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

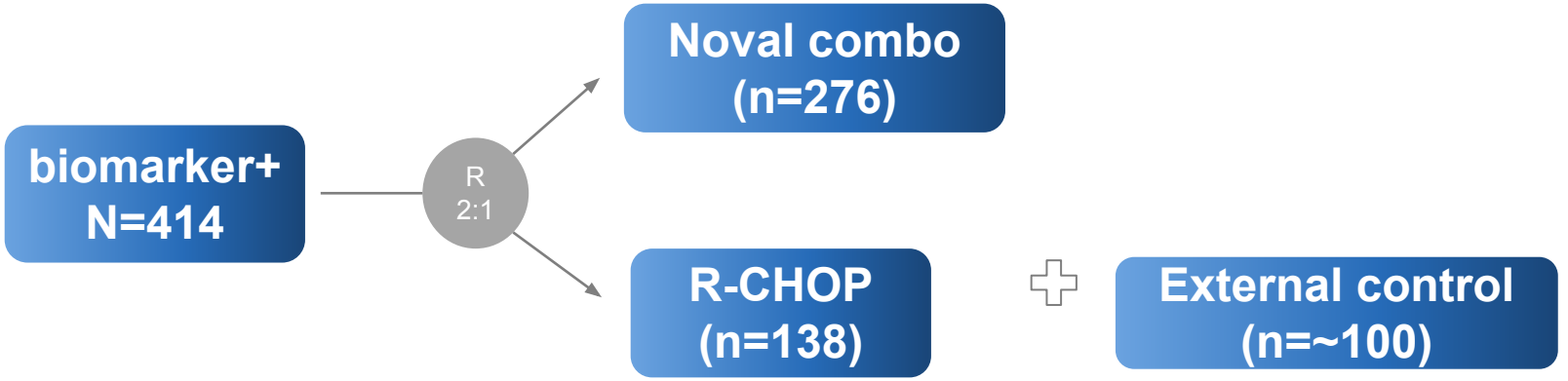
Unmet medical need in certain subgroup of DLBCL patients

- 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

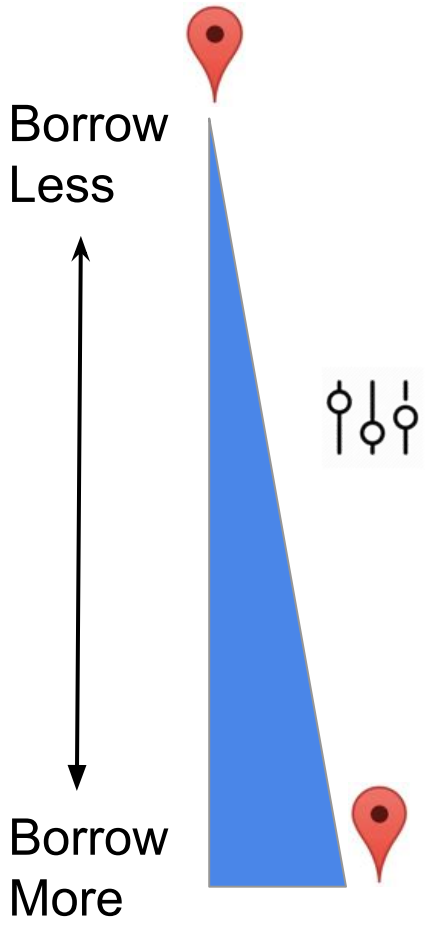


Primary Endpoint:
- **PFS Investigator Assessed**

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

- 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

Borrowing approaches



- **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

Method reference: Lewis, CJ, Sarkar, S, Zhu, J, Carlin, BP et al. Borrowing from historical control data in cancer drug development:A cautionary tale and practical guidelines.Statistics in Biopharmaceutical Research, 11(1):67–78, 2019

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

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

*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

Roche

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

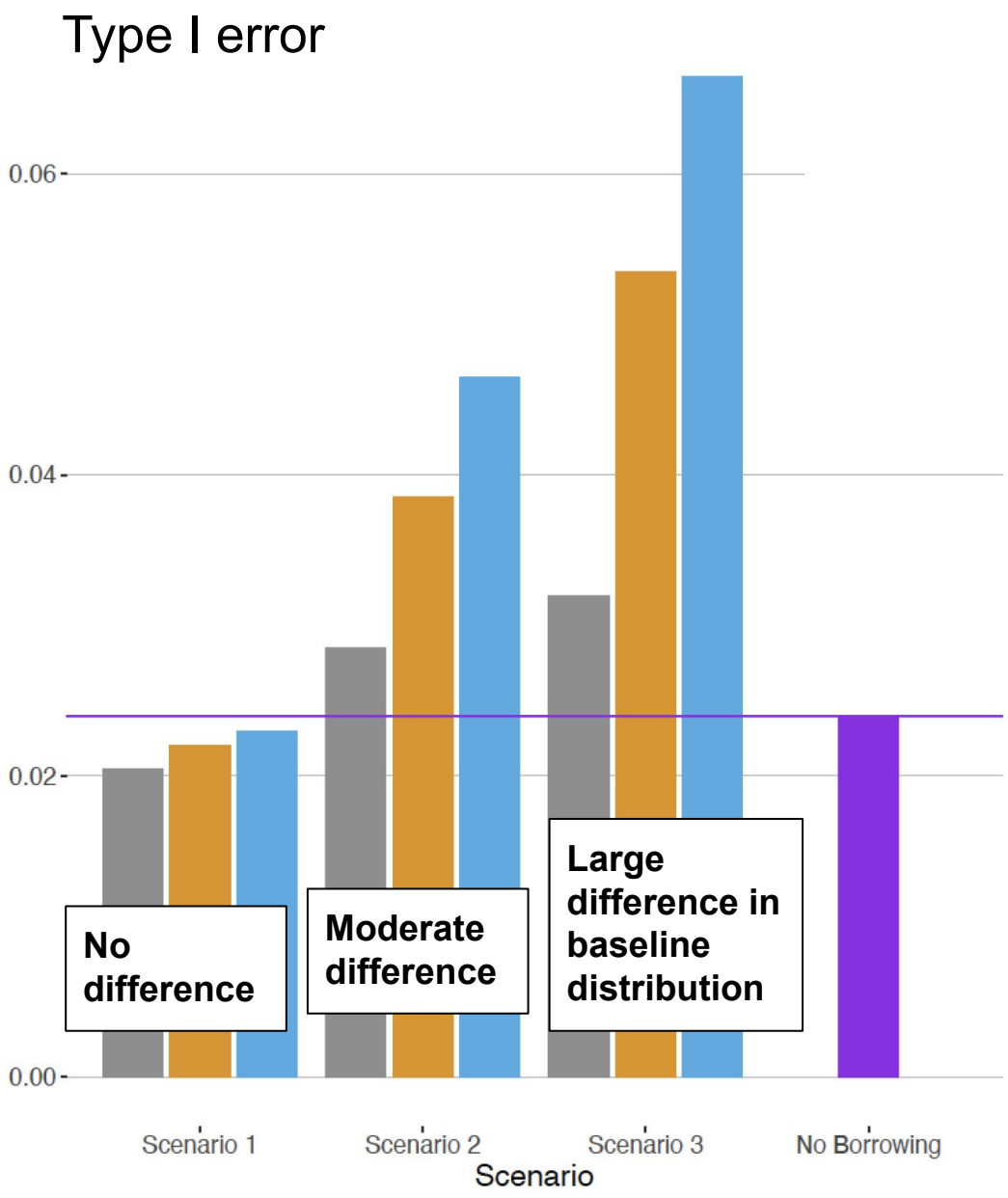
* In the rare case of missing data, those data for prognostic factors will be accounted for by using nearest neighbor (NN) imputation under a missing at random (MAR) assumption

Simulation scope and objective

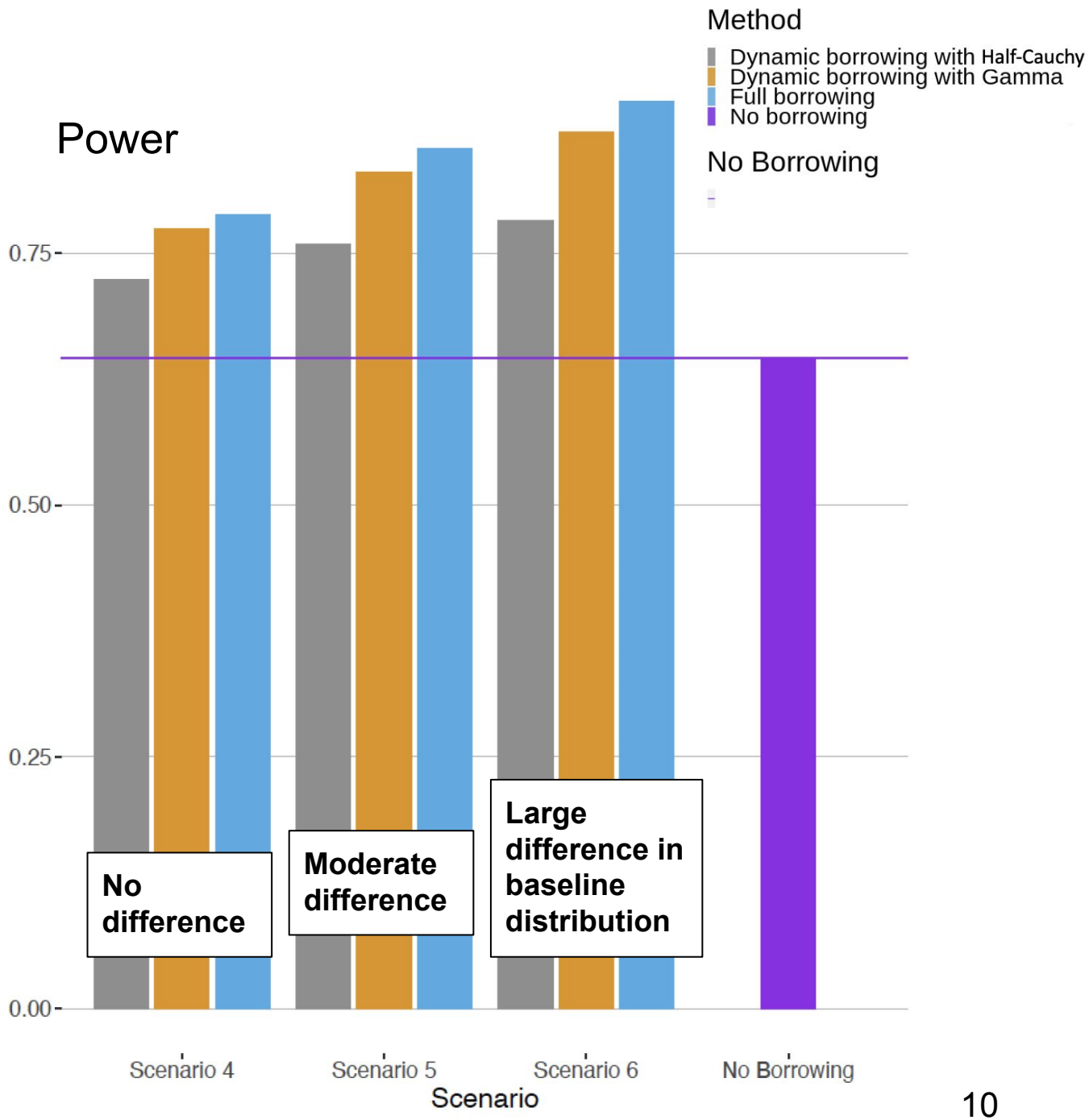
- Focused on the evaluation of the proposed statistical method (PS matching and the Bayesian commensurate prior approach)

- 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



HR=hazard ratio; OS=overall survival.



HR=hazard ratio; OS=overall survival.

Simulation results discussion

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

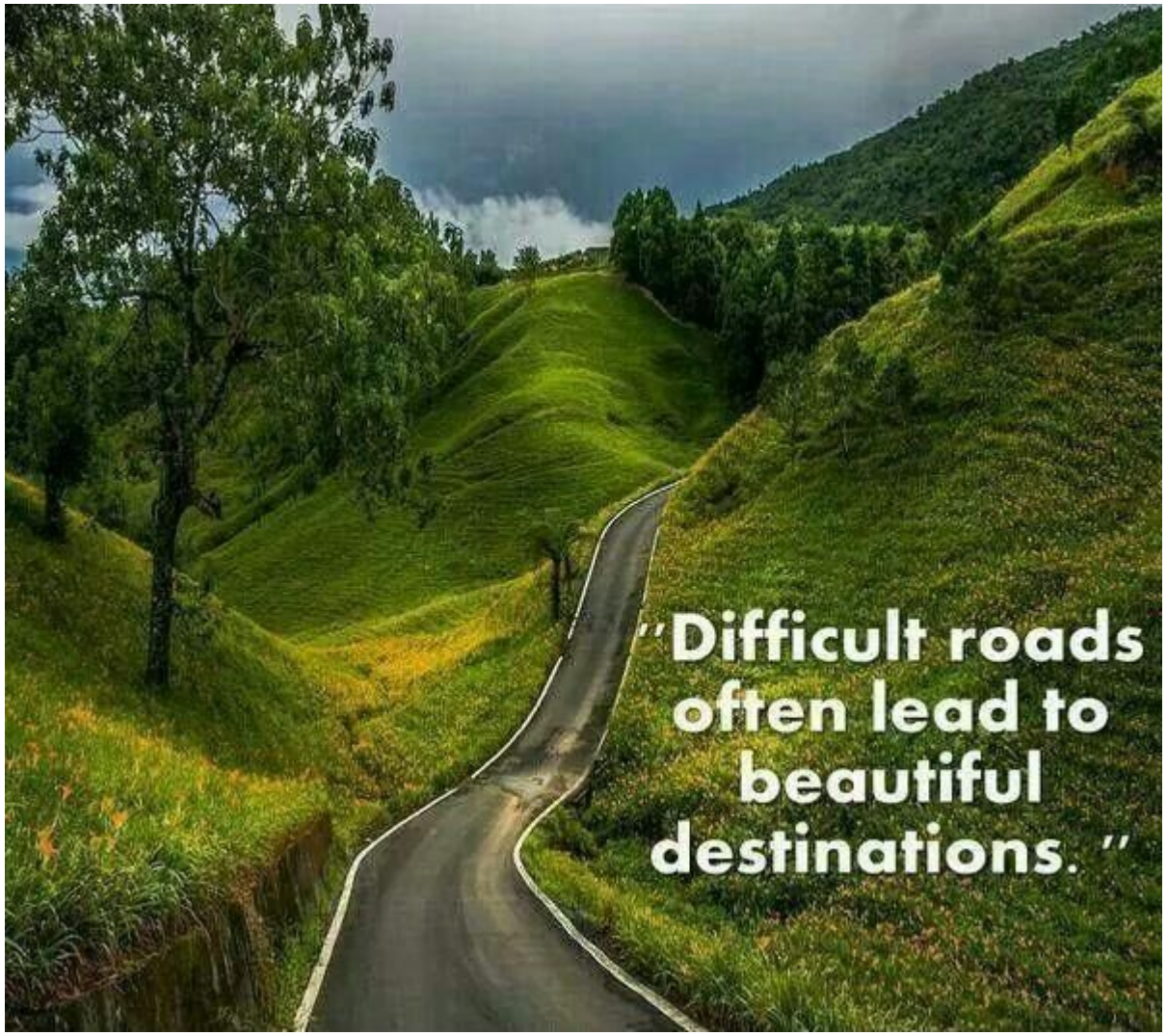
Approaches		Average Error Rate	Weighted Type I Error Rate*	Max Type I Error Rate
No borrowing (only RCT data)		0.024	0.024	0.024
Dynamic borrowing (with external control)	Conservative prior	0.023	0.023	0.032
	Aggressive prior	0.028	0.026	0.054
Full borrowing (pooling two control arms)		0.033	0.029	0.067

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.

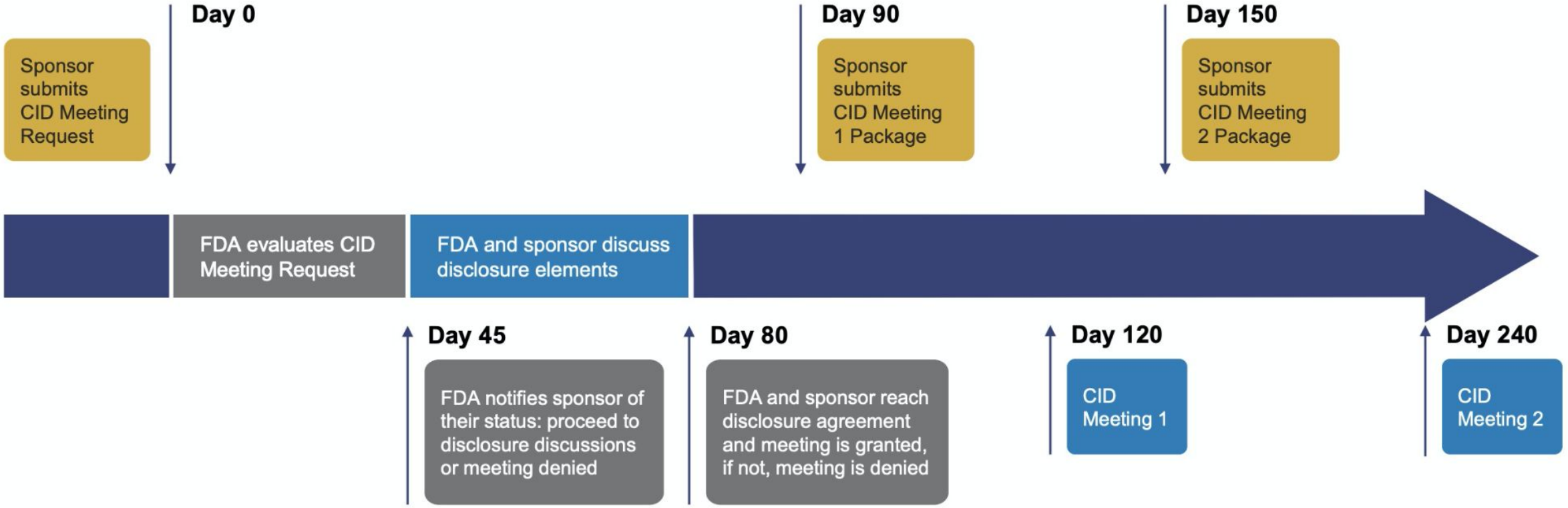
Acknowledgements

- 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



List of simulation scenarios for violation of assumptions (subset)

Scenario	OS HR	Assumed Parameters Between Internal Subjects and External Controls	Violation of Assumptions	Borrowing Approaches
13	1	The same	Unmeasured confounding	No borrowing, conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing
14	1	Moderate difference	Unmeasured confounding	Conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing
19	0.67	The same	Unmeasured confounding	No borrowing, conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing
20	0.67	Moderate difference	Unmeasured confounding	Conservative prior (Half Cauchy), aggressive prior (Gamma), full borrowing

Simulation results highlights