Inclusion of Special Populations in Clinical Trials, and Inference

Jyotirmoy Dey (AbbVie Inc.)

Sep, 2018
Disclosure

The support of the presentation was provided by AbbVie. AbbVie participated in the review and approval of the content.

I, Jyotirmoy Dey, am an employee of AbbVie Inc.

The views expressed here are my own and not of AbbVie as a company.
Outline

• Problem statement and inference issues

• Examples and sampling considerations
  – Cardiovascular (CV), Neuroscience, Oncology

• Inference under disproportionate stratified sampling
Special Populations

- Vulnerable and under-studied groups of patients
  - Women (pregnant and non-pregnant)
  - Pediatric
  - Geriatric / Frail
  - Subjects with significant comorbidities
  - Patients with rare diseases
  - Minorities and other vulnerable populations

- **Problem:** Special populations are often under-sampled in clinical trials (CT)\(^1\-^7\)

- **Goal:** Studied population should be representative of the intended (or actual) use population\(^2\-^7\)
Inclusion of Special Population in Clinical Trials

• Challenges\textsuperscript{2,4,6,8}
  – Safety concerns
  – Lack of Compliance
  – Difficulties with recruitment
  – Heterogeneity of biology and treatment effects
  – Missing data and difficulty in measuring endpoints
  – Confounding and competing risks
  – Difficulties in assessing frailty
  – Unnecessary exclusion (e.g., age-related exclusion criterion)
Inclusion of Special Population in Clinical Trials

• Opportunities$^{2,3,4,8}$
  – Improve generalizability of trial results
  – Consideration of innovative/adaptive trial designs
  – Regulatory and payer carrots (exclusivity, other incentives) and sticks (label/coverage consistent with evidence)
  – Procedures similar to the Pediatric Written Request has been suggested for the Geriatric population$^5$
  – Understand how a disease changes due to biology and age/frailty
  – Understand how treatment effects change with biology
  – Assessment of baseline frailty – Comprehensive Geriatric Assessment (CGA), Short Physical Performance Battery (SPPB), Gait Speed
  – Improve sensitivity of treatment effect through enrichment
  – Utilize real-world evidence (RWE)
  – Master protocols for best treatment selection
Statistical Inference Issues due to Under-sampling

• Assess correlation between subgroups
  – Extrapolation of results to special populations needs additional clinical information and justification (ICH E7)
  – Mediated through physiological or metabolic factors
• Plan adequate and feasible sample-size
  – Uncertain efficacy and safety due to large variance when sub-sample is of inadequate size for meaningful inference
  – Question of “pool-ability” for overall treatment effect needs to be addressed by assessing if large differences between subgroups exist
• Avoid bias in pooled inference to ensure generalizability
  – Disproportionate stratified sampling may provide reasonable answers for each subgroup; Questions about appropriate overall effect estimates and significance remains (naïve pooling introduces bias)
**Cardiovascular Safety Non-Inferiority Study**

- Post-marketing study with design specs as proposed by the US FDA
  - Key Stipulations:
    - Time to MACE as primary endpoint
    - Non-inferiority (NI) margin: HR = 1.5; 256 MACE need to be observed
    - Study duration: *Approximately 5 years*
  - Assumptions:
    - 90% power; 20% “loss to follow-up”
    - 1.0% to 1.5% annual background rate of MACE

- FDA estimated 32000 Patient-years (PY) of exposure needed with 1% event rate; 21334 PY needed for 1.5% event rate

- Feasibility necessitates risk-based enrichment
**Real World Evidence for Risk Factors**

- Two large healthcare databases, Truven (a.k.a. MarketScan) and Optum explored for enrichment opportunities
  - Subjects < 44 years excluded; low background CV risk, user population is mostly middle-aged

![Diagram showing data on real-world evidence for risk factors](image-url)
### RWE: Baseline Factors

<table>
<thead>
<tr>
<th></th>
<th>Truven (146,936)</th>
<th>Optum (68,691)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Post index</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI or stroke</td>
<td>9,132 (6.2%)</td>
<td>5,865 (8.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>At Baseline</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-existing Diabetes</td>
<td>40,630 (27.7%)</td>
<td>18,881 (27.5%)</td>
</tr>
<tr>
<td>Pre-existing Dyslipidemia</td>
<td>102,867 (70.0%)</td>
<td>53,635 (78.1%)</td>
</tr>
<tr>
<td>Pre-existing Hypertension</td>
<td>91,956 (62.6%)</td>
<td>44,280 (64.5%)</td>
</tr>
<tr>
<td>Pre-existing Stage 3 CKD</td>
<td>2,985 (2.0%)</td>
<td>1,143 (1.7%)</td>
</tr>
<tr>
<td>Prior MI, stroke or PVD</td>
<td>23,175 (15.8%)</td>
<td>12,101 (17.6%)</td>
</tr>
<tr>
<td>Age ≤ 55</td>
<td>74,029 (50.4%)</td>
<td>38,393 (55.9%)</td>
</tr>
<tr>
<td>Age &gt; 65</td>
<td>15,125 (10.3%)</td>
<td>4,779 (7.0%)</td>
</tr>
<tr>
<td>Mean (SD); Range</td>
<td>56.33 (7.8); 45 – 98</td>
<td>55.16 (7.0); 45 – 89</td>
</tr>
</tbody>
</table>
RWE: Classification Tree for Truven and Optum Data

Preexisting MI, stroke or PVD?

Yes

- Age > 57?
  - High risk for MI or Stroke
    - Age > 65?
      - High risk for MI or Stroke
    - Age > 65?
      - Low risk for MI or Stroke

No

- Age > 54?
  - Low risk for MI or Stroke
    - Age > 65?
      - Low risk for MI or Stroke
    - Age > 65?
      - High risk for MI or Stroke

Preexisting MI, stroke or PVD?
# RWE: Evaluating the Classification Tree

<table>
<thead>
<tr>
<th>Node</th>
<th>Prediction</th>
<th>Truven # Subjects in node</th>
<th>Truven # Subjects with MI or stroke</th>
<th>Optum # Subjects in node</th>
<th>Optum # Subjects with MI or stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low risk</td>
<td>79,996</td>
<td>1,803 (2.3%)</td>
<td>39,513</td>
<td>1,655 (4.2%)</td>
</tr>
<tr>
<td>2</td>
<td>Low risk</td>
<td>34,860</td>
<td>1,567 (4.5%)</td>
<td>14,355</td>
<td>1,038 (7.2%)</td>
</tr>
<tr>
<td>3</td>
<td>High risk</td>
<td>89,05</td>
<td>890 (10%)</td>
<td>2,722</td>
<td>373 (13.7%)</td>
</tr>
<tr>
<td>4</td>
<td>High risk</td>
<td>23,175</td>
<td>4,872 (21%)</td>
<td>12,101</td>
<td>2,799 (23.1%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Node</th>
<th>Prediction</th>
<th>Optum tree applied to Truven # Subjects in node</th>
<th>Optum tree applied to Truven # Subjects with MI or stroke</th>
<th>Truven tree applied to Optum # Subjects in node</th>
<th>Truven tree applied to Optum # Subjects with MI or stroke</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Low risk</td>
<td>61,272</td>
<td>1,231 (2%)</td>
<td>31,135</td>
<td>1,189 (3.8%)</td>
</tr>
<tr>
<td>2</td>
<td>Low risk</td>
<td>53,584</td>
<td>2,139 (4%)</td>
<td>22,733</td>
<td>1,504 (6.6%)</td>
</tr>
<tr>
<td>3</td>
<td>High risk</td>
<td>8,905</td>
<td>890 (10%)</td>
<td>2,722</td>
<td>373 (13.7%)</td>
</tr>
<tr>
<td>4</td>
<td>High risk</td>
<td>23,175</td>
<td>4,872 (21%)</td>
<td>12,101</td>
<td>2,799 (23.1%)</td>
</tr>
</tbody>
</table>
Treatment for Advanced Parkinson’s Disease

- Duopa® (levodopa+ carbidopa) enteral suspension is a drug device combo used for treatment of advanced Parkinson’s disease
  - Robust effect seen in a randomized, double-blind, double-dummy study
  - 49% of patients were ≥ 65 years old; 8% were ≥ 75 years old

LCIG = Levodopa-Carbidopa (LC) Intestinal Gel system
LC-IR = LC Instant Release (oral formulation)

Figures as presented by Antonini and colleagues at the 65th American Academy of Neurology Annual Meeting in San Diego, CA, United States, in March, 2013.
**Treatment for Advanced Parkinson’s Disease**

- Geriatric use (Duopa® label): Increased risk for elevated BUN and CPK in ≥ 65 yr old compared to < 65 yr olds.
- In ≥ 75 yr olds, RCT participation was low. Participation in an ongoing registry study is much higher.
- Age-related condition, and hence no pediatric requirement
- Trial design considerations
  - Open-label trial was acceptable to EMA for approval.
  - Randomized double-blind, double-dummy trial requested by FDA.
    - Enrollment was very slow (it took 31 months to enroll 71 subjects)
    - Two pivotal studies (Study Nos. S187.3.001/S187.3.002) combined as a single pivotal mainly due to enrollment challenge. Analysis of pooled data as single pivotal accepted by FDA.
  - What would have been a more reasonable design? Are there better options?
**Geriatric representation in Glioblastoma**

- Let us compare age distribution of GBM patients in an epidemiologic study compared to two real RCTs (Phase 2 and Phase 3)
  - Distributions similar between phases, but lower than epidemiologic incidence. Is there a selection bias?
  - RCT subjects must have surgery and EGFR-amplified tumors; Brain surgery is rare at 80 years or beyond.
  - Naïve comparison is misleading.

**Epidemiology**

![Incidence Rates by Gender and Age Groups, 2000-2010](image)

**PHASE 2**

**PHASE 3**
**Targeted Therapy and Precision Medicine**

- Easier to extrapolate efficacy as long as target is present; biomarker threshold can be the key; with precision/personalized medicine, other biological differences may matter less
- More emphasis on Ph-1 work: PK profile of and dose selection for pediatric and frail patients
- Prognostic dependence\(^{12}\) is likely present across all ages; pts who are expected to live longer, benefit more from longer tx or a delayed tx effect

---

**CHECKMATE-141: SCCHN\(^{10-11}\)**

![Graph of Overall Survival vs Probability of Survival showing comparison between OPDIVO and Inv Choice](image1)

**CHECKMATE-057: Non-Sq.NSCLC\(^{10}\)**

![Graph of Overall Survival vs Probability of Survival showing comparison between OPDIVO and Docetaxel](image2)
Statistical Inference under Disproportionate Stratified Sampling for Time to Event
**Disproportionate Stratified Sampling**

- Disproportionate stratified sampling, or stratification with non-proportional allocation\(^1\) (cf. Cochran 1977), is a sampling scheme whereby the sample proportions of two or more strata are different from their proportions in the population.
Inference under Disproportionate Sampling

Proportion of younger subjects in
• Unenriched sample of size \( n = p = n_Y/n \)
• Enriched sample of size \( m = p' = n_Y/m \)

\( \beta(p, n) \): parameter from sample of \( n \) patients with proportion of young = \( p \).
Statistical problem: Estimate \( \beta(p, m) \) unbiasedly when the sampled proportion is \( p' \neq p \). We need a relationship between \( \beta(p, m) \) and \( \beta(p', m) \).
Estimation of Survival Functions

• A weighted Kaplan-Meier estimator suffices for estimation of the survival function

\[ \hat{S}(t) = \hat{p} \hat{S}_Y(t) + (1 - \hat{p}) \hat{S}_O(t) \]

– \( \hat{p} \) is an estimate of \( p \) (the true proportion of young in the general population)

– Its variance may be computed as:

\[
\text{Var}\{\hat{S}(t)\} = E\{\text{Var}\{\hat{S}(t)|\hat{p}\}\} + \text{Var}\{E\{\hat{S}(t)|\hat{p}\}\} \\
= E\{\hat{p}^2\} \text{Var}\{\hat{S}_Y(t)\} + E(1 - \hat{p})^2 \text{Var}\{\hat{S}_O(t)\} + (E\{\hat{S}_Y(t)\} - E\{\hat{S}_O(t)\})^2 \text{Var}\{\hat{p}\}
\]

– Point-wise inference follows using a Wald-type test statistic

– Procedure needed for overall inference, i.e., comparison of survival distributions, between treatment arms in an proportionate setting
Inference using the Hazard Ratio

• The hazard ratio (HR), henceforth denoted by \( \theta \), is a traditional measure for quantifying the treatment effect of TTE endpoints
  — Cox proportional hazards regression is the conventional method

• Can inference be drawn about a fixed HR in unenriched setting given stratum-specific estimates of the HR (\( \theta_k \) for k-th strata)?
  — Can we reduce this to a problem of comparison of means using asymptotic normality of log HR?
    • Note \( \log(\hat{\theta}_k) \sim N\left(\log(\theta_k), \frac{4}{d_k}\right) \) where \( d_k \) is the # of events in stratum k

• Comparisons using other measures are not currently accepted
  • E.g., Restricted Mean Survival Time (RMST)
Issue with Inference using Hazard Ratios

• Under prognostic and/or predictive stratification, statistical inference based on the HR cannot provide a clear answer about treatment effect under disproportionate sampling

• HR varies with time even under simple assumptions
  – Suppose we assume the effects to be as follows:

<table>
<thead>
<tr>
<th>Arm</th>
<th>Young</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>( \lambda_{C,Y} = \lambda_Y )</td>
<td>( \lambda_{C,O} = \lambda_O )</td>
</tr>
<tr>
<td>Treatment</td>
<td>( \lambda_{T,Y} = \lambda_Y \theta_Y )</td>
<td>( \lambda_{T,O} = \lambda_O \theta_O )</td>
</tr>
</tbody>
</table>

  – Hazard functions for combined population under T and C are:

\[
\lambda_T(t) = \theta \lambda_{C,Y} + \left( \lambda_{C,O} - \lambda_{C,Y} \right) (1 - p) e^{-\theta t \lambda_{C,O}} \\
\lambda_C(t) = \lambda_{C,Y} + \left( \lambda_{C,O} - \lambda_{C,Y} \right) (1 - p) e^{-\theta t \lambda_{C,O}}
\]

  – Therefore, \( \theta(t) = \lambda_T(t) / \lambda_C(t) \) also varies with time
Inference using the Stratified Log-rank Test

• Decompose the log-rank statistic linearly as follows:
  
  - Test: \( Z = \frac{U}{\sqrt{V}} \sim N, \quad U = \sum_{j=1}^{r} (d_1 - e_1) \quad V = \sum_{j=1}^{r} \frac{n_1 j n_2 j d_j (n_j - d_j)}{n_j^2 (n_j - 1)} \approx \frac{d}{4} \)

• We rewrite the stratified log-rank test statistic as:

\[
  z = \frac{U_Y + U_0}{\sqrt{V_Y + V_0}} = \frac{\sqrt{V_Y}}{\sqrt{V_Y + V_0}} Z_Y + \frac{\sqrt{V_0}}{\sqrt{V_Y + V_0}} Z_O \approx \frac{\sqrt{d_Y}}{\sqrt{d_Y + d_0}} Z_Y + \frac{\sqrt{d_0}}{\sqrt{d_Y + d_0}} Z_O \tag{1}
\]

• While \( Z_Y, Z_O \) in each subgroup are not observed due to disproportionate sampling, unbiased stratum-specific estimates may be obtained \( Z'_Y, Z'_O \)

• If \( p \) is the proportion of young subjects in the general population and \( p' \) is their proportion in an enriched sample, then it can be derived that

\[
  Z_Y \approx Z'_Y \frac{p}{\sqrt{p'}} \quad \text{and} \quad Z_O \approx Z'_O \frac{1 - p}{\sqrt{1 - p'}}
\]
Conclusions

• Unbiased and generalizable statistical inference is still possible under disproportionate stratified sampling, i.e. when sample proportions are different from population proportions

• Naïve comparison of population distributions in clinical trials with epidemiological distributions can be quite misleading; the evaluation needs to be contextual and nuanced

• When considering studies in special populations, emphasizing feasibility of enrollment and pragmatism in design selection may be as important as avoiding unjustified exclusions.

• Innovative designs and disproportionate sampling can be beneficial in terms of trial conduct.
References


References


