Seeking Harmony: Estimands and Sensitivity Analyses for Randomized Clinical Trials

Devan V. Mehrotra*

* devan_mehrotra@merck.com

ASA-BIOP Regulatory-Industry Statistics Workshop
Washington DC, 9-13-18
Outline

- ICH E9/R1: why and what?
- Continuous endpoint
  - Case study (diabetes)
  - Case study (depression)
- Time-to-event endpoint
  - Hazard ratio estimands
  - Example
- Conclusions
ICH E9/R1: Why?

- Feedback from regulatory statisticians on several clinical trial protocols & new drug applications:
  - Insufficient clarity in objectives and related treatment effect parameters (i.e., estimands) of interest
  - Lack of logical connectivity between trial objectives, design, conduct, analysis and interpretation
  - Misalignment between “missing data” analysis methods and estimands of interest

2011 FDA advisory committee for dapagliflozin

- **Sponsor**: Remove data after initiation of rescue medication
  - Switch to rescue medication

- **FDA**: Include all data regardless of rescue medication
  - Switch to rescue medication
ICH E9/R1: A Structured Framework

For a given trial objective: aligning target of estimation, design, method of estimation and sensitivity analysis

**Trial Objective**

**Estimand**

**Main Estimator**

**Main Estimate**

**Design**

**Data**

**WHAT to estimate?**

**HOW to estimate?**

**Sensitivity Estimator 1**

**Sensitivity Estimate 1**

**Sensitivity Estimator 2**

**Sensitivity Estimate 2**

Sensitivity analysis (assess key assumptions)
ICH E9/R1: Inputs for Defining an Estimand

A Population

B Variable (Endpoint)

C Intercurrent Event(s)

D Population-level Endpoint Summary

- Treatment discontinuation due to adverse events
- Treatment discontinuation due to lack of efficacy
- Use of rescue medication
- Treatment switch
- Death
Case Study: Diabetes (HbA1c; asymptomatic)

Primary Objective: assess whether drug is more effective than placebo in lowering HbA1c without rescue medication (the latter was allowed, but to address a different objective)

**Construction of Primary Estimand**

**Population**: adults with type II diabetes (per intended label)

**Endpoint**: HbA1c change from baseline at 24 weeks

**Intercurrent event**: rescue medication

Treatment effect of interest is based on envisioned endpoint under complete follow-up even if the assigned treatment is discontinued, but without rescue medication at any time

**Population-level summary**: mean of the endpoint
Case Study (continued)

- **Estimand**: between-treatment difference in target population endpoint means for the treatment effect of interest ($\delta$)

- **Statistical objectives**
  - Deliver acceptable point estimate and 95% CI for $\delta$
  - Test $H_{null}: \delta=0$ vs. $H_{alt}: \delta<0$ (with type 1 error rate $\leq \alpha$)

- **Tackling rescue medication in the analysis**
  - Given estimand of interest, HbA1c values after initiation of rescue medication can be discarded, resulting in “missing” endpoint data for such patients (and dropouts)

- **Analysis challenge**: all randomized patients need to be included in the analysis (per the estimand), so how do we tackle the missing endpoint data problem?
Case Study (continued)

% missing endpoint: 20% (33/165) Drug  
21% (34/164) Placebo

1 patient assigned to drug and 2 patients assigned to placebo were dropouts before week 6
Control-based mean imputation is one of many options that can be considered in the pre-specified analysis plan.

- Obs = endpoint observed, miss = endpoint missing
- $\pi_i^{\text{miss}} = \text{true Pr(endpoint missing under trt } i) = 1 - \pi_i^{\text{obs}}$

<table>
<thead>
<tr>
<th>Placebo</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_P = \pi_P^{\text{obs}} \mu_P^{\text{obs}} + \pi_P^{\text{miss}} \mu_P^{\text{miss}}$</td>
<td>$\mu_D = \pi_D^{\text{obs}} \mu_D^{\text{obs}} + \pi_D^{\text{miss}} \mu_D^{\text{miss}}$</td>
</tr>
<tr>
<td>$\hat{\mu}_P = \hat{\pi}_P^{\text{obs}} \hat{\mu}_P^{\text{obs}} + \hat{\pi}_P^{\text{miss}} \hat{\mu}_P^{\text{miss}}$</td>
<td>$\hat{\mu}_D[c] = \hat{\pi}_D^{\text{obs}} \hat{\mu}_D^{\text{obs}} + \hat{\pi}_D^{\text{miss}} (\hat{\mu}_P + c)$</td>
</tr>
</tbody>
</table>

$\hat{\mu}_P^{\text{miss}} = \text{estimate of } \mu_P \text{ assuming missing endpoints are MAR for placebo}$

Estimand: $\delta = \mu_D - \mu_P$

Estimation: $\hat{\delta}[c] = \hat{\mu}_D[c] - \hat{\mu}_P$

**Primary Analysis:** use $c = 0$

**Sensitivity Analysis:** increase $c$ until Tipping Point is reached

**Details:** Mehrotra, Liu, Permutt (2017; Pharmaceutical Statistics)
Case Study (continued)

- MMRM: mixed model repeated measures assuming MAR dropout for drug and placebo [shown for historical reference only]
- Stretch: imputed mean for drug dropouts matches estimated mean for placebo dropouts assuming MAR dropout for placebo
- Note: tipping point after stretch imputation ⇒ robust evidence of drug effect
Case Study #2: Depression (HAMD17; symptomatic)

% missing endpoint: **24%** (20/84) Drug, **26%** (23/88) Placebo

![Graphs showing the study for depression with MMRM, Primary, Stretch, and Tipping Point analyses.](image)
HR Estimands for Time-to-Event Endpoints

Estimand for a homogeneous population
• $S_j(t) = \text{true proportion event-free at time } t \text{ under trt } j$

  Under proportional hazards: $\theta(t) = \frac{\log\{S_A(t)\}}{\log\{S_B(t)\}} = \theta$ for all $t$

  $\theta$ is the time-invariant hazard ratio under PH

Estimand for a mixture of homogeneous subpopulations
• Assume PH holds within each subpopulation (“stratum”)
  Define $\bar{\beta} = \sum_j f_j \beta_j$

  $f_j = \text{prevalence and } \beta_j = \log(\theta_j)$ for stratum $j$

  $\bar{\theta} = \exp(\bar{\beta})$ is the time-invariant average hazard ratio
Example: Simulated Dataset (N=400)

<table>
<thead>
<tr>
<th>Method</th>
<th>Hazard Ratio (HR) Estimate</th>
<th>95% CI</th>
<th>2-tailed p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox PH model</td>
<td>0.82</td>
<td>(0.62, 1.07)</td>
<td>.1398</td>
</tr>
</tbody>
</table>

Are results reliable?
Example (continued): Check PH assumption

Using Cumulative Sums of Martingale Residuals (Lin et al 1993; PHREG/ASSESS)

Checking Proportional Hazards Assumption for trt
Observed Path and First 20 Simulated Paths

Pr > MaxAbsVal: 0.0450 (1000 Simulations)

Not PH
Example (continued): Patients in Baseline Risk “Stratum 1”

<table>
<thead>
<tr>
<th>No. of patients</th>
<th>HR in Stratum 1</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.74</td>
<td>(0.49, 1.13)</td>
</tr>
</tbody>
</table>

**Product-Limit Survival Estimates**

With Number of Subjects at Risk

- **Trt A** (test)
- **Trt B** (control)

**PH**

**Checking Proportional Hazards Assumption for trt**

Observed Path and First 20 Simulated Paths

Pr > MaxAbsVal: 0.5500 (1000 Simulations)
Example (continued): Patients in Baseline Risk “Stratum 2”

<table>
<thead>
<tr>
<th>No. of patients</th>
<th>HR in Stratum 2</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td></td>
<td>0.74</td>
<td>(0.52, 1.07)</td>
</tr>
</tbody>
</table>

Product-Limit Survival Estimates
With Number of Subjects at Risk

Trt A (test)
Trt B (control)

PH

Checking Proportional Hazards Assumption for trt
Observed Path and First 20 Simulated Paths

Pr > MaxAbsVal: 0.1510 (1000 Simulations)
Example (continued): Summary of Results

### Analysis Method

<table>
<thead>
<tr>
<th>Analysis Method</th>
<th>Hazard Ratio (HR)</th>
<th>2-tailed p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstratified Cox PH (pooled)</td>
<td>0.82 (0.62, 1.07)</td>
<td>.1398</td>
</tr>
<tr>
<td>Stratified analysis (SSIZE wts)*</td>
<td>0.74 (0.56, 0.98)</td>
<td>.0362</td>
</tr>
</tbody>
</table>

* Mehrotra et al (2012; Statistics in Medicine)
Conclusions

- **ICH E9/R1 estimand & analysis framework is intended to:**
  - Enable better planning of clinical trials and application dossiers for new drugs/vaccines/biologics
  - Strengthen understanding of decision-making by regulatory authorities and advisory committees

- **Open items for estimands and sensitivity analyses:**
  - Time-to-event endpoints under PH/non-PH settings
  - Non-inferiority trials: different from superiority?
  - Treatment effects in well-defined but non-identifiable populations (principal stratification)
  - Others
Back-Up Slide
Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial


A: subgrp 1
(N=405, HR=0.75)

B: subgrp 1 + subgrp 2
(N=658, HR=0.83)

C: subgrp 1 + subgrp 2 + subgrp 3
(N=1314, HR=0.94)