

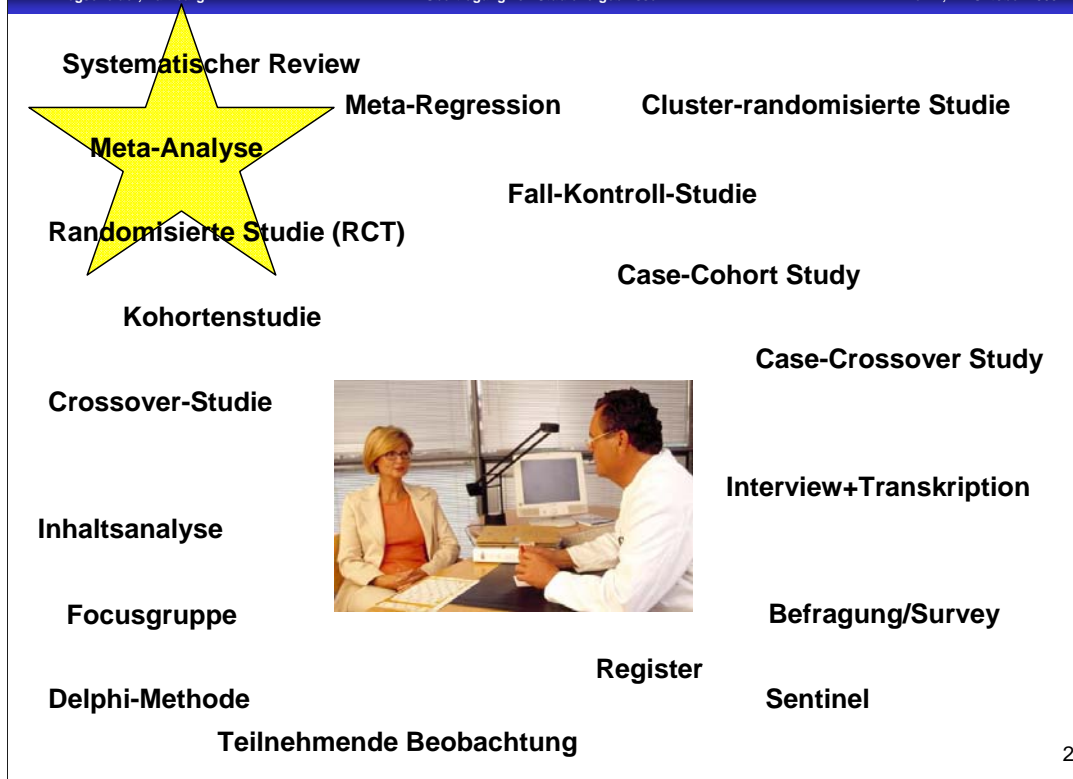
*2. Wissenschaftliches Diskussionsforum zur Nutzenbewertung im Gesundheitswesen
Berlin, 21. Oktober 2008*

*Übertragung von Studienergebnissen auf den
Versorgungsalltag:*

*Beitrag unterschiedlicher qualitativer und
quantitativer Forschungsansätze*

Prof. Dr. Karl Wegscheider

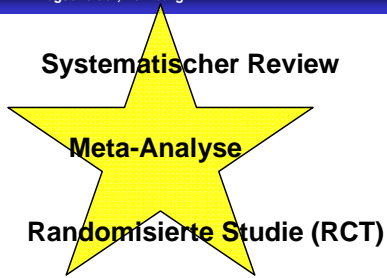
**Institut für Medizinische Biometrie und Epidemiologie
Universitätsklinikum Hamburg-Eppendorf**



Meine Damen und Herren,

sind wir nicht in einer herrlichen Lage? Verglichen mit früheren Generationen haben wir ein riesiges Angebot an Studienergebnissen, die mit den unterschiedlichsten Methoden erzielt wurden, eine Vielzahl von Studientypen und Forschungsansätzen. Es scheint für jedes Gesundheitsproblem eine passende Studienlösung zu geben. Von den Lösungsansätzen sind viele hochspezialisiert, einige auch umstritten, es finden sich aber auch unsere inzwischen allgemein anerkannten Goldstandards der Evidenzklasse 1 darunter, die randomisierten kontrollierten Studien (RCTs), die Meta-Analysen derselben und die systematischen Reviews.

Erstaunlicherweise sind Arzt und Patienten dennoch nicht zufrieden. Einerseits können sie das Überangebot an Studienergebnissen kaum bewältigen, andererseits fühlen sie sich häufig in den Studien nicht gebührend repräsentiert. Sie haben beide das Gefühl, dass ihre persönliche Wirklichkeit in den Studien nicht gut widerspiegelt wird. Und sie fühlen sich oft mit ihren Problemen allein gelassen. Immer da, wo man gerade eine konkrete Frage hat, scheint sich eine Lücke in der Evidenz aufzutun. Die klinische Wirklichkeit, der man sich aktuell ausgesetzt sieht, scheint sich in den Studien nicht so recht abzubilden. Woher rührt diese verbreitete häufig unterschwellige Unzufriedenheit?



Eine Meinung, die ich dazu öfter höre und die vielleicht auch hier im Raum Anhänger findet, ist, dass sich in dieser Studienvielfalt eigentlich viel zu viel Minderwertiges verberge. Man solle sich auf die erwiesenermaßen besten Studien beschränken, d.h. auf die Goldstandard-Studien. Man solle von diesen Studien dann nicht viele kleine, sondern wenige große besonders saubere machen und den Rest einfach vergessen, davon hätten auch Arzt und Patient am meisten.

Meine Damen und Herren, ich bin der festen Überzeugung, dass diese Vorgehensweise vom Grundsatz her falsch ist und zu einer Verschärfung der Probleme führen würde. Ich bin der Meinung, dass wir ganz im Gegenteil die Studienvielfalt weiter verstärken müssen und uns auch wieder stärker um die Pflege der nicht-randomisierten Studien kümmern müssen, die wir in den letzten Jahren vernachlässigt haben.

Gliederung

- 1) Wie wir aus klinischen Studien (nicht) lernen**
- 2) Welche Informationen brauchen Patient und Arzt?**
- 3) Welche Informationen bekommen Patient und Arzt aus Studien? (Forschungs-Bias)**
- 4) Wie sollte eine zukünftige Forschungslandschaft aussehen?**

In meinem Vortrag will ich versuchen, diese Position zu begründen. Ich will zunächst etwas dazu sagen, wie man, unabhängig vom Studientyp, aus klinischen Studien prinzipiell lernen kann bzw. nicht lernen kann. Dazu werde ich einige grundsätzliche Ausführungen zur Natur des Wissenschaftsprozesses machen. Dann werde ich mich der Frage zuwenden, welche Information und welche Studientypen Patient und Arzt für ihre therapeutische Entscheidungsfindung benötigen. Ich werde anschließend prüfen, ob Arzt und Patient in der aktuellen Forschungslandschaft tatsächlich mit dem versorgt werden, was sie brauchen. Ich werde dabei einen Forschungs-Bias konstatieren und dazu zwei Zeugen aufrufen. Und schließlich werde ich einige Hinweise geben, in welche Richtung sich die Forschungslandschaft vielleicht entwickeln sollte. Aber damit laufe ich schon Gefahr, meinen Nachrednern vorzugreifen, die konkrete Vorschläge unterbreiten werden.

- 1 -

Lernen aus klinischen Studien

5

Die folgende Folie setzen wir im Unterricht bei der Behandlung von klinischen Studien ein. Wir haben zuvor die Vorteile randomisierter Studien besprochen und müssen nun auch über die Grenzen reden, die allen Studien unabhängig vom Studientyp gesetzt sind. Es handelt sich dabei im weiteren Sinne ausnahmslos um Übertragungsprobleme.

Begrenzungen von Studien

- 1) Aus Studien folgen statistische Aussagen (Kollektiv-Aussagen), keine individuellen Aussagen.**
- 2) Statistische Aussagen sind von Natur aus unscharf.**

6

Aus Studien folgen statistische Aussagen (Kollektiv-Aussagen), keine individuellen Aussagen.

Von Florence Nightingale stammt der Satz: „Wenn wir wissen wollen, was Gott mit uns vorhat, müssen wir Statistik lernen.“ Leider stimmt dieser Satz nicht. Wenn überhaupt, dann erfahren wir durch Statistik, was Gott mit der Menschheit vor hat, aber nichts über Individuen (und das ist vielleicht auch gut so, wenn es um den Tod geht). Patient und behandelnder Arzt befinden sich aber in einer Situation, in der es hauptsächlich um das Individuum gehen sollte. Leider können aber Studienergebnisse nicht so ohne weiteres auf jedes Individuum übertragen werden.

Statistische Aussagen sind von Natur aus unscharf.

Wie immer wir es anstellen, die quantitativen Ergebnisse einer klinischen Studie werden unvermeidbar mit einer Unsicherheit belegt sein, die sich aus der Stichprobenziehung ergibt. Diese Unsicherheit kann beträchtlich sein und bis zur Nutzlosigkeit von Aussagen führen. Konfidenzbereiche sind ein Versuch, sich diese Unsicherheit vor Augen zu führen. Das kann sehr desillusionierend sein. Wir gucken da oft nicht so gern hin und ziehen es daher vor, p-Werte zu verwenden, die uns die Unsicherheit von Aussagen weniger aufdrängen und uns anschließend einen sorgloseren Umgang mit den Mittelwerten ermöglichen.

Mittelwerte sind aber ohnehin nur ein Teil der Wahrheit. Interessant ist die ganze Bandbreite biologischer Reaktionen auf eine Therapie, ausgedrückt in der gesamten Verteilung oder zumindest der Varianz. Varianzen sind aber noch schwieriger zu schätzen als Mittelwerte. Meist sind unsere Studien daher für viele klinisch wichtige Fragen nicht trennscharf genug.

Begrenzungen von Studien

- 1) **Aus Studien folgen statistische Aussagen (Kollektiv-Aussagen), keine individuellen Aussagen.**
- 2) **Statistische Aussagen sind von Natur aus unscharf.**
- 3) **Studienaussagen gelten nicht notwendig für andere Settings.
(Externe Validität)**
- 4) **Studienaussagen gelten nicht notwendig für alle Untergruppen.**

7

Studienaussagen gelten nicht notwendig für andere Settings. (Externe Validität)

Dieser Gedanke ist inzwischen vielen vertraut. Die meisten Autoren denken an die externe Validität, wenn sie von Übertragbarkeitsproblemen sprechen. Tatsächlich ist jedoch die externe Validität nur eines von mehreren Übertragungsproblemen, die wir bei klinischen Studien haben.

Studienaussagen gelten nicht notwendig für alle Untergruppen.

Ein besonders tückisches Problem ist die möglicherweise fehlende Übertragbarkeit nach innen. Wir sehen in unseren Studien die Patientenkollektive häufig als homogene Einheiten an. Tatsächlich ist jedoch keineswegs gewährleistet, dass die Gesamtaussage der Studien auch für alle Untergruppen gilt. Inhomogenitäten (Subgruppeneffekte) sind sogar zu erwarten. Meist gelingt es in einer Studie jedoch nicht, abweichende Subgruppen zu identifizieren, da die Power nur für die Analyse der Gesamtgruppe ausreicht.

Studienaussagen müssen nicht ewig gelten.

Tatsächlich ist auch die Übertragung von Studienergebnissen von einer Periode auf eine andere äußerst problematisch. Die Bevölkerung, die Lebensumstände wie das medizinische Umfeld entwickeln sich. Wenn eine Therapie gestern hochwirksam war, muss dies heute nicht mehr gelten. Ein solcher schleichender Wirksamkeitsverlust wird „Creeping“ genannt. Die FDA hat diesen Effekt inzwischen entdeckt. Sie und andere Agenturen verlangen bei Nichtunterlegenheitsstudien daher einen direkten Nachweis, dass das Vergleichspräparat noch wirksam ist.

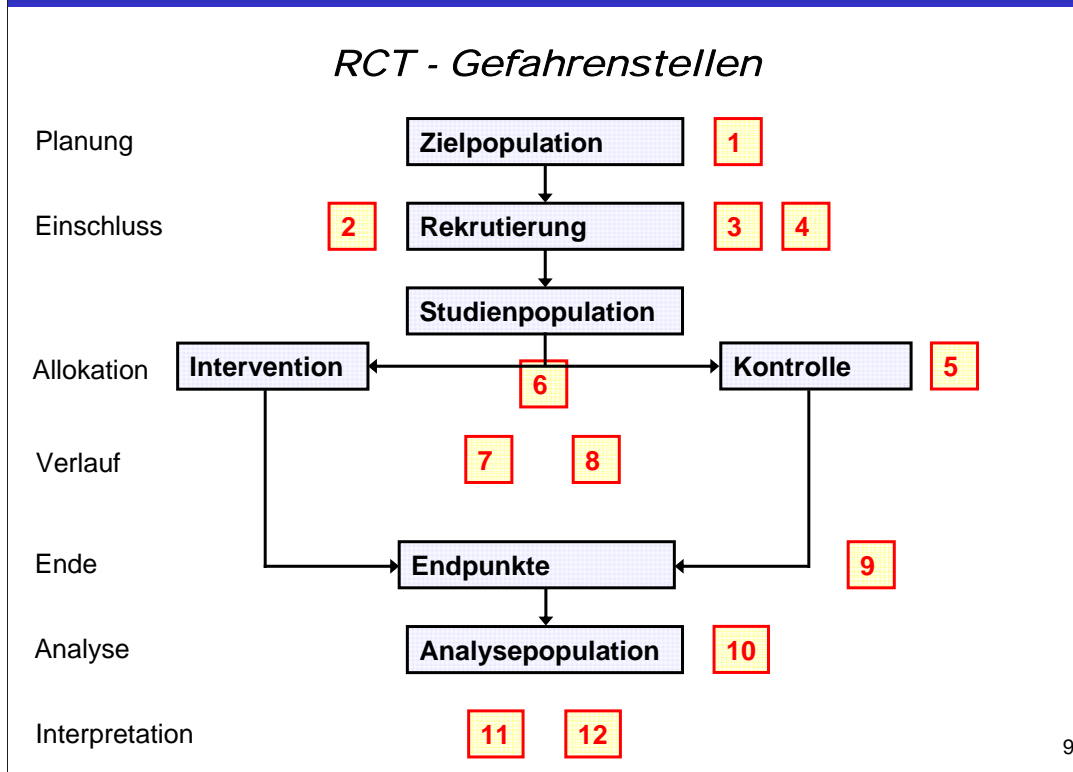
Begrenzungen von Studien

- 1) **Aus Studien folgen statistische Aussagen (Kollektiv-Aussagen), keine individuellen Aussagen.**
- 2) **Statistische Aussagen sind von Natur aus unscharf.**
- 3) **Studienaussagen gelten nicht notwendig für andere Settings.
(Externe Validität)**
- 4) **Studienaussagen gelten nicht notwendig für alle Untergruppen.**
- 5) **Studienaussagen müssen nicht ewig gelten. („Creeping“)**

8

In der Summe ist festzuhalten, dass es nicht nur ein Übertragungsproblem gibt, sondern viele. Sie haben gemeinsam, dass es keine empirische Methode gibt, um festzustellen, ob die Übertragbarkeit gerechtfertigt oder zweifelhaft ist. Um herauszufinden, ob ein Studienergebnis auf ein anderes Studiensetting, eine Untergruppe, eine zukünftige Kohorte übertragbar ist, müsste ich eine ausreichend große neue Studie in diesem Setting, dieser Untergruppe und dieser Kohorte durchführen. Dann muss ich aber die Übertragbarkeit gar nicht mehr beurteilen, weil ich ja direkte Beobachtungen aus dem Zielkollektiv habe. **Es folgt, dass die Übertragbarkeit von Studienergebnissen grundsätzlich nicht empirisch prüfbar ist.**

Es sei im übrigen darauf hingewiesen, dass dieses Problem den Pionieren der modernen klinischen Forschung durchaus bereits in seiner ganzen Problematik vertraut war. Ein Versuch, mit diesem Problem pragmatisch umzugehen, stellt das Prinzip der doppelten Wiederholung dar, dass seit langer Zeit als eine der Richtlinien bei der Bewertung der Evidenzlage in der Arzneimittelzulassung Verwendung findet: Ergebnisse müssen sowohl auf Patientenebene reproduziert werden (ausreichende Fallzahl) als auch auf Studienebene (in der Regel mindestens zwei randomisierte Studien in verschiedenen Settings).



Eine weitere wichtige Begrenzung ergibt sich daraus, dass es in der Praxis außerordentlich schwierig ist, Studien fehlerfrei durchzuführen. In Zulassungsverfahren gibt es praktisch keine Studien, die ohne kritische Relativierungen der Methodiker durch den Review kommen. Das gilt auch für randomisierte Studien. In unseren Prüfarztkursen zeigen wir den Studierenden eine Folie mit 12 Schwachpunkten, an denen auch randomisierte Studien fundamental scheitern können (es gibt zu jedem dieser Punkte Beispiele). Fehler bei der Randomisation sind nur einer dieser Punkte (Nr. 6).

Potentielle Trugschlussquellen nach Studientypen

Quelle	Meta-analyse Review	RCT	Kohortenstudie	Fall-/Kontr.-studie	Register	Qualitat. Studie
Overt Bias	☛	-	-	-	-	☛
Hidden Bias	☛	-	☛	☛	☛	☛
Zufall	>5%	5%	5%	5%	5%	nicht bezifferbar
Mangelnde Verblindung	☛	☛	☛	☛	☛	☛
Differential missings	☛	☛	☛	☛	☛	☛
Mangelnde Repräs.	☛	☛	☛	☛	-	☛
Heterogenität	☛	☛	☛	☛	-	☛
Ungeeigneter Outcome	☛	☛	☛	☛	☛	☛

10

Die Abbildung zeigt die Schwachstellen verschiedener Studientypen im Vergleich. Man erkennt, dass die randomisierte Studie zwar die einzige Studie ist, die es erlaubt, den Hidden Bias (nicht erkennbare Verfälschungen) statistisch zu kontrollieren. In Bezug auf alle anderen Punkte aber ist die randomisierte Studie genau so gefährdet wie andere Studientypen auch.

*Warum machen wir bei diesen Unsicherheiten
dennoch Studien?*

- 1) **Studien sind bei einer Vielzahl von medizinischen Problemen unsere einzige Chance, zu lernen und besser zu werden.**
- 2) **Es funktioniert!**

Studien geben unserem klinischen Verständnis wertvolle Impulse, obwohl sie selten klare und einfache Schlüsse zulassen. Sie „erden“ unsere Abbilder, die wir im Kopf tragen. – Dafür ist eine saubere Methodik unerlässlich.

11

Warum machen wir, wenn das alles denn so schwierig ist, überhaupt noch Studien? Nun, ich meine, auf diese nahe liegende Frage gibt es eine defensive und eine offensive Antwort:

Dass wir trotz der genannten Generalisierbarkeitsprobleme und trotz der grundsätzlichen Zweifel an der Aussagekraft aus Studien viel lernen können, verdanken wir den Besonderheiten des Wissenschaftsprozesses.

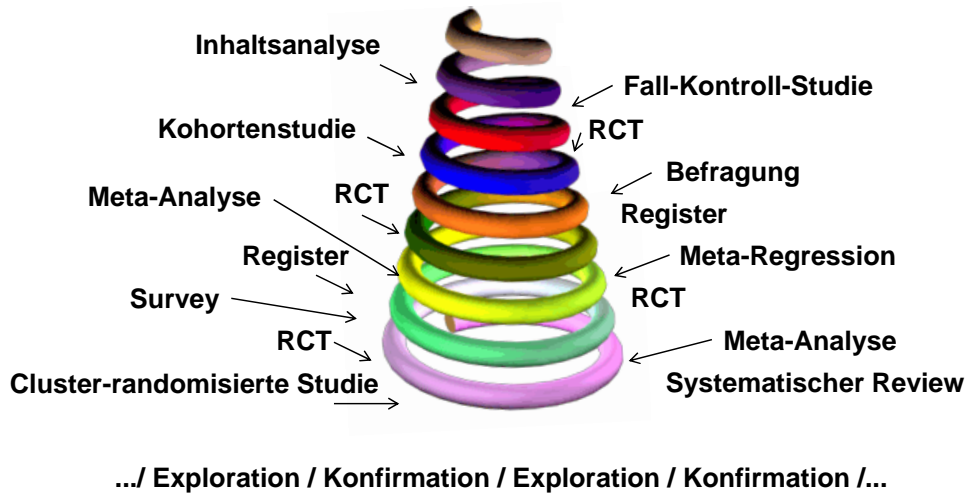
Der Wissenschaftsprozess



Dass wir trotz der genannten Generalisierbarkeitsprobleme und trotz der grundsätzlichen Zweifel an der Aussagekraft aus Studien viel lernen können, verdanken wir den Besonderheiten des Wissenschaftsprozesses.

Das Fortschreiten des wissenschaftlichen Erkenntnisprozesses lässt sich gut durch eine Spirale veranschaulichen, die nach unten, d.h. in Richtung der Erkenntnis, immer breiter wird. Die Bewegungsrichtung verläuft hauptsächlich in Schleifen. Die Wissenschaft kreist für eine längere Zeit beständig um die gleichen Fragen. Mal dominiert die eine, mal die andere Auffassung im ununterbrochen heftig geführten Diskurs, ein Markt der Meinungen und Einschätzungen, der von keinem wirklich beherrscht wird, obwohl es viele versuchen. Kurzfristig betrachtet scheint sich die Diskussion nur in die Breite zu bewegen. Längerfristig bewegt sich der Prozess aber in Richtung größerer Erkenntnis, in die immer mehr Gesichtspunkte einbezogen werden. Dabei werden zu keinem Zeitpunkt endgültige Entscheidungen getroffen. Alles ist vielmehr vorläufig und kann jederzeit relativiert werden. Strittige Fragestellungen werden jedoch im Laufe der Zeit wieder verlassen. Eine bestimmte Auffassung setzt sich dann durch, ohne jemals offiziell zur Wahrheit erklärt worden zu sein, und wird zur Grundlage weiterer Überlegungen. Eigentlich wird dabei nie eine Frage wirklich entschieden. Diskurse gehen häufig erst dadurch zu Ende, dass eine neue Generation Wissenschaftler die restliche Uneinigkeit nicht mehr so wichtig findet.

Der Wissenschaftsprozess



Lückenschluss: - Hypothesen
- Analog-Schlüsse

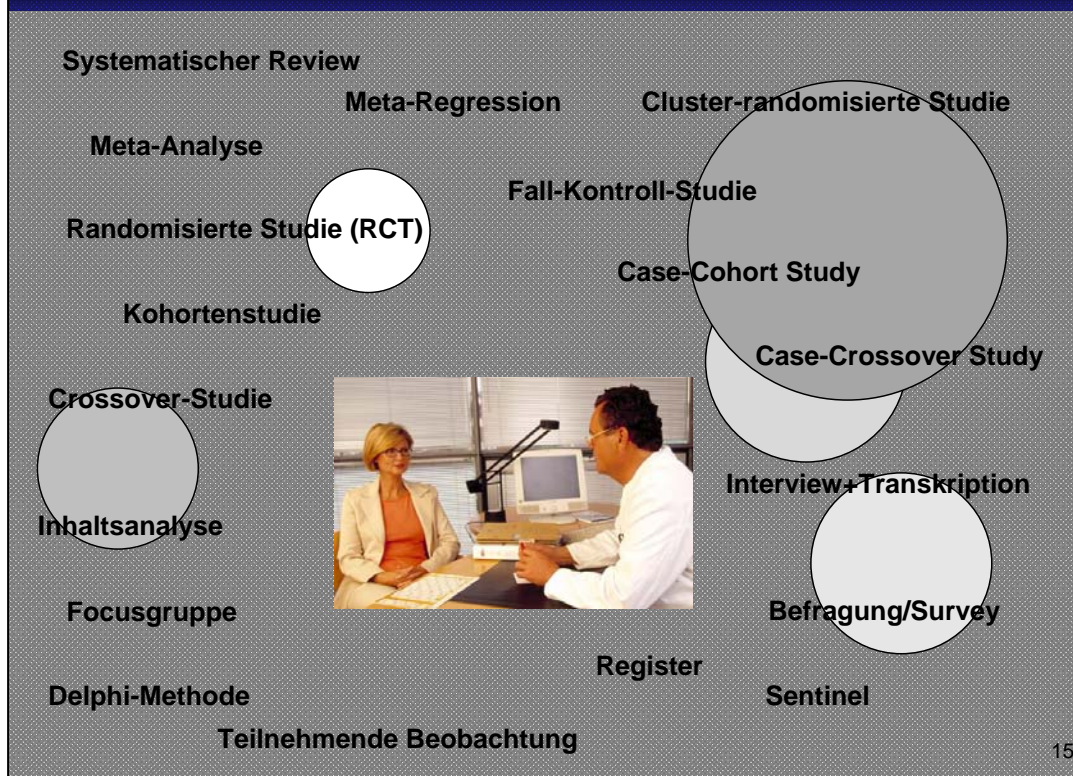
13

Der Prozess wird genährt durch ständige Impulse einer Vielzahl von empirischen Studien unterschiedlicher Qualität. Dabei wechseln sich Hypothesengenerierung (Exploration) und Hypothesenvalidierung (Konfirmation) in schneller Folge ab. In der Exploration liegt die Kreativität, in der Konfirmation die Anbindung an die Realität. Wissenslücken sind dabei ein normaler Zustand und werden durch Hypothesenbildung oder Analogie-Schlüsse gefüllt. Die Fähigkeit zu empirisch nicht oder noch nicht abgesicherten Analogieschlüssen ist dabei das eigentliche Charakteristikum und eine Stärke, nicht eine Schwäche der Wissenschaft, ein Charakteristikum, das als solches nur dem Menschen eigen ist. Im Zusammenwirken mit dem kritischen Geist wird der Mensch durch diese Fähigkeit in die Lage versetzt, Gesetzmäßigkeiten und ihre Grenzen zu erkennen. Der Umstand, dass das grundsätzlich nur unscharf und fehlerbehaftet möglich ist, ändert nichts an der beispiellosen Erfolgsgeschichte von Wissenschaft, ohne die menschliches Überleben in der heutigen Form kaum denkbar wäre.

die (RCT)



Eine Erfolgsgeschichte ist auch die Entwicklung der Instrumente ‚Klinische Studien‘ und ‚Klinische Epidemiologie‘, ohne die ein Großteil der heutigen medizinischen Praxis nicht in der Form existieren würde, die wir inzwischen als selbstverständlich hinnehmen. Dabei wird allerdings, solange man sich auf RCTs beschränkt, die therapeutische Wirklichkeit nur sehr ungleichmäßig ausgeleuchtet. Neben therapeutischen Fragestellungen, auf die große randomisierter Studien ein scharfes Schlaglicht richten, finden sich teilweise in unmittelbarer Nachbarschaft alltägliche Fragestellungen, für die es keine Studien der Evidenzklasse 1a gibt.



Nicht-randomisierte Studien können dieses Dunkel zwar nicht hell ausleuchten, sie schaffen aber weitere schwächer erleuchtete Stellen. Die restlichen Lücken müssen vom behandelnden Arzt ausgefüllt werden. Er füllt sie ebenfalls mit Analogieschlüssen, das heißt er unterstellt Übertragbarkeit auf der Basis von Plausibilitätsüberlegungen. Dies ist zwar nicht im engeren Sinne evidenzbasiert, aber dennoch eine wissenschaftliche Vorgehensweise. Damit liegt dieser Bereich zwar zweifellos eher in einem Dämmerlicht, und wir haben Mühe, klar zu erkennen, was sich dort verbirgt. Dieser Zustand ist aber immer noch besser als die völlige Dunkelheit, die an diesen Stellen zu finden wäre, wenn wir uns nicht-randomisierte Studien und Analogieschlüsse verbieten würden.

Problem:
Entscheidungen auf wissenschaftlicher Basis

Regulator:

- schafft Regelwerk
- verlangt Evidenznachweise
- konzentriert sich auf überschaubare Maßzahlen:
Mittelwerte, Signifikanzen, Meta-Mittel, Meta-Signifikanzen
- fordert Nutzenbericht an

16

Mit dieser Situation können die meisten Akteure gut leben.

Ein Problem mit dieser Situation haben hingegen die Regulatoren, die sich zum Ziel gesetzt haben, Therapiebewertungen hauptsächlich auf der Basis hochrangiger Evidenz vorzunehmen, sei es bei der Zulassung oder bei der Kostenerstattung. Für sie ist der freie Analogieschluss unzureichend, weil schwer objektivierbar und interpretationsabhängig. Man kann sich zunächst das Problem eine Zeitlang vom Leib halten, indem man den Antragstellern, die eine Zulassung oder Kostenerstattung anstreben, hohe Nachweishürden setzt. Auf diese Weise schafft man sich eine idealisierte Subwelt, in der man sich der Illusion der zumindest statistischen Kontrolle von Schlussfehlern hingeben kann. Spätestens mit der Ausweitung des Bewertungsauftrages über den Wirksamkeitsnachweis hinaus auf eine vergleichende Nutzenbewertung ist diese Position nicht mehr durchzuhalten. Die überall lauenden Evidenzlücken machen die Frage der Übertragbarkeit und den Wunsch nach einer wissenschaftlichen Methodik hierfür dringend. Der Regulator (z.B. der GBA) kann (und sollte vernünftigerweise) den Bewertungsauftrag an ein wissenschaftliches Institut verschieben (z.B. indem man das IQWiG mit einem Nutzenbericht beauftragt). Damit ist das Problem aber noch nicht gelöst. Dieser Tatsache verdanken wir denn auch wohl die heutige Einladung.

*Beispiel: Die TORCH-Studie***The NEW ENGLAND
JOURNAL of MEDICINE**

ESTABLISHED IN 1812

FEBRUARY 22, 2007

VOL. 356 NO. 8

**Salmeterol and Fluticasone Propionate and Survival
in Chronic Obstructive Pulmonary Disease**

Peter M.A. Calverley, M.D., Julie A. Anderson, M.A., Bartolome Celli, M.D., Gary T. Ferguson, M.D., Christine Jenkins, M.D., Paul W. Jones, M.D., Julie C. Yates, B.S., and Jørgen Vestbo, M.D., for the TORCH investigators*

ABSTRACT

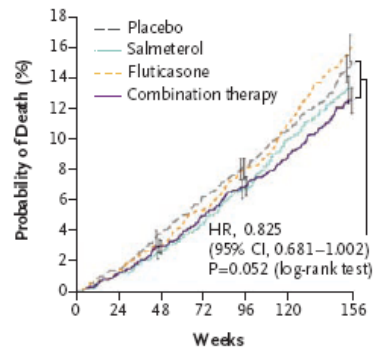
BACKGROUND

Long-acting beta-agonists and inhaled corticosteroids are used to treat chronic obstructive pulmonary disease (COPD), but their effect on survival is unknown.

From University Hospital Aintree, Liverpool, United Kingdom (P.M.A.C.); Glaxo-SmithKline Research and Development,

17

Eine Folge der Konzentration auf Studien höchster Evidenzklassen (RCTs) und darauf aufbauende Meta-Analysen ist die Einengung des Blickwinkels auf stark zusammenfassende Maßzahlen (Mittelwerte, Signifikanzen, Meta-Mittelwerte, Meta-Signifikanzen). Damit kann eine vom Wesentlichen ablenkende Reduktion der Information verbunden sein. Als Beispiel sei hier die TORCH-Studie angeführt (N Engl J Med 2007;356:775-89). In dieser Studie wurde eine fixe Kombination (C) von zwei Asthma-Medikamenten mit unterschiedlichen Wirkpfaden (Salmeterol (S) und Fluticasone propionate (F)) bei COPD-Patienten gegen Placebo und gegen die einzelnen Komponenten getestet. Hauptzielgröße war die Gesamt-Mortalität.

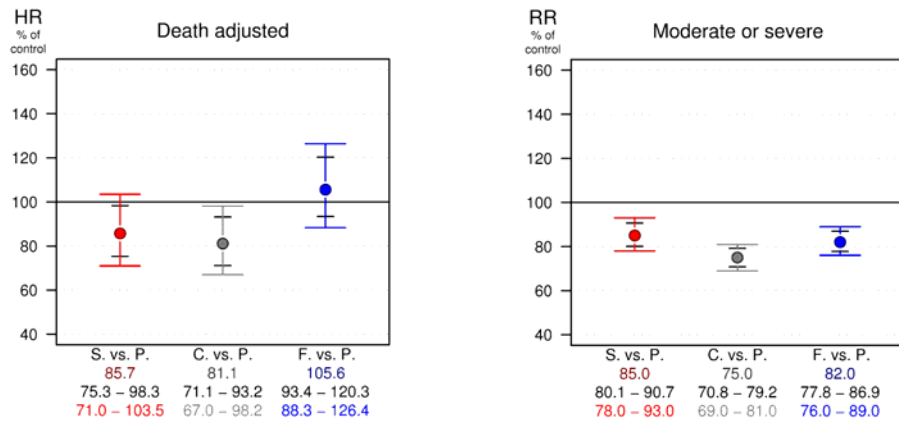
B Death from Any Cause**Editorial:**

In the end, the trial failed to meet its goal: the P value for death from any cause was 0.052, which was higher than the prespecified value of 0.050. **All clinical trials are a gamble, and the TORCH investigators came close to winning but did not win. Thus, the results of this trial are difficult to interpret.**

Der Sponsor strebte die Zulassung der Kombination zum Zweck der Mortalitätsreduktion an. Entsprechend wurde die Studie auf den Vergleich zu Placebo optimiert. Hierzu wurde ein gruppen-sequentielles Design verwendet mit dem Ziel, die erforderliche Patientenzahl zu minimieren. Im Ergebnis wurde die formale Signifikanz knapp verfehlt.

Der Kommentar des Editor spiegelt die Blickverengung der Primäranalyse. Das Ergebnis der klinischen Studie wird auf einen p-Wert reduziert. Die Frage, ob der p-Wert knapp oberhalb oder unterhalb von 0,05 liegt, wird zum dominierenden Gegenstand der Betrachtung. Die klinische Studie wird hier zum erklärten Spielball der Zocker. Die aktuelle Bankenkrise führt uns warnend vor Augen, wo wir hinkommen, wenn wir das Feld den Zockern überlassen.

TORCH aus der Perspektive der Wissenschaft



Die wirkliche Bedeutung der TORCH-Studie liegt hingegen auf einer anderen Ebene. Als erste große randomisierte Mortalitätsstudie in ihrem klinischen Setting gab sie der wissenschaftlichen Diskussion über die richtige Behandlung von COPD-Patienten einen kräftigen Impuls. Während man vorher vermutet hatte, dass die Mortalitätsreduktion eher durch die Kortikosteroide (F) als durch langwirkende Beta-Agonisten (S) bewirkt wird, ließ die Studie erkennen, dass tatsächlich nur Beta-Agonisten die Sterblichkeit reduzieren (linkes Bild). Beide Substanzen wirken hingegen symptomlindernd (rechtes Bild). Die Medikamente interagieren nicht und können frei kombiniert werden. Diese für die Behandlungspraxis wichtigen Aspekte ließen sich jedoch aus der Studie nicht sicher schlussfolgern, da die Power dafür nicht ausreichte. Vielmehr müssen supportiv weitere Studien hinzutreten, die diese Schlüsse erhärten oder relativieren können. - Der tatsächliche Wert der Studie liegt in der Diskussion dieser Aspekte durch die Fachwelt, nicht in der evidenzbasierten Beurteilung der Fixkombination.

Problem:
Entscheidungen auf wissenschaftlicher Basis

Regulator:

- schafft Regelwerk
- verlangt Evidenznachweise
- konzentriert sich auf überschaubare Maßzahlen:
Mittelwerte, Signifikanzen, Meta-Mittel, Meta-Signifikanzen
- fordert Nutzenbericht an

Patient und Arzt?

20

Mit den Evidenzlücken hat nun aber nicht nur der Regulator, sondern insbesondere auch der Arzt und sein Patient Probleme. Sie stehen wie der Regulator und anders als die Wissenschaft unter Entscheidungszwang. In gewisser Weise ist ihre Situation noch viel dramatischer als die des Regulators: während dieser zur Not wegen schlechter Evidenzlage auch Zuflucht zu einer pragmatischen politischen Entscheidung nehmen könnte, müssen Arzt und Patient eine Therapie-Entscheidung (die auch Zuwarten sein kann) treffen, die weit reichende Konsequenzen haben kann.

- 2 -

**Welche Informationen brauchen Arzt und Patient aus
klinischen Studien?**

Die Aufgabe des Arztes

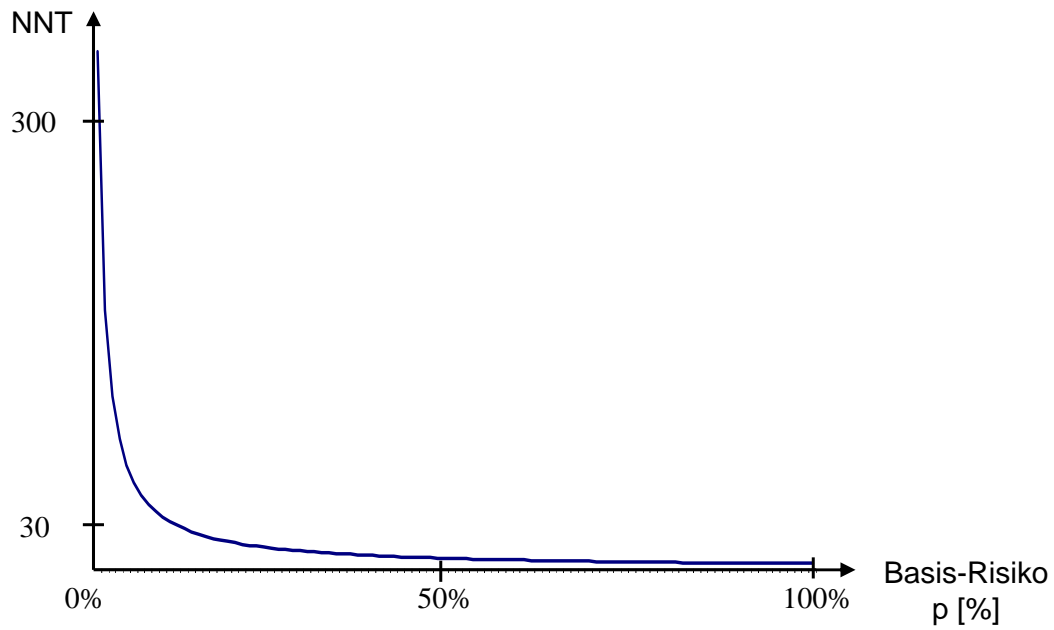
Die Aufgabe des behandelnden Arztes geht über die Evidenzfindung hinaus.

Er muss die Eignung der Therapie für den Patienten prüfen (individueller Nutzen).

Er muss „weiche“ Entscheidungen mit dem Patienten zusammen treffen.

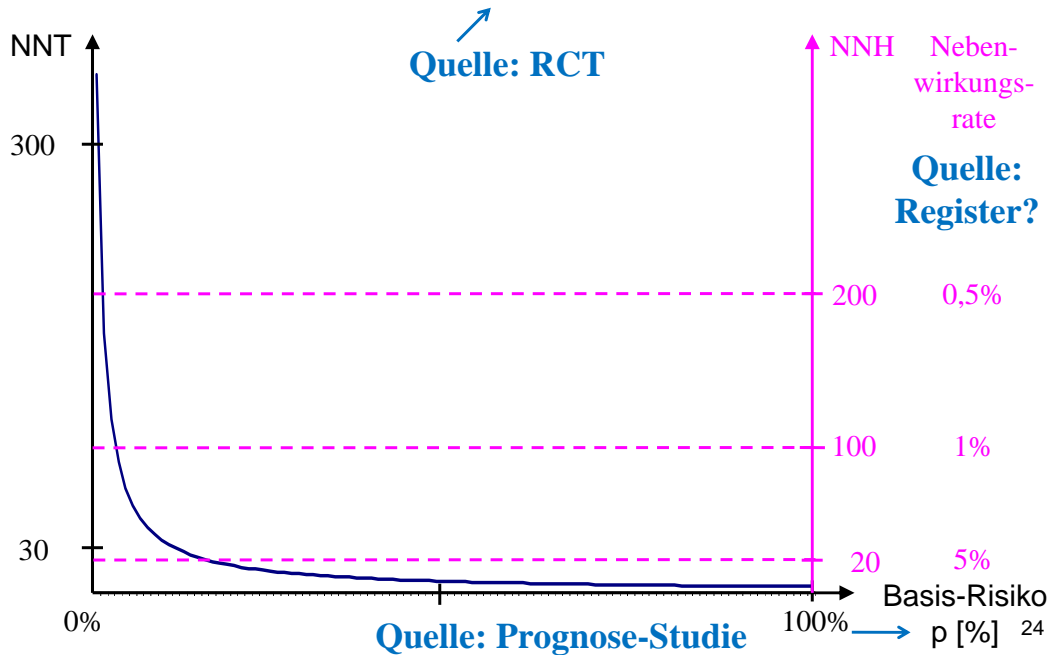
Und der behandelnde Arzt hat zudem eine weitaus schwerere Aufgabe zu lösen als der Wissenschaftler oder der Regulator. Er muss sich nicht nur Gewissheit über die Wirksamkeit einer Therapie verschaffen, sondern darüber hinaus einschätzen, ob die Therapie seinem individuellen Patienten nutzen wird und ob sie ihm schaden kann. Dabei geht es anders als in den IQWiG-Nutzenberichten nicht um einen kollektiven Nutzen, sondern um einen ganz persönlichen Nutzen.

Beispiel Primär- oder Sekundärprävention: Therapie-Entscheidung bei 30% Risikoreduktion



Die entscheidende Schlüsselzahl hierfür ist nicht die Risikoreduktion, die wir zurecht zur Beurteilung der Wirksamkeit heranziehen, sondern die sog. ‚number needed to treat (NNT)‘, die die Zahl der Patienten angibt, die man behandeln muss, damit einer von ihnen profitiert. Die NNT hängt entscheidend vom Basis-Risiko des Patienten (dem Risiko ohne Behandlung) ab, muss also als Funktion dieses Basisrisikos abgetragen werden, weil sie für jeden Patienten anders ist. Die Kurve hat einen bemerkenswerten Verlauf. Bei großem Risiko ist die individuelle NNT so niedrig, dass eine Behandlung sehr empfehlenswert erscheint. Bei sehr kleinem Risiko erscheint eine Behandlung bei sehr hohen NNT-Werten unwirtschaftlich. In dem Bereich dazwischen hängt die NNT extrem vom genauen Zahlenwert des Risikos ab. Für die Behandlungsentscheidung ist in diesem Bereich, in dem häufig ein großer Teil der Patienten zu finden ist, die Beurteilung des Basis-Risikos des Patienten wichtiger als die genaue Bezifferung der Risikoreduktion.

Beispiel Primär- oder Sekundärprävention: Therapie-Entscheidung bei 30% Risikoreduktion



Bei der Behandlungsentscheidung müssen Arzt und Patient darüber hinaus berücksichtigen, welche Nebenwirkungen unter der Therapie zu erwarten sind. Das geschieht am besten durch die zusätzliche Betrachtung der ‚number needed to harm (NNH)‘, die häufig, aber nicht immer, unabhängig vom Basis-Risiko ist und zur NNT in Relation gesetzt werden muss.

Woher erhält der Arzt die Angaben, die er für den beschriebenen Abwägungsprozess braucht? Die Risikoreduktion erfährt er aus RCTs und Meta-Analysen. Das Basis-Risiko hingegen stammt eher aus Prognosestudien, die überwiegend als nicht-randomisierte Kohorten-Studien organisiert sind (Beispiel: Framingham-Studie). Nebenwirkungsdaten können aus den RCTs vor Zulassung meist nicht realitätsnah und ausreichend präzise geschätzt werden. Hierzu sind qualitativ hochwertige Registerdaten und Angaben aus Pharmakovigilanzdatenbanken erforderlich.

Welche Informationen brauchen Arzt und Patient aus Studien?

	Frage	Aspekt	Studientyp
<u>Patient/Arzt</u>	Was wird helfen?	Wirksamkeit	RCT
	Was wird, wenn ich das Risiko nicht eingehe?	natürlicher Verlauf	Kohorten-Studie
	Welche Risiken gehe ich dafür ein?	Sicherheit	RCT, Kohorten-Studie, Register
<u>Patient</u>	Wie ist meine Krankheit einzuordnen?	Krankheits- verständnis	Qualitative Studie
	Wie wird es mir unter der Therapie gehen?	Lebensqualität	Kohorten-Studie, RCT

25

Der Patient will darüber hinaus häufig noch mehr wissen, z.B., wie es ihm unter der Therapie gehen wird. Fragen nach der Lebensqualität können am besten aus im Gefolge von Interventions- oder Kohortenstudien durchgeführten wiederholten Befragungen (Surveys) beantwortet werden. Insbesondere bei chronischen Erkrankungen hat der Patient zudem ein Bedürfnis, seine Krankheit in sein Leben einzuordnen. Für die Aufgabe der Krankheitsbewältigung sind häufig um Verständnis bemühte qualitative Studien hilfreicher als quantitative Studien. Damit wird deutlich, dass wir verschiedene methodische Zugänge brauchen, um in Bezug auf Krankheiten und Therapien alles das aus Daten zu lernen, was man aus Daten lernen kann.

Insgesamt ist somit festzuhalten, dass randomisierte Studien nur einen Teil der Informationen bereit stellen können, die Patient und Arzt für eine Therapie-Entscheidung benötigen. Andere Studientypen müssen hinzutreten und können für die individuelle Entscheidung sogar bedeutsamer sein als die Quantifizierung der Wirksamkeit durch eine RCT. Bei der reinen Wirksamkeitsbeurteilung sind diese Studien den RCTs in Bezug auf die interne Validität zwar unterlegen und werden deshalb in niedrigeren Evidenzklassen geführt. Sie tragen jedoch andere für Therapie-Entscheidungen unverzichtbare Informationen bei. **Zu betonen ist in diesem Zusammenhang auch, dass die beschriebene individuelle Nutzenabwägung originäre Aufgabe des behandelnden Arztes ist und nicht durch eine kollektive Nutzenentscheidung einer zentralen Institution ersetzt werden kann.** Es ist hingegen sehr wohl vorstellbar, dass bei guter Studienlage eine Bestimmung des kollektiven Nutzens entfallen kann, weil der individuelle Nutzen gut beurteilbar ist.

- 3 -

**Welche Informationen bekommen Patient und Arzt aus
klinischen Studien?**

- oder:

Der Bias in der medizinischen Forschung

Unterschiede bei der Bewertung von Wirksamkeit und Sicherheit von Medikamenten

Wirksamkeitsprüfung:

- vor Markteinführung
- aktive Datensammlung
prospektiv
- systematische Suche
nach Wirkungen
- ausfinanziert auf Sponsoren-
und FDA-Seite

Sicherheitsprüfung:

- nach Markteinführung
- passive Datensammlung
retrospektiv
- zufälliges Entdecken
von Nebenwirkungen
- Finanzierung unregelt

Institute of Medicine, USA: Recommendations on Drug Safety

- Neuregulierung des gesamten Safety- Bereiches

27

Bekommen Patient und Arzt nun die Studien, die sie brauchen? Über diese Frage haben sich die Mitglieder des Institute of Medicine (IoM), eines großen virtuellen Institutes, dessen Mitglieder Verantwortung tragende Wissenschaftler aus den USA sind, Gedanken gemacht. Sie untersuchten die Hintergründe für die spektakulären Warn- und Rücknahme-Aktionen aus Sicherheitsgründen, die wir in den letzten Jahren im Arzneimittelbereich erlebt haben. Sie konstatierten einen kräftigen Bias in der Forschungslandschaft. Während die Wirksamkeit eines neuen Medikamentes generalstabsmäßig mit aktiver Datensammlung und sorgfältig abgestimmten und finanziell gut ausgestatteten Studienprogrammen bereits vor der Markteinführung erfolgt und dann auf eine gut ausgestattete und präparierte Behörde mit Sanktionsmöglichkeiten trifft, wird die Sicherheit nach der Zulassung nur mit passiven Datensammlungen und unterfinanziert untersucht. Nebenwirkungen werden nur zufällig entdeckt. Die unterausgestattete Behörde hat kein abgestuftes Instrumentarium zur Erkundung von und zum differenzierten Umgang mit Risiken. Patient, Arzt und die Gesellschaft als Ganzes bekommen so ein lückenhaftes und verfälschtes Bild von der Nutzen-Risiko-Bilanz von Medikamenten.

Entsprechend fordert das IoM die Neuregulierung des gesamten Bereiches.

Vorschläge Institute of Medicine :

- Gleichmäßige Verteilung der FDA-Gelder zwischen Efficacy und Safety
- Autorisierung der FDA, Post-Marketing-Studien zu verordnen und Konsequenzen aus den Ergebnissen zu ziehen
- Ex-ante-Registrierung aller Studien
- Verbesserung der Risikokommunikation mit Ärzten und Patienten
- Kennzeichnung neuer Medikamente mit einem schwarzen Dreieck
- Zwei Jahre Direct-to-consumer-Werbeverbot

28

Hierzu werden vom IoM eine Reihe von Maßnahmen empfohlen, die in den USA mittlerweile Gegenstand der politischen Diskussion sind.

Efficacy vs. Efficiency

Spitzenforschung vs. Versorgungswirklichkeit

<u>Wirksamkeit (Efficacy)</u>		<u>Nützlichkeit / Wirtschaftlichkeit (Efficiency)</u>
unter idealen Bedingungen	Schauplatz	Im Feld
Phase II / III	Arzneimittelentwicklung	Phase IV
forschungsorientiert, Karriereweg vorhanden	Verantwortlicher Arzt	Grauhaarig, weise
	Wissenschaftliches Interesse	Nicht vorhanden
Abschnitt im Methodenteil	Rolle des Patienten	Hauptperson
Das volle Leben	Rolle der Technologie	Ein kleiner Teil des Lebens
veröffentlicht, aber wahr?	Ergebnisse	wahr, aber nicht veröffentlicht
vorhanden	Impuls für wirtschaftliche Entwicklung	nicht vorhanden

(nach Picano)

Während das IoM in der ungeeigneten Regulation die Hauptursache für den Bias in der Forschungslandschaft sieht, weist der italienische Kardiologe Picano, ein Spezialist für bildgebende Verfahren, auf die unterschiedliche Gewichtung von Studien unter Idealbedingungen zur Untersuchung der Wirksamkeit und Studien im Feld zur Untersuchung der Nützlichkeit und Wirtschaftlichkeit unter Routinebedingungen hin. Wissenschaftlich attraktiv und karrierefördernd sind nur ideale Studien. Sie schüren Hoffnungen und erhöhen die Chancen auf wirtschaftliche Verwertung. Patientennahe Studien zur Abbildung der Versorgungswirklichkeit werden entsprechend zu wenig durchgeführt. Gerade diese Studien werden aber für Therapie-Entscheidungen im Einzelfall dringend benötigt.

- 4 -

Wie könnte eine Studienlandschaft aussehen, die Arzt und Patient besser dient?

Wie kann man allgemein die Situation verbessern?

- **Wegkommen von der ausschließlichen Orientierung an Wirksamkeitsnachweisen, Versorgungsaspekte stärker in Studien gelten lassen**
- **Weniger ergebnisorientiert denken, Komplexität würdigen, Lernprozesse würdigen, die Bedeutung der Rezeption erkennen**
- **Weniger auf schnelle Signifikanz optimierte, dafür mehr pragmatische RCTs durchführen.**
- **Integrierte Studien durchführen**

31

Welche Empfehlungen können aus den vorstehenden Überlegungen abgeleitet werden? Sicherlich ist es wünschenswert, dass wir uns von der ausschließlichen Orientierung an Wirksamkeitsnachweisen lösen und Versorgungsaspekte stärker in Studien zur Geltung kommen lassen. Insgesamt wäre es vorteilhaft, wir würden weniger ergebnisorientiert denken, Komplexität und wissenschaftliche Lernprozesse stärker würdigen und die Bedeutung der diskursiven Rezeption von Studienergebnissen erkennen. Wir sollten weniger auf Signifikanz orientierte, mehr pragmatische Studien durchführen. Randomisierung bleibt ein wesentliches Element, das die Vergleichbarkeit innerhalb von Studien verbessert. Dennoch ist es sinnvoll, eine RCT zu ergänzen, indem man sie in Beobachtungsstudien oder Register integriert, die ähnlich hohen Qualitätsstandards genügen sollten. Von hoffnungsvollen Ansätzen für derart integrierte Forschungslandschaften wird zunehmend berichtet.