Control of False Positives in Phase III Randomized Clinical Trials

Changyu Shen
Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology
Beth Israel Deaconess Medical Center
Harvard Medical School
09/27/2017
Background

• Phase III randomized clinical trials (RCTs) serve as the gold-standard for the evaluation of the efficacy of a medical intervention; it is the last layer of the evaluation process before a product is on market.

• Although rigorous statistical examination assures a small probability of false positive for a given trial, the large number of Phase III RCTs from the biopharmaceutical industry in the United States could substantially inflate the chance of false positives.
Objective

• Control of expected number of false positives among all Phase III trials during a fixed time period overall
• Allow false positive control of individual trial to be more flexible
Proportion of Null or Negative (PNN)

- PNN ($\rho$): proportion of comparisons in a population of trials where the intervention has a null or negative efficacy
Idea 1: fixed type I error rate

• If we know $\rho$, the expected number of false positives (ENFP) under a constant type I error control at $\alpha$ for $n$ trials is no more than $n\rho\alpha$.
• If we can obtain an estimate of $\rho$, $\hat{\rho}$, then
  \[
  \text{ENFP} = n\hat{\rho}\alpha
  \]
• Control ENFP means control of the number of trials
Idea 2: flexible type I error rate

• Suppose $X$ represents a vector of measurements with respect to the medical intervention, the disease and other factors that could affect our decision on the level of type I error, e.g. earlier phase results

• Let $\rho(X) = \Pr[\text{Null or negative efficacy}|X]$ and $\alpha(X)$ be the type I error rate as a function of $X$, $\alpha$ spending function

• $ENFP = nE[\rho(X)\alpha(X)]$
Idea 2 Cont.

• If \( \text{Corr}[\rho(X), \alpha(X)] \leq 0 \), then
  \[
  E[\rho(X)\alpha(X)] \leq E[\rho(X)] E[\alpha(X)] = \rho E[\alpha(X)]
  \]
  \[
  \text{ENFP} \leq n \rho E[\alpha(X)] \approx \hat{\rho} \sum_{i=1}^{n} \alpha_i
  \]
• Control of ENFP means control of \( \sum_{i=1}^{n} \alpha_i \)
• \( \sum_{i=1}^{n} \alpha_i \) can be viewed as the “error budget” that can be un-uniformly spent among trials
  – If we want to control ENFP in a 5-year period to be lower than 1, then the error budget is \( 1/\hat{\rho} \).
Idea 2 Cont.

- **Advantage**
  - Only require $\text{Corr}[\rho(X), \alpha(X)] \leq 0$
  - Flexible type I error control for each individual trial
  - No need of the knowledge of $\rho(X)$
Idea 2 Cont.

• Why flexible type I error control?
  – Earlier phase studies offer different evidence of efficacy. It is not efficient to stick to the same level of type I error rate for all trials
  – Potential savings in cost and time with relaxed type I error rate
  – Ethical considerations: less patients in the control arm?
  – More stringent type I error control to prevent false positive with substantial public health impact?
How to estimate $\rho$?
Formulation of the estimation of $\rho$

- The treatment effect $\beta$ (e.g. mean difference, logarithm of odds ratio or hazard ratio etc., where positive value means efficacy) can be estimated by $\hat{\beta}$ that has (approximately) a normal distribution with mean $\beta$ and standard error $\sigma$
- $Z = \hat{\beta} / \sigma$ is (approximately) normally distribution with mean $\theta = \beta / \sigma$ and standard deviation 1
- $\theta$ is amplified effect size; $Z$ is a noisy version of $\theta$
- $\text{PNN}=\rho=\Pr(\theta \leq 0)$
- Objective is to estimate the distribution of $\theta$ using data on $Z$ (i.e. deconvolution)
Deconvolution methods

• Nonparametric method has a very slow convergence rate
• Parametric method is sensitive to model misspecification
• Efron (2015) proposed a semi-parametric method, where the distribution of $\theta$ (at the logarithm scale) is in the class of natural cubic spline with various degrees of freedom; Maximum likelihood estimation (MLE) or penalized maximum likelihood estimation (PMLE) can be used for estimation
Clinicaltrials.gov

• Section 801 of the Food and Drug Administration Amendments Act (FDAAA) mandates registration at clinicaltrials.gov (ct.gov) for all Phase III clinical trials on drugs, medical devices and biologics that were initiated after September 27, 2007, or were ongoing as of December 26, 2007 (Anderson et al., 2015; Zarin et al., 2011, 2015).

• Trials registered at ct.gov offers the most unbiased coverage of trials (registration is required REGARDLESS of trial results)
Selection of trials

• Inclusion criteria
  – Randomized Phase III trials
  – Primary completion date in 2008-2012
  – Funded at least in part by an industrial organization
  – FDA as the oversight authority
Exclusion criteria

• Exclusion criteria
  – Primary endpoint(s) is(are) not clinical efficacy (e.g. safety, compliance, pharmacokinetics or pharmacodynamics ) (EC1)
  – Primary objective is for non-inferiority/equivalence or at least the trial is powered to test non-inferiority/equivalence (EC2)
  – Comparative effectiveness studies where both interventions are treated the same way (no control) (EC3)
  – Withdrawn prior to enrollment (EC4)
  – Unavailability of meaningful analysis due to early termination of the trial that led to overly small sample size, low enrollment, inability to collect primary endpoints for most enrolled subjects, decision of not doing efficacy analysis, or poor compliance (EC5) (note that this criterion does not include those trials terminated after a formal interim analysis)
  – No between-arm comparison was planned by trial design (EC6)
  – Other (EC7)
Is comparison result available from CT.gov?
  Yes → Record result
  No → Is comparison result available from publications?
    Yes → Record result
    No → Is comparison result computable based on outcome measure on CT.gov?
      Yes → Compute result*
      No → Is comparison result available or computable from other sources, e.g. abstract, press release etc.?
        Yes → Record/compute result
        No → Is comparison result available from sponsor?
          Yes → Record result
          No → Missing value
1,985 trials meet the inclusion criteria as of 03/03/2015 based on CT.gov query

583 trials excluded

1,401 trials

8 trials excluded

1,393 trials included in the study

EC7:
- 3 not completed
- 1 completed before 2008
- 2 longer follow-up studies
- 2 combined with other trials for merged analysis

1,221 trials with results

172 trials without results
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Trials with result (N=1221)</th>
<th>Trials without result (N=172)</th>
<th>Total (N=1393)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary purpose of trial—no. (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1147 (93.9)</td>
<td>161 (93.6)</td>
<td>1,308 (93.9)</td>
</tr>
<tr>
<td>Prevention</td>
<td>61 (5.0)</td>
<td>11 (6.4)</td>
<td>72 (5.2)</td>
</tr>
<tr>
<td>Supportive care</td>
<td>7 (0.6)</td>
<td>0</td>
<td>7 (0.5)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>6 (0.5)</td>
<td>0</td>
<td>6 (0.4)</td>
</tr>
<tr>
<td><strong>Intervention type—no. (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug</td>
<td>1061 (86.9)</td>
<td>150 (87.2)</td>
<td>1,211 (86.9)</td>
</tr>
<tr>
<td>Biological</td>
<td>101 (8.3)</td>
<td>8 (4.7)</td>
<td>109 (7.8)</td>
</tr>
<tr>
<td>Device</td>
<td>57 (4.7)</td>
<td>14 (8.1)</td>
<td>71 (5.1)</td>
</tr>
<tr>
<td>Procedure</td>
<td>2 (0.2)</td>
<td>0</td>
<td>2 (0.1)</td>
</tr>
<tr>
<td><strong>Allocation arms—no. (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>756 (61.9)</td>
<td>113 (65.7)</td>
<td>869 (62.4)</td>
</tr>
<tr>
<td>3</td>
<td>277 (22.7)</td>
<td>37 (21.5)</td>
<td>314 (22.5)</td>
</tr>
<tr>
<td>&gt;3</td>
<td>188 (15.4)</td>
<td>22 (12.8)</td>
<td>210 (15.1)</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Trials with result (N=1221)</td>
<td>Trials without result (N=172)</td>
<td>Total (N=1393)</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>----------------------------</td>
<td>--------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Primary endpoint—no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>935 (76.6)</td>
<td>138 (80.2)</td>
<td>1073 (77.0)</td>
</tr>
<tr>
<td>2</td>
<td>193 (15.8)</td>
<td>31 (18.0)</td>
<td>224 (16.1)</td>
</tr>
<tr>
<td>&gt;2</td>
<td>93 (7.6)</td>
<td>3 (1.7)</td>
<td>96 (6.9)</td>
</tr>
<tr>
<td>Positive result—no. (%)‡</td>
<td>789 (64.6)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Positive with two-sided</td>
<td>561 (45.9)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>p&lt;0.001—no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size—median (IQR §)</td>
<td>471 (250-796)</td>
<td>324 (177-654)</td>
<td>455 (240-787)</td>
</tr>
<tr>
<td>Primary completion year—no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>321 (26.3)</td>
<td>45 (26.2)</td>
<td>366 (26.3)</td>
</tr>
<tr>
<td>2009</td>
<td>273 (22.4)</td>
<td>38 (22.1)</td>
<td>311 (22.3)</td>
</tr>
<tr>
<td>2010</td>
<td>211 (17.3)</td>
<td>23 (13.4)</td>
<td>234 (16.8)</td>
</tr>
<tr>
<td>2011</td>
<td>215 (17.6)</td>
<td>26 (15.1)</td>
<td>241 (17.3)</td>
</tr>
<tr>
<td>2012</td>
<td>201 (16.5)</td>
<td>40 (23.3)</td>
<td>241 (17.3)</td>
</tr>
</tbody>
</table>
Sensitivity Analysis

• Trials without results represent missing data that may bias the analysis

• In the sensitivity analysis, we assume that the distribution of the Z values for trials without results follow the same empirical distribution of the observed Z values less than 1.96 (or p-value > 0.05)
  – Rationale: publication bias
  – Conservative
## Result

<table>
<thead>
<tr>
<th>Parameter estimate (95% CI)</th>
<th>Missing Complete at Random</th>
<th>Missing Not at Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNN</td>
<td>7.2% (5.6-8.8%)</td>
<td>9.0% (6.8-11.1%)</td>
</tr>
<tr>
<td>Expected # of false positives among the 1393 selected comparisons</td>
<td>2.5 (1.9-3.1)</td>
<td>3.1 (2.4-3.8)</td>
</tr>
<tr>
<td>PIT</td>
<td>18.1% (14.1-22.1%)</td>
<td>21.8% (16.5-26.9%)</td>
</tr>
<tr>
<td>Expected # of trials with at least one false positive comparison among the 1393 selected trials</td>
<td>6.3 (4.9-7.7)</td>
<td>7.6 (5.7-9.4)</td>
</tr>
</tbody>
</table>
Questions

• Is this a big deal?
• Is this practical?
• Validation, more data?
• FDA?
• Anything else?
Thank you!