Incorporate Non-concurrent Control using Gaussian Processes in Platform Trials

Presenter: Yan Peng
Mentor: Yuhan Qian

Outline

01

Platform Trials

Definition and features

02

Population

Introduction of ECE population;
Concurrent vs Non-concurrent Data

03

Gaussian Processes

Uni-task and Multi-task GP

04

Variance Reduction

How GP works?

01

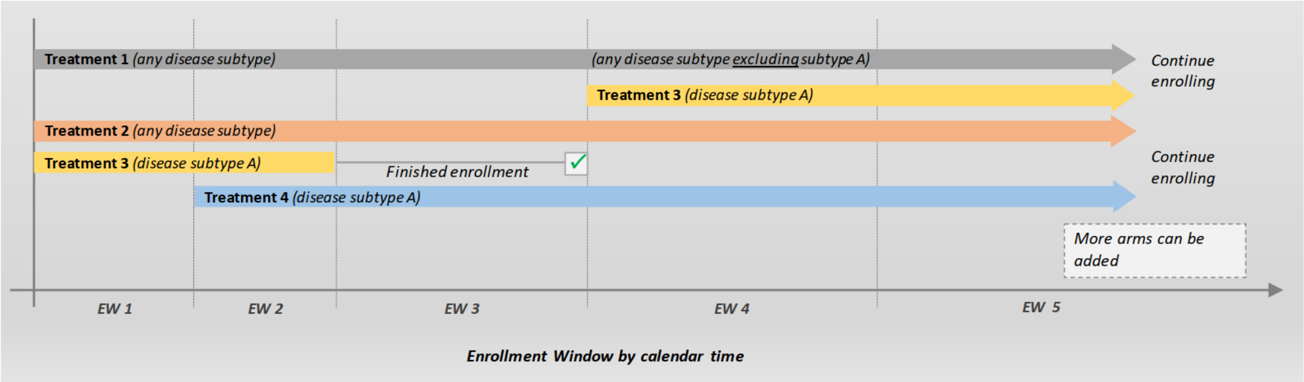
Platform Trial

Definition and features

What a Platform Trial is?

- Platform trials:
  - They can add or drop the treatments
- Introduce challenges:
  - How to define populations
  - How to estimate treatment effects reliably

(b) Multi-Arm Platform Trials



02

Population

Introduction of ECE population;
Concurrent vs Non-concurrent Data

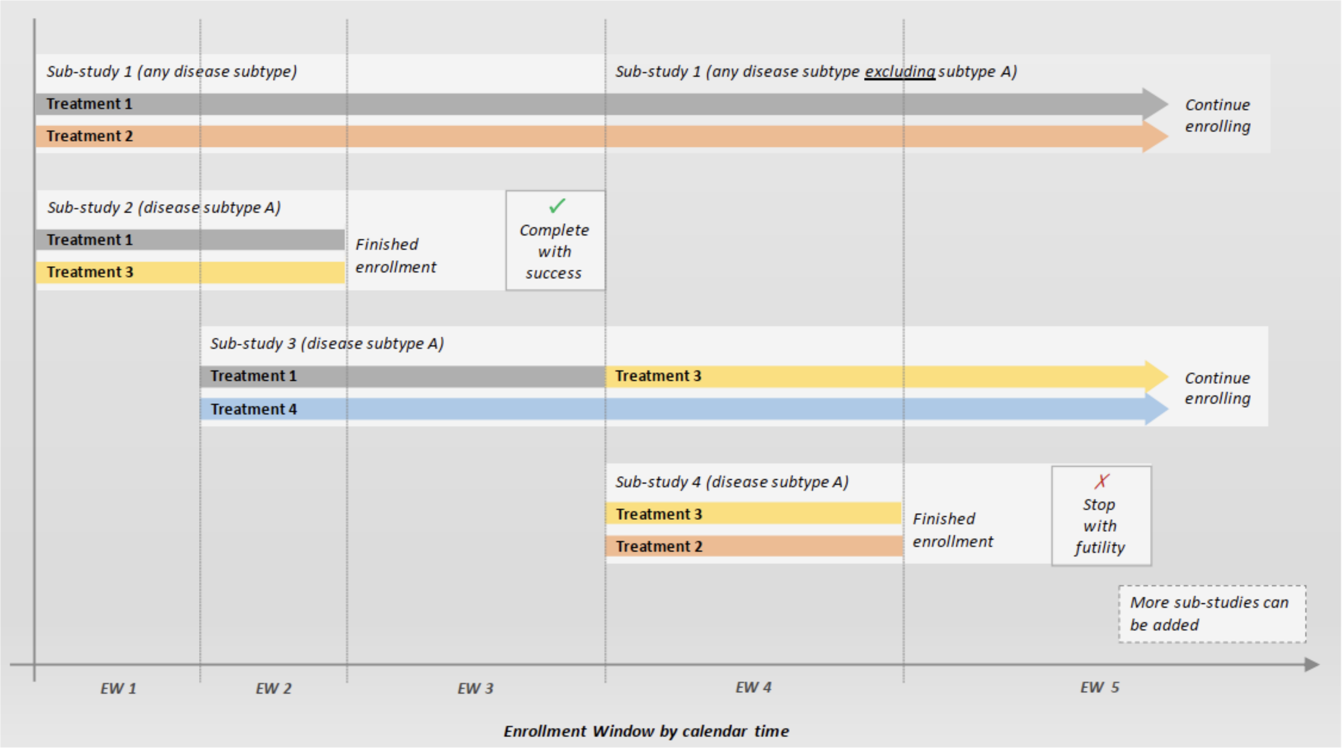
How to define populations

- Entire Concurrently Eligible (ECE) Population:
  - The population of all individuals who meet the eligibility criteria for both treatment j, k and could potentially be enrolled during a time period when both treatments are available.
  - Formally, the ECE population is a population of all individuals with $\pi_j(Z) > 0$ and $\pi_k(Z) > 0$
  - Z: observed baseline variable, eg. Enrollment window
  - $\pi_j(Z)$: probability that individual is assigned to treatment j given Z

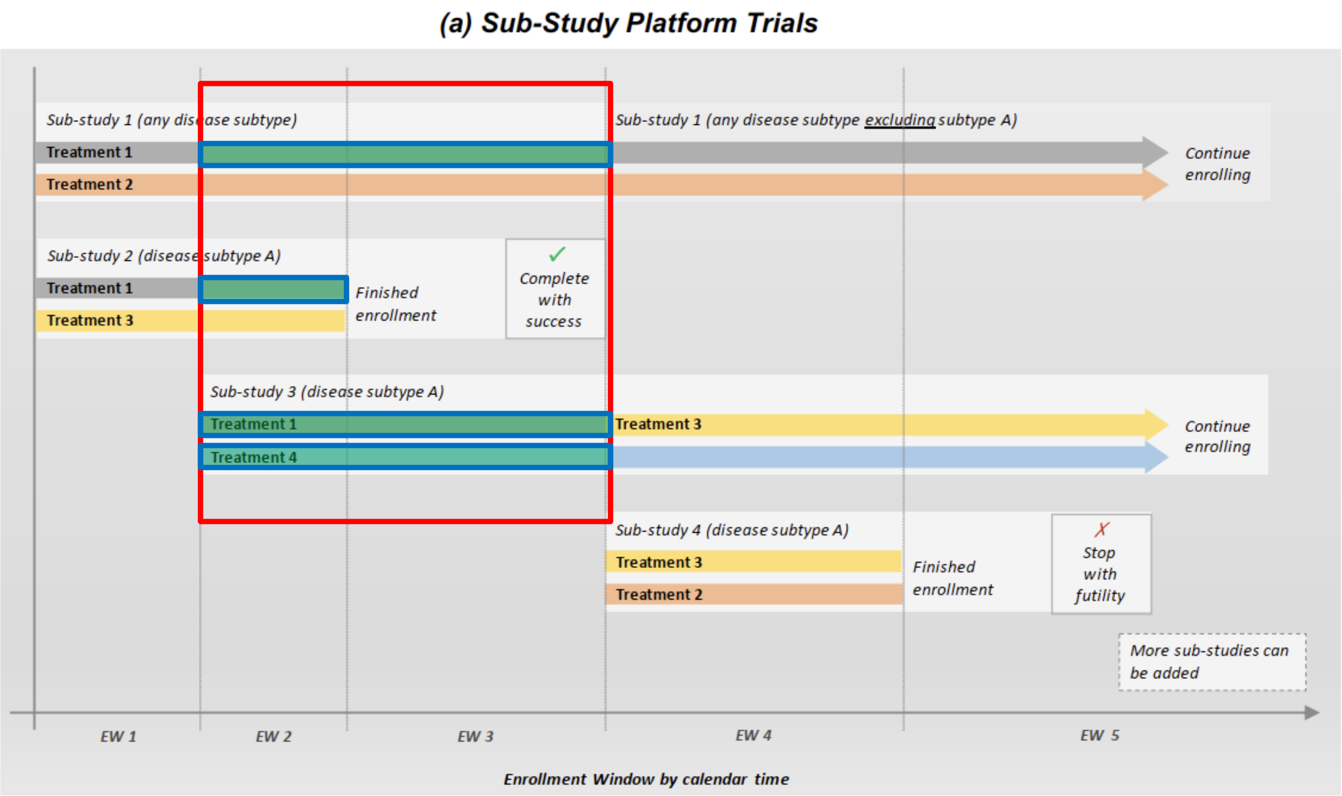
"From Estimands to Robust Inference of Treatment Effects in Platform Trials." *arXiv preprint arXiv:2411.12944* (2024).

Concurrent Data

(a) Sub-Study Platform Trials

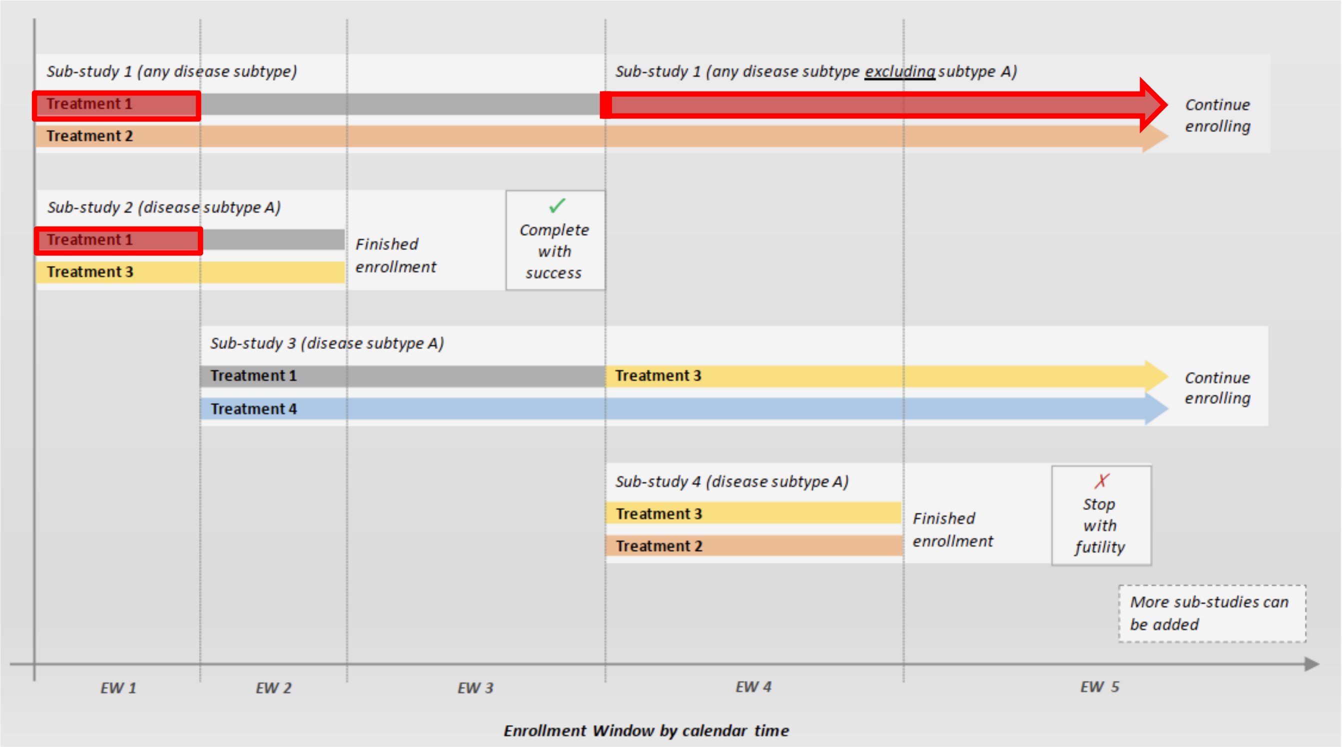


Concurrent Data



Non-Concurrent Control

(a) Sub-Study Platform Trials



03

Gaussian Processes

Uni-task and Multi-task GP

Uni-task Gaussian Processes

- Gaussian Processes model of treatment a = j and k separately:

$$Y_i^a = \mu_a + f_a(E_i) + \epsilon_{i,a}$$

  - GP prior for f_a, flexible modeling of enrollment time(E_i) effects.
  - A non-parametric Bayesian method for regression
  - Models complex relationships and quantifies uncertainty naturally.

Multi-Task Gaussian Process

$$\begin{pmatrix} Y_i^{(j)} \\ Y_i^{(k)} \end{pmatrix} = f(\begin{pmatrix} E_i \\ E_i \end{pmatrix}) + \epsilon_i$$

$$f(\begin{pmatrix} \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \end{pmatrix}) = \begin{pmatrix} f_j(\boldsymbol{E}_1) \\ f_k(\boldsymbol{E}_2) \end{pmatrix} \sim \mathcal{GP}(0, \begin{pmatrix} \boldsymbol{k}_j(\boldsymbol{E}_1, \boldsymbol{E}_1^T) & \boldsymbol{k}_{jk}(\boldsymbol{E}_1, \boldsymbol{E}_2^T) \\ \boldsymbol{k}_{jk}(\boldsymbol{E}_2, \boldsymbol{E}_1^T) & \boldsymbol{k}_k(\boldsymbol{E}_2, \boldsymbol{E}_2^T) \end{pmatrix})$$

$$\epsilon_i \sim N(0, \begin{pmatrix} \sigma_j^2 & \\ & \sigma_k^2 \end{pmatrix})$$

- $Y_i^{(j)}$ and $Y_i^{(k)}$: Potential outcomes
- f(E_i): Share information between treatments

- Joint model of multiple treatments, including control arms.
- Captures relationships across groups using block covariance matrix.

04

Variance Reduction

How GP works?

Variance Reduction in Gaussian Processes

- We assume the following multi-task GP model:

  $$Y^{(k)} = f(E) + \epsilon_k, \qquad Y^{(j)} = f(E) + \Delta(E) + \epsilon_j$$

  - f(E) is the baseline GP for control
  - $\Delta(E) \sim \mathcal{GP}(0, k_\Delta(E, E^T))$
    - captures the difference between treatment and control outcomes.

Variance Reduction in Gaussian Processes

- Theorem:
  - Under the previous model, the posterior covariance matrix of $\Delta(E^*)$ is denoted by $\bar{\Sigma}_\Delta$, incorporating non-concurrent controls leads to:

$$\bar{\Sigma}_{\Delta,CC} - \bar{\Sigma}_{\Delta,NC} \geq 0.$$

- Theorem proof:
  - $\Delta(E^*)|Y,E \sim N(\overline{m}_\Delta(E^*), \overline{k}_\Delta(E^*))$
  - $\overline{k}_\Delta(E^*) = k_\Delta(E^*, E^*) - k_\Delta(E^*, E^j)A_{11}^{-1}k_\Delta(E^j, E^*)$
  - where $A_{11} = k_{f,jj} + k_{\Delta,jj} + \sigma_j^2 I_{n,j} - k_{f,jk}(k_{f,kk} + \sigma_n^2 I_{nk})^{-1}k_{f,kj}$

  - For $\overline{\Sigma}_{\Delta,CC}$ - $\overline{\Sigma}_{\Delta,NC}$, we have:

$\overline{\Sigma}_{\Delta,CC}$ - $\overline{\Sigma}_{\Delta,NC}$ $= (k_\Delta(E^*, E^*) - k_\Delta(E^*, E^j)A_{11,CC}^{-1}k_\Delta(E^j, E^*))$-$((k_\Delta(E^*, E^*) - k_\Delta(E^*, E^j)A_{11,NC}^{-1}k_\Delta(E^j, E^*))$ = $k_\Delta\left(E^*, E^j\right)(A_{11,NC}^{-1} - A_{11,CC}^{-1})k_\Delta(E^j, E^*)$

  - $A_{11,CC}^{-1} \leq A_{11,NC}^{-1}$ $\Longrightarrow$ $\overline{\Sigma}_{\Delta,CC}$ - $\overline{\Sigma}_{\Delta,NC} \geq 0$.

Conclusion & Future Plan

Conclusion & Future Plan

- ECE population ensures unbiased estimation
- Incorporating non-concurrent controls with GP model shows efficiency again

- Future plan:
  - Check the potential bias when incorporating non-concurrent data with GP.

Thank you ！

Paper

- Qian, Y., Yi, Y., Shao, J., Yi, Y., Levin, G., Mayer-Hamblett, N., Heagerty, P. J., & Ye, T. (2024). From estimands to robust inference of treatment effects in platform trials. arXiv preprint arXiv:2411.12944. https://arxiv.org/abs/2411.12944

Q&A