Workshop on subgroup analysis

Subgroup analyses: from nasty business to stratified medicine – an HTA perspective

Stefan Lange, MD, PhD
Deputy Director
Institute for Quality and Efficiency in Health Care (IQWiG)
COMMMENTARY

Standards for subgroup analyses are needed?—we couldn’t agree more

J. Hasford\textsuperscript{a,\ast}, P. Bramlage\textsuperscript{b}, G. Koch\textsuperscript{c}, W. Lehmacher\textsuperscript{d}, K. Einhäupl\textsuperscript{e}, P.M. Rothwell\textsuperscript{f}

\textsuperscript{a}Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, University of München, Germany
\textsuperscript{b}Institute for Cardiovascular Pharmacology and Epidemiology, Mahlow, Germany
\textsuperscript{c}Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA
\textsuperscript{d}Institute of Medical Statistics, Informatics and Epidemiology, University of Cologne, Köln, Germany
\textsuperscript{e}Department of Neurology, Medical School, Charité, Berlin, Germany
\textsuperscript{f}Stroke Prevention Research Unit, Department of Clinical Neurology, John Radcliffe Hospital, Oxford, UK

Accepted 4 October 2010

INVITED COMMENTARY

The assessment of heterogeneity is mandatory in clinical trials and systematic reviews

Ralf Bender\textsuperscript{a,\ast}, Armin Koch\textsuperscript{b}, Guido Skipka\textsuperscript{a}, Thomas Kaiser\textsuperscript{c}, Stefan Lange\textsuperscript{d}

\textsuperscript{a}Department of Medical Biometry, Institute for Quality and Efficiency in Health Care (IQWiG), Dillenburger Str. 27, D-51105 Cologne, Germany
\textsuperscript{b}Institute for Biometry, Centre for Biometry, Medical Informatics and Medical Technology, Hanover Medical School, D-30625 Hanover, Germany
\textsuperscript{c}Department of Drug Assessment, Institute for Quality and Efficiency in Health Care (IQWiG), Dillenburger Str. 27, D-51105 Cologne, Germany
\textsuperscript{d}Institute Management, Institute for Quality and Efficiency in Health Care (IQWiG), Dillenburger Str. 27, D-51105 Cologne, Germany

Accepted 4 October 2010
‘Subgroups kill people’


‘Not doing subgroup analyses has very probably killed more people.’

Ambivalence

The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 19 September 2002
CPMP/EWP/908/99

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS
(CPMP)

POINTS TO CONSIDER ON MULTIPLICITY ISSUES IN CLINICAL TRIALS
‘Additional claims on statistically significant and clinically relevant findings based on … subgroups are possible only after the primary objective of the clinical trial has been achieved, and if the respective questions were pre-specified, and were part of an appropriately planned statistical analysis strategy.’

‘It is highly unlikely that claims based on subgroup analyses would be accepted in the absence of a significant effect for the overall study population.’

‘Subgroups kill people’
‘Additional claims on statistically significant and clinically relevant findings based on ... subgroups are possible only after the primary objective of the clinical trial has been achieved, and if the respective questions would be accepted in the absence of a significant effect for the overall study population.’

‘Not doing subgroup analyses has very probably killed more people.’

‘Restriction of a license to certain subgroups is also possible, if a large variety of sub-populations are investigated without proper plans to deal with this situation in the protocol. From the regulatory perspective an overall positive result (statistically and clinically) in the whole study population may not lead to valid claims for all sub-populations if there is a reason to expect heterogeneity of the treatment effect in the respective sub-populations.’
§ 139a (3) SGB V (IQWiG)

2. ‘Preparation of scientific reports and expert opinions on quality and efficiency issues of services of the statutory health insurance, taking age, gender, and personal circumstances into account …’
§ 35a (1) SGB V (Assessment of the benefit of pharmaceuticals with new active ingredients)

‘The Federal Joint Committee assesses the benefit of reimbursable pharmaceuticals with new active ingredients. This includes in particular the assessment of the added benefit compared with the appropriate comparator therapy, the extent of added benefit and its therapeutic relevance. The benefit assessment is conducted on the basis of evidence provided by the pharmaceutical company, …and which must include in particular the following information: …

→ (even retrospective) Subgroup-analyses mandatory!

4. Number of patients and patient groups for whom there is a therapeutically relevant added benefit, …’
‘The main argument against subgroup analysis is that qualitative heterogeneity of relative treatment effect (defined as the treatment effect being in different directions in different groups of patients, ie, benefit in one subgroup and harm in another) is very rare.\textsuperscript{2–5} However, this observation is much less reassuring than it seems.’

Stratified medicine

**Fig 1.** Marker by Treatment Interaction
Design to test a predictive factor question; same treatments in both prognostic groups.
Qualitative interaction

Population

Subgroup

Treatment

Outcome

Effect

Interaction

Qualitative interaction

Population

Subgroup

Treatment

Outcome

Effect

Interaction

Qualitative interaction

Population

Subgroup

Treatment

Outcome

Effect

Interaction

Large quantitative

Approximately qualitative?
RESEARCH METHODS & REPORTING

Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses

Xin Sun, Matthias Briel, Stephen D Walter, Gordon H Guyatt

How can we tell the difference between spurious and real subgroup effects? This article identifies new criteria and proposes a checklist for judging the credibility of subgroup analyses

BMJ 2010; 340: 850-4
Criteria

‘An approach … is to place the likelihood that a subgroup effect is real on a continuum from “highly plausible” to “extremely unlikely”, possibly by using a visual analogue scale. The question is then a decision of where on this continuum a putative subgroup effect lies.’

BMJ 2010; 340: 850-4

Criteria to assess the credibility of subgroup analyses

<table>
<thead>
<tr>
<th>Criteria to assess the credibility of subgroup analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design</strong></td>
</tr>
<tr>
<td>• Is the subgroup variable a characteristic measured at baseline or after randomisation?*</td>
</tr>
<tr>
<td>• Is the effect suggested by comparisons within rather than between studies?</td>
</tr>
<tr>
<td>• Was the hypothesis specified a priori?</td>
</tr>
<tr>
<td>• Was the direction of the subgroup effect specified a priori*</td>
</tr>
<tr>
<td>• Was the subgroup effect one of a small number of hypothesised effects tested?</td>
</tr>
<tr>
<td><strong>Analysis</strong></td>
</tr>
<tr>
<td>• Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?</td>
</tr>
<tr>
<td>• Is the significant subgroup effect independent?*</td>
</tr>
<tr>
<td><strong>Context</strong></td>
</tr>
<tr>
<td>• Is the size of the subgroup effect large?</td>
</tr>
<tr>
<td>• Is the interaction consistent across studies?</td>
</tr>
<tr>
<td>• Is the interaction consistent across closely related outcomes within the study?*</td>
</tr>
<tr>
<td>• Is there indirect evidence that supports the hypothesised interaction (biological rationale)?</td>
</tr>
</tbody>
</table>

*New criteria.
Criteria

Criteria to assess the credibility of subgroup analyses

Design
• Is the subgroup variable a characteristic measured at baseline or after randomisation?*
• Is the effect suggested by comparisons within rather than between studies?
• Was the hypothesis specified a priori?
• Was the direction of the subgroup effect specified a priori*

Analysis
• Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?
• Is the significant subgroup effect independent?*
  • Is the size of the subgroup effect large?
  • Is the interaction consistent across studies?
  • Is the interaction consistent across closely related outcomes within the study?*
  • Is there indirect evidence that supports the hypothesised interaction (biological rationale)?

*New criteria.

BMJ 2010; 340: 850-4
## Criteria to assess the credibility of subgroup analyses

**Design**
- Is the subgroup variable a characteristic measured at baseline or after randomisation?*
- Is the effect suggested by comparisons within rather than between studies?
- Was the hypothesis specified a priori?
- Was the direction of the subgroup effect specified a priori*
- Was the subgroup effect one of a small number of hypothesised effects tested?

**Analysis**
- Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?
- Is the significant subgroup effect independent?*

**Context**
- Is the size of the subgroup effect large?
- Is the interaction consistent across studies?
- Is the interaction consistent across closely related outcomes within the study?*
- Is there indirect evidence that supports the hypothesised interaction (biological rationale)?

*New criteria.

---

**Overall effect?**

*BMJ 2010; 340: 850-4*
Conclusion Industry funded randomised controlled trials, in the absence of statistically significant primary outcomes, are more likely to report subgroup analyses than non-industry funded trials. Industry funded trials less frequently prespecify subgroup hypotheses and less frequently test for interaction than non-industry funded trials. Subgroup analyses from industry funded trials with negative results for the primary outcome should be viewed with caution.

Sadeesh K Srinathan, assistant professor,14 Philipp Dahm, associate professor,15 Bradley C Johnston, postdoctoral fellow,1 Pablo Alonso-Coello, researcher,16 Basil Hassouneh, research fellow,1 Jessica Truong, undergraduate student,17 Neil D Dattani, medical student,18 Stephen D Walter, professor,1 Diane Heels-Ansdell, statistician,1 Neera Bhatnagar, librarian,19 Douglas G Altman, professor,20 Gordon H Guyatt, professor1

BMJ 2011; 342: d1569. doi: 10.1136/bmj.d1569
Ambivalence

The European Agency for the Evaluation of Medicinal Products
Evaluation of Medicines for Human Use

London, 22 May 2003
CPMP/EWP/2863/99

COMMITTEE FOR PROPRIETARY MEDICINAL PRODUCTS (CPMP)

POINTS TO CONSIDER ON ADJUSTMENT FOR BASELINE COVARIATES
‘Conversely, an interaction cannot be considered as relevant on the sole basis of a significant test for interaction. Assessment of interaction terms based on statistical significance tests is therefore of little value.’

‘If some interaction turn out to be large from a clinical point of view or significant from a statistical point of view, this provides evidence that the effect of treatment may vary across subgroups.’
## Theses I

<table>
<thead>
<tr>
<th>Requirement / Problem</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplicity</td>
<td>Role in Benefit-risk or benefit-harm assessment questionable (e.g. multiple endpoints, pre-cautionary principle)</td>
</tr>
</tbody>
</table>

### Regulatory agencies
- Identify patient groups with negative benefit-risk ratio
- Restrictions due to safety concerns

### HTA agencies
- Identify patient groups with patient-relevant (additional) benefit (of a certain magnitude)
- Restriction due to too small benefit
<table>
<thead>
<tr>
<th>Requirement / Problem</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplicity</td>
<td>Role in Benefit-risk or benefit-harm assessment questionable (e.g. multiple endpoints, pre-cautionary principle)</td>
</tr>
<tr>
<td>Confirmatory strategy</td>
<td>Challenge! Weighing: Low power of interaction test vs. insisting in 5% level</td>
</tr>
</tbody>
</table>
Confirmatory strategies

Intervention Effects in the Context of Heterogeneity between Three Subgroups
Assessment within the Framework of Systematic Reviews

G. Skipka; R. Bender
Institute for Quality and Efficiency in Health Care, Cologne, Germany

Conclusions: By increasing the significance level for the pairwise tests to 0.20, non-transitive relations are virtually avoidable. The proposed hierarchical testing procedure represents a clear practical guidance to perform subgroup analyses in the framework of systematic reviews.

Methods Inf Med 2010; 49: 613–617

<table>
<thead>
<tr>
<th>Study pool</th>
<th>OR (95%-CI)</th>
<th>p&lt;sub&gt;Interaction&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>High dose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium dose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low dose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EMA – Workshop on subgroup analysis, London, HTA perspective, 18.11.2011, SL
Future trend?

Applying meta-analytical thinking to subgroup-analysis within (single) clinical trials

Thinking not in homogeneity but in heterogeneity
Test first on interaction
Apply closed testing procedure to the analysis of multiple effect modifiers (subgroups)
### Theses III

<table>
<thead>
<tr>
<th>Requirement / Problem</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplicity</td>
<td>Role in Benefit-risk or benefit-harm assessment questionable (e.g. multiple endpoints, pre-cautionary principle)</td>
</tr>
<tr>
<td>Confirmatory strategy</td>
<td>Challenge! Weighing: Low power of interaction test vs. insisting in 5% level</td>
</tr>
<tr>
<td>Pre-planning</td>
<td>At certain circumstances of questionable relevance (e.g. strong biological plausibility, stratified randomization, generally accepted possible effect modifiers [e.g. age, gender, disease stage, performance state, approval status, ...])</td>
</tr>
<tr>
<td>Presence of overall effect</td>
<td>Of questionable relevance within the context of ‘Personalized‘ (better: stratified) medicine (goal: identifying ‘qualitative’ interaction)</td>
</tr>
</tbody>
</table>
Institute for Quality and Efficiency in Health Care (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

Thank you very much!