankung zahlreiche schwer quantifizierbare Einflussgrößen existieren.

Ein Arzt kann

- einem Testergebnis weniger vertrauen als seiner klinischen Erfahrung,
- dem Testergebnis vertrauen, aber dennoch eingespielten Handlungsabläufen verhaftet sein, die keiner externen Kontrolle mehr unterliegen oder
- ein Testergebnis falsch interpretieren.

Patienten können

- die Empfehlungen ihres Arztes ablehnen oder
- entgegen den Testergebnissen auf weitere Diagnostik oder eine bestimmte Therapie bestehen.

1.2. Besonderheiten der Diagnostik Schwerverletzter

Diagnostische Tests nehmen in der Schwerverletztenversorgung naturgemäß einen anderen Stellenwert ein als bei der elektiven Klärung von Symptomen und Befindlichkeitsstörungen. Ihr Ergebnis entscheidet innerhalb eines engen Zeitfensters über die adäquate Auswahl lebensrettender Interventionen.

Durch die Neustrukturierung des Rettungs- und Notarztwesens, die Einrichtung spezialisierter Traumazentren und die Etablierung von Behandlungsalgorithmen konnte die Letalität in der Versorgungsphase zwischen Erstbehandlung am Unfallort, Transport und Aufnahme im Krankenhaus drastisch gesenkt werden. Die verfügbaren Zahlen legen eine Reduktion der Sterblichkeit von 40% auf unter 20% in den letzten 20 Jahren nahe.⁷

Die die Klinik lebend erreichende Patientenklientel ist allerdings ungleich schwerer verletzt und stellt hohe Ansprüche an die Logistik und Belastbarkeit

⁷ Regel G, Lobenhofer P, Lehmann U. Ergebnisse in der Behandlung Polytraumatisierter. Eine vergleichende Analyse von 3406 Fällen zwischen 1972-1991. *Unfallchirurg* 1993;96:350-362.

multiprofessioneller Teams. Von 2983 im Jahr 2002 durch das Traumaregister der Deutschen Gesellschaft für Unfallchirurgie (DGU) erfassten Patienten erfüllten 67% mit einem Injury Severity Score (ISS) ≥ 16 die Definition des Polytraumas.⁸

Der Arzt handelt in dieser Situation, bewusst oder unbewusst, nach Algorithmen, da ihm keine Zeit für eine Abwägung des für und wider möglicher Teststrategien, im Idealfall im Konsens mit dem Betroffenen oder Angehörigen bleibt („shared decision making“).

Moderne Schockraum-Algorithmen können in die Phasen Alpha (erste Minute, lebensrettende Sofortmaßnahmen wie z.B. Thoraxdrainagen), Bravo (fünf Minuten, dringliche Sofortmaßnahmen), Charlie (bis 30 Minuten, dringliche obligate Maßnahmen) und Delta (Komplettierung der Diagnostik) aufgliedert werden.⁹

Ein präziser Eindruck vom physiologischen und anatomischen Status muss in den ersten beiden ersten Phasen gewonnen werden, um eine adäquate weitere Diagnostik und Therapie einzuleiten.

Wesentliche Anforderungen an diagnostische Verfahren in der Akutversorgung sind

- Möglichkeit der Point of Care (POC) Testung,
- einfache Handhabung und geringe Invasivität sowie
- hohe Sensitivität¹⁰.

Der Arzt muss sich darauf verlassen können, dass sowohl seine klinische Untersuchung als auch die in Algorithmen vorgegebenen diagnostischen und therapeutischen Schritte gesichert effektiv sind.

Für die Therapieebene trifft dies in der Regel zu. Ihre Aktualisierung unterliegt stark einer kontinuierlichen Kontrolle durch die wissenschaftliche Gemeinschaft. Ebenso wie für Leitlinien wächst die Zahl systematischer Übersichten und Meta-Analysen bzw. Gesundheitstechnologieberichte als Quelle bester verfügbarer Information stetig; potenzielle Innovationen werden rasch in die klinische Praxis übernommen.

Demgegenüber ist die Kultur der Auseinandersetzung mit dem verfügbaren Repertoire diagnostischer Tests noch unzureichend entwickelt.

⁸ Arbeitsgemeinschaft Polytrauma der Deutschen Gesellschaft für Unfallchirurgie. Jahresbericht 2002 des Traumaregisters. http://www.traumaregister.de/downloads/Jahresbericht_2002.pdf

⁹ Waydhas C, Kanz KG, Ruchholtz S, Nast-Kolb D. Algorithmen in der Traumaversorgung. *Unfallchirurg* 1997;100:913-921.

¹⁰ Die Diagnosefindung beim Akutkranken bzw. Schwerstverletzten läuft in der Regel im Ausschlussverfahren ab- es ist hilfreicher, zunächst bestimmte Erkrankungen oder Verletzungen sicher auszuschließen, als sich auf offensichtliche Zustandsbilder zu konzentrieren.

Auch „alt gedienten“ Klinikern fällt die Bewertung von Therapieverfahren häufig leichter als die Bewertung eines diagnostischen Tests. Dies liegt zum großen Teil an der Eindimensionalität therapeutischer Maßnahmen: sie wirken oder sie wirken nicht.

Diagnostische Tests haben wenigstens zwei Dimensionen: ihr Ergebnis dient entweder der Erhärtung oder dem Ausschluss einer Erkrankung oder eines Merkmals. Nur wenige Tests werden beiden Funktionen in gleicher Weise gerecht.

Das diagnostische Spektrum in der Schwerverletztenversorgung umfasst

- klinische Tests (körperliche Untersuchung, äußere Verletzungszeichen, insbesondere sog. „Indikatorverletzungen“ und klinische Entscheidungsregeln)
- bildgebende Verfahren (Ultraschall, Röntgen- und CT-Diagnostik) und
- operative Interventionen (Peritoneallavage [im Europäischen Raum bedeutungslos], Laparoskopie/ Laparotomie, Thorakotomie).

Eine besondere Stellung nehmen laborchemische Methoden ein, da sie ein Hybrid aus diagnostischem Instrument und prognostischem Element darstellen¹¹. Neben den etablierten Tests wie Blutbild, Elektrolyt-Bestimmung und Gerinnungsanalysen wurden zahlreiche biologische Parameter wie C-reaktives Protein (CRP), prä-pro-Calcitonin, Interleukine, Tumor-Nekrose-Faktor (TNF) alpha und TNF-alpha-Rezeptor sowie Genpolymorphismen als diagnostische und prognostische Variablen postuliert.

Die gezielte Zusammenführung von klinischen, bildgebenden und laborchemischen Untersuchungen zu einer effektiven und effizienten diagnostischen Strategie könnte die Qualität der Schwerverletztenversorgung verbessern. Dies erfordert eine methodische Prüfung einzelner Komponenten auf unterschiedlichen Bewertungsebenen.

1.3. Zielsetzung

Die in dieser Schrift zusammengefassten Arbeiten überprüfen die Effektivität (d.h., die diagnostische Genauigkeit) und die Effizienz (d.h., ihren Einfluss auf Entscheidungsfindung und Outcome) diagnostischer Konzepte in der Schwerverletztenversorgung.

¹¹ Der diagnostische Wert einer Methode ergibt sich aus der Möglichkeit, eine therapeutische Konsequenz abzuleiten (Individualisierung von Therapieentscheidungen), die möglicherweise zu einem besseren Outcome führt. Eine ungünstige Prognose ist hingegen kaum zu beeinflussen.

Die wissenschaftlichen Fragestellungen ergaben sich aus alltäglichen klinischen Problemen- ein wesentliches Ziel bestand damit auch darin, komplexe statistische Informationen in weithin verständliche und akzeptable Lösungsstrategien umzusetzen.

Neben Handlungsempfehlungen für die Praxis sollten Daten gewonnen werden, die für evidenzbasierte Leitlinien der wissenschaftlichen Fachgesellschaften, aber auch für gesundheitspolitische Entscheidungen genutzt werden können.

Die hier präsentierten Ergebnisse folgen der bereits genannten Hierarchie der diagnostischen Forschung von

- der Genauigkeit (Sensitivität und Spezifität) über
- den Einfluss der Testresultate auf die ärztliche Entscheidungsfindung hin zu
- durch die Testanwendung erzielte patientenzentrierte Outcomes.

Eine zweite Rangfolge ergibt sich aus typischen Abläufen der Diagnosekaskade

- vom klinischen Zustandsbild über
- körperliche Untersuchungsbefunde hin
- zu radiologischen Verfahren und Labormethoden.

Im Einzelfall wurden folgende Fragen formuliert, die aufgrund des aktuellen Umbruchs sowohl im Schockraummanagement (Einführung von Advanced Trauma Life Support [ATLS®] des American College of Surgeons durch die Deutsche Gesellschaft für Unfallchirurgie) als auch reger baulicher Veränderungen in bundesdeutschen Traumazentren (Computertomografie im Schockraum bzw. in unmittelbarer Schockraumnähe) bedeutsam sind:

1. Wie viel und welche Information benötigt der Arzt, um eine therapeutische Entscheidung zu fällen?
2. Bedingen sog. Indikatorverletzungen tatsächlich eine höhere Wahrscheinlichkeit für das Vorliegen anderer Verletzungen, bzw. schließt ihr Fehlen auch begleitende Traumafolgen aus (beeinflussen bestimmte Verletzungsmuster somit die klinische Vortest-Wahrscheinlichkeit?)
3. Wie effektiv und effizient sind allgemein als unverzichtbar angesehene diagnostische Tests, und wie hängt ihre angenommene Genauigkeit von der Einhaltung methodischer Standards in häufig zitierten Publikationen ab? Kann eine noch unzureichende Effektivität durch einfache technische Modifikationen verbessert werden?

Alle Arbeiten beschäftigen sich mit der Mehrfachverletzung bzw. dem Polytrauma und dem stumpfen Bauchtrauma. Diese Krankheitsbilder wurden gewählt, weil

1. sie mit hoher Inzidenz auftreten und ihr Management mit erheblichen finanziellen und personellen Ressourcen vergesellschaftet ist,
2. ihre präzise Diagnose weiterhin ein methodisch und klinisch anspruchsvolles, dabei aber in vielen Facetten noch unbefriedigend gelöstes Problem darstellt,
3. die rasche (im Idealfall innerhalb der sog. „golden hour of shock“) diagnostische Abklärung unmittelbaren Einfluss auf den weiteren klinischen Verlauf und die Prognose der Betroffenen nimmt und
4. international unterschiedliche, z.T. auch widersprüchliche Handlungsempfehlungen existieren.

Für die wissenschaftliche Bearbeitung der Fragenkomplexe wurden im Unfallkrankenhaus Berlin Kohortenstudien durchgeführt, die sich durch die vom klinischen Befund unabhängige Prüfung durch einen diagnostischen Goldstandard (hier: der kontrastmittelverstärkten Ganzkörper-Computertomografie) und damit einer Vermeidung eines sog. verification bias auszeichnen.

Den Forschungsschwerpunkt der letzten sechs Jahre bildeten theoretisch-methodische Grundlagen und praktische Anwendungen von Meta-Analysen. Unter anderem wurde, erstmalig seit ihrer Einführung zu Beginn der 80er Jahre, die internationale Literatur zur fokussierten Ultraschalluntersuchung des Abdomens mit Hilfe von Meta-Analyse- und Meta-Regressionsmethoden quantitativ zusammengeführt.

Im Nebenschluss wurden ein neuartiger Ansatz zur Messung kritischer Informationsmengen für die ärztliche Entscheidungsfindung und eine grafische Methode zur Darstellung der Ergebnisse aus Diagnosestudien entwickelt.