


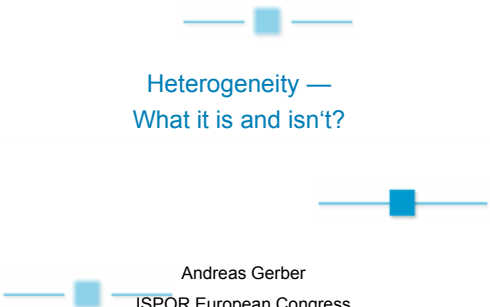

Speaker

Third Plenary Session
Heterogeneity in the Cost-effectiveness of Medical Interventions: The Challenge of Matching Patients to Appropriate Care




Andreas Gerber, PhD, MD
 Head, Department of Health Economics
 Institute of Quality and Efficiency
 in Health Care
 Cologne, Germany





Heterogeneity —
What it is and isn't?


Andreas Gerber
 ISPOR European Congress
 Madrid, Nov. 5th-8th, 2011



What any patient wants to know is „Would this treatment elicit a better response than the control for *me in particular*?“ That is a question no one can answer.

Kraemer et al. 2006

Andreas Gerber - ISPOR Madrid - Nov. 8th, 2011



What any patient wants to know is „Would this treatment elicit a better response than the control for *me in particular*?“ That is a question no one can answer.

Kraemer et al. 2006

Rephrase:
 Would this treatment elicit a better/worse expected response than the control for me as a member of a subgroup?

Andreas Gerber - ISPOR Madrid - Nov. 8th, 2011

Content

- Get the terms right
- Are we approaching the question from the same perspective?
- IQWiG's way of dealing with heterogeneity?
- What does this mean for health economic evaluation?

A potentially serious error is interpreting the effect size from an RCT as the effect of treatment on each individual within the population studied.

Kraemer et al. 2006

What is heterogeneity?

- In systematic reviews or meta-analyses:
- Homogeneity/ heterogeneity (simply) denotes whether the effects in the studies included are similar (homogeneous) or differing (heterogeneous), that is
- VARIABILITY BETWEEN STUDIES
- With statistical tests for heterogeneity one can detect whether the differences between studies are larger than could be expected by chance
- Reasons/causes for heterogeneity: differences in patients' characteristics, interventions or endpoints
- Meta-analyses performed including heterogeneous studies are „problematic“

THEREFORE: Heterogeneity is not

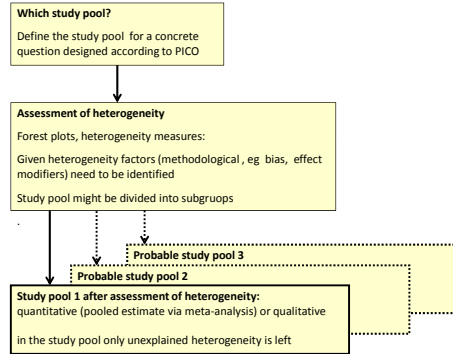
- Personalized medicine
- Individualized medicine as each study may represent a different population, but not the aspect of personalized medicine
- Could be stratified medicine
- What are you looking for:
Anything needs a sound statistical basis before making assumptions on subgroups

Causes for heterogeneity

- Studies differ in terms of
- patients (eg age, severity, genotype ...)
 - interventions (eg dosage)
 - endpoints (eg quality of life scores)
 - time frame
 - study design (eg blinding)
 - study conduct (eg premature termination)
 - analysis (eg ITT)
 - selective reporting (eg publication bias)
 - ...
 - study site ...
- and: approach of the meta-analysis (eg distance of the mean)
- clinical relevance
(PICO, question)

study design
(bias)

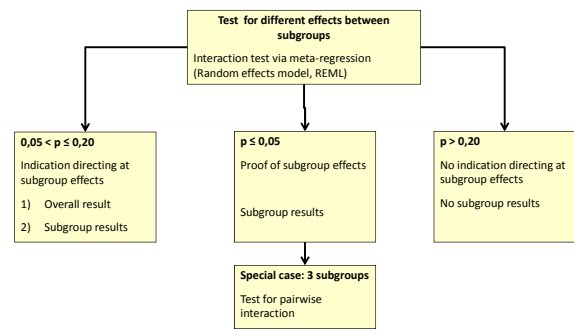
Which studies to pool?



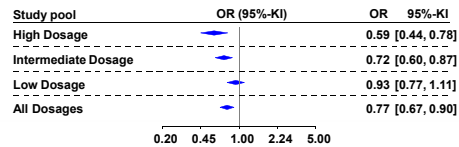
Preliminary conclusions or as always: Who knows the true answer?

- When to pool estimates?
 - „high“ heterogeneity ($p < 0,05$ or $I^2 > 70\%$) → pooling is not appropriate
 - „low“ heterogeneity ($p > 0,20$ or $I^2 < 25\%$) → pooling is appropriate
 - „intermediate“ heterogeneity
 - Fixed thresholds for heterogeneity tests? → simple and transparent, but probably oversimplifying and too much across-the-board
 - p-values or I^2 or ... ?
 - Thresholds depending on numbers of studies or ... ?
 - Always pool, but keep heterogeneity in mind
- Even with heterogeneity in the studies conclusions can be drawn!!!

How to proceed with subgroup analyses?



Example: Effect modifier → Dosage

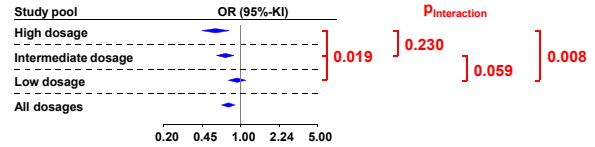


Heterogeneity between study pools: $Q=7.88$, $df=2$, $p=0.019$, $I^2=74.6\%$

Situation

- Statistically significant effect ($\alpha = 0.05$) for all dosages
- Dosage is statistically significant effect modifier ($\alpha = 0.05$)
- Question: Can two of the three dosage regimes be merged for higher precision?

Pairwise test for interaction



Problem often occurring with $\alpha = 0.05$:

- There is a difference between high and low dosage, but neither between high and intermediate nor between intermediate and low (not transitive relation)
- Possible solution: Lift the level ($\alpha > 0.05$) in pairwise tests
 → HOW MUCH? $\alpha = 0.10$ or 0.20 or 0.30 ?

"Prognostic markers classically identify patients with differing risks of a specific outcome, such as progression or death. ...

A prognostic marker can distinguish populations into groups where different treatment options are appropriate (possibly including no treatment), but it cannot guide the choice of a particular therapy. ...

A predictive marker is a marker that predicts the differential efficacy (benefit) of a particular therapy based on marker status (eg, only patients expressing the marker will respond to the specific treatment or will respond to a greater degree than those without the marker). A predictive marker could, therefore, guide the choice of therapy."

Sargent DJ et al. Clinical Trial Designs for Predictive Marker Validation in Cancer Treatment Trials. J Clin Oncol 2005; 23:2020-2027.

If the hypothesis of moderation is primary, the study must be designed to have a sample size adequate to generate sufficient power to detect a clinically significant interaction.

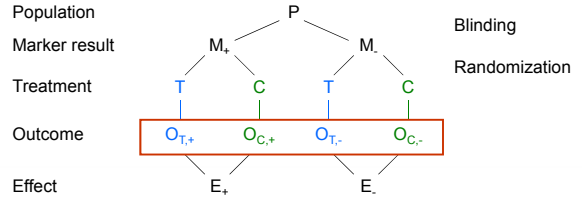
Kraemer et al. 2006

A marker has a benefit if – and only if – there is an interaction between the information (marker result) and the treatment effect (in general: the effect of a consequence).

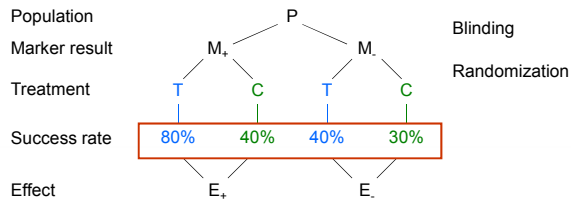
→ Marker x treatment interaction

An effect is the difference in outcomes between two (or more) treatments (in general: options) which can causally attributed to the difference in treatments.

An effect can usually be proven only in a randomized controlled trial (RCT).

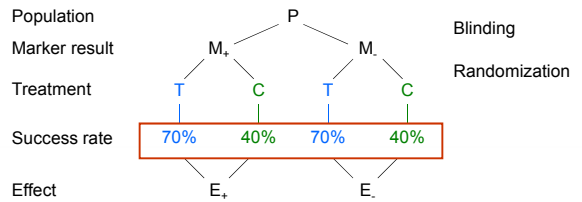


Benefit	$E_+ > 0$	\wedge	$E_- \leq 0$	Interaction (qual.)
Benefit	E_+	\gg	E_-	Interaction (quant.)
Benefit ?	E_+	$>$	E_-	Interaction (small)
No benefit	E_+	$=$	E_-	No interaction



Benefit $E_+ \gg E_-$ Interaction (quant.)

However: certain patients don't get access to an effective intervention ☹️



No benefit $E_+ = E_-$ No interaction

"Research methods for biomarker evaluation lag behind those for evaluating therapeutic treatments. a coherent and comprehensive set of guidelines for study design has not been delineated. We describe a nested case – control study design that involves **prospective collection** of specimens before outcome ascertainment from a study cohort that is relevant to the clinical application. The biomarker is assayed in a **blinded fashion** on specimens from **randomly selected** case patients and control subjects in the study cohort. The design can be applied to studies of biomarkers intended for use in disease diagnosis, screening, or prognosis. Common biases that pervade the biomarker research literature would be eliminated if these rigorous standards were followed."

Pepe MS et al. Pivotal Evaluation of the Accuracy of a Biomarker Used for Classification or Prediction: Standards for Study Design. J Natl Cancer Inst 2008; 100: 1432 – 1438

What are the effects of AMNOG?

Status quo:

1 dossier assessed by IQWiG and currently in the appraisal process with FJC

What kind of subgroup analyses would influence reimbursement?

Health economic evaluation under AMNOG?

What kind of subgroup analyses would further influence the reimbursable price to be recommended on the basis of a health economic analysis?

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG)

- Dillenburger Straße 27
- D-51105 Köln

- Telefon +49-221/3 56 85-0
- Telefax +49-221/3 56 85-1

- info@iqwig.de
- www.iqwig.de




Speaker

Third Plenary Session

Heterogeneity in the Cost-effectiveness of Medical Interventions: The Challenge of Matching Patients to Appropriate Care



Andreas Gerber, PhD, MD
 Head, Department of Health Economics
 Institute of Quality and Efficiency
 in Health Care
 Cologne, Germany