Exploratory Subgroup Analyses: Why Do We Need Particular Caution?

Thomas R. Fleming, Ph.D.
Professor, Dept. of Biostatistics
University of Washington

* Fleming TR “Clinical Trials: Discerning Hype from Substance”
  • Annals of Internal Medicine 2010; 153:400-406
Interest in “Positive” Results in Clinical Trials

- **Industry Sponsors**
  - Company profits, ↑ value of stock options, promotion

- **Government Sponsors**
  - Claims of success in advancing health care
  - Leverage for ↑ in federal funding

- **Journal Editors** (Publication bias)

- **Academic Investigators / Caregivers**
  - Increased ability to publish results
    - ↑ professional stature, earlier promotion, ↑ salary
  - Desire to offer more therapeutic options to patients

...Result: Wide Spread & Significant Conflicts of Interest
An Illustration of Exploratory Analyses: Post-hoc Subgroup Analyses

Surgical Adjuvant Therapy of Colorectal Cancer

5-FU + Levamisole
Levamisole
Control
Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial

Years from randomization

Surviving, %

- 5-FU+LEV n=81
- LEV n=85
- Control n=81
NORTH CENTRAL TREATMENT GROUP STUDY
Looking at Treatment Effect on Overall Survival

Females Only

<table>
<thead>
<tr>
<th>Years from Registration</th>
<th>At Risk</th>
<th>Death</th>
<th>5-Yr Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Males Only

<table>
<thead>
<tr>
<th>Years from Registration</th>
<th>At Risk</th>
<th>Death</th>
<th>5-Yr Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial

- 5-FU+LEV n=81
- LEV n=85
- Control n=81

Years from randomization
Surviving, %
Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial

- 5-FU+LEV n=81
- LEV n=85
- Control n=81

Cancer Intergroup Trial

- 5-FU+LEV n=304
- LEV n=310
- Control n=315

Years from randomization
Surviving, %
INTERGROUP STUDY 0035
Looking at Treatment Effect on Overall Survival

### Females Only

<table>
<thead>
<tr>
<th>At Risk</th>
<th>Death</th>
<th>5-Yr Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>163</td>
<td>74</td>
<td>58%</td>
</tr>
<tr>
<td>149</td>
<td>77</td>
<td>54%</td>
</tr>
</tbody>
</table>

### Males Only

<table>
<thead>
<tr>
<th>At Risk</th>
<th>Death</th>
<th>5-Yr Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>141</td>
<td>47</td>
<td>70%</td>
</tr>
<tr>
<td>166</td>
<td>91</td>
<td>51%</td>
</tr>
</tbody>
</table>

**5-FU+Levamisole**

**Follow-Up Only**

Years from Registration
Duke’s C Colon Cancer Adjuvant

Percent ↓ in Death Rate: 5-FU + Levamisole
Control

<table>
<thead>
<tr>
<th>Analysis Group</th>
<th>North Central Treatment Group Study (n = 162)</th>
<th>Intergroup Study # 0035 (n = 619)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>28%</td>
<td>33%</td>
</tr>
<tr>
<td>Female</td>
<td>43%</td>
<td>15%</td>
</tr>
<tr>
<td>Male</td>
<td>9%</td>
<td>50%</td>
</tr>
<tr>
<td>Young</td>
<td>40%</td>
<td>23%</td>
</tr>
<tr>
<td>Old</td>
<td>13%</td>
<td>41%</td>
</tr>
</tbody>
</table>
Confirmatory vs. Exploratory Analyses

• Hyp. Confirmation vs. Hyp. Generation

~ Post-hoc analyses & *Random High Bias*
(new endpoints, new analyses, interim analyses
subgroup analyses, covariate adjustments)

Illustrations and Motivation:
*Baseball* & Clinical Research
An Illustration of Exploratory Analyses: Post-hoc Subgroup Analyses

Radiation Treatment in Rectal Cancer
Princess Margaret Hospital

Pre-operative R.T.
Control
Survival of Patients with Rectal Carcinoma
Princess Margaret Hospital, Toronto (1977)

Survival %

Years

Pre-operative Irradiation
Control

# = no. at risk
Survival of Patients with Rectal Carcinoma
Exploratory Subgroup: Dukes’ Stage C Disease

Survival %

2-sided \( p = 0.01 \)

# = no. at risk

Pre-Operative Irradiation

Control

John Oliver  May 8, 2016  “P-hacking”
www.youtube.com/watch?v=0Rnq1NpHdmw
Medical Research Council (MRC) Confirmatory Trial

![Survival rate graph with different treatment groups]

- No XRT (275)
- Single fraction (277)
- Multiple fractions (272)

Time, mo

Survival rate, %
MRC Subgroup Analysis: Dukes’ C Cases

Survival rate, %

Time, mo

- No XRT (111)
- Single fraction (110)
- Multiple fractions (79)
Some Important Observations

• $P$-values are only interpretable when you understand the sampling context from which they were derived

• Random High bias is real

• Exploratory Analyses usually should be viewed to be “Hypothesis Generating”

• Confirmatory Trials greatly enhance the reliability of conclusions

Fleming TR “Clinical Trials: Discerning Hype from Substance” *Annals of Internal Medicine* 2010; 153:400-406
Survival of Patients with Rectal Carcinoma
Exploratory Subgroup: Dukes’ Stage C Disease

Survival of Patients with Rectal Carcinoma
Exploratory Subgroup: Dukes’ Stage C Disease

2-sided $p = 0.01$

John Oliver  May 8, 2016  “P-hacking”
www.youtube.com/watch?v=0Rnq1NpHdmw

# = no. at risk

Pre-Operative Irradiation

Control
MRC Subgroup Analysis: Dukes’ C Cases

- No XRT (111)
- Single fraction (110)
- Multiple fractions (79)

Survival rate, %

Time, mo

0 6 12 18 24 30 36 42 48 54 60 66
“It isn’t so much the things we *don’t know* that get us in trouble.

It’s the things we *know* that aren’t so”.

—Artemus Ward (1834-1867)
Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet ’86)
  - SK reduces mortality by 20%
    confined to:
      anterior MI
      < 65 years
      < 6 hours from symptom onset
Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet ’86)
  - SK reduces mortality by 20%
    confined to:
    - anterior MI
    - < 65 years
    - < 6 hours from symptom onset
- Subset restriction not confirmed by ISIS-2, ASSET, AIMS
- While in ISIS-2:
  Aspirin beneficial overall…
Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet ’86)
  - SK reduces mortality by 20%
    confined to:
      anterior MI
      < 65 years
      < 6 hours from symptom onset
  - Subset restriction not confirmed by ISIS-2, ASSET, AIMS
  - While in ISIS-2:
    Aspirin beneficial overall…
    … yet harmful to patients with
    astrological signs Libra and Gemini
Bias for “Positive” Results in Clinical Trials

- Protocol Specified Clinical Trial “Primary Objective”
  - Very frequent wording:
    - “To establish that the experimental regimen is safe and effective”
Bias for “Positive” Results in Clinical Trials

➢ Protocol Specified Clinical Trial “Primary Objective”:

• Very frequent wording:
  ~ “To establish that the experimental regimen is safe and effective”

• Scientifically unbiased wording:
  ~ “To determine whether the experimental regimen is safe and effective”

• Regulatory Industry Statistics Workshop: (9/22/2011)
  …Credibility of exploratory analyses…
  …a Paradox…
Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus
- **Trial #1**: Time to renal flare: Minimal effect, $(2p = 0.51)$
Interest in “Positive” Results in Clinical Trials

• Abetimus Sodium: Reducing Renal Flare Rate in Lupus

• **Trial #1**: Time to renal flare: Minimal effect, ($2p = 0.51$)
  …exploratory *high affinity* subgroup: $2p = 0.007$

• **Trial #2** conducted in *high affinity* subgroup:
  Time to renal flare:
Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus

- **Trial #1**: Time to renal flare: Minimal effect, \((2p = 0.51)\)
  ...exploratory *high affinity* subgroup: \(2p = 0.007\)

- **Trial #2** conducted in *high affinity* subgroup:
  Time to renal flare: Minimal non-significant effect
Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus

- **Trial #1**: Time to renal flare: Minimal effect, \( (2p = 0.51) \)
  …exploratory high affinity subgroup: \( 2p = 0.007 \)

- **Trial #2** conducted in high affinity subgroup:
  Time to renal flare: Minimal non-significant effect
  …exploratory truncation at 12 months is favorable

- **Trial #3** conducted in high affinity subgroup with prespecified truncation at 12 months follow-up:
Interest in “Positive” Results in Clinical Trials

- **Abetimus Sodium**: Reducing Renal Flare Rate in Lupus

- **Trial #1**: Time to renal flare: Minimal effect, \( (2p = 0.51) \)
  …exploratory *high affinity* subgroup: \( 2p = 0.007 \)

- **Trial #2** conducted in *high affinity* subgroup:
  Time to renal flare: Minimal non-significant effect
  …exploratory *truncation at 12 months* is favorable

- **Trial #3** conducted in *high affinity* subgroup with prespecified *truncation at 12 months follow-up*:
  …early termination by DMC for futility.
“If you Torture Data Long Enough, They will Confess”

* Fleming TR “Clinical Trials: Discerning Hype from Substance” Annals of Internal Medicine 2010; 153:400-406
Some Conclusions

• Recognize bias resulting from strong interest to achieve “positive” results

• When refereeing journal publications, request:
  ➢ the clinical trial protocol
  ➢ the statistical analysis plan (SAP)
  ➢ the clinical study report (CSR)

• The only $P$-values presented in CSRs & publications should be for $\alpha$-spending analyses pre-specified in the SAP

• Recognize unreliability of Exploratory Analyses… …generating hypotheses, but with “random high” bias

• Exploratory subgroup analyses should be presented descriptively, for example using forest plots
Some Conclusions

• For reliable evidence regarding effects by subgroups, such as evaluating effects in biomarker positive vs negative subgroups, it is important to have pre-specified hypotheses (potentially with alpha spending)

• Cautionary Note: “When it is prespecified that biomarker-negative patients should not be included in the primary analysis of treatment effect in biomarker-positive patients because of the likelihood that treatment effects would differ between the 2 subgroups, it is logically inconsistent to include biomarker-positive patients in the primary analysis of treatment effect in biomarker-negative patients.” *
