

# Konfirmatorische Nutzenbewertung? HTA meets EMA - Methodische Spannungsfelder aus Sicht der Industrie

M Hennig (GlaxoSmithKline), D Knoerzer (Roche), K Kupas (Bristol-Myers Squibb), F Leverkus (Pfizer), C Nicolay (Eli Lilly)

IQWiG im Dialog, 12. Juni 2015

# Disclaimer

- Die im Folgenden dargestellten Ansichten sind diejenigen der Autoren.
- Diese Ansichten müssen nicht notwendigerweise mit denen der Firmen übereinstimmen, bei denen die Autoren beschäftigt sind.

# Übersicht

- **EMA und HTA – 2 Welten**
  - Ziel, Endpunkte, Methodik
- **Ist die frühe Nutzenbewertung confirmatorisch?**
- **Spannungsfelder**
  - Evidenz
  - Subgruppenanalyse
- **Lösungsvorschläge**
- **Zusammenfassung**

# Anforderungen der Zulassung: Elemente eines prospektiven konfirmatorischen Studiendesigns

- **Ziel**
  - Zulassung erfolgt primär auf Basis eines günstigen Nutzen-Risiko-Verhältnisses
    - unter Berücksichtigung der Qualität, Wirksamkeit und Unbedenklichkeit
  - ICH Guidelines und CHMP „Points to consider“: Vorgaben für das Design der Studie, Auswahl der Endpunkte und Analysen
- **Endpunkte**
  - primärer Endpunkt → konfirmatorische Aussage
  - sekundäre Endpunkte → explorativ oder supportiv
  - Fallzahlschätzung basiert auf primärem Endpunkt
- **Methodik**
  - *prospektive* Studien
  - (statistisch) signifikante Überlegenheit oder Nicht-Unterlegenheit
  - a priori geplant, konfirmatorisch, auf der Basis primärer Zielgröße(n) mit Hypothesentest (ICH-E9)

# Elemente der Evidence-based Medicine (EbM)

- **Ziel**  
Verbesserung der wissenschaftlichen Grundlagen für Entscheidungen im Gesundheitssystem. Dieses Ziel wird vor allem durch die Erstellung, Aktualisierung und Verbreitung systematischer Übersichtsarbeiten angestrebt.
- **Endpunkte**
  - gleichwertig
- **Methodik**
  - *retrospektive* Analyse vorhandener Studien
  - Literaturrecherche und Meta-Analysen
  - **PICO-Kriterien** für systematische Reviews  
(**P**opulation - **I**ntervention - **C**ontrol (**C**omparator) – **O**utcome)

Quelle: Cochrane Collaboration (<http://www.cochrane.org/>)

# EMA und HTA – Zwei Welten

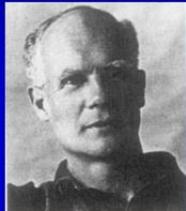
## EMA, Klinische Prüfungen

### „Neyman-Pearson“-Welt

- P-Wert => dichotome Entscheidung
- Fokus auf formalem, konfirmatorischen Nachweis der Wirksamkeit
- Fokus auf primärem(n) Endpunkt(en) – konfirmatorisch  
Sekundäre Endpunkte explorativ / unterstützend
- $\alpha$ -/ $\beta$ -Fehler kontrolliert



Jerzy Neyman

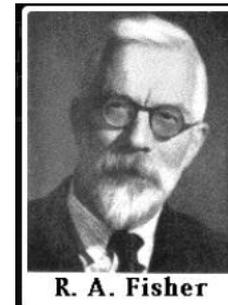


Egon Pearson

## EbM, HTA

### „Fisher“-Welt

- P-Wert => Grad der Evidenz gegen die Nullhypothese
- Konfirmatorik spielt keine Rolle
- Kein Unterschied zwischen primären und sekundären Endpunkten
- $\alpha$ -/ $\beta$ -Fehler nicht kontrolliert



R. A. Fisher

# Zulassung und HTA – Forderung an Evidenz

## Zulassung

- schaut sich Studien einzeln an, Meta-Analysen eher supportiv
- Replikation der Ergebnisse
  - „Ich brauche zwei einzelne pivotale Studien“



London, 31 May 2001  
CPMP/EWP/2330/99

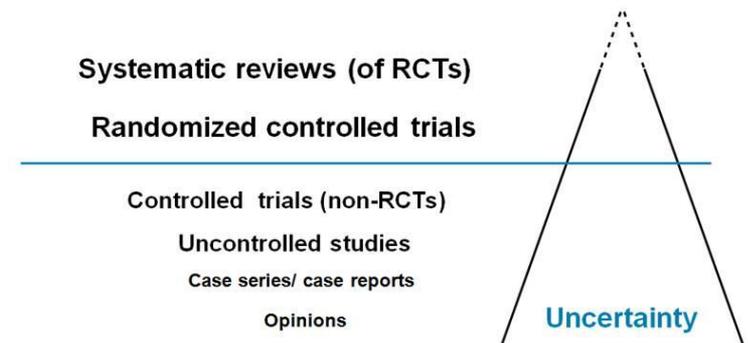
COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS  
(CPMP)

POINTS TO CONSIDER ON APPLICATION WITH  
1. META-ANALYSES; 2. ONE PIVOTAL STUDY

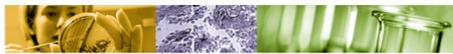
## HTA

- Meta-Analyse gefordert, wenn  $\geq 2$  Studien vorhanden sind, die sich kombinieren lassen  
→ Höchste Evidenz
- Grad der Evidenz
  - 1A (Meta-Analyse) besser als 1B (RCT)

## Reliable results – Hierarchy of evidence



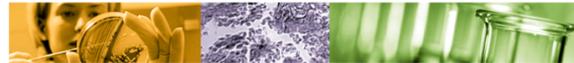
Pharma Pricing & Market Access Outlook Europe 2011, Germany: Latest Developments in Pricing and Reimbursement Policy



## Konfirmatorische Nutzenanalyse unter besonderer Berücksichtigung des Zwei-Studien-Prinzips

**IQWiG im Dialog**  
**Köln, 30.05.2008**

Friedhelm Leverkus, Julia Schiffner, Hans-Jürgen Lomp,  
Marlis Herbold  
**VfA UA Biostatistik**



### Zwischenfazit (2)

Konzepte der Zulassung unterscheiden sich von denen der Nutzen-Bewertung :

- Zulassung:
  - Behörde bewertet und entscheidet über benefit-risk-ratio
  - Erstmaliger Nachweis der Wirksamkeit
  - Dichotome, konfirmatorische Entscheidung
- Nutzenbewertung:
  - Institut bewertet Zusatznutzen von zugelassen Medikamenten
  - Beschreibung der Effektgröße und der statistischen Unsicherheit
  - Bewertung der Stärke der Evidenz
  - p-Werte allenfalls deskriptives Maß für die Evidenz gegen  $H_0$

# Interpretation der p-Werte: Ist die frühe Nutzenbewertung nach AMNOG confirmatorisch?

- Es ist **kein Entscheidungsproblem**  
→ Das Medikament ist ja bereits zugelassen und wird damit auch erstattet
- Es ist **ein Schätzproblem**  
→ Es soll das Ausmaß und die Wahrscheinlichkeit des Zusatznutzens gegenüber ZVT bezüglich patienten-relevanter Endpunkte dargestellt werden, als Basis für die Preisverhandlungen

# Spannungsfelder

# Gesamte Evidenz vs. bestmögliche (höchste) Evidenz

- **Ganze Studien**
  - Evidenz aus Studien, die nicht gegen die ZVT testen, wird nicht berücksichtigt
  - Studienprogramm vs. Finales Label
  - Zulassung berücksichtigt auch Nicht RCT-Evidenz (z.B. Phase 2, conditional approval)
- **Teile von Studien**
  - Studienpopulation vs Finales Label
  - Studienpopulation vs Deutsche Zulassung (hinsichtlich der ZVT / Standard of Care)
- **Endpunkte**
  - Einige Endpunkte, die mit den Zulassungsbehörden zum Nachweis der Wirksamkeit vereinbart wurden, werden nicht anerkannt (z.B. primärer Endpunkt HbA1c bei Diabetes, PFS in der Onkologie)
  - Patientenrelevanz
  - Validität von Surrogatendpunkten

# Subgruppen (G-BA)

*„... Ergebnisse von Subgruppenanalysen können die Identifizierung von Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen unterstützen. ...“*

**Effektmodifikatoren** = klinische Faktoren, die die Effekte beeinflussen, z.B.

- direkte Patientencharakteristika (Subgruppenmerkmale)
- Spezifika der Behandlungen (z. B. die Dosis)

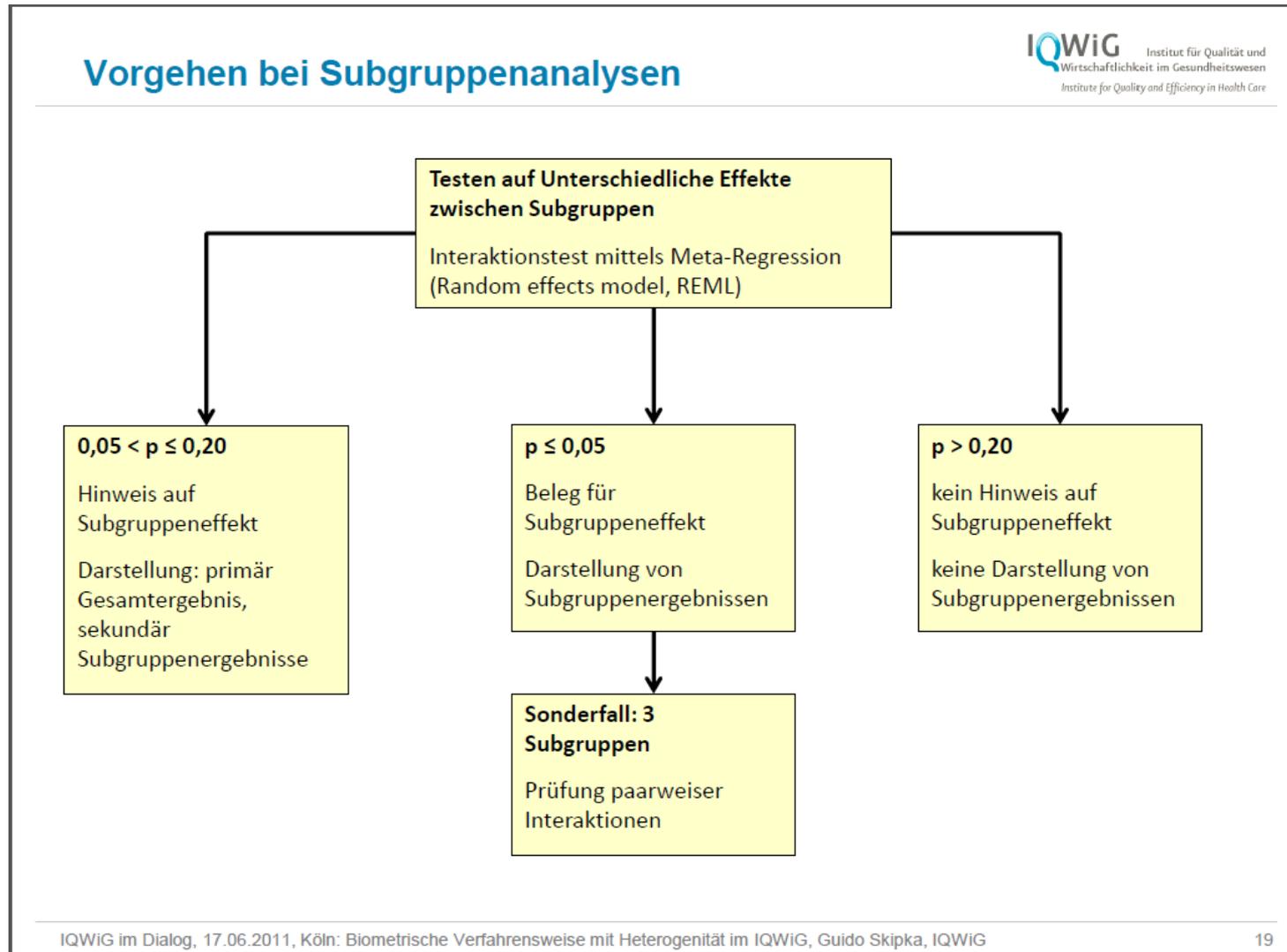
**Ziel:** Aufdeckung möglicher Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika

**Methode:** Homogenitäts- bzw. Interaktionstests

Insbesondere folgende Faktoren sollen bezüglich einer möglichen Effektmodifikation berücksichtigt werden:

- Geschlecht
- Alter
- Krankheitsschwere bzw. –stadium
- Zentrums- und Ländereffekte
- Prä-spezifizierte Subgruppen und Stratifikationskriterien

# Subgruppenanalyse gemäß IQWiG-Methodik



# Aussagekraft von Subgruppenanalysen - 10 Kriterien -

## Was ist erfüllt?

- ✓ *a priori* spezifiziert
- ✓ SG-Variable =  
Baseline Variable
- ✓ SG-Variable =  
Stratifizierungsfaktor für  
Randomisierung
- ✓ Interaktionstest statistisch  
signifikant ( $p < 0,05$ )

## Raum für Verbesserung

- Multiple signifikante Interaktionen:  
Unabhängig voneinander?
- Anzahl der Subgruppen ( $\leq 5$ )

### SG-Effekte: Kontext

- Richtung korrekt präspezifiziert?
- Konsistent mit Evidenz vorheriger Studien?
- Konsistent über mehrere Endpunkte hinweg?
- Gibt es unterstützende, indirekte Evidenz,  
z.B. biologische Plausibilität, Labortests,  
Tierstudien?

Quelle: Sun X, Briel M et al. BMJ 2012; 314:e1553

# Simulationsstudie

- **Annahmen**
  - 1 RCT, Endpunkt: time-to-event → HR
  - Fallzahl: HR=0,85;  $\alpha=0,05$ ;  $1-\beta=80\%$  → N=597 pro Arm
- **Simulation (10.000 Wiederholungen)**
  - Simulation von einer Studie mit N=597 pro Arm
  - Simulierte Behandlungseffekte
    - (1) kein Zusatznutzen → HR=1**
    - (2) beträchtlicher Zusatznutzen → HR=0.85**
  - Simulierte Subgruppeneffekte
    - (1) kein Subgruppeneffekt**
  - Anzahl der Subgruppen: 10 binäre Subgruppen  
Das Verhältnis der Untergruppen innerhalb der Subgruppen ist 50% zu 50%.
  - Test auf Subgruppeneffekt: Interaktionstest  
( $p < 0,05$  → Beleg für Interaktion Effekt \* Subgruppe)

# Simulationsstudie Szenario 1:

## Kein Zusatznutzen - Kein Subgruppeneffekt

Ausmaß des ZN	Haupteffekt
“schlechter“	2,33%
kein	<b>95,37%</b>
gering	2,15%
beträchtlich	0,15%
erheblich	0,00%

Ausmaß des ZN in Subgruppen: Ausprägung 1 / Ausprägung 2 z.B. (männlich / weiblich)		Subgruppen (Wahrscheinlichkeit für mind. ein Auftreten)
W'keit Interaktionstest $p < 0,05$		<b>40,66%</b>
Bei $p < 0,05$		
Kein ZN	Kein ZN	15,11%
<b>Kein ZN</b>	<b>“schlechter“</b>	<b>13,27%</b>
Geringer ZN	“schlechter“	1,03%
Geringer ZN	Kein ZN	9,35%
Beträchtlicher ZN	“schlechter“	0,24%
Beträchtlicher ZN	Kein ZN	4,33%
Beträchtlicher ZN	geringer ZN	-
Erheblicher ZN	“schlechter“	-
Erheblicher ZN	kein ZN	0,08%
Erheblicher ZN	geringer ZN	-
Erheblicher ZN	Beträchtlicher ZN	-

# Simulationsstudie Szenario 2:

## Beträchtlicher Zusatznutzen - Kein Subgruppeneffekt

Ausmaß des ZN	Haupteffekt
“schlechter“	0,00%
kein	20,75%
gering	30,49%
<b>beträchtlich</b>	<b>46,55%</b>
erheblich	2,21%

Ausmaß des ZN in Subgruppen: Ausprägung 1 / Ausprägung 2 z.B. (männlich / weiblich)		Subgruppen (Wahrscheinlichkeit für mind. ein Auftreten)
W'keit Interaktionstest $p < 0,05$		<b>41,01%</b>
Bei $p < 0,05$ :		
Kein ZN	Kein ZN	0,53%
Kein ZN	“schlechter“	0,04%
Geringer ZN	“schlechter“	0,01%
Geringer ZN	Kein ZN	3,31%
Beträchtlicher ZN	“schlechter“	0,02%
Beträchtlicher ZN	Kein ZN	24,50%
Beträchtlicher ZN	geringer ZN	-
Erheblicher ZN	“schlechter“	-
Erheblicher ZN	kein ZN	14,01%
Erheblicher ZN	geringer ZN	0,57%
Erheblicher ZN	Beträchtlicher ZN	0,09%

# Simulation - Zusammenfassung

- Liegt in Wirklichkeit **kein Zusatznutzen** vor:  
Gleiche Anzahl von „schlechter“ und Nutzen („besser“)  
→ Nullsummenspiel
- Liegt in Wirklichkeit ein **beträchtlicher Zusatznutzen** vor, wird in ca. 30% der Bewertung die Population mit Zusatznutzen fälschlicherweise eingeschränkt!  
→ kein Nullsummenspiel, sondern Verlustspiel
- Vielen Dank an Johannes Hengelbrock für die Simulation!\*

\* Publikation geplant

# SG-Analysen: Was tun? - Lösungsvorschläge

- **System-Ebene**
  - Methodische Problematik bei Entscheidern adressieren
  - Einbeziehung des IQWiG in die frühe Beratung + Dossierberatung
- Untersuchung der **Stabilität der Subgruppen-Effekte**
  - Über Endpunkte hinweg
  - Verschiedene Operationalisierung der Subgruppen nutzen/untersuchen
  - Biologische Plausibilität stärker berücksichtigen
- **Weiterentwicklung der Methodik**
  - Interaktionstest geeignet bei klinischen Prüfungen mit geringer Anzahl von Subgruppen
  - Bei Vielzahl von Endpunkten mit Vielzahl von Subgruppen andere Gebieten geeignete Methoden?  
z.B. schrittweise Regressionsverfahren, multivariate Modelle, CART, .....

# Allgemeine Zusammenfassung

- Neyman-Pearson und Fisher sind legitime Ansätze
- Nutzenbewertung ist nicht konfirmatorisch
- Gesamtevidenz sollte berücksichtigt werden
- Vielzahl von Subgruppen -Analysen kann fälschlicherweise zu einer Einschränkung der Population mit Zusatznutzen führen
- Stärkere Betrachtung der Stabilität und biologischen Plausibilität
- Untersuchung der Tauglichkeit von ergänzenden methodischen Ansätzen
- Einbeziehung des IQWiG und der Zulassungsbehörden in eine frühe, gemeinsame Beratung mit dem G-BA

# Literatur (1/2)

- Brookes ST, Whitely E, Egger M et al.** Subgroup analysis in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. *J Clin Epi* 2004; 57:229-236
- Gonnermann A, Kottas M, Koch A.** Biometrische Entscheidungsunterstützung in Zulassung und Nutzenbewertung am Beispiel der Implikationen von heterogenen Ergebnissen in Untergruppen der Studienpopulation. *Bundesgesundheitsbl* 2015; 58:274-282
- Hager W.** The statistical theories of Fisher and of Neyman and Pearson: A methodological perspective. *Theory & Psychology* 2013; 23(2): 251–270
- Kaiser T, Vervölgyi V, Wieseler B.** Nutzenbewertung von Arzneimitteln. *Bundesgesundheitsbl* 2015; 58:232-239
- Lehman EL.** The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two? *JASA* 1993; 88:1242-1249
- Lenhard J.** Models and Statistical Inference: The Controversy between Fisher and Neyman–Pearson. *Brit. J. Phil. Sci.* 2006; 57:69–91
- Paget MA, Chuang-Stein C, Christine Fletcher C, Reid C.** Subgroup analyses of clinical effectiveness to support health technology assessments. *Pharmaceut. Statist.* 2011; DOI: 10.1002/pst.531 ([wileyonlinelibrary.com](http://wileyonlinelibrary.com))

# Literatur (2/2)

- Rothwell PM.** Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005; 365:176-86
- Sun X, Briel M, Busse JW et al.** Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ* 2012; 344:e1553 (published 15 March 2012)
- vfa –Stellungnahme.** „Das AMNOG im vierten Jahr“. Erfahrungsbericht und Regelungsvorschläge der forschenden Pharmaunternehmen. 2014  
([http://www.vfa.de/de/download-manager/\\_amnog-4tes-jahr-lang.pdf](http://www.vfa.de/de/download-manager/_amnog-4tes-jahr-lang.pdf))
- Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM.** Statistics in medicine—reporting of subgroup analyses in clinical trials. *New Engl J Med* 2007; **357**:2189–2194.
- Wang R, Schoenfeld DA, Hoepfner B, Evins AE.** Detecting treatment-covariate interaction using permutation methods. *Statist Med* 2015; DOI: 10.1002/sim.6457
- Willke RJ, Zheng Z, Subedi P et al.** From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: a primer. *BMC Medical Research Methodology* 2012, 12:185

# Links

**Cochrane-Collaboration (Homepage):** <http://www.cochrane.org/>

**G-BA (frühe Nutzenbewertung):** <https://www.g-ba.de/informationen/nutzenbewertung/>

**EMA (Homepage):** <http://www.ema.europa.eu/ema/>

**EMA (guidelines Biostatistics):**

[http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general\\_content\\_000602.jsp&mid=WC0b01ac05807d91a4](http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000602.jsp&mid=WC0b01ac05807d91a4)

**IQWiG (Homepage):** <https://www.iqwig.de/>

**vfa (Homepage):** <https://www.vfa.de/>