



# **Nutzen und Schaden von Arzneimitteln – zwei Seiten einer Medaille**

**Überlegungen zum Wann, Wer und Wie  
der Bewertung nach der Zulassung**

**Dr. Dr. Norbert Banik  
GlaxoSmithKline GmbH & Co. KG  
München**

## Disclaimer

***Die Ausführungen geben die Meinung des Autors wieder und nicht notwendigerweise die von GlaxoSmithKline.***

***Die Aussagen des Vortrags sind keinesfalls so zu interpretieren, dass die generelle Wertigkeit von konfirmatorischen RCT (z.B. im Zulassungsprozess) vom Autor als abgeschwächt gesehen oder in Frage gestellt würde.***

# Gliederung

- **Begriffsbestimmungen „Nutzen“ und „Schaden“**
  - Konsequenzen für die Bewertung
- **Welche Erkenntnisquellen („Evidenz“) berücksichtigen?**
  - Exkurs: Unverzerrte Schätzung in RCT
  - externe Validität in RCT
  - Evidenzhierarchie(n)
  - Evidenz in vergleichenden Nutzenbewertungen
- **Welche Sicherheit für Empfehlungen erforderlich und erreichbar?**
- **Vergleichende „Nutzen-Schaden-Bewertungen“**
- **Zuständigkeiten für vergleichende Nutzenbewertung**
- **Ausblick: Verbesserungsmöglichkeiten**
- **Zusammenfassung**



# **“Nutzen” & “Schaden” - Begriffe**

# „Nutzen“ & „Schaden“ - Begriffe

## ● SGB V

- §§ 2, 12, 35, 72, 92, 95, 139b
- § 35 (1b) „Ein höherer Nutzen nach Satz 1 kann auch **eine Verringerung der Häufigkeit oder des Schweregrads therapierelevanter Nebenwirkungen** sein.“
- §35b (1) „Beim Patienten-Nutzen sollen insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine **Verringerung der Nebenwirkungen** sowie eine Verbesserung der Lebensqualität, bei der wirtschaftlichen Bewertung ... angemessen berücksichtigt werden.“

## ● G-BA → Verfahrensordnung

- 4. Kapitel „Bewertung von AM und MP“
  - 2. Abschnitt „Bew. d. therapeut. Nutzens“: § 6 (1) „... therapeutischer Nutzen“; (2) „... Ausmaß der Beeinflussung patientenrelevanter Endpunkte ... Morbidität, Mortalität und Lebensqualität.“
  - 3. Abschnitt „Festbetragsgruppenbildung“ § 23 „Geringere Nebenwirkungen als therapeut. Verbesserung“; § 24 „Nachweis der ther. Verbesserung“: (2) „Der allgemein anerkannte Stand der medizin. Erkenntnisse wird auf der Basis der arznei-gesetzlichen Zulassung festgestellt.“

# „Nutzen“ & „Schaden“ - Begriffe

## ● IQWiG, Allgemeine Methoden, 3.0 (5/2008):

- „Mit dem Begriff **„Nutzen“** werden kausal begründete positive Effekte, mit dem Begriff **„Schaden“** kausal begründete negative Effekte einer medizinischen Intervention auf patientenrelevante Endpunkte (s. u.) bezeichnet.“ (S.31)
- „Das Schadenpotenzial beschreibt das Risiko für das Auftreten unerwünschter Wirkungen bei der Anwendung einer medizinischen Intervention. **Die Darstellung des Schadenpotenzials ist wesentlicher und gleichberechtigter Bestandteil der Nutzenbewertung einer Intervention.** Sie gewährleistet eine informierte, populationsbezogene, aber auch individuelle Nutzen-Schaden-Abwägung“. (S. 35)
- „Über eine Nutzen-Schaden-Abwägung im Sinne einer Gegenüberstellung von Nutzen- und Schadenpotenzial hinaus besteht die Möglichkeit, Nutzen und Schaden vergleichend zu gewichten. In diesem Fall würden die Aussagen des IQWiG **für jeden einzelnen patientenrelevanten Endpunkt** durch die Gewichtung von Nutzen und Schaden z. B. unter Verwendung eines Summenscores berichtet.“ (S. 38)
- „In der Regel sind RCTs zur Evaluation von Arzneimitteln möglich und praktisch durchführbar. **Das Institut lässt daher in der Regel RCTs in die Nutzenbewertung von Arzneimitteln einfließen. Nur in begründeten Ausnahmefällen wird auf nicht randomisierte Interventionsstudien oder Beobachtungsstudien zurückgegriffen.**“ (S. 45) ... „Die besondere Begründungspflicht für ein nicht randomisiertes Design bei der Prüfung von Arzneimitteln findet sich auch im Rahmen des Arzneimittelzulassungsrechts in den Arzneimittelprüfrichtlinien (Richtlinie 2001/83/EG, Abschnitt 5.2.5.“ (S.46)

## ● IQWiG, Methodik ... Verhältnissen zwischen Nutzen und Kosten ..., 1.1 (10/2008)

# „Nutzen“ & „Schaden“ - Begriffe

- **„Nutzen“** (teilweise nach Raspe):
  - Gesamtheit der erwünschten Effekte
  - hat „Nutznießer“ und Dimensionen („Qualitäten“, „Endpunkte“)
    - „Nutzenqualitäten“ können z.B. direkte Verbesserungen des Gesundheitszustandes oder geringere Belastungen, bessere Prognose sein
  - kann quantitativ abgeschätzt werden
  - tritt mit Wahrscheinlichkeit innerhalb bestimmter „settings“ ein
  - besitzt zeitliche Charakteristika
- **„Schaden“**:
  - Gesamtheit der unerwünschten Effekte
  - hat „Geschädigte“ und Dimensionen („Qualitäten“, „Endpunkte“)
  - kann quantitativ abgeschätzt werden
  - tritt mit Wahrscheinlichkeit innerhalb bestimmter „settings“ ein
  - besitzt zeitliche Charakteristika

# „Nutzen“ & „Schaden“ - Begriffe

## Wie kann man „Schäden“ typisieren?:

- negative Effekte → „Wirksamkeit mit negativem Vorzeichen“
- unerwünschte (Arzneimittel-)Wirkungen („unvermeidbare unerwünschte Wirkungen“)
- vermeidbare unerwünschte Wirkungen (Medikations-/Applikationsfehler)

also:

- **„Netto-Nutzen“** (nach Raspe): Das, was vom Nutzen bleibt, wenn man aktuelle und zukünftige Belastungen und Schäden abzieht



# „Nutzen“ & „Schaden“ - Begriffe

- **Vorschläge zur Sprachregelung:**

- „Nutzen“ → (bezieht sich auf) erwünschte Effekte
- „Schaden“ → (bezieht sich auf) unerwünschte Effekte
- Netto-Nutzen → **NUTZEN**

- deshalb ja auch „**Nutzenbewertung**“

(statt: „Nettonutzenbewertung“, die es eigentlich ist!)

- **Nutzen** - im Sinne von Netto-Nutzen verwendet, enthebt uns von der Notwendigkeit, von „Nutzen“ und „Schaden“ zu sprechen

- nicht verwirren lassen: „Zusatznutzen“ sei weiterhin der „zusätzliche Nutzen einer Intervention im Vergleich zu einer anderen Intervention“, also in bisherigen Begriffen der „Zusatz-Nettonutzen“

# Konsequenzen für die Bewertung

- **Erwünschte und unerwünschte Effekte („Nutzen“ und „Schaden“)** sind also erkenntnistheoretisch sehr **vergleichbare Konstrukte**, manchmal „nur“ durch ein entgegengesetztes Vorzeichen separiert
- Kein Grund, weshalb deren Bewertung nach der Zulassung prinzipiell unterschiedlichen methodischen Anforderungen genügen sollte
- Und deshalb auch kein Grund, weshalb die jeweiligen Teil-Aspekte auf unterschiedlicher Evidenz beruhen sollten
- Bewertung von „erwünschten Effekten“ und „unerwünschten Effekten“ erfordert auch deshalb gleiche Methoden, weil jede methodische Imbalance (bei erwünschten Effekten konservativ, bei unerwünschten Effekten aber antikonservativ heranzugehen) nach der Zulassung eines AM die Rate falsch negativer Nutzenbewertungen erhöhen würde
- **Schlußfolgerungen:**
  - **Prinzipiell gleiche Methoden für erwünschte und unerwünschte Effekte anwenden**
  - **Einheitliche Evidenzbasis für gesamte Nutzenbewertung zu Grunde legen**

# Konsequenzen für die Bewertung

## Fazit

Wenn etwas vergleichend beurteilt wird und es besteht „Strukturgleichheit“, dann soll auch „Beobachtungsgleichheit“ herrschen!

(Ivo Saliger „Das Parisurteil“)





**Welche Erkenntnisquellen (“Evidenz”)  
berücksichtigen?**

## Exkurs: Unverzerrte Schätzung in RCT (interne Validität)

- Randomisierung stellt „Austauschbarkeit“ (exchangeability) auf Populationsniveau her (eine Exposition hat bei Vertretern der Gruppe 1 den gleichen Effekt wie bei Vertretern der Gruppe 2)
- Austauschbarkeit (die in einem ideal durchgeführten RCT gilt) garantiert aber nur dann einen unverzerrten Effektschätzer, wenn\*:
  - **Outcome für alle Subjekte gemessen wird**
  - **Keine Non-Compliance vorliegt**
  - **Keine Kombination der randomisierten Exposition mit anderen Einflüssen vorliegt (insbes. kein ‚unblinding‘).**
- **Also nur in idealen RCT ist der Effektschätzer auch ein unverzerrter Schätzer für die Behandlungsdifferenz**
- CAVE: hier spielt wieder das Paradigma des „primären Endpunkts“ hinein; bei post-hoc Analysen von sekundären Endpunkten ist das Risiko merklich erhöht, dass wenigstens einer der drei Bedingungen verletzt ist!
- **Aber aus der „Austauschbarkeit“ folgt bei RCT immerhin, dass das Risiko für Verzerrungen (relativ zu nonRCT!) am niedrigsten ist.**
- **Das aktuelle Ausmaß der Verzerrung der Effektschätzung bei konkreten RCT (oder nonRCT) bleibt jedoch immer unbekannt!**

\* Hernán, J Epidem Comm Health 2004; 58:265-271

# externe Validität

- Auf welche Patienten ist ein Studienergebnis anwendbar (applicability) oder übertragbar (generalizability)?
- Mächtiges Thema – hier heute mit dem spezifischen Interesse an Studien zu „unerwünschten Effekten“ besprochen.
- Thema ist in Hintergrund getreten gegenüber der Diskussion um „Evidenzhierarchien“ (interne Validität), aber für die verantwortliche Bewertung von unerw. Effekten ist es von mindestens vergleichbarer Relevanz!

## Panel 1: Evidence of the neglect of consideration of external validity of RCTs and systematic reviews

- Research into internal validity of RCTs and systematic reviews far outweighs research into how results should best be used in practice.<sup>44,6</sup>
- Rules governing the performance of trials, such as good clinical practice,<sup>44</sup> do not cover issues of external validity.
- Drug licensing bodies, such as the US Food and Drug Administration, do not require evidence that a drug has a clinically useful treatment effect, or a trial population that is representative of routine clinical practice.<sup>46</sup>
- Guidance on the design and performance of RCTs from funding agencies, such as that from the UK Medical Research Council,<sup>49</sup> makes virtually no mention of issues related to external validity.
- Guidance from ethics committees, such as that from the UK Department of Health,<sup>47</sup> indicates that clinical research should be internally valid, and raises some issues that relate to external validity, but makes no explicit recommendations about the need for results to be generalisable.
- Guidelines on the reporting of RCTs and systematic reviews focus mainly on internal validity and give very little space to external validity.<sup>45,50</sup>
- None of the many scores for judging the quality of RCTs address external validity adequately.<sup>31</sup>
- There are no accepted guidelines on how external validity of RCTs should be assessed.

# externe Validität in RCT – einige illustrative Beispiele

## Osteoporose-Patienten

Table 3. Reasons patients not being eligible for study<sup>a</sup>

Reason	%
Too young <sup>b</sup>	28.3
Too old <sup>b</sup>	7.5
Disease too severe <sup>c</sup>	19.2
Comorbid conditions	60.0
Cancer or cancer history	14.2
Gastrointestinal disorders	11.7
Rheumatic disorders	8.3
Scoliosis	7.5
Other conditions	25.8
Medications	60.0
Estrogen	30.8
Anti-osteoporosis agents	25.0
Other medications	25.8
All other	2.5

<sup>a</sup>Percentages total to more than 100 because multiple conditions and/or medications apply in many patients.

Dowd et al. Osteoporos Int 2000

## Schizophrenie-Patienten

TABLE 1. The Effect of Study Inclusion and Exclusion Criteria on Reducing the Size of the Target Population

Criteria	Target Population	Reduction by %	Remaining Population
Only schizophrenia	3.6 million	99	36,000
Monodrug therapy	36,000	92	2880
Only males	2880	50	1440
No drugs/alcohol	1440	35	936
No obesity	936	30	655
No hepatitis B/HIV	655	3.5	632

General available population, 3,600,000; Kansas City MSA, 1,073,725; Oklahoma City MSA, 1,038,999; Tulsa MSA, 776,906; Wichita MSA, 544,343; Topeka MSA, 165,348 [From the United States Census Bureau of 1998].

Khan et al. J Clin Pharm 2005

# externe Validität in RCT – einige illustrative Beispiele

## Asthma-Patienten

Table 3 No. of asthma patients remaining after stepwise introduction of selection criterias.

	Asthma patients (VAS <2.5)					
	GP		Specialist		Total	
	No	%	No	%	No	%
All patients	199		135		334	
FEV <sub>1</sub> 50–85% of pred	72	36.2	52	38.5	124	37.1
Rev 12% last year	26	13.1	24	17.8	50	15.0
No co-morbidity	16	8.0	16	11.9	32	9.6
Pack year <10	11	5.5	7	5.2	18	5.4
Regular use of ICS	10	5.0	5	3.7	15	4.5
Symptomatic asthma	6	3.0	5	3.7	11	3.3

Herland et al. Resp Med. 2005



# externe Validität in RCT – einige illustrative Beispiele

## „Cox2-Patienten“

Table 1. Characteristics of patients included in the main RCTs and of patients using Cox-2 inhibitors in UK clinical practice

	RCT rofecoxib <sup>4</sup>	GPRD rofecoxib (N = 72 096)	RCT celecoxib <sup>5</sup>	GPRD celecoxib (N = 67 346)
Mean age (years)	58	65	61	66
Women (%)	79.6	65.6	68.5	65.4
Use of NSAIDs at study entry (%)	82.1	27.6*	81.4	25.9*
Indication (%)				
Osteoarthritis	0	35.0	72.7	35.5
Rheumatoid arthritis	100	4.8	27.3	4.8
Other	0	61.8	0	61.4
Use of oral glucocorticoids (%)	55.8	6.0	30.6	5.6
Daily dose (mean, mg)	50	22	800	224
Percentage of patients with continuous Cox-2 use	71 (11 months)	17 (11 months)	57 (6 months)	24 (6 months)
Number of patients using Cox-2 therapy beyond trial duration	—	10 807	—	14 746
Rate of upper GI events	2.1	0.7	2.2	0.5
Rate of MI	0.7	0.7	0.7	0.6

van Staa et al. Pharmacoepi Drug Safety 2008

# externe Validität als Faktor der Ergebnissicherheit

## Applicability

- Like publication bias, applicability is regarded as separate from but related to strength of evidence. **Low applicability can reduce the strength of evidence for a particular decision maker.** CERs should summarize characteristics that decision makers may need or want to consider in assessing the **generalizability** (sometimes denoted **external validity**) of the evidence. Decision makers may take into account how well the evidence maps to the patient populations, settings, diseases or conditions, interventions, comparators, and outcomes which are most relevant to their decisions. Because these factors differ with the perspective of decision makers, **we have chosen to consider applicability as separate from but related to strength of evidence.**
- ***CERs should record information about applicability for the outcomes and comparisons for which they specify an overall strength of evidence rating. As described in Chapter 6, EPCs should summarize this information in a separate table that decision makers can review along with the strength of evidence table.***

Agency for Healthcare Research and Quality (AHRQ):  
Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, 2008

# Eingeschränkte Eignung von ausschließlich RCT zur *post-hoc* Bewertung von unerwünschten Effekten

Weitere Gründe, weshalb Informationen zu unerwünschten Effekten aus RCT nicht generell als zuverlässig eingeschätzt werden:

- Literatur-Datenbanken unvollständig (*Derry 2001*)
- Reporting nicht selten inkomplett (*z.B. Ioannidis 2001*)
- Standardisierung des Berichtens, der Begriffe (Codierung!) nicht immer gegeben (Standards bei Pharma-Studien hier i.d.R. sehr hoch!)
- übliche „Effektmaße“ für UE nicht immer nützlich (rel. Risiko vs. absolutes Risiko) – (*van Staa 2008*)
- Onset- / Offset-Phänomene in RCT kaum abbildbar (*van Staa 2008*)
- (multivariate) Verteilung der unerwünschten Effekte und der erwünschten Effekte über die Patientenpopulation kaum erkennbar, nur „marginale“ Effekte (*van Staa 2008*)
- **außerdem:** Wahrscheinlichkeit für subjektiven Selektions-Bias ist bei unerwünschten Effekten in nicht-randomisierten und/oder nicht-verblindeten Studien deutlich reduziert (im Vergleich zu Wirksamkeits-Endpunkten bei diesen Designs)

# Evidenzhierarchie(n)

- Einige Autoren gehen so weit, dass sie für Aussagen zu unerwünschten Effekten oder generell für Bewertungen, bei denen es nicht um den a priori geplanten Beweis einer Hypothese geht, eine Umkehrung der üblichen „Evidenzhierarchie“ für Wirksamkeitsnachweise vorschlagen.
- Andere räumen ein, dass „Evidenz-Hierarchien“ nicht zielführend sind, wo z.T. widerstrebende Anforderungen basierend auf einer möglichst breiten Evidenzbasis zu beantworten oder vergleichend zu bewerten sind:

## Benefits and Risks of Drug Treatments

How to Combine the Best Evidence on Benefits  
With the Best Data About Adverse Effects

Vandenbroucke, MD, PhD  
Vandenbroucke, MD, PhD

the new drug. The sample size precludes  
detection of adverse effects unless they occur a  
certain number of person-years.

What is the best evidence for determining harms  
of medical treatment?

Vandenbroucke

Comparison of evidence on harms of medical interventions  
in randomized and nonrandomized studies

Panagiotis N. Papanikolaou, Georgia D. Christidi, John P.A. Ioannidis

Observational Versus Experimental Studies: What's the  
Evidence for a Hierarchy?

John Concato

Observational Research, Randomised  
Trials, and Two Views of Medical Science

Jan P. Vandenbroucke

RANDOMIZED, CONTROLLED TRIALS, OBSERVATIONAL STUDIES,  
AND THE HIERARCHY OF RESEARCH DESIGNS

JOHN CONCATO, M.D., M.P.H., NIRAV SHAH, M.D., M.P.H., AND RALPH I. HORWITZ, M.D.

# Evidenzhierarchie(n)

Randomized trials and observational studies: still mostly similar results, still crucial differences

Regina Kunz\*

*Basel Institute for Clinical Epidemiology, University Hospital Basel, Hebelstrasse 10, 4031 Basel, Switzerland*

Accepted 21 May 2007

Although methodologists have long recognized that randomized trials provide stronger evidence than nonrandomized observational studies, for management decisions in which randomized trials are unavailable, clinicians and guideline developers must still rely on observational studies. Indeed, for a great many questions of clinical importance, observational studies constitute the best available evidence. In these situations, clinicians, guideline developers, and decision makers rightly expect guidance from methodologists in the use and interpretation of those studies and their integration into guideline recommendations.

## Evidenzhierarchie(n)

**It is often argued that observational data are not ‘robust’ enough for regulatory decision-making and risk assessment and that observational data should only be used if causality is firmly established and that one should wait until evidence of highest quality appears.**

**Unfortunately, RCTs are rarely conducted to specifically study safety concerns.**

von Staa et al. Int J Epi, 2008

**The observational data are a better reflection of the frequency of harms experienced by patients in actual practice. Clinical trial data and observational evidence really complement each other.**

Vandenbroucke & Psaty JAMA 2008

# Evidenz in vergleichenden Nutzenbewertungen

## Core Principles on relative effectiveness

### Good practice principles for relative effectiveness assessment

...

6. **RE assessments should be capable of addressing transparently uncertainty in the evidence base**, and the methodological challenge of translating evidence on relative efficacy and other appropriate available data into conclusions on relative effectiveness.
7. The sources of evidence which are to form the relevant RE input should be specifically discussed among the identified key stakeholders, who should each be able to submit evidence or argumentation for appraisal.

...

EC Pharmaforum, Relative Effectiveness WG, April 2009

# Evidenz in vergleichenden Nutzenbewertungen

- zahlreiche (weitere) Definitionen und Guidances (US, Canada, UK; Cochrane, NICE, AHRQ) – *siehe auch back-up Slides* -
- „Conceptualizing and Combining Evidence for Health System Guidance“ (Canadian Health Services Research Foundation, May 2005):  
**„Evidence is inherently uncertain, dynamic, complex, contestable, and rarely complete.“**
- ein letztes Zitat (ISPOR retrospective database analysis task force- Part I, 2009):  
**„Different types of policy decisions may present different tradeoffs in the tension between the quality of available evidence and the need to make a decision.**  
**At one end of the spectrum are regulatory decisions such as drug approval. ... At the other end of the spectrum, the risk of acting ‚too late‘ may look greater than the risk of acting on imperfect evidence.“**





**Welche Empfehlungssicherheit  
erforderlich und erreichbar?**

# Entscheiden unter „Unsicherheit“

- Welche „Unsicherheit“ akzeptieren **die Entscheider** im Gesundheitssystem, konkret: der **G-BA** als Auftraggeber von Nutzenbewertungen?

## Welche Möglichkeiten nutzt der G-BA, um mit fehlender oder schwacher Evidenz umzugehen?

### Methodisches Vorgehen

- Analyse der Endberichte bzw. der Beschlussbegründungen von Entscheidungen zur Kostenübernahme nach der Methodenbewertung
- im Bereich der vertragsärztlichen Versorgung
- im Zeitraum vom 1. Januar 1998 bis 31. Oktober 2008
- Fokus jeweils auf:
  - Wie lautet die Entscheidung zur Kostenübernahme?
  - Auf welchem Evidenzniveau wurde die Entscheidung getroffen?

# G-BA: Entscheidet unter „Unsicherheit“!

## Ergebnisse

**21 von 36 (58.3%) Beschlüssen ohne Evidenz 1**

➤ Analyse von insgesamt 36 Beschlüssen der Richtlinie  
Methoden der vertragsärztlichen Versorgung

- 12 positive Kostenübernahmeentscheidung (Anlage I)
  - bei fünf Beschlüssen keine Unterlagen der Evidenzstufe 1 vorhanden

**→ 5 von 12 (41.7%) der pos. Entsch. ohne Evidenz 1**

- 22 negative Kostenübernahmeentscheidung (Anlage II)
  - bei 16 Beschlüssen keine Unterlagen Evidenzstufe 1 vorhanden

**→ 16 von 22 (72,7%) der neg. Entsch. ohne Evidenz 1**

- zwei Entscheidung Kostenübernahme ausgesetzt (Anlage III)
  - bei beiden Beschlüssen Unterlagen der Evidenzstufe 1 vorhanden

<http://www.g-ba.de/informationen/beschluesse/zur-richtlinie/7/>

**FAZIT: Der G-BA als Auftraggeber akzeptiert „unsicherere“ Evidenz und entscheidet, auch wenn Aussagen auf der Evidenzstufe 1 fehlen.**

# Sicherheit der Empfehlungen

Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan A07-01    Version 1.0  
ICS/LABA-Fixkombinationen bei Asthma bronchiale – Ergänzungsauftrag    19.11.2007

Dr. Norbert Banik:

Zweitens kommt immer wieder das Argument hinzu, das ich vorhin schon sagte, in dem konkreten Fall haben wir keine andere Evidenz. Dann möchte ich eine mich irgendwie überzeugende Argumentation hören, warum wir dann die vorhandene Evidenz nicht nehmen wollen und immer wieder darauf verwiesen werden, ihr hättet es ja machen können. Wir stehen aber heute hier am 31. August 2007 und haben diese Daten nicht, und Sie wollen eine Bewertung machen.

Moderator Prof. Dr. Peter T. Sawicki: Richtig. Vielleicht kann ich Ihnen darauf antworten, weil es um eine allgemeine Frage geht, die auch auf andere Bereiche der Institutsarbeit ausstrahlt. Wenn man nicht eine Frage aufgrund valider Daten beantworten kann, dann muss man das so sagen. Man ist ja nicht gehalten, mit schlechten Daten eine Frage zu beantworten, sondern die Frage muss so beantwortet werden, dass sie eine ausreichende Verlässlichkeit zulässt. Wenn man das nicht kann, weil zuverlässige Daten nicht vorliegen, dann geht man nicht auf die Ebene von schlechten Daten, sondern man sagt, man kann es nicht beantworten. Das zu Ihrem zweiten Teil der Frage.

# Sicherheit der Empfehlungen

Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan A07-01      Version 1.0  
ICS/LABA-Fixkombinationen bei Asthma bronchiale – Ergänzungsauftrag      19.11.2007

Moderator Prof. Dr. Peter T. Sawicki: Wir wollen sichere Empfehlungen an den Gemeinsamen Bundesausschuss weitergeben und nicht unsichere Empfehlungen. Man könnte natürlich - das wäre eine mögliche Methodik - Empfehlungen abgeben und sagen, dass die unsicher sind und es auch anders sein könnte. Diesen Weg wollen wir nicht beschreiten, und zwar vor dem Hintergrund, dass wir mehr an Versorgungsforschung in Deutschland in Zukunft fordern wollen und fordern. Wir glauben nicht, dass der Weg ist, schlechtere Evidenz bei der Versorgung unserer Patienten zu berücksichtigen, sondern wir glauben, der Weg ist, bessere Evidenz zu generieren. Vor diesem Hintergrund werden wir uns nicht auf schlechte Daten verlassen und darauf eine Entscheidung über die Patientenversorgung in Deutschland aufbauen und dies dem Gemeinsamen Bundesausschuss empfehlen, sondern - im Gegenteil - wir werden sagen, wir haben hier keine verlässliche Evidenz, weswegen wir fordern, dass wir in diesem Bereich möglichst schnell - das wird zum Teil auch umgesetzt - zuverlässige Studien durchführen. Das ist nicht unbedingt die Aufgabe der Hersteller. Das kann und muss meiner Meinung nach die Aufgabe der gesetzlichen Versicherungssysteme sein.

## „Unsicherheit“ ...

- **Zusammenfassung: Unsicherheit ist inhärent; kritischer Umgang mit Unsicherheit erforderlich; Abwägen zwischen „Relevanz“ und „Verzerrungsrisiko“!**

Greater openness about uncertainty would challenge the prevailing health care culture of uncontrolled experimentation on the many and controlled experimentation on the few. Reliable answers to many important therapeutic questions require large scale randomised controlled trials.<sup>9</sup> Acknowledging uncertainty is a prerequisite for such trials, and the “uncertainty principle” can be used to simplify trial entry criteria and make large trials possible. According

Alderson/Roberts BMJ 2000

Now and then, clinicians will have to accept and explain that uncertainty is an inherent facet of the uniqueness of human nature. Evidence helps to quantify that uncertainty, but cannot remove it.

Summerskill Lancet 2005

- We need better syntheses of evidence about the effects of interventions to influence the wider determinants of health
- Some have questioned whether selecting evidence according to a rigid, unidimensional hierarchy based on study design—for example, only including randomised controlled trials—is appropriate in this field
- We lack an accepted, evidence based methodology for selecting useful evidence for inclusion in evidence synthesis.

Ogilvie et al. J Epi Comm H 2005

# „Unsicherheit“ ...

- Motto des IQWiG – Allg. Methoden 1.0 bis 3.0:

*Eine Hauptursache der Armut in den Wissenschaften ist meist eingebildeter Reichtum. Es ist nicht ihr Ziel, der unendlichen Weisheit eine Tür zu öffnen, sondern eine Grenze zu setzen dem unendlichen Irrtum.*

Bertolt Brecht. Leben des Galilei. Frankfurt: Suhrkamp. Uraufführung, erste Version 1943,

- Wie kann man dem unendlichen Irrtum am besten eine Grenze setzen?  
In dem man ....  
nur die „besten Früchte“ auswählt – oder in dem man einbezieht, was es wertvolles gibt?





# **Vergleichende Nutzen-Schaden- Bewertung**



# Vergleichende Nutzen-Schaden-Bewertung

- Bemühungen, die „klassischen risk-benefit-analyses“ transparenter und – soweit und wo möglich – auch quantifizierbar zu machen, sind seit über 10 Jahren vorhanden:
  - CIOMS IV 1998: Benefit-Risk Balance for Marketed Drugs: Evaluating Safety Signals
  - FDA *Critical Path Initiative* („*Innovation vs. Stagnation*“), ab 2004
  - EMEA *New Safe Medicines Faster*, ab 1999
  - Methodische Workshops DIA, Int. Soc. Pharmacoeconomics...
- Steht im Zusammenhang mit Überlegungen zu „signal generation“, „signal detection“, „risk management“ – EMEA Guidelines und points to consider zu den Themen sind seither erschienen
- Thema umfaßt: „benefit evaluation“, „risk evaluation“, „benefit-risk evaluation“ (BRE; quantification of BR) und „options analysis“
- Dies spiegelt das bis dato typische Vorgehen wider:
  - separate und „marginale“ Analyse und Zusammenfassung der „erwünschten Effekte“ sowie der „unerwünschten Effekte“
  - danach Versuche, diese beiden marginalen Maße zu aggregieren (unterschiedliche Vorschläge)

# Vergleichende Nutzen-Schaden-Bewertung

- ~~TWiST / QTWiST~~
  - ~~survival weighted risk-benefit difference~~
  - QALY-Modellierungen
  - Principle of Three
  - TURBO
  - Ratio
  - Multicriteria Decision Analysis (MCDA)
  - global benefit-risk ratio (GBR)
  - NNTH / NNTB
  - Benefit-Less-Risk-Measure
  - Benefit-Risk-Index (Therapeutischer Index)
  - Incremental Net Health Benefit
  - Minimum-Clinical Efficacy (MCE)
  - Stated Preferences for Risk Benefit
- } um diese Maße anwenden zu können, prospektive Erhebung erforderlich

# Vergleichende Nutzen-Schaden-Bewertung

- Hier (nach der Zulassung): keine „akute“ Bewertungssituation wegen eines neuen „Signals“; bei Nutzenbewertung keine individualisierten Bewertungen sondern auf Populationsniveau (im Unterschied zu Zulassungsbehörden)
- Bisher sind dies nur Bewertungen von Einzelsubstanzen, aber Quantifizierungsbemühungen der BRE stehen im Zeichen, vergleichen zu können.
- Wie ist jetzt die Situation bei vergleichender Nutzenbewertung?

„Über eine Nutzen-Schaden-Abwägung im Sinne einer **Gegenüberstellung** von Nutzen- und Schadenpotenzial hinaus besteht die Möglichkeit, **Nutzen und Schaden vergleichend zu gewichten**. In diesem Fall würden die Aussagen des IQWiG für **jeden einzelnen patientenrelevanten Endpunkt** durch die Gewichtung von Nutzen und Schaden z. B. unter Verwendung eines Summenscores berichtet.“

- IQWiG: siehe Allg. Methoden 3.0, S.38: „Gegenüberstellung“; „vergleichend zu gewichten“ (? ist evtl. gemeint: *gewichtet zu vergleichen?*); **je Endpunkt!**

# Vergleichende Nutzen-Schaden-Bewertung

- Relevant ist auch die Frage der Reihenfolge der Aggregation:
  1. zuerst erwünschte Effekte und unerwünschte Effekte je Intervention feststellen, daraus Nutzen-Schaden-Relation je Intervention bestimmen und abschließend diese vergleichend bewerten?

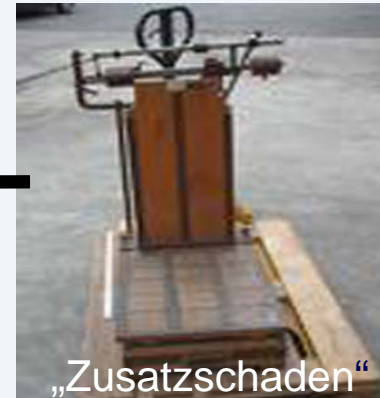
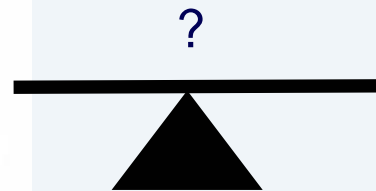


- oder

2. zuerst zusätzliche erwünschte Effekte und zusätzliche unerwünschte Effekte feststellen, daraus „Zusatz-Nutzen-Schaden-Relation“ ableiten?



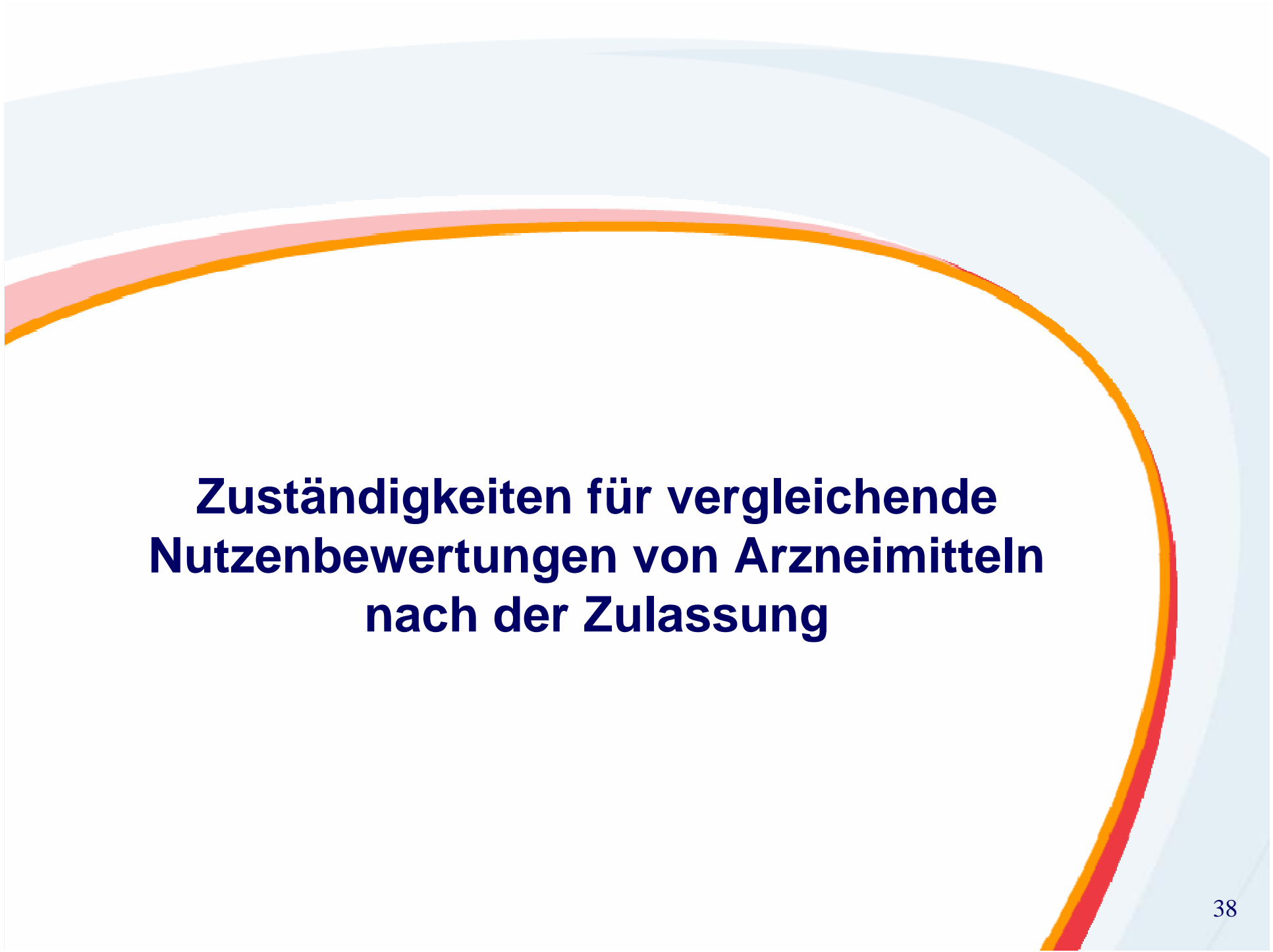
„Zusatznutzen“



„Zusatzschaden“

# Vergleichende Nutzen-Schaden-Bewertung

- Die Varianten 1 und 2 sind nur dann theoretisch identisch, wenn die Populationen, in denen Intervention A und Intervention B zur Anwendung kommen, homogen und vergleichbar sind, d.h. es gibt keine „risk-benefit-interaction“
- Was ist eine „risk-benefit-interaction“? Die Nutzen-Schaden-Bewertung geht in gewissen Untergruppen unterschiedlich aus. Problematik ist dann relevant, wenn zusätzlich die Zusammensetzung der Patientenpopulation zwischen den Komparatoren bezüglich dieser Untergruppen unterschiedlich ist!
- In der **Zusatz**-Nutzen-Schaden-Bewertung liegen also noch neue, bisher nicht adressierte oder sogar ungelöste methodische Probleme, die über die eigentlichen (methodischen) Probleme der Nutzen-Schaden-Bewertung hinausgehen



## **Zuständigkeiten für vergleichende Nutzenbewertungen von Arzneimitteln nach der Zulassung**

# Zuständigkeiten für vergleichende Nutzenbewertungen

- Bis zur Zulassung und für laufende Bewertungen nach der Zulassung ist die Zuständigkeit der Zulassungsbehörden für die Bewertung des Risk-Benefit-Ratio (RBR) soweit bekannt (→ u.a. heutiger Vortrag von U. Hagemann)
- AM-Hersteller haben diese Verpflichtungen ebenso!
- Datenbasis und fachliche Voraussetzungen sind bei Zulassungsbehörde optimal für RBR geeignet:
  - Änderungen des RBR sind vom Zulassungsinhaber mitzuteilen;
  - PSUR's;
  - It. GCP-Verordnung und AMG erhält Zulassungsbehörde auch wesentliche Informationen über klinische Prüfungen (Änderungen von Nutzenaspekten);
  - Behörde kann einzelne Studien anordnen (PASS);
  - Aktivitäten mit MAH geplant (Risk-Management Pläne);
  - Signal detection tools zur laufenden Bewertung vorhanden;

# Zuständigkeiten für vergleichende Nutzenbewertungen

Core Principles on relative effectiveness

Good practice principles for relative effectiveness assessment

- 4. RE assessment processes should remain separate from product market authorisation procedures (though this does not mean that they are necessarily performed by different organisations).**

EC Pharmaforum, Relative Effectiveness WG, April 2009



# Zuständigkeiten für vergleichende Nutzenbewertungen

- G-BA/IQWiG mit den bekannten Aufgaben lt. SGB V...
- Die Erkenntnisse der Zulassungsbehörde während des Zulassungsprozesses und schließlich die Zulassungsentscheidung (*RBR favourable*) sollten als a priori Information bei der post-hoc Nutzenbewertung einfließen.
- Enger Informationsaustausch zwischen diesen Institutionen scheint also zumindest sinnvoll und erfolgversprechend
- Die jeweiligen Hoheits- /Aufgaben-Gebiete sind aber auch rechtlich gegeneinander abgegrenzt:

Urteil des BSG vom 31.5.2006, B 6 KA 13/05R, Nr. 71:

*„Im Rahmen der Bewertung von Arzneimitteln an Hand der in der GKV geltenden Maßstäbe ist allerdings - im Unterschied zur Methodenanerkennung - die arzneimittelrechtliche Zulassung eines Wirkstoffs zu beachten, bei der gemäß § 21 Abs 2 AMG **Qualität, Wirksamkeit und medizinische Unbedenklichkeit des Wirkstoffs für die vorgesehenen Indikationen geprüft und abschließend bewertet werden** ( näher BSGE 89, 184, 185 = SozR 3-2500 § 31 Nr 8 S 29 ).*

***Diese Kriterien darf der GBA unter dem Aspekt des "medizinischen Nutzens" eines Arzneimittels oder Wirkstoffs nicht abweichend von der Beurteilung der für die Zulassung nach dem AMG zuständigen Behörde bewerten...**“*

# Zuständigkeiten für vergleichende Nutzenbewertungen

- Frage des Zeitpunktes der vergleichenden Nutzenbewertung von Arzneimitteln:
  - je früher, desto vorhersehbarer kann die Datenbasis nicht unterschiedlich von der der Zulassungsbehörde sein (RCT); Beobachtungsstudien können ja per Def. erst nach der Zulassung durchgeführt werden
  - je früher, desto weniger praktische Erfahrung liegt für die Anwendungsbreite (Population) als auch für eine größere Anwendungsdauer vor
  - je früher, desto schwerer ist es in den meisten Fällen, Zusatznutzen nachweisen zu können, vor allem bei chronischen Erkrankungen (patientenrelevante Endpunkte!)
  - je früher, desto schlechter sind Studien planbar/durchführbar
- Es muß also eine faire Balance aus diesen (und anderen) Einschränkungen und dem Wunsch nach frühzeitiger Beurteilung gefunden werden.

***Evidenz-Vergleich Ist Tatsächlich Anspruchsvoll!***



## **Ausblick: Verbesserungsvorschläge**

# Verbesserungsvorschläge

## 1. bezüglich Studiendesigns / Analyse-Methodik:

- objective observational study design (Rubin 2007)
- Peer reviewed observational studies (Psaty & Vandembroucke 2008)
- cluster randomized~, practical clinical trials (large clinical trials)
- propensity scores, instrumental variable models
- prior event rate ratio (Tannen 2009)

# Verbesserungsvorschläge

TABLE 1 Ten key points established in the Brussels Declaration from the European Summit for Change in Asthma Management

## 2. bezüglich Prozess-/Struktur-Verbesserungen

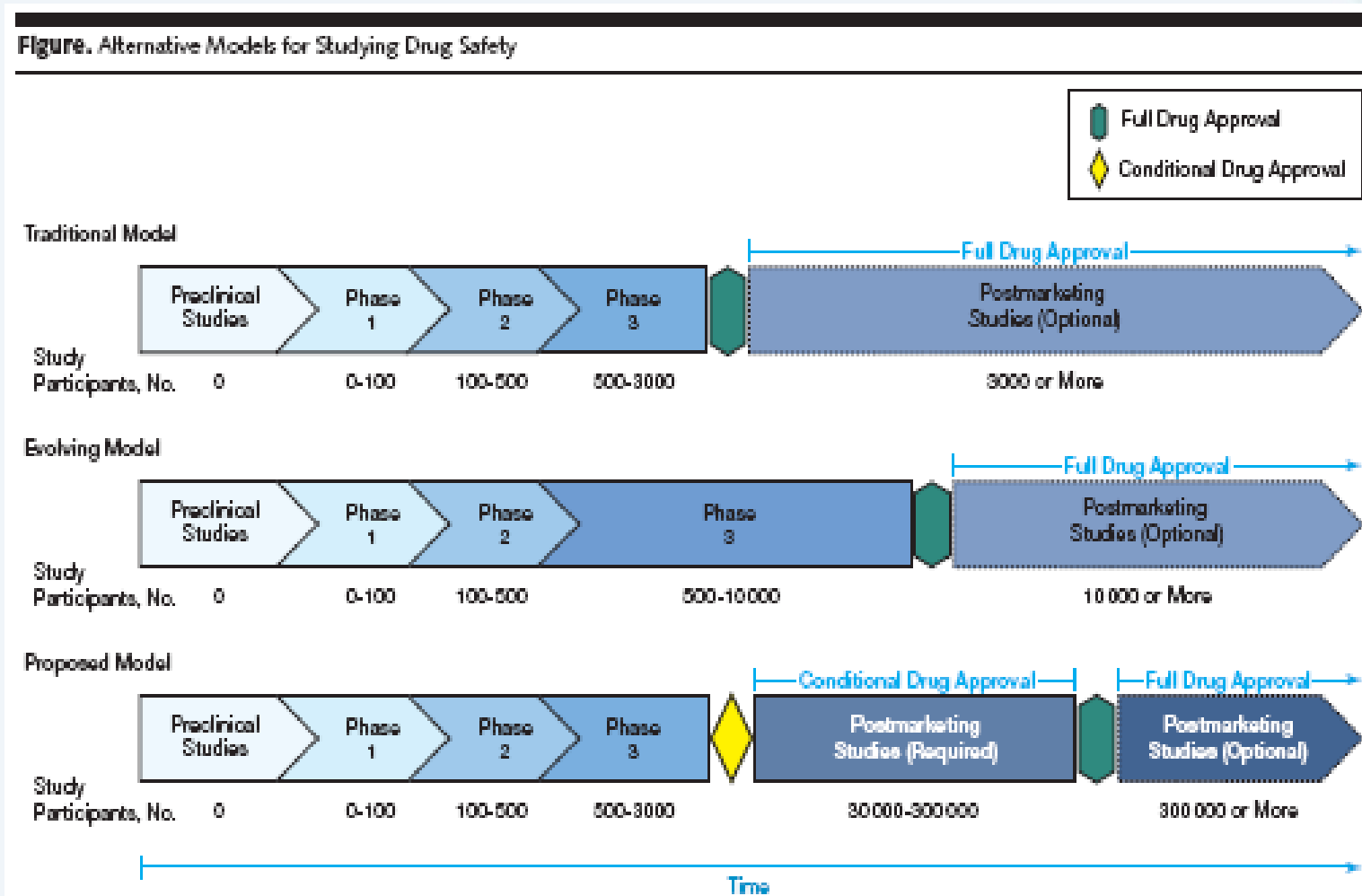
### ● „The Brussels Declaration“

3. Ensure rapid responses to the most current scientific understanding of asthma	Asthma presents differently in adults and children and there are also differences between various ethnic groups. The medical community must communicate the latest research clearly and effectively to allow both clinicians and patients to make informed management decisions and to allow policy makers to shape new asthma-related initiatives.
4. Update the EMEA regulatory guidance notes on asthma	The EMEA guidance note must reflect the understanding of asthma as a respiratory manifestation of systemic inflammatory processes, with the control of inflammation at the core. It must also emphasise the importance of control and recognise a broader range of markers for control beyond FEV <sub>1</sub> , reflecting changes to guidelines. Of particular importance is the development of new paediatric guidelines, which are currently lacking. The guidance note must be reviewed yearly in the future and all stakeholders should be kept abreast of changes.
5. Include evidence from real-world studies in treatment guidelines	Asthma treatment guidelines are primarily based on evidence from large clinical trials that frequently assess lung function as the primary outcome. However, inflammatory biomarker levels, asthma exacerbations and other outcomes may worsen regardless of lung function status. In addition, due to strict inclusion criteria, clinical trial patients are not typical of patients seen in everyday clinical practice, and children are rarely included. The inclusion of real-world data exploring a wider range of patients and end-points should allow clinicians to make more informed treatment decisions.
6. Provide funding for real-world studies	Both the EU and Member States should fund real-world studies conducted in primary-care settings that include patients from different ethnic groups, smokers and people with different asthma phenotypes. Inclusion of patients that are more typical of the average patients seen in the clinic should allow a better assessment of the impact of comorbidities and the development of strategies to promote adherence to treatment.
7. Explore variations in asthma care across Europe	There are inconsistencies in diagnosis, management and outcome in asthma between and within countries in Europe. Surveys show both undertreatment and over-treatment. For example, a survey of 10 European countries with the Finnish Asthma Programme show

Holgate et al. ERJ 2008

# Verbesserungsvorschläge

- conditional approval (Strom 2006)



# Zusammenfassung

## Zusammenfassung (1)

- Nutzenbewertung besteht definitionsgemäß in einer gemeinsamen Bewertung von erwünschten und unerwünschten Effekten
- Diese gipfelt in einer (möglichst quantifizierten) Bewertung der Relation zwischen diesen beiden
- Die Evidenzbasis, aus der die für diese beiden Effekte relevante Quellen identifiziert und extrahiert werden, muss einheitlich sein
- Sie sollte sehr umfassend sein (zumindest nicht a priori auf RCT eingeschränkt); besonders die Bewertung der unerwünschten Effekte und die zu fordernde externe Validität untermauern diese Anforderung
- Unsicherheit der Bewertungen/Empfehlungen wird sich nie beseitigen lassen, diese sollte bewertet, einbezogen und beschrieben werden; „sichere Empfehlungen“ wird es nicht geben können!



## Zusammenfassung (2)

- Abstimmung zwischen AM-Hersteller, Zulassungsbehörde und Nutzenbewertungsinstitution sollte verstärkt werden → zukünftig eventuell ein gemeinsamer Prozeß?
- Verbesserungen des status quo bei Nutzenbewertungen sind vor allem hinsichtlich von Studiendesigns möglich, die helfen können, die Evidenzbasis zu optimieren (höhere externe Validität bei niedrigerer Verzerrungs-Wahrscheinlichkeit und patientenrelevanten Endpunkten)
- Modifikationen beider Elemente – sowohl des „klassischen Zulassungsverfahrens“ (z.B. „conditional approval“ - z.T. ja schon praktiziert), als auch des bisherigen Nutzenbewertungsverfahrens – im Sinne einer „Annäherung“ könnten ebenfalls positive Veränderungen im Bewertungsprozess und der prospektiven Erzeugung einer adäquate(re)n Evidenzbasis bewirken



**Back-up**

# Vergleichende Nutzenbewertung

## Draft Definition of Comparative Effectiveness Research for the Federal Coordinating Council

Comparative effectiveness research is the conduct and synthesis of systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions. **The purpose of this research is to inform patients, providers, and decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.** To provide this information, comparative effectiveness research must assess a comprehensive array of health-related outcomes for diverse patient populations. Defined interventions compared may include medications, procedures, medical and assistive devices and technologies, behavioral change strategies, and delivery system interventions. **This research necessitates the development, expansion, and use of a variety of data sources and methods to assess comparative effectiveness.**

US Dept. Health & Human Services, Federal CER Council, May 2009

# Vergleichende Nutzenbewertung

## 8.1. Key Points (Harms)

- Assess all important harms, whenever possible.
- Use multiple sources of information to identify harms.
- Gather evidence on harms from a broad range of sources, including observational studies, particularly when clinical trials are lacking; when generalizability is uncertain; or when investigating rare, long-term, or unexpected harms.
- Do not assume studies adequately assess harms because methods used to assess and report benefits are appropriate; rather, evaluate how well studies identify and analyze harms.
- Avoid inappropriate combining of data of harms, and thoroughly investigate inconsistent results.
- Be cautious about drawing conclusions on harms when events are rare and estimates of risk are imprecise. When describing evidence on harms, avoid using terms implying causality when causality is uncertain.
- Do not draw conclusions about equivalence and non-inferiority for harms unless there is appropriate data justifying such statements.
- Avoid assuming that class effects on harms are present for two or more interventions unless there is a strong pathophysiologic rationale as well as supporting clinical outcomes data (such as similar estimates of risk). Include analyses of inconsistency when combining data on harms from two or more interventions.
- Avoid implicit indirect comparisons when judging comparative risk of adverse events. Rather, evaluate whether different sets of trials meet assumptions for similarity of treatment effects, and if so, perform “formal” indirect comparisons if possible.

Agency for Healthcare Research and Quality:  
Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews, 2008