



HTA

Health Technology Assessment



INSTITUT FÜR
TECHNIKFOLGEN-
ABSCHÄTZUNG

-Newsletter

Evaluation medizinischer Interventionen

Inhalt

Sondernummer: Methoden

Schlaglicht auf Deutsche EBM-Jahrestagung

Bias-Spotting: Studieninterpretation für Fortgeschrittene....1

Fallstudien

Methoden-Beurteilung.....2

Metaanalyse

diagnostischer Studien.....2

Outcome

Patientenorientierte Ergebnismessung.....3

RCT vs. Nicht-RCT

Vergleich der Behandlungseffekte.....4

Vorausschau April

- Vorhersagemodelle für chronische Herzinsuffizienz
- Screening auf abdominales Aorten Aneurysma
- Screening auf Depression
- Insulinpumpe vs. Multiple Injektionen

SCHLAGLICHT AUF DEUTSCHE EBM-JAHRESTAGUNG

Bias-Spotting: Studieninterpretation für Fortgeschrittene

EBM/Evidence based Medicine und HTA/Health Technology Assessment als Methoden der Bewertung der tatsächlichen Wirksamkeit der zahlreichen auf den Gesundheitsmarkt drängenden Interventionen/Technologien werden von der Politik immer ernster genommen. Dementsprechend werden auch die Methoden (selbstreflexiv) ständig weiterentwickelt. Die diesjährige Deutsche EBM-Jahrestagung zeugte – obwohl Deutschland (wie auch Österreich) erst lange nach anderen westlichen Ländern die Relevanz von EBM erkannte – von Aufbruchstimmung, von Lebhaftigkeit und von einer Kraft, die nur von gemeinsamer Bewegung ausgeht.

Nur EIN Schlaglicht: Dass EBM und HTA nicht nur Wissenschaftsdisziplinen sind, sondern auch etwas mit Investigativem Journalismus zu tun haben, zeigte ein Workshop zu „Bias Spotting – Studieninterpretation für Fortgeschrittene“. Dass in RCTs häufig gerade jene Patientengruppen ausgeschlossen werden, bei denen gültige Ergebnisse hochinteressant wären, dagegen der Einschluss von Patienten mit guter Prognose weniger Aussagekraft hat, oder dass Endpunkte gewählt werden und statistische Signifikanz zeigen, die für die Patienten und ihr Wohlbefinden irrelevant sind, oder dass allein die

Kenntnis und die Präferenz der untersuchten Therapie die Ergebnisse beeinflussen, all das ist wohlbekannt, verbreitet und kann als „Bias Spotting für Anfänger“ durchgehen.

Wenn aber das Prinzip, dass in der Versuchsgruppe die „neue“ Therapieform und in der Kontrollgruppe die Standardtherapie verglichen werden, aufgehoben wird und statt dessen in der Kontrollbehandlung

- ein wirkungsloses Mittel oder
- eine zu niedrige Dosis oder
- eine fragwürdige Galenik oder
- die Standardbehandlung gar nicht erst verglichen wird,

dann muss von Manipulation gesprochen werden.

Die beiden Vortragenden M. Lelgemann (ÄZQ/ärztliches Zentrum für Qualitätssicherung) und N. Donner-Banzhoff (Abt. für Allgemeinmedizin Univ. Marburg) brachten in dem Workshop noch zahlreiche weitere Beispiele, die „alert“ machen, was gemeinhin so an medizinischen Studien den Markt und damit das medizinische Handeln vorbereitet. Die Manipulationsmöglichkeiten sind tatsächlich reichhaltig.

6. Jahrestagung des Dt. Netzwerks EbM <http://www.ebm2005.de/>.

Termine

7./8. April 2005

The World Health Congress
Best Practice for improved
delivery and outcomes
Chantilly/FR

<http://www.worldcongress.com/europe/index.cfm>

13.–15. April 2005

10th European Forum
on Quality Improvement in
Health Care
London

<http://www.healtheconomics.org/conferences/2005/04/13/10th-european-forum-on.html>

7./8. Juni 2005

Ökonomie und Gesundheit:
Lösbare Herausforderung oder
dauernder Widerspruch?
St. Pölten

<http://www.oegkk.at>

20.–22. Juni 2005

2nd HTAi Jahrestagung
Bringing HTA into Practice
Rom

<http://www.prex.it/congressi/hta/index.html>

13.–15. Juli 2005

5th iHEA Jahresweltkonferenz
Investing in Health
Barcelona

<http://www.healtheconomics.org>

3./4. November 2005

6. Dt. HTA-Symposium
Schwerpunkt Biomedizin
Köln

<http://www.dimdi.de/de/hta/symposien/index.htm>

10.–12. November 2005

EUPHA/European Public Health
Conference
Promoting the Public's Health
Graz

http://www.eupha.org/html/menu3_2.html

Fallstudien

Methoden-Beurteilung

Randomisierte Studien über die Wirksamkeit einer Intervention stehen für Assessments nicht immer zur Verfügung. Fallstudien zählen in der Hierarchie klinischer Studien zu einer schwachen Form von Evidenz. Ein britisches Methodenassessment (2005) versuchte die Validität unterschiedlicher Fallstudien-Designs zu vergleichen, kommt aber nur zu allgemeinen, nicht Design-spezifischen Aussagen: Einzelne Studien können in der Qualität weit voneinander abweichen und zu gleichen Interventionen sehr verschiedene Schätzungen der Ergebnisse hervorbringen. Fallstudien sind höchst anfällig für Fehlerquellen (Bias) und Verzerrungen.

Fallstudien werden zu einem bedeutenden Anteil in HTAs berücksichtigt, vor allem werden sie sehr häufig in der Beurteilung chirurgischer Interventionen herangezogen. Bei NICE etwa werden sie in 30 % der dortigen Assessments verwendet, obwohl es keinen Konsens darüber gibt, welche Fallstudien in HTAs zu inkludieren sind, wie sie verwendet werden oder wie ihre Qualität zu bewerten ist. Die Autoren des Methodenassessments schränkten sich in ihrer Methodenbeurteilung auf Beobachtungsstudien mit Kontrollgruppen ein. Sie fanden widersprüchliche Ergebnisse in Fallstudien zur Wirksamkeit derselben Interventionen, die entweder größere, ähnliche oder schwächere Behandlungseffekte aufwiesen. Die Gültigkeit/Validität jeder einzelnen Studie, welcher Form auch immer, hängt von der Qualität des Designs, der Ausführung und der Interpretation ab. Die Autoren fanden geringe Evidenz hinsichtlich der Verwendung von Faktoren, die in die methodische Qualitätsbeurteilung einbezogen werden sollen: sie fanden kein Verhältnis zwischen Studiengröße und Ergebnis bei vier untersuchten Beispielen (funktionale endoskopische Sinuschirurgie bei Nasenpolypen, Rückenmarkstimulation bei Rückenschmerz, Perkutane transluminale koronare Angioplastie (PTCA), koronarer Arterien Bypass

(CABG). Im Vergleich etwa mit der Evidenz aus RCTs, die keinen Unterschied zwischen PTCA und CABG aufwies, zeigten die Schätzungen von Fallstudien einen 1–2 % höheren Anstieg der Mortalität bei CABG. Wie in RCTs ist aber die Standardisierung beeinflussender Faktoren auch in Fallstudien von großer Bedeutung. Interessant für „Assessoren“ sind die Checklisten im Anhang des Assessment zur Beurteilung von Fallstudien.

NCCHTA/GB 2005: Do the findings of case series studies vary significantly according to methodological characteristics?
<http://www.ncchta.org/execsumm/summ902.htm>.

Metaanalyse

diagnostischer Studien

HTA hat die Aufgabe medizinische Technologien möglichst vollständig, systematisch und transparent zu bewerten. Systematische Übersichtsarbeiten/Reviews (qualitative Assessments) und Metaanalysen (quantitative Assessments) leisten einen wertvollen Beitrag, indem sie die vorliegende wissenschaftliche Evidenz übersichtlich machen. Im Vergleich zur Bewertung therapeutischer Interventionen ist die Beurteilung diagnostischer Maßnahmen wesentlich schwieriger. Ein deutsches Methodenassessment (2003) hatte zum Ziel, die derzeitigen Ansätze für die Metaanalyse diagnostischer Genauigkeitsstudien hinsichtlich ihrer Tauglichkeit zu bewerten.

Systematische Reviews sind Übersichtsarbeiten auf Basis von Primärstudien zu einer klar formulierten Fragestellung, bei denen systematisch und anhand strenger Kriterien relevante Literatur identifiziert und kritisch bewertet werden und einer qualitativen Analyse unterzogen werden. Ob darüber hinaus ev. auch eine quantitative Metaanalyse durchgeführt werden kann, hängt von der Anzahl und der methodischen Qualität der eingeschlossenen Primärstudien und dem Grad der Heterogenität der Schätzer der diagnostischen Genauigkeit ab.

Ergebnis: Diagnostische Studien mit binären Ergebnissen, d. h. Test positiv oder negativ, finden sich am häufigsten. Daraus lassen sich Sensitivität (TPR/true positive rate), Spezifität (FPR/false positive rate) und die Likelihood-Ratio entsprechend bestimmen. Eine häufig benutzte Maßzahl ist außerdem die diagnostische Odds-Ratio/DOR, die über die diskriminatorische Fähigkeit eines Tests Aussage trifft. Es gibt nun verschiedene Ansätze, die die diagnostische Genauigkeit von Primärstudien zusammenzufassen: z. B. Studien unterscheiden sich weder im Grenzwert noch in der Genauigkeit – hier werden gemeinsame Parameter für FPR und TPR berechnet; oder Studien besitzen gleiche Genauigkeit, aber unterschiedliche Grenzwerte – werden mittels SPROC-Kurve zusammengefasst. Für die Heterogenität zwischen den Studien sind eine Reihe von Kovariablen verantwortlich. Dazu gehören Qualität, Publikationsjahr, Untersuchungstechnik. **Fazit:** Systematische Übersichtsarbeiten und Metaanalysen ermöglichen es, schlechte oder nutzlose Tests zu eliminieren, bevor sie weitverbreitet Anwendung finden. Sie stellen eine verbesserte Information über diagnostische Tests zur Verfügung, indem sie das Patienenspektrum oder relevante Subgruppen analysieren.

DAHTA/D 2003: Methoden der Metaanalyse von diagnostischen Genauigkeitsstudien, http://gripsdb.dimdi.de/de/hta/hta_berichte/hta025_bericht_de.pdf.

Outcome

Patientenorientierte Ergebnismessung

Unter patientenorientierten Ergebnismessungen versteht man eine Reihe von Erhebungen, die Gesundheit, Krankheit und Nutzen von Interventionen aus der Patientenperspektive bewerten. Sie werden als primäre oder sekundäre Endpunkte in klinischen Studien verwendet und dienen zu Erhebung des Gesundheitsstatus und der subjektiv erlebten Lebensqualität von Patienten. Die Beurteilung von medizinischen Interventio-

nen aus Patientenperspektive wird immer häufiger gefordert: was das genau heißen mag, welche Instrumente vorliegen und wie diese beurteilt werden können, war Inhalt eines Methodenassessments (1998).

Standardisierte Messinstrumente werden in klinischen Studien oft verwendet, um für Interventionen wesentliche Erkenntnisse aus der Patientenperspektive zu ziehen. Ein britisches Methodenassessment beschreibt die Vielfalt von patientenorientierten Ergebnismessungen und evaluiert die Vor- und Nachteile der Messinstrumente: In der Literatur finden sich sieben Arten von Fragebögen: krankheitsspezifische, Körperteil-spezifische, nach Dimensionen des Befindens (physisch, psychisch, kognitiv), allgemeine, summarische, individualisierte oder Utility/Nutzenorientierte. Die Definition dieser Messinstrumenten wird in der realen Anwendung im Allgemeinen nicht klar und konsistent verwendet. So gibt es keine Übereinkunft bei den Begriffen Lebensqualität (Quality of life) und gesundheitsbezogene Lebensqualität. In der Anwendung variiert der beabsichtigte Zweck und der Inhalt der Fragebögen beträchtlich und bedarf stärkerer Standardisierung. Als Richtlinien für die Anwendung und Entwicklung neuer Messinstrumente dienen acht Kriterien, die angewendet werden sollen, um Patienten orientierte Ergebniswerte für klinische Studien zu erarbeiten:

1. *Angemessenheit* des Messinstruments (entsprechen die Fragen überhaupt den Intentionen der klinischen Studie)?
 2. *Verlässlichkeit* (sind die Ergebnisse des Messinstruments reproduzierbar?),
 3. *Gültigkeit* (misst das Messinstrument, was es messen will?),
 4. *Responsiveness/Empfindlichkeit* (entdeckt das Messinstrument für den Patienten bedeutsame Veränderungen?),
 5. *Präzision/Genauigkeit* (wie präzise ist das Messinstrument?),
 6. *Interpretierbarkeit* (sind die Ergebnisse interpretierbar?),
 7. *Annehmbarkeit* (wird das Messinstrument von den Patienten akzeptiert?) und
 8. *Durchführbarkeit/Umsetzbarkeit* (ist das Messinstrument überhaupt umsetzbar?).
- Weitere Untersuchungen zu Evaluation von führenden Messinstrumenten in

Rezente Assessments

BCBS/Blue Cross Blue Shield Technology Evaluation Center
<http://www.bcbs.com/tec/tecassessments.html>

In Press:

Artificial Vertebral Disc

Cardiac Computed Tomographic Angiography

Electrostimulation and Electromagnetic Therapy for the Treatment of Chronic Wounds

Gene Expression Profiling to Select Node-Negative Breast Cancer Patients for Adjuvant Chemotherapy

März 2005:

Extracorporeal Shock Wave Treatment for Chronic Plantar Fasciitis

Use of Implantable Cardioverter-Defibrillators for Prevention of Sudden Death in Patients at High Risk for Ventricular Arrhythmia

February 2005:

Extracorporeal Shock Wave Treatment for Chronic Tendinitis of the Elbow (Lateral Epicondylitis)

Angioplasty and Stenting of the Cervical Carotid Artery with Distal Embolic Protection of the Cerebral Circulation

Special Report: Intra-Articular Hyaluronan for Osteoarthritis of the Knee

BCBS: Laufende Assessments

Magnetic Resonance-Guided Ultrasound

Newer Bariatric Surgery Techniques for Morbid Obesity

T-Wave Alternans

Vagus Nerve Stimulation for Refractory Depression

HTA-Projekte am ITA

Chronische Hepatitis C
in Österreich:
Entscheidungsanalytische
Modellierung

Polymorphismen Screening
– Begleitung von Leitlinien
Entwicklung

Impressum

Redaktion: Claudia Wild
Saskia Schlichting

Graphik: Manuela Kaitna

Eigentümer und Herausgeber:
ITA – Institut für Technikfolgen-
Abschätzung der Österreichischen
Akademie der Wissenschaften,
A-1030 Wien, Strohgasse 45/3
<http://www.oew.ac.at/ita>

Der HTA-Newsletter erscheint
10 x pro Jahr und ausschließlich
auf der ITA-homepage

ISSN: 1680-9602

Wir freuen uns über
Reaktionen und Hinweise
insbes. auf für Österreich
relevante Themen!
Kommunikation an:
cwild@oew.ac.at

verschiedenen Gesundheitsbereichen
erfolgen laufend. Das – nicht ganz jun-
ge – Methodenassessment gibt einen
guten Überblick über patientenorien-
tierte Ergebnismessung.

NCCHTA/GB 1998: Evaluating patient-
based outcome measures for use in clini-
cal trials, [http://www.hta.nhsweb.nhs.uk/
execsumm/summ214.htm](http://www.hta.nhsweb.nhs.uk/execsumm/summ214.htm).

RCT vs. Nicht-RCT

Vergleich der Behandlungseffekte

Für eine sorgfältige und aussage-
kräftige Bewertung von Interventio-
nen stellen randomisierte kontrol-
lierte Studien (RCT) das ideale De-
sign dar. Dennoch gibt es Gründe,
weshalb für bestimmte Fragestellun-
gen keine RCT zur Entscheidungsfin-
dung bereit stehen: Ethische Grün-
de, Fehlen von regulierenden Vor-
schriften, sowie bei Krankheiten mit
niedriger Prävalenz sind RCTs schwer
durchführbar. Da zur Überprüfung
der Wirkung und des Nutzens neuer
Verfahren kontrollierte Studien we-
sentlich sind, hatte ein Methoden-
assessment (1998) zum Ziel, die Er-
gebnisse prospektiver nicht-rando-
misierten Studien zu vergleichen
und die Frage nach systematischen
Effektverzerrungen zu stellen: Im Ver-
gleich von RCT mit Nicht-RCT fanden
sich keine eindeutigen Ergebnisun-
terschiede.

Eine Randomisierung bedeutet eine zu-
fällige Zuteilung von Patienten einer
klinischen Studie in eine Behandlungs-
(neue Therapie) oder eine Vergleichs-
gruppe (Kontrollarm). Sie sichert zu,
dass sich die vergleichenden Patien-
tengruppen nur hinsichtlich Therapie
unterscheiden. Bekannte und unbe-
kannte Risikofaktoren und prognosti-
sche Parameter sind in beiden Grup-
pen gleich verteilt. Weichen die Ergeb-
nisse der Behandlungsarme ab, so las-
sen sich diese mit großer Wahrschein-
lichkeit auf die Intervention zurückfüh-
ren. Kritiker des Studiendesigns RCT
betonen, dass viele Individuen, für die

die Ergebnisse eigentlich bestimmt sind
(Alte, Minoritäten, Multimorbide, Kin-
der), aus methodischen Gründen aus-
gegrenzt werden. Andererseits werden
Einflussfaktoren wie persönliche Präfe-
renzen des Arztes oder des Patientes,
die die Ergebnisse verfälschen und zur
Täuschung führen, herausgehalten.
Diese Kritiker befürworten daher nicht-
randomisierte Studien. Das Methoden-
assessment versuchte die Auswirkun-
gen auf die Effektschätzung zusammen-
zufassen. *Ergebnis:* Aus 18 Artikeln,
die den direkten Vergleich von RCT
und prospektiven Nicht-RCT versuch-
ten, zeichneten sich keine eindeutigen
Muster ab. Weder RCT noch Nicht-RCTs
ergaben durchwegs größere oder klei-
nere Schätzungen des Behandlungsef-
fekts. Ein größerer Effekt wird erreicht,
da Patienten unter einer neuen Thera-
pie oftmals besonders sorgfältig be-
treut werden. Die Ergebnisse von RCT
und Nicht-RCT können sich unter-
scheiden, tun dies aber nicht immer.
Richtung und Ausmaß der Effektver-
zerrung sind nicht allein aufgrund des
Designs vorhersagbar. Eine nicht un-
erhebliche Rolle spielt die sonstige me-
thodische Qualität der Studienplanung
und -durchführung. RCT wie Nicht-
RCT besitzen gleichermaßen mögliche
Fehlerquellen: ein gut geplanter und
durchgeführter Nicht-RCT ist einer klei-
nen und schlecht geplanten RCT vor-
zuziehen.

NCCHTA/GB 1998: Choosing between
randomised and non-randomised studies:
a systematic review,
[http://www.hta.nhsweb.nhs.uk/
execsumm/summ213.htm](http://www.hta.nhsweb.nhs.uk/execsumm/summ213.htm).

Richtigstellung Hyperthermie

Nicht die KBV/Kassenärztliche Bundes-
vereinigung, sondern der G-BA/Ge-
meinsame Bundesausschuss ist He-
rausgeber des (noch unveröffentlic-
hten) Hyperthermie-Assessments. Die-
ser wird dort (www.g-ba.de) veröffent-
licht. Darüberhinaus bezieht sich der
G-BA Beschluss nur auf den ambulan-
ten Sektor, nicht – wie missverständlich
berichtet – auch auf den stationären.