Label-enabling dynamic borrowing with external control for
OS - FDA Complex Innovative Designs Pilot Program
6th EFSPi regulatory statistics workshop, Sept 14, 2021

Jiawen Zhu
Senior Principal Statistical Scientist, Genentech/Roche
zhu.jiawen@gene.com
Why innovative design was needed

Unmet medical need in certain subgroup of DLBCL patients

• Diffuse large B-cell lymphoma (DLBCL) is the most common non-Hodgkin’s lymphoma (NHL) worldwide, with 25,000 newly diagnosed patients in the United States (US) annually

• Standard of care for 1L DLBCL patients established over 20 years ago: it is well characterized and well understood

• Patients in certain subgroup of DLBCL have a poorer prognosis and consequently a high unmet medical need

“Borrowing” patients from the control arm of another study helps us

• Having fewer ‘new’ patients treated with a control regimen that is well established and that we know well

• Shorten our study

• Conducting more efficient trials by sharing control data between trials
Proposed Phase 3 Study Design in 1L DLBCL

- Analysis of primary endpoint (PFS) based on the randomized patients, designed to provide 80% power at the 5% significance level to detect a target HR of 0.6, one IA at 80% of events
- External control patients to be selected from a contemporary, ongoing internal clinical trial
- External control arm intended to support early OS analysis at the time of the primary PFS analysis
- Randomized study with external control arm using matched external controls through Bayesian dynamic borrowing

Primary Endpoint:
- PFS Investigator Assessed

Key Secondary Endpoints
- OS, based on randomized patients & matched external control

- N=414
  - biomarker+
  - R:1
  - Noval combo (n=276)
  - R-CHOP (n=138)
  - External control (n=100)
No borrowing
only RCT data is used to estimate treatment effect

Dynamic borrowing: Conservative prior (Half Cauchy)
Skeptical on external control

Dynamic borrowing: Aggressive prior (Gamma)
Optimistic on external control

Full borrowing
Two controls are pooled together when estimating treatment effect

CID Pilot Program Process & Our Experience

● The program lasts for 240 day counting from submitting meeting request, and includes two 1.5hr meetings
● FDA is very collaborative, open to discussion, interested in our proposal, and willing to do extra research on their own; This is a pilot program, FDA is also learning as they go
● The opportunity for 2 separate meetings really helped to reach alignment on the statistical methodology.
  ○ Preliminary method proposal and simulation was included as early as the program application package
  ○ We were able to clarify design and analysis in CID #1 and provide updated analysis plan and simulation before CID #2
  ○ FDA accepts e-mail clarifications outside of the designated two meetings
  ○ Additional requirements (simulations) will require more time, while FDA is flexible with extensions, it will also push timelines out
● FDA agreed that updated statistical methodology and new simulations is acceptable for the analysis of OS as the first secondary endpoint, which has the potential to be included in labeling
● Overall, wonderful experience on the FDA CID pilot
### Novel designs – Making it happen

<table>
<thead>
<tr>
<th>Typical design</th>
<th>vs.</th>
<th>Hybrid Bayesian dynamic borrowing</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Decide on parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Fixed scenario</td>
<td>&lt;Front-loading&gt;</td>
<td>- Extensive simulations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Many scenarios (~20+ for each FDA meeting)</td>
</tr>
<tr>
<td>Implications</td>
<td></td>
<td>- Plan early</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Allocate time/resources</td>
</tr>
<tr>
<td>Solutions</td>
<td></td>
<td>- CRAN R Software available: <strong>psborrow</strong>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Roche statistics method group and method experts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Learnings from CID program</td>
</tr>
<tr>
<td>Methods R&amp;D</td>
<td></td>
<td>- FDA U01 grant (ongoing work)</td>
</tr>
</tbody>
</table>

*psborrow: Bayesian Dynamic Borrowing with Propensity Score https://cran.r-project.org/web/packages/psborrow/index.html
What was the FDA looking for?

- **Model-assumptions assessment**
  - Standard analysis typically requires few assumptions
  - Borrowing: more assumptions and less standard; FDA provided valuable input on where and how to make assessments

- **Pre-specification**

- **What could hamper inclusion of OS in label (similar to traditional designs)?**
  - Examples:
    - Whether the model assumptions appear to be met
    - Any outlying subgroup effects
    - The endpoint was credibly captured or not
    - Overall conduct of the study
    - Missing data
    - Baseline characteristics are the same

- **Non-statistical considerations:**
  - Is the summary of analysis clear?
  - Interpretable by clinicians?
  - Provides valuable information?

Along with these considerations, ultimately, the FDA requires the final data from such a novel design to gain confidence in the ability to utilize external controls more readily.
Final Analysis Flow Diagram

Control comparability evaluation
- Apply inclusion/exclusion criteria
- Flag baseline factors with significant difference between internal and external trials

Propensity score matching
- Match patient population between internal and external trials using propensity score matching (PSM)
- Enhance covariates balance by filtering out unmatched patients

Bayesian dynamic borrowing
A method to:
- Automatically downweight external control data based on internal/external control agreement
- Provide inference of treatment effect with hybrid control (i.e. OS analysis)

Sensitivity analysis follows main analysis

* In the rare case of missing data, those data for prognostic factors will be accounted for by using nearest neighbor (NN) imputation under a missing at random (MAR) assumption
Simulation scope and objective

- Focused on the evaluation of the proposed statistical method (PS matching and the Bayesian commensurate prior approach)

- Examined the trial operating characteristics (OC) under:
  - Varying magnitude of **differences in baseline characteristics**
  - Different **choices of the commensurate prior** which influences the degree of borrowing
  - **Violation** of various **assumptions**
Simulation results highlights

Type I error

- Scenario 1: No difference
- Scenario 2: Moderate difference in baseline distribution
- Scenario 3: Large difference in baseline distribution
- No Borrowing

Power

- Scenario 4: No difference
- Scenario 5: Moderate difference in baseline distribution
- Scenario 6: Large difference in baseline distribution
- No Borrowing

Note: HR = hazard ratio; OS = overall survival.
Simulation results discussion

Table 19: Summary Table to Compare Method Performance for Differences in Baseline Characteristics Investigations

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Average Error Rate</th>
<th>Weighted Type I Error Rate</th>
<th>Max Type I Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>No borrowing (only RCT data)</td>
<td>0.024</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>Dynamic borrowing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(with external control)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservative prior</td>
<td>0.023</td>
<td>0.023</td>
<td>0.032</td>
</tr>
<tr>
<td>Aggressive prior</td>
<td>0.028</td>
<td>0.026</td>
<td>0.054</td>
</tr>
<tr>
<td>Full borrowing (pooling two</td>
<td>0.033</td>
<td>0.029</td>
<td>0.067</td>
</tr>
<tr>
<td>control arms)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RCT = randomized controlled trial

* Weighted Type I Error Rate is calculated based on the assumed probability on the various scenarios: The probability for “The same” is assumed to be 62.5%, “moderate”, 20%, “large” 5%, “moderate reverse” 10%, and “large reverse” 2.5%.
Complex Innovative Designs: remaining challenges and questions

- Study design assessment is less standard given the nature of CID. What will be the type I error control consideration for regulatory decision making?
- Room for non-traditional decision making framework, e.g. Bayesian inference?
- Plans of other HAs to establish similar pilot efforts or leverage existing programs?
- What does it take for designs to graduate from a pilot and become normal practice?
Because of the destination, it needs us to work together to find a path.

"Difficult roads often lead to beautiful destinations."
Acknowledgements

- Laura Wong
- Sofia Khan
- Zac Taylor
- Emma Clark
- Hannah Douthwaite-Billing
- Sarah Kirk
- Michelle Boyer
- Alex Bazeos
- Herb Pang
- Kaspar Rufibach
- Victor Huang
- Aijing Lin
- YJ Choi
- Yichen Lu (intern)
- Jiaheng Qiu
- Mark Yan
- Yanwen Jiang
- Gracie Lieberman
- Jane Fridlyand
Doing now what patients need next
CID Timelines

Venetoclax 1L DLBCL became the 1st Roche/GNE program to be accepted into the FDA CID Pilot Meeting Program

- **Day 0**: Sponsor submits CID Meeting Request
- **Day 45**: FDA evaluates CID Meeting Request
  - FDA notifies sponsor of their status: proceed to disclosure discussions or meeting denied
- **Day 90**: FDA and sponsor discuss disclosure elements
  - FDA and sponsor reach disclosure agreement and meeting is granted, if not, meeting is denied
- **Day 120**: CID Meeting 1
- **Day 150**: Sponsor submits CID Meeting 2 Package
- **Day 240**: CID Meeting 2

© 2020 F. Hoffmann-La Roche Ltd | v1 | 01OCT2020
<table>
<thead>
<tr>
<th>Scenario</th>
<th>OS</th>
<th>HR</th>
<th>Assumed Parameters Between Internal Subjects and External Controls</th>
<th>Violation of Assumptions</th>
<th>Borrowing Approaches</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>The same</td>
<td>Unmeasured confounding</td>
<td>No borrowing, conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>Moderate difference</td>
<td>Unmeasured confounding</td>
<td>Conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing</td>
</tr>
<tr>
<td>19</td>
<td>0.67</td>
<td></td>
<td>The same</td>
<td>Unmeasured confounding</td>
<td>No borrowing, conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing</td>
</tr>
<tr>
<td>20</td>
<td>0.67</td>
<td></td>
<td>Moderate difference</td>
<td>Unmeasured confounding</td>
<td>Conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing</td>
</tr>
</tbody>
</table>
Simulation results highlights

Type I error

Power

<table>
<thead>
<tr>
<th>Method</th>
<th>Scenario 13</th>
<th>Scenario 1</th>
<th>Scenario 14</th>
<th>Scenario 2</th>
<th>No Borrowing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic borrowing with Half-Cauchy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic borrowing with Gamma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full borrowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No borrowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| No Borrowing                  |     |     |     |     |              |