Strengthening the Interpretation of Clinical Trial Data

Jitendra Ganju
Ganju Clinical Trials, LLC
jganju@yahoo.com

Joint work with J Ma, X Yu, K Zhou, Y Lin
Convention

- We prespecify the endpoint and **analysis method**
- **Formal interpretation** relies on **ONE method**
  
  What if we guessed wrong?
Example

Same data, two models, different results

<table>
<thead>
<tr>
<th>Model</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.2274</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Data from chapter 9 (lots 1 and 2) of *An introduction to the bootstrap*, Efron and Tibshirani (1993)
When single method is risky

- Rare disease, small N
- Complex clinical trials
- Risk in method tied to our experience with endpoint
  HbA1c, FEV1, …, 6MWT, time to event, recurrent event, new PRO, days hospitalized

less experience
Assumption violation

Figure 2. Kaplan–Meier Estimates of the Incidence of Death from Coronary Heart Disease and Nonfatal Myocardial Infarction in the Gemfibrozil and Placebo Groups. The relative risk reduction was 22 percent (P=0.006), as derived from a Cox model.

Rubins et al. NEJM 1995
Proposal

• Prespecify more than one method

• Combine p-values. Control alpha

Robust, more power, flexible
## Robustness

### Covariate transformation?

<table>
<thead>
<tr>
<th>Model</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(X)</td>
<td>0.2274</td>
</tr>
<tr>
<td>No transformation</td>
<td>0.0004</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td><strong>0.0040</strong></td>
</tr>
</tbody>
</table>

Data from chapter 9 (lots 1 and 2) of *An introduction to the bootstrap*, Efron and Tibshirani (1993)
Robustness

Endpoint transformation?

<table>
<thead>
<tr>
<th>Model</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Y)</td>
<td>0.02</td>
</tr>
<tr>
<td>No transformation</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td><strong>0.03</strong></td>
</tr>
</tbody>
</table>

Data from Edwards, Stat Med 1999. Full model fit in each case
Robustness

Different metrics and analysis methods

% change from baseline or raw scores?

**Power %**

- T-test: 27%
- Wilcoxon raw scores: 86%
- Combined: 78%

Endpoint is count data
Data from a mixture of Poisson distributions
More Power

Small N, many covariates

N = 20, covariates = 16

Combined method gives more power than any single method

Combined includes 3 methods: one with lowest power, and the other 2 include different subsets of covariates
Versatility

• Group sequential trials
  
  Convention: Same single method at each interim analysis

  Combined methods more flexible

• It’s not just interpretation, trial may stop earlier
### Versatility

Different methods at interim and final, and multiple methods at each time

<table>
<thead>
<tr>
<th></th>
<th>Convention</th>
<th>New 1</th>
<th>New 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interim</strong></td>
<td>LR</td>
<td>wLR</td>
<td>LR, wLR</td>
</tr>
<tr>
<td><strong>Final</strong></td>
<td>LR</td>
<td>LR</td>
<td>Cox1, Cox2</td>
</tr>
</tbody>
</table>

As before, combined methods robust

LR = logrank, wLR = weighted logrank, Cox1 and Cox2 are different Cox models
Remarks

- **Limitation**: combining p-values method doesn’t give estimate of treatment effect

- To build experience, can start using as complementary method
  Method applies to efficacy or safety endpoints

- Many ways to combine: e.g. min p-value, Fisher’s combination
  - Alpha control is via permutations
Robust inference

More power

Group sequential trials
Assumption violation

Figure 3. Kaplan–Meier Analysis of the Time to a Definite Nonfatal Myocardial Infarction (Panel A), Death Definitely from Coronary Heart Disease (Panel B), Death from All Cardiovascular Causes (Panel C), Death from Noncardiovascular Causes (Panel D), and Death from Any Cause (Panel E), According to Treatment Group.

PREVENTION OF CORONARY HEART DISEASE WITH PRAVASTATIN IN MEN WITH HYPERCHOLESTEROLEMIA

NEJM, Shepherd et al, 1995
Assumption violation

B  Analysis of All-Cause Mortality

Probability of Survival

Hazard ratio, 0.70 (95% CI, 0.51–0.96)

Months since First Dose

Pooled tafamidis

Placebo

NEJM, Maurer et al, 2018
Robustness

**Endpoint:** % change from baseline in Disability Index of Health Assessment Questionnaire in patients with rheumatoid arthritis

<table>
<thead>
<tr>
<th>Model</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test</td>
<td>0.14</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td><strong>0.04</strong></td>
</tr>
</tbody>
</table>

Data from RCT using subset of trial data. N ≈ 60/group