Clinical trials: Discerning Hype from Substance

May 5, 2011
Thomas R. Fleming, Ph.D.
Professor of Biostatistics
University of Washington
fleming@seattlebiostat.com

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

Confirmatory vs. Exploratory Analyses

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

Illustrations and Motivation:

- 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: @12 wk, 16 wk, 18 wk, etc.
    ~ Time to Clinical Worsening
      ✓ Death, PAH Hosp, L.T., (NYHA↑ & 6MWT↓ & Rx Δ)
  - Analysis Methods
    ~ Normally distributed: T-test, ANOVA, Wilcoxon
    ~ Time to event: Logrank, Cox Regression
    ~ Dichotomous: Fisher’s Exact Test, Pearson χ2
Confirmatory vs. Exploratory Analyses

- Biomarker Endpoints (Hemodynamic parameters)
  - Pulmonary & Systemic BP
  - Systolic & Diastolic Pulmonary Arterial Pressure
  - Systolic & Diastolic Systemic Arterial Pressure
  - Systemic & Pulmonary Vascular Resistance
  - Heart Rate
- Analyses over Calendar Time
- ~ Normally distributed: T-test, ANOVA, Wilcoxon
- ~ Time to event: Logrank, 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

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

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

Females Only

- 5-FU+levamisole
- Follow-Up Only

<table>
<thead>
<tr>
<th>At Risk</th>
<th>Death</th>
<th>5-Yr Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>21</td>
<td>57%</td>
</tr>
<tr>
<td>43</td>
<td>29</td>
<td>40%</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>37</td>
<td>25</td>
<td>51%</td>
</tr>
<tr>
<td>38</td>
<td>24</td>
<td>47%</td>
</tr>
</tbody>
</table>
Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial  
Cancer Intergroup Trial

INTERGROUP STUDY 0035  
Looking at Treatment Effect on Overall Survival

Females Only  
Males Only

Duke’s C Colon Cancer Adjuvant

Relative Risk of Survival:  5-FU + Levamisole  
Control

<table>
<thead>
<tr>
<th>Analysis Group</th>
<th>North Central Group Study (n = 162)</th>
<th>Intergroup Study # 0035 (n = 619)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td>0.72</td>
<td>0.67</td>
</tr>
<tr>
<td>Female</td>
<td><strong>0.57</strong></td>
<td><strong>0.85</strong></td>
</tr>
<tr>
<td>Male</td>
<td><strong>0.91</strong></td>
<td><strong>0.50</strong></td>
</tr>
<tr>
<td>Young</td>
<td>0.60</td>
<td>0.77</td>
</tr>
<tr>
<td>Old</td>
<td>0.87</td>
<td>0.59</td>
</tr>
</tbody>
</table>

How Reliable are Exploratory Subgroup Analyses?

- Examples:
  - 5-FU + Levamisole in Adjuvant Colon Cancer  
    NCCTG; Cancer Intergroup 0035
  - Pre-operative RT in Rectal Cancer  
    Princess Margaret Hospital & M.R.C.
Survival of Patients with Rectal Carcinoma in Control and Irradiated Groups

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

Survival by Treatment Allocated

Survival by Treatment for Dukes’ C Cases
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

---

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

Survival by Treatment for Dukes’ C Cases

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

  - 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., Tysabri & PML)
  - Biologically plausible effect
    - White Paper Illustration
  - Confirmed by external results

Surgical Adjuvant Therapy Of Colorectal Cancer

Pre-trial probability the intervention is effective is $\pi = 0.04$

Then, even if $\alpha = 0.025$ and $1-\beta = 0.90$,

Probability a trial positive will be a true positive is $36 \div 60 = 0.60$

RESULT OF TRUTH

<table>
<thead>
<tr>
<th>EXPERIMENT</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>36</td>
<td>24</td>
</tr>
<tr>
<td>Negative</td>
<td>4</td>
<td>936</td>
</tr>
</tbody>
</table>

| 40       | 960      | 1000     |
Pre-trial probability the intervention is effective is \( \pi = 0.60 \) (when 1st trial is positive)

Then, if \( \alpha = 0.025 \) and \( 1-\beta = 0.90 \),

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>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>540</td>
</tr>
<tr>
<td>Negative</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>1000</td>
</tr>
</tbody>
</table>

---

“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)
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”

~ 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 reliably answers the questions the trial was designed to address”
Hazards of evaluating interventions
Using observational databases

Mega doses of Vitamin C:
What is the effect on duration of survival in pre-terminal cancer patients?

- Linus Pauling: Loch Lomanside, Scotland
  Cameron, Pauling. Proc Natl Acad Sci 1976; 1978
  Median Survival: 50 vs. 210 days; 38 vs. 293 days

- Mayo Clinic sponsored randomized trial
  Moertel, Fleming, Creagan et. al. NEJM 1985; 312: 137-141

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:
  - 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.

Some Initial Conclusions

- *P*-values are only interpretable when you understand the sampling context from which they were derived
- Point estimates and confidence intervals are preferable to reporting *P*-values when findings about treatment effects from exploratory analyses are presented.
- Random High bias is real
- Exploratory Analyses usually should be viewed to be “Hypothesis Generating”
- Confirmatory Trials greatly enhance the reliability of conclusions

Principles & Insights

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

The Mystery
Principles & Insights

The Substantial Consequences
Of Misleading Exploratory Analyses

Illustration: Confirmatory vs. Exploratory Analyses

Actimmune vs Placebo in IPF: 9/00 – 8/02
(Progression/Death: Target 20% vs 40% at 1 year)
Data Monitoring Committee Meeting: 8/19/02

<table>
<thead>
<tr>
<th>GIPF-001</th>
<th>N</th>
<th>Progression</th>
<th>Death/Prog</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actimmune</td>
<td>162</td>
<td>68 (42.0%)</td>
<td>75 (46.3%)</td>
<td>16 (18)</td>
</tr>
<tr>
<td>Placebo</td>
<td>168</td>
<td>75 (44.6%)</td>
<td>87 (51.8%)</td>
<td>28 (28)</td>
</tr>
<tr>
<td>RR</td>
<td>0.942</td>
<td>RR = 0.894</td>
<td>2p = 0.084</td>
<td>2p = 0.53</td>
</tr>
</tbody>
</table>

Safety Profile

<table>
<thead>
<tr>
<th>Safety Profile</th>
<th>Pulmonary SAEs</th>
<th>Pneumonia SAEs</th>
<th>Vascular Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actimmune</td>
<td>41</td>
<td>20</td>
<td>7</td>
</tr>
<tr>
<td>Placebo</td>
<td>34</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

Actimmune in IPF

- 8/19/02: DMC Meeting, releasing data to sponsor leadership…
  Recommendation for follow-up of secondary endpoints to 11/02
- 8/27/02: Sponsor—FDA Meeting: FDA recognized the trial did not establish efficacy; FDA: OK to conduct confirmatory trial

InterMune GIPF #001
Idiopathic Pulmonary Fibrosis (IPF)

Primary Endpoint: Progression-free Survival
✓ FVC ↓ by >10%
✓ A-a Gradient ↑ by > 5 mmHg
✓ Death

Enrollment: 9/00 - 9/01; Follow-up to 8/02

Secondary Endpoints:
✓ 10 listed, by “order of importance”
  …Overall Survival was 7th
Actimmune in IPF

- 8/19/02: DMC Meeting, releasing data to sponsor leadership… Recommendation for follow-up of secondary endpoints to 11/02
- 8/27/02: Sponsor—FDA Meeting: FDA recognized the trial did not establish efficacy; FDA: OK to conduct confirmatory trial
- 8/28/02: Sponsor released News Release: “InterMune Announces Phase III Data Demonstrating Survival Benefit of Actimmune in IPF…Reduces mortality by 70% in patients with mild to moderate disease, (p = 0.004). …The mortality benefit is very compelling and represents a major breakthrough in this difficult disease.” …interim data base with 16 vs. 28 deaths used …exploration of secondary endpoint, overall survival …exploratory subgroup of patients with “mild to mod disease”
- 9/02: Stock value rises; off label use of Actimmune in IPF soars …projected by sponsor to reach $400-$500 million/year.

- 9/5/02: DMC Chair Letter to Sponsor… “The claims for established survival benefit in the mild to moderate disease subgroup…are so fallacious that they would provide a humorous illustration of an absurd misrepresentation of exploratory statistical analyses if not for the serious consequences to patients, caregivers, and the investment community who might be mislead in their therapeutic and financial decision-making processes…I am calling on our colleagues at InterMune to identify corrective actions that could effectively address these serious misrepresentations…”

- 10/02-11/02: DMC Chair holds meetings with sponsor leadership… …authorities will be notified…preferable to be done by sponsor.

- 10/17/02: DMC discovers sponsor is conducting illegal off-label advertising (through Priority Healthcare Corporation)… “InterMune Announces Phase III Data Demonstrating Survival Benefit of Actimmune in IPF…Reduces Mortality by 70% in Patients with Mild to Moderate Disease”
Actimmune in IPF

- 10/02-11/02: DMC Chair holds meetings with sponsor leadership…authorities will be notified…preferable to be done by sponsor.
- Subsequently…
  ✓ Sponsor reports acknowledging inappropriate activities to authorities
  ✓ Company statistician resigns…
  ✓ More balanced presentation given on 11/05/02 at Am College of Chest Physicians Meeting

- ‘04-’07: Sponsor conducts GIPF-007 trial, evaluating Actimmune effect on survival in 826 IPF patients with mild-to-moderate disease

Actimmune in Mild-to-Moderate IPF

- GIPF #007 Trial: 2/04 to 10/07
  Primary Endpoint: Survival, Target RR = 0.50
- Data Monitoring Committee Meeting: 2/28/07
<table>
<thead>
<tr>
<th></th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>GIPF-007</td>
<td>551</td>
</tr>
<tr>
<td>Placebo</td>
<td>275</td>
</tr>
</tbody>
</table>

- ‘04-’07: Sponsor conducts GIPF-007 trial, evaluating Actimmune effect on survival in 826 IPF patients with mild-to-moderate disease
- 1/05: FBI Agents approach UW SPHCM Dean, seeking T. Fleming…subpoena issued to appear before attorneys for US Dept of Justice…Deferred Prosecution Agreement
Actimmune in Mild-to-Moderate IPF

- GIPF #007 Trial: 2/04 to 10/07
  Primary Endpoint: Survival, Target RR = 0.50

- Data Monitoring Committee Meeting: 2/28/07
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actimmune</td>
<td>551</td>
<td>80 (14.5%)</td>
</tr>
<tr>
<td>Placebo</td>
<td>275</td>
<td>35 (12.7%)</td>
</tr>
</tbody>
</table>

- 3/5/07: Sponsor Press Release: “the DMC found the overall survival result crossed a predefined stopping boundary for lack of benefit of Actimmune® relative to placebo” and where overall mortality was “14.5% in the Actimmune group as compared to 12.7% in the placebo group.”

Some Initial Conclusions

- P-values are only interpretable when you understand the sampling context from which they were derived
- Point estimates and confidence intervals are preferable to reporting P-values when findings about treatment effects from exploratory analyses are presented.
- Random High bias is real
- Exploratory Analyses usually should be viewed to be “Hypothesis Generating”
- Confirmatory Trials greatly enhance the reliability of conclusions

Actimmune in IPF

- 8/28/02: Sponsor released News Release:
  “InterMune Announces Phase III Data Demonstrating Survival Benefit of Actimmune in IPF...Reduces mortality by 70% in patients with mild to moderate disease, (p = 0.004)....The mortality benefit is very compelling and represents a major breakthrough in this difficult disease.”

- 3/5/07: Sponsor Press Release:
  “the DMC found the overall survival result crossed a predefined stopping boundary for lack of benefit of Actimmune® relative to placebo”... overall mortality was “14.5% in the Actimmune group as compared to 12.7% in the placebo group.”

Principles & Insights

“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 Additional Conclusions

- There should be a clear specification of and focus on the pre-specified primary analyses of the primary and secondary end points when submitting results for peer review and when disseminating results in press releases and journal publications.
- Protocols should have at most 3 or 4 pre-specified secondary analyses to further address multiplicity.
- Recognize strong bias for achieving “positive” results
- When refereeing journal publications, request:
  - the clinical trial protocol
  - the statistical analysis plan (i.e., the SAP)
  - the clinical study report

Some Additional Conclusions

- The criteria used by journal editors and reviewers in evaluating manuscripts should be based on the importance of the questions that the studies are designed to address and the quality of study conduct rather than on the level of positivity of study results.
- All clinical trials should be registered with ClinicalTrials.gov to reduce publication bias.
- It should be recognized that bias will persist if meta-analyses include the hypothesis-generating trial.

Principles & Insights

Evaluating Safety

Illustration: Cancer Risk with Vytorin in Slowing progression of Aortic-Valve Stenosis

- **SEAS Trial**
  - Relative Risk: 1.55
  - 95% C.I.: (1.13, 2.12)

Challenge:
- Interpreting safety signals from exploratory analyses
Can Efficacy or Safety Signals Discovered in Exploratory Analyses Be Viewed to be Reliable Results?

- Criteria to be simultaneously satisfied:
  - $< < P$-values (e.g., Tysabri & PML)
  - Biologically plausible effect
    - Ezetimibe blocks the absorption of phytosterols and other phyt nutrients linked to protection against cancer; which provides some biologic plausibility that the drug could have an effect on the growth of cancer cells
  - Confirmed by external results

Illustration: Cancer Risk with Vytorin in Slowing progression of Aortic-Valve Stenosis

- **SEAS Trial**
  - N
  - CA. Incidence
  - CA. Deaths
  - Vytorin: 944, 101, 37
  - Placebo: 929, 65, 20
  - Relative Risk: 1.55, 1.78
  - 95% C.I.: (1.13, 2.12), (1.03, 3.11)

- **IMPROVE-IT & SHARP Trials**
  - N
  - CA. Incidence
  - CA. Deaths
  - Vytorin: 10,391, 313, 97
  - Control: 10,298, 326, 72
  - Relative Risk: 0.96, 1.34
  - 95% C.I.: (0.82, 1.12), (0.98, 1.84)

Interpreting the SEAS, IMPROVE-IT & SHARP Trials Regarding Cancer Risk with Vytorin

- **Peto et. al. (NEJM, 2008)**
  - The available results from these 3 trials do not provide credible evidence of any adverse effect of ezetimibe on rates of cancer.

- However, safety is established by ruling out unacceptable increases in safety risks...
  - ...i.e. by what you can say, not what you can’t say...
  - Fleming (NEJM, 2008)
"Absence of Evidence Is not Evidence of Absence"


Characteristics Integral to being Effective as a Biostatistical Collaborator

- Statistical Science: Strong Training in Theory
- Insights into the Art: Appropriately applying Scientific Methods to effectively address real world problems
- Communication Skills, Oral and Written: Ability to collaborate effectively with non-statistical scientists
  - Passionate and Organized

An Ideal We Should Strive to Achieve

- The Clinician’s job description for a Biostatistician
- Not SSE (ref, Mary Foulkes)… rather, True Collaboration

The question initially posed by our colleagues usually is not what most needs to be answered

- Why aren’t Biostatisticians more widely engaged as true collaborators? Is it on us?
An Ideal We Should Strive to Achieve

✓ Rather than “Statistical” or “Clinical” Issues, collectively we and our collaborators face “Scientific” Issues

✓ Susan Ellenberg: lessons from an ideal Data Monitoring Committee…

An Ideal We Should Strive to Achieve

✓ The “Conscience of the Research Team” (William Taylor, Mayo Clinic, 1977)

…Keeping in mind the Principles & Insights…

…Bringing Objectivity, Seeking to “Determine Whether”…

…Having the Courage To Advocate for the Truth.