How to understand subgroup analysis in clinical studies

Massimo Di Maio
SCDU Medical Oncology
AO Ordine Mauriziano, Torino, Italy
Department of Oncology,
University of Turin
massimo.dimaio@unito.it

Claudia Cardone
Massimo Di Maio

How to understand subgroup analysis in clinical studies

5 April
18:30 CEST
IN DEFENSE OF CURIOUSITY

By MRS. FRANKLIN D. ROOSEVELT

A SHORT time ago a cartoon appeared depicting two miners looking up in surprise and saying with undisguised horror, “Here comes Mrs. Roosevelt!”

In strange and subtle ways, it was indicated to me that I should feel somewhat ashamed of that cartoon, and there certainly was something the matter with a woman who wanted to see so much and to know so much.

Somehow or other, most of the people who spoke to me, or wrote to me, about it, seemed to feel that it was unbecoming in a woman to have a variety of interests. Perhaps that arose from the old inherent theory that woman’s interests must lie only in her home. This is a kind of blindness which seems to make people feel that interest in the home stops within the four walls of the house in which you live. Few seem capable of realizing that the real reason that home is important is that it is so closely tied, by million strings, to the rest of the world. That is what makes it an important factor in the life of only upon the buying power of people like herself but upon the buying power of the great mass of agricultural people throughout the country. The farm housewife must realize, too, that her interests are tied up with those of the wage earner and his employer throughout the nation, for her husband’s products can only find a ready market when the city dweller is prosperous.

There is ever present, of course, the economic question of how to keep balanced the cost of living and the wages the man receives. The theory of low wages and low living costs has been held by many economists to be sound, for they centered what money one has will provide as much as high wages do in countries where living costs are also high.

We have gone, as a rule, on the theory, in this country, particularly in areas of prosperity, that high wages and high costs make for a higher standard of living, and that we really obtain more for our money, even though our prices are higher.

This notion is annual back and forth, and the quilting pattern or recipe in the neighborhood. Isn’t that better than waiting days for a letter which may never come?

To the city or suburban dweller, the price of a subway ride is of great importance, for if it costs ten cents a day to come and go from work, he may have enough left at the end of the week to take his wife to a movie, but twenty cents a day may mean that he has nothing left for entertainment. The city dweller could also do much for the price of milk, if he realized the dairy farmer’s plight and how important the consumption of milk and its price is to general prosperity.

This corollation of interests is something that every woman would understand if she had the curiosity to find out the reasons for certain conditions instead of merely accepting them, usually with rather bad grace.

Curiosity is the Mother of Opinion
...even more than before, in the era of personalized medicine and precision oncology, **subgroup analysis** seems a valuable tool for optimizing treatment choices.
Subgroup analysis may impact regulatory decisions: Durvalumab in locally advanced NSCLC

Faivre-Finn C, ESMO 2020
Annals of Oncology (2020) 31 (suppl_4): S1142-S1215
Subgroup analyses: why?

• Within a study with overall positive conclusions, subgroup analyses might help to better identify patients who benefit more, patients who benefit less or patients who don’t benefit at all.
The famous example of the IPASS trial: qualitative interaction!

The famous example of the IPASS trial: qualitative interaction!

Subgroup analyses: why?

• Within a study with overall positive conclusions, subgroup analyses might help to better identify patients who benefit more, patients who benefit less or patients who don’t benefit at all.

• Within a study with overall negative conclusions, subgroup analyses might help to avoid «throwing the baby out with the bath water», by identifying certain groups of patients in whom the experimental treatment appears to work.
“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.”

J W Tukey, 1962
In defense of curiosity…

…but remember that curiosity can be dangerous!
Let’s make an example outside oncology!

**ISIS-2: Second International Study of Infarct Survival**

Vascular mortality over 35 days: individual therapies

ISIS-2 trial: Aspirin vs Placebo
Mortality 1 month after myocardial infarction

N. of deaths
A vs P

P

All cases
804 vs 1016
<0.0001
### ISIS-2 trial: Aspirin vs Placebo
Mortality 1 month after myocardial infarction

<table>
<thead>
<tr>
<th>Zodiac sign</th>
<th>N. of deaths</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cases</td>
<td>804 vs 1016</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Other signs</td>
<td>654 vs 869</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Libra or Gemini</td>
<td>150 vs 147</td>
<td>0.5 (ns)</td>
</tr>
</tbody>
</table>
Should I suspect a risk of false negative result in a subgroup?

Phase III Trial of Bevacizumab in Non-Squamous NSCLC: ECOG 4599

**Eligibility:**
- Non-squamous NSCLC
- No Hx of hemoptysis
- No CNS metastases

**Stratification Variables:**
- RT vs no RT
- Stage IIIB or IV vs recurrent
- Wt loss <5% vs ≥5%
- Measurable vs non-measurable

(PC)
Paclitaxel 200 mg/m²
Carboplatin AUC = 6
(q 3 weeks) × 6 cycles

(PCB)
PC × 6 cycles + Bevacizumab (15mg/kg q 3 wks) to PD

No crossover to Bevacizumab permitted

Sandler AB et al., ASCO 2005, abstract 4
Should I suspect a risk of false negative result in a subgroup?

Sandler AB et al., ASCO 2005, abstract 4
Should I suspect a risk of false negative result in a subgroup?

**Exploratory Subgroup Analyses**

- Subgroups analyzed
  - Stage
  - Weight loss
  - Prior RT
  - Race
  - PS
  - Age
  - Gender

- These were not pre-specified analyses

- Survival benefit was seen across all treatment subgroups except for gender

Sandler AB et al., ASCO 2005, abstract 4
Should I suspect a risk of false negative result in a subgroup?

Sandler AB et al., ASCO 2005, abstract 4
Should I suspect a risk of false negative result in a subgroup?

Possible Explanations for Survival Differences by Gender?

• Use of second and third-line treatment
  – EGFR-TKI’s
  – chemotherapy
• Imbalance in unmeasured baseline prognostic factors
  – Demographic
  – Molecular
• Statistical chance alone
• True difference

Sandler AB et al., ASCO 2005, abstract 4
The risk of «belief bias»...

If a conclusion supports your existing beliefs, you'll rationalize anything that supports it.
...and the risk of HARKing

HARKing: Hypothesizing After the Results areKnown

Norbert L. Kerr
Department of Psychology
Michigan State University

This article considers a practice in scientific communication termed HARKing (Hypothesizing After the Results are Known). HARKing is defined as presenting a post hoc hypothesis (i.e., one based on or informed by one’s results) in one’s research report as if it were, in fact, an a priori hypotheses. Several forms of HARKing are identified and survey data are presented that suggests that at least some forms of HARKing are widely practiced and widely seen as inappropriate. I identify several reasons why scientists might HARK. Then I discuss several reasons why scientists ought not to HARK. It is conceded that the question of whether HARKing’s costs exceed its benefits is a complex one that ought to be addressed through research, open discussion, and debate. To help stimulate such discussion (and for those such as myself who suspect that HARKing’s costs do exceed its benefits), I conclude the article with some suggestions for deterring HARKing.

Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives

ST Brookes
EW Whitely
TJ Peters
PA Mulheran
M Egger
G Davey Smith
FIGURE 22 Summary of results for the simplest case (overall test result significant). This figure combines the results from data simulated with no overall treatment effect and with a true overall treatment effect detectable at nominal powers of 50, 80, 90 and 95%
Should I suspect a false positive result in a subgroup?

**REACH: Study Design**

- Prior Sorafenib
- BCLC stage B/C
- Child-Pugh A
- ECOG PS 0 or 1

**Randomize (1:1)**

- Ramucirumab (8 mg/kg) q2wks per cycle and BSC N=272
- Placebo q2wks per cycle and BSC N=272

Treatment until disease progression or unacceptable toxicity

**Primary endpoint:** Overall Survival

**Secondary endpoints:** PFS, TTP, ORR, safety, patient-reported outcomes

Stratification factors:
- Geographic Regions
  - North and South America
  - Europe
  - Asia
- Etiology of Liver Disease
  - Hepatitis B
  - Hepatitis C
  - Other etiologies

Abbreviations: BCLC=Barcelona Clinic Liver Cancer; BSC=best supportive care; ECOG PS=Eastern Cooperative Oncology Group performance status; ORR=objective response rate; PFS=progression-free survival; q2wks=every 2 weeks; TTP=time-to-progression.

Zhu A et al, ESMO 2014
Should I suspect a false positive result in a subgroup?

Overall Survival of ITT Population

<table>
<thead>
<tr>
<th></th>
<th>Ramucirumab</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median, months</td>
<td>9.2 (8.1, 10.6)</td>
<td>7.6 (6.0, 9.3)</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>0.866 (0.717, 1.046)</td>
<td>0.1391</td>
</tr>
<tr>
<td>P-value (log-rank)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Patients at Risk

<table>
<thead>
<tr>
<th></th>
<th>Ramucirumab</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients at Risk</td>
<td>263 261 214 175 149 122 101 78 61 43 32 27 20 15 11 5 4 2 1</td>
<td>282 255 189 151 129 110 83 63 54 35 30 23 18 12 9 4 3 1 1</td>
</tr>
</tbody>
</table>

Abbreviations: CI=confidence interval; HR=hazard ratio; ITT=intent to treat; OS=overall survival.

Zhu A et al, ESMO 2014

Zhu et al, ESMO 2014
Should I suspect a false positive result in a subgroup?

**Overall Survival in Patients With Baseline Alpha-fetoprotein ≥ or <400 ng/mL**

**AFP ≥400 ng/mL**

<table>
<thead>
<tr>
<th>Patients</th>
<th>Ramucirumab (N=119)</th>
<th>Placebo (N=131)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median, months</td>
<td>7.8</td>
<td>4.2</td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(5.8, 9.3)</td>
<td>(3.7, 4.8)</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>0.674 (0.508, 0.895)</td>
<td></td>
</tr>
<tr>
<td>P-value (log-rank)</td>
<td>0.0059</td>
<td></td>
</tr>
</tbody>
</table>

**AFP <400 ng/mL**

<table>
<thead>
<tr>
<th>Patients</th>
<th>Ramucirumab (N=160)</th>
<th>Placebo (N=150)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median, months</td>
<td>10.1</td>
<td>11.8</td>
</tr>
<tr>
<td>(95% CI)</td>
<td>(8.7, 12.3)</td>
<td>(9.9, 13.1)</td>
</tr>
<tr>
<td>HR (95% CI)</td>
<td>1.093 (0.836, 1.428)</td>
<td></td>
</tr>
<tr>
<td>P-value (log-rank)</td>
<td>0.5059</td>
<td></td>
</tr>
</tbody>
</table>

Zhu et al, ESMO 2014
FIGURE 21  Summary of results for the simplest case (overall test result not significant). This figure combines the results from data simulated with no overall treatment effect and with a true overall treatment effect detectable at nominal powers of 50, 80, 90 and 95%
Subgroup analysis can be hypothesis-generating for a subsequent trial!

Subgroup analysis can be hypothesis-generating for a subsequent trial!
How to correctly interpret subgroup analyses?

1: Treatment effects in subgroups of men and women in three hypothetical trials

- **Study 1**
  - Men: 0.001
  - Women: 0.72

- **Study 2**
  - Men: 0.001
  - Women: 0.86

- **Study 3**
  - Men: 0.001
  - Women: 0.95

Overall relative risk: 0.75 for Study 1; 0.80 for Study 2; 0.87 for Study 3; represented by the vertical dashed line in each case.

Simes RJ et al, MJA 2004
How to correctly interpret subgroup analyses?

Study 1

<table>
<thead>
<tr>
<th>Group</th>
<th>P value for treatment</th>
<th>P value for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.001</td>
<td>0.95</td>
</tr>
<tr>
<td>Women</td>
<td>0.72</td>
<td></td>
</tr>
</tbody>
</table>

Simes RJ et al, MJA 2004
How to correctly interpret subgroup analyses?

In cases like this, please **DO NOT CLAIM** that experimental treatment is significantly effective in men but not in women!

Simes RJ et al, MJA 2004
How to correctly interpret subgroup analyses?

Study 2

<table>
<thead>
<tr>
<th>Group</th>
<th>P value for treatment</th>
<th>P value for heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.001</td>
<td>0.20</td>
</tr>
<tr>
<td>Women</td>
<td>0.86</td>
<td></td>
</tr>
</tbody>
</table>

Simes RJ et al, MJA 2004
In cases like this, it is legitimate to suspect that treatment efficacy could be different…

…unfortunately, we cannot exclude that the difference we are observing is due to chance!
How to correctly interpret subgroup analyses?

Study 3

Men

Women

$P$ value for treatment

0.001

0.95

$P$ value for heterogeneity

0.01

Relative risk

Treatment better

Control better

Simes RJ et al, MJA 2004
In cases like this, it is legitimate to discuss the heterogeneity of treatment effect between men and women. Interaction test tells us that this difference is unlikely to be due to chance.

Simes RJ et al, MJA 2004
Scenario n.1

<table>
<thead>
<tr>
<th></th>
<th>Number of patients</th>
<th>Median overall survival (months)</th>
<th>Unstratified hazard ratio for death (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nivolumab plus ipilimumab group (n=303)</td>
<td>Chemotherapy group (n=302)</td>
</tr>
<tr>
<td>All randomly assigned</td>
<td>605</td>
<td>18.1</td>
<td>14.1</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>467</td>
<td>17.5</td>
<td>13.7</td>
</tr>
<tr>
<td>Female</td>
<td>138</td>
<td>21.4</td>
<td>18.0</td>
</tr>
</tbody>
</table>

Baas et al, Lancet 2021
**Scenario n.1**

- Interaction test ($p=0.91$) is NOT significant: heterogeneity of efficacy between men and women is NOT demonstrated.

<table>
<thead>
<tr>
<th></th>
<th>Number of patients</th>
<th>Median overall survival (months)</th>
<th>Unstratified hazard ratio for death (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nivolumab plus ipilimumab group (n=303)</td>
<td>Chemotherapy group (n=302)</td>
</tr>
<tr>
<td>All randomly assigned</td>
<td>605</td>
<td>18.1</td>
<td>14.1</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>467</td>
<td>17.5</td>
<td>13.7</td>
</tr>
<tr>
<td>Female</td>
<td>138</td>
<td>21.4</td>
<td>18.0</td>
</tr>
</tbody>
</table>

Di Maio M, Tagliamento M.  
Heterogeneity of treatment effects in malignant pleural mesothelioma.  
Lancet. 2021 Jul 24;398(10297):301-302..
**Scenario n.2**

<table>
<thead>
<tr>
<th>Tumour histology</th>
<th>Number of patients</th>
<th>Median overall survival (months)</th>
<th>Unstratified hazard ratio for death (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nivolumab plus ipilimumab group (n=303)</td>
<td>Chemotherapy group (n=302)</td>
</tr>
<tr>
<td>All randomly assigned</td>
<td>605</td>
<td>18.1</td>
<td>14.1</td>
</tr>
<tr>
<td>Epithelioid</td>
<td>456</td>
<td>18.7</td>
<td>16.5</td>
</tr>
<tr>
<td>Non-epithelioid</td>
<td>149</td>
<td>18.1</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Baas et al, Lancet 2021
**Scenario n.2**

<table>
<thead>
<tr>
<th>Tumour histology</th>
<th>Number of patients</th>
<th>Median overall survival (months)</th>
<th>Unstratified hazard ratio for death (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All randomly assigned</td>
<td>605</td>
<td>18.1</td>
<td>0.75 (0.62-0.91)</td>
</tr>
<tr>
<td>Epithelioid</td>
<td>456</td>
<td>18.7</td>
<td>0.86 (0.69-1.08)</td>
</tr>
<tr>
<td>Non-epithelioid</td>
<td>149</td>
<td>18.1</td>
<td>0.46 (0.31-0.68)</td>
</tr>
</tbody>
</table>

Interaction test is significant (p=0.007)
Heterogeneity of efficacy between epithelioid and non epithelioid tumors is demonstrated

Di Maio M, Tagliamento M.
Subgroup analyses: take home messages
Subgroup analyses: take home messages

• Caution!
Subgroup analyses: take home messages

• Caution!

• Hypothesis generation
Subgroup analyses:
take home messages

- Caution!
- Hypothesis generation
- Multiplicity: risks of false positive and false negative
Subgroup analyses: take home messages

- Caution!
- Hypothesis generation
- Multiplicity: risks of false positive and false negative
- Look at consistency among studies
Subgroup analyses: take home messages

- Caution!
- Hypothesis generation
- Multiplicity: risks of false positive and false negative
- Look at consistency among studies
- Plausibility (but beware of belief bias!)
Subgroup analyses: take home messages

- Caution!
- Hypothesis generation
- Multiplicity: risks of false positive and false negative
- Look at consistency among studies
- Plausibility (but beware of belief bias!)
- Look at the interaction test!
Subgroup analyses in randomized phase III trials of systemic treatments in advanced solid tumours: a systematic review of trials published between 2017 and 2020
Selected references

• Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. The Lancet, 2005; 365 (9454), 8–14: 176-186
How to understand subgroup analysis in clinical studies

Massimo Di Maio
SCDU Medical Oncology
AO Ordine Mauriziano, Torino, Italy
Department of Oncology, University of Turin
massimo.dimaio@unito.it

Claudia Cardone
Massimo Di Maio

5 April
18:30 CEST