Unplanned versus pre-specified subgroup analysis reporting

Doug Altman

Centre for Statistics in Medicine
University of Oxford

ESMO, Copenhagen, 10 October 2016
Conflicts of interest - None
Subgroup analyses

“Properly performed, analysis of subgroups can yield useful insights into therapy; unfortunately, many commonly used approaches are often uninformative or misleading.”

[Yusuf et al, JAMA 1991]
Some questions

- What are (in)appropriate statistical approaches?
- How often is an appropriate method used?
- How good is current reporting?
- How does oncology compare with other specialties?
- What is good practice?
Main subgroup analysis strategies in RCTs

- **Treatment effect in subset of the participants (e.g. diabetics)**
  - Disregard the others (e.g. non-diabetics)

- **Treatment effects separately for 2 or more complementary subsets of participants (e.g. by diabetes; by cancer stage)**
  - Separate analyses (2+ P values)

- **Compare treatment effects across complementary subgroups**
  - **test of interaction**
    - One analysis (one P value)
Main subgroup analysis strategies in RCTs

- Treatment effect in subset of the participants (e.g. diabetics)
  - Disregard the others (e.g. non-diabetics)

- Treatment effects separately for 2 or more complementary subsets of participants (e.g. by diabetes; by cancer stage)
  - Separate analyses (2+ P values)

- Compare treatment effects across complementary subgroups
  - test of interaction
  - One analysis (one P value)
Published subgroup analyses

- **Subgroup analyses should be pre-planned**
  - Frequent discrepancies between trial protocols and subsequent publications

- **Even prespecified subgroup analyses lack power**
  - So most significant results will be false positives

- **Results of all subgroup analyses should be reported**
  - Publishing only significant results magnifies problems
  - Indicate whether pre-specified
Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications

[Kasenda et al, BMJ 2014]
## Comparison of protocol and publication

<table>
<thead>
<tr>
<th>Clinical discipline</th>
<th>No. of trials</th>
<th>No. (%) pre-planned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oncology</td>
<td>155</td>
<td>42 (27%)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>108</td>
<td>49 (45%)</td>
</tr>
<tr>
<td>Infectious disease</td>
<td>87</td>
<td>27 (31%)</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>62</td>
<td>15 (24%)</td>
</tr>
<tr>
<td>Neurology</td>
<td>61</td>
<td>24 (39%)</td>
</tr>
<tr>
<td>Other</td>
<td>421</td>
<td>95 (23%)</td>
</tr>
</tbody>
</table>

[Kasenda et al, BMJ 2014]
An example of subgroup discrepancies

Outcome: time to progression or death

Subgroup analyses:

**Protocol:** baseline disease severity

**Publication:** duration of previous treatment*, type of previous treatment*, blood count*, disease severity

* Described explicitly as pre-specified despite not appearing in the protocol
No improvement in the reporting of clinical trial subgroup effects in high-impact general medical journals

Nicole B. Gabler, Naihua Duan, Eli Raneses, Leah Suttner, Michael Ciarametaro, Elizabeth Cooney, Robert W. Dubois, Scott D. Halpern and Richard L. Kravitz
Fig. 2 Percentage of trials reporting subgroup analysis utilizing appropriate statistical methods
<table>
<thead>
<tr>
<th></th>
<th>No. of trials</th>
<th>Reported subgroup analyses</th>
<th>Used appropriate method</th>
</tr>
</thead>
<tbody>
<tr>
<td>All trials</td>
<td>437</td>
<td>270 (62%)</td>
<td>185 (69%)</td>
</tr>
<tr>
<td></td>
<td>No. of trials</td>
<td>Reported subgroup analyses</td>
<td>Used appropriate method</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>All trials</td>
<td>437</td>
<td>270 (62%)</td>
<td>185 (69%)</td>
</tr>
<tr>
<td>Cancer</td>
<td>59</td>
<td>45 (76%)</td>
<td>24 (53%)</td>
</tr>
</tbody>
</table>

Cancer was 2\text{nd} worst of 9 medical areas studied
<table>
<thead>
<tr>
<th></th>
<th>No. of trials</th>
<th>Reported subgroup analyses</th>
<th>Used appropriate method</th>
</tr>
</thead>
<tbody>
<tr>
<td>All trials</td>
<td>437</td>
<td>270 (62%)</td>
<td>185 (69%)</td>
</tr>
<tr>
<td>Cancer</td>
<td>59</td>
<td>45 (76%)</td>
<td>24 (53%)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>101</td>
<td>73 (72%)</td>
<td>63 (86%)</td>
</tr>
</tbody>
</table>

Cancer was 2nd worst of 9 medical areas studied
Bad approaches

- Separate P values for each subset (or ignoring one subset)
- Post hoc comparison of subgroups based on observed results
  - i.e. not pre-planned
- Comparison of subset of active group with whole comparison group
- Subgroup defined by variable not known at randomisation and possibly influenced by treatment (“improper subgroups”)
  - e.g. compliance, early clinical response

- All the above methods can be extremely misleading
- Pre-specifying doesn’t guarantee an appropriate analysis
Heterogeneity of treatment effect - a better approach [Kent et al, Trials 2010]

- Conventional subgroup analyses do not account for the fact that patients have multiple characteristics simultaneously that affect the likelihood of treatment benefit

- A good approach is to estimate treatment effect in relation to multivariable risk score (risk stratification)
Reporting subgroup analyses

**Methods**
- Prespecified subgroup analyses (as in protocol)
- Postulated direction of effect
- Statistical method

**Results**
- Estimated effect size (with CI) in each subgroup
- Test of interaction (estimate of relative effect; P value)
Carboplatin and weekly paclitaxel doublet chemotherapy compared with monotherapy in elderly patients with advanced non-small-cell lung cancer: IFCT-0501 randomised, phase 3 trial

Elisabeth Quoix, Gérard Zalcman, Jean-Philippe Oster, Virginie Westeel, Eric Pichon, Armelle Lavolé, Jérôme Dauba, Didier Debieuvre, Pierre-Jean Souquet, Laurence Bigay-Game, Eric Dansin, Michel Poudenx, Olivier Molinier, Fabien Vaylet, Denis Moro-Sibilot, Dominique Herman, Jaafar Bennouna, Jean Tredaniel, Alain Ducoloné, Marie-Paule Lebitasy, Laurence Baudrin, Silvy Laporte, Bernard Milleron, on behalf of Intergroupe Francophone de Cancérologie Thoracique

Summary
Background Platinum-based doublet chemotherapy is recommended to treat advanced non-small-cell lung cancer (NSCLC) in fit, non-elderly adults, but monotherapy is recommended for patients older than 70 years. We compared a carboplatin and paclitaxel doublet chemotherapy regimen with monotherapy in elderly patients with advanced NSCLC.
Figure 3: Forest plot for subgroup analysis of overall survival
Survival is for doublet chemotherapy vs monotherapy. Data are derived from Cox's analysis without covariates.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>HR (95% CI)</th>
<th>p value for interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n=451)</td>
<td>0.64 (0.52–0.78)</td>
<td>0.557</td>
</tr>
<tr>
<td>PS 0–1 (n=327)</td>
<td>0.63 (0.49–0.81)</td>
<td></td>
</tr>
<tr>
<td>PS 2 (n=123)*</td>
<td>0.63 (0.43–0.91)</td>
<td></td>
</tr>
<tr>
<td>Age ≤80 years (n=337)</td>
<td>0.68 (0.53–0.86)</td>
<td>0.299</td>
</tr>
<tr>
<td>Age &gt;80 years (n=114)</td>
<td>0.53 (0.36–0.80)</td>
<td></td>
</tr>
<tr>
<td>Stage IV (n=364)</td>
<td>0.58 (0.46–0.73)</td>
<td>0.098</td>
</tr>
<tr>
<td>Stage IIIB–III (n=87)</td>
<td>0.94 (0.59–1.50)</td>
<td></td>
</tr>
</tbody>
</table>

Quoix et al, Lancet 2011
Good practice

- **Pre-specify a (very) few planned subgroup analyses in trial protocol**
  - Preferably with rationale and direction of postulated difference
- **Use only variables known at baseline**
- **Use interaction analysis or multivariable risk score**
- **Indicate all subgroups analyses undertaken**
  - and whether prespecified or post hoc
- **Interpret subgroup findings very cautiously**
  - Exploratory analyses are good for hypothesis generating
  - Even pre-planned analyses may be misleading
Some key references


