Exploratory Analyses: Why Do We Need Particular Caution?

July 25, 2018

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
### Data Driven Hypothesis for the Cancer Risk with Vytorin in Aortic-Valve Stenosis

<table>
<thead>
<tr>
<th>Trial</th>
<th>N</th>
<th>CA. Incidence</th>
<th>CA. Deaths</th>
<th>Relative Risk</th>
<th>95% C.I.</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEAS Trial</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vytorin</td>
<td>944</td>
<td>101</td>
<td>37</td>
<td>1.55</td>
<td>(1.13, 2.12)</td>
<td></td>
</tr>
<tr>
<td>Placebo</td>
<td>929</td>
<td>65</td>
<td>20</td>
<td>1.78</td>
<td>(1.03, 3.11)</td>
<td></td>
</tr>
<tr>
<td><strong>IMPROVE-IT &amp; SHARP Trials</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vytorin</td>
<td>10,391</td>
<td>313</td>
<td>97</td>
<td>0.96</td>
<td>(0.82, 1.12)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>10,298</td>
<td>326</td>
<td>72</td>
<td>1.34</td>
<td>(0.98, 1.84)</td>
<td></td>
</tr>
</tbody>
</table>
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*
What is the definition of a successful clinical trial?

A very common response:

“A clinical trial that achieves a positive result”
What is the definition of a successful clinical trial?

- A very common response: “A clinical trial that achieves a positive result”

- The proper scientific response: “A clinical trial that addresses a clinically important issue, and that reliably answers the questions it was designed to address”
Confirmatory vs. Exploratory Analyses

  - Post-hoc analyses & Random High Bias
    (new endpoints, new analyses, interim analyses, subgroup analyses, covariate adjustments)
Confirmatory vs. Exploratory Analyses

- Clinical Endpoints in Pulmonary Arterial Hypertension
  - Overall survival
  - Quality of Life: SF-36 (8 domains), Borg Dyspnea Score
  - NYHA Functional Class
  - 6MWT: @18 wk, 24 wk, 48 wk, etc.
  - Time to Clinical Worsening
    - Death, PAH Hosp, L.T., (NYHA↑ & 6MWT↓)

- Analysis Methods
  - Normally distributed: T-test, ANCOVA, Wilcoxon
  - Time to event: Log-rank, Cox Regression
  - Dichotomous: Fisher’s Exact Test, Pearson $\chi^2$
Confirmatory vs. Exploratory Analyses

• Biomarker Endpoints (Hemodynamic parameters)
  ~ Pulmonary Arterial Pressure
  ~ Systolic & Diastolic Systemic Arterial Pressure
  ~ Systemic & Pulmonary Vascular Resistance
  ~ Heart Rate & Cardiac Output

• Analyses over Calendar Time
  • ~ Normally distributed: T-test, ANOVA, Wilcoxon
  • ~ Time to event: Log-rank, Cox Regression
  • ~ Dichotomous: Fisher’s Exact Test, Pearson $\chi^2$
Confirmatory vs. Exploratory Analyses

• Subgroup Analysis & Prognostic Covariate Adjustment

~ WHO PAH Functional Class: I v II v III v IV
~ Etiology: Idiopathic PAH, Assoc w CTD, SLE, Other
~ Baseline Walking Distance: < 325 v > 325 meters
~ Gender: male v female
~ Age: By decade
~ Ethnicity: White v Black v Asian v Other
~ mean PAP: < 50 v > 50

Epoprostenol +/- Sildenafil
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:
Confirmatory vs. Exploratory Analyses

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

Illustrations and Motivation:

Maternity Wards, Baseball & Clinical Research

20 vs 2: (.71, .99), 2p = 0.0001
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

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

Years from randomization

Surviving, %
NORTH CENTRAL TREATMENT GROUP STUDY
Looking at Treatment Effect on Overall Survival

<table>
<thead>
<tr>
<th>Females Only</th>
<th>Males Only</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>At Risk</strong></td>
<td><strong>37</strong></td>
</tr>
<tr>
<td><strong>Death</strong></td>
<td><strong>25</strong></td>
</tr>
<tr>
<td><strong>5-Yr Estimate</strong></td>
<td><strong>51%</strong></td>
</tr>
<tr>
<td><strong>At Risk</strong></td>
<td><strong>38</strong></td>
</tr>
<tr>
<td><strong>Death</strong></td>
<td><strong>24</strong></td>
</tr>
<tr>
<td><strong>5-Yr Estimate</strong></td>
<td><strong>47%</strong></td>
</tr>
</tbody>
</table>

- **FU+Levamisole**
- **Follow-Up Only**

Years from Registration
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>

Years from Registration
## Duke’s C Colon Cancer Adjuvant

### Percent ↓ in Death Rate:

<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>
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 in Control and Irradiated Groups

PMH--Toronto Study

# = no. at risk

Survival %

Years
Survival of Patients with Dukes’ Stage C Rectal Carcinoma in Control and Irradiated Groups

Survival of Patients with Dukes’ Stage C Rectal Carcinoma in Control and Irradiated Groups

PMH--Toronto Study
# = no. at risk

2p = 0.01
Survival by Treatment Allocated

Survival rate, %

Med. Research Council Study

Time, mo

No XRT (275)
Single fraction (277)
Multiple fractions (272)
Survival by Treatment for Dukes’ C Cases

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

Survival rate, %

Time, mo

Med. Research Council Study
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:
Maternity Wards, Baseball & Clinical Research
Survival of Patients with Dukes’ Stage C Rectal Carcinoma in Control and Irradiated Groups

Survival %

PMH--Toronto Study

# = no. at risk
Survival by Treatment for Dukes’ C Cases

Survival rate, %

Time, mo

0 6 12 18 24 30 36 42 48 54 60 66

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

Med. Research Council Study
Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet ’86)
  - SK reduces mortality by 20%
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
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**
Can Efficacy or Safety Signals Discovered in Exploratory Analyses Be Viewed to be Reliable Results?

- Criteria to be simultaneously satisfied:
  - $< < P$-values (e.g., Natalizumab & PML & Carvedilol in Heart Failure)
  - Biologically plausible effect
  - Confirmed by external results

White Paper Illustration
Survival of Patients with Dukes’ Stage C Rectal Carcinoma in Control and Irradiated Groups

PMH--Toronto Study
# = no. at risk
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 Of Colorectal Cancer

NCCTG Trial

Cancer Intergroup Trial

Years from randomization

Surviving, %

Years from randomization

5-FU+LEV  n=91
Levamisole  n=85
Control  n=86

5-FU+LEV  n=304
Levamisole  n=310
Control  n=315
Of all experimental interventions studied in colon adjuvant, suppose only 4% are truly positive & 96% are truly negative.

Suppose the "false negative error rate" is $\beta = 0.10$

(so the "statistical power" is $1 - \beta = 0.90$)

& Suppose the "false positive error rate" is $\alpha = 0.025$

Then, the probability a trial positive will be a true positive is $\frac{36}{60} = 0.60$

<table>
<thead>
<tr>
<th>RESULT OF EXPERIMENT</th>
<th>TRUTH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
<td>36</td>
</tr>
<tr>
<td>Negative</td>
<td>Positive</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>40</td>
</tr>
</tbody>
</table>
Of all experimental interventions studied, suppose 60% are truly positive & 40% are truly negative

Suppose the “false negative error rate” is $\beta = 0.10$
(since the “statistical power” is $1-\beta = 0.90$)
& Suppose the “false positive error rate” is $\alpha = 0.025$

Then, the probability a trial positive will be a true positive is $\frac{540}{550} = 0.98$

<table>
<thead>
<tr>
<th>RESULT OF EXPERIMENT</th>
<th>TRUTH</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>540</td>
<td>10</td>
</tr>
<tr>
<td>Negative</td>
<td>60</td>
<td>390</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>400</td>
</tr>
</tbody>
</table>
Surgical Adjuvant Therapy Of Colorectal Cancer

NCCTG Trial

- 5-FU+LEV n=91
- LEV n=85
- Control n=86

Years from randomization

Cancer Intergroup Trial

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

Years from randomization
“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)
Some Conclusions

• 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
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:

Maternity Wards, Baseball & Clinical Research

20 vs 2: (.71, .99), 2p = 0.0001
Meta-Analysis: 31 vs 13: (.55, .83), 2p = 0.0096
Bias for “Positive” Results in Clinical Trials

➢ Protocol Specified Primary Objective of the Clinical trial:

• Very frequent wording:
  ~ “To establish that the experimental regimen is safe and effective”
Bias for “Positive” Results in Clinical Trials

➢ Protocol Specified Primary Objective of the Clinical trial:

• 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”

…building a story with supportive analyses…
Bias for “Positive” Results in Clinical Trials

…Andrew Fleming’s insight from Psychology…

“Cognitive Dissonance”

…The Harvard Professor’s Course…

…The Apparent Lack of Benefit in Males…
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
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
Principles & Insights

“The Goal of Clinical Research:
“The Goal of Clinical Research:
To Determine Whether,
Not to Establish,
the Experimental Regimen
Is Safe and Effective”

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