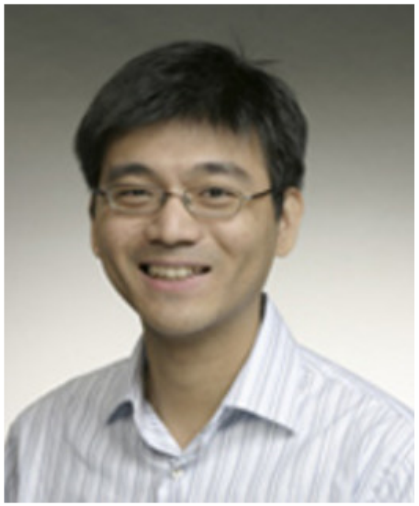
A Gate-keeping Approach for Selecting Adaptive Interventions under General SMART Designs

Tony Zhong, DrPH
Icahn School of Medicine at Mount Sinai
(Feb 20, 2019)

Workshop on Design of mHealth Intervention Studio, 2019
Institute for Mathematical Science
National University of Singapore



Ken Cheung

Bin Cheng

Min Qian

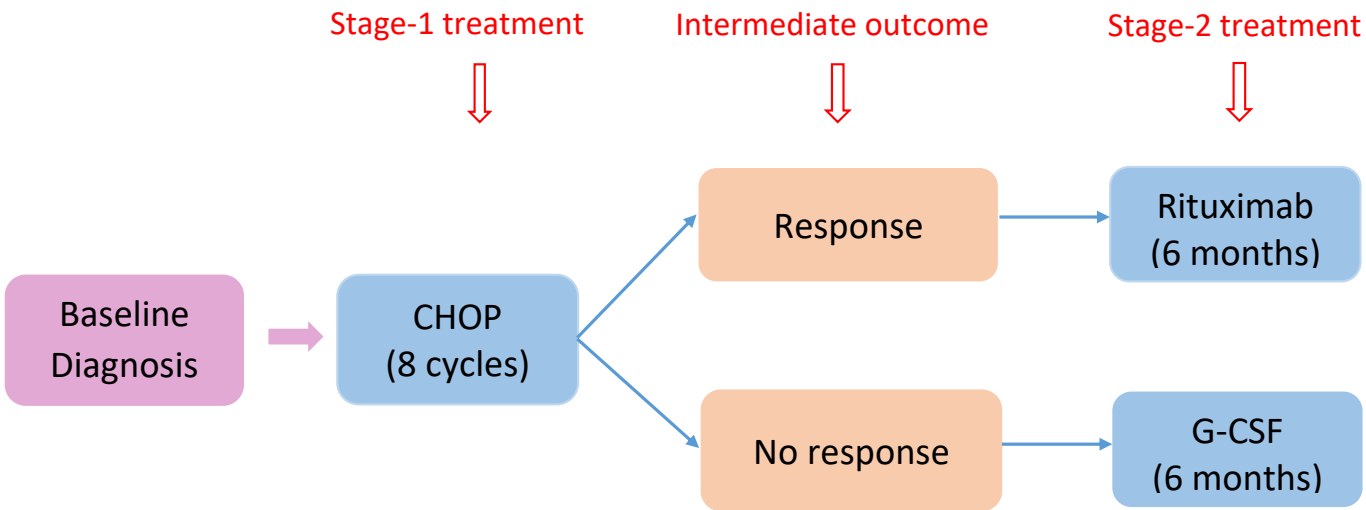
Tony Zhong

Adaptive Intervention

- For chronic diseases (e.g. cancer, depression)
  - Clinicans often need to make a sequence of treatment decisions throughout the course of a patient's clinical care
  - Such decisions are usually adapted to the patient's individual response to previous treatments
  - **Adaptive intervention** formalizes the mechanism of such a sequence of decision makings.
  - It provides an opportunity to optimize a patient's long-term response

Example 1: Oncology Intervention

- Example of adaptive intervention for Lymphoma patients[1]



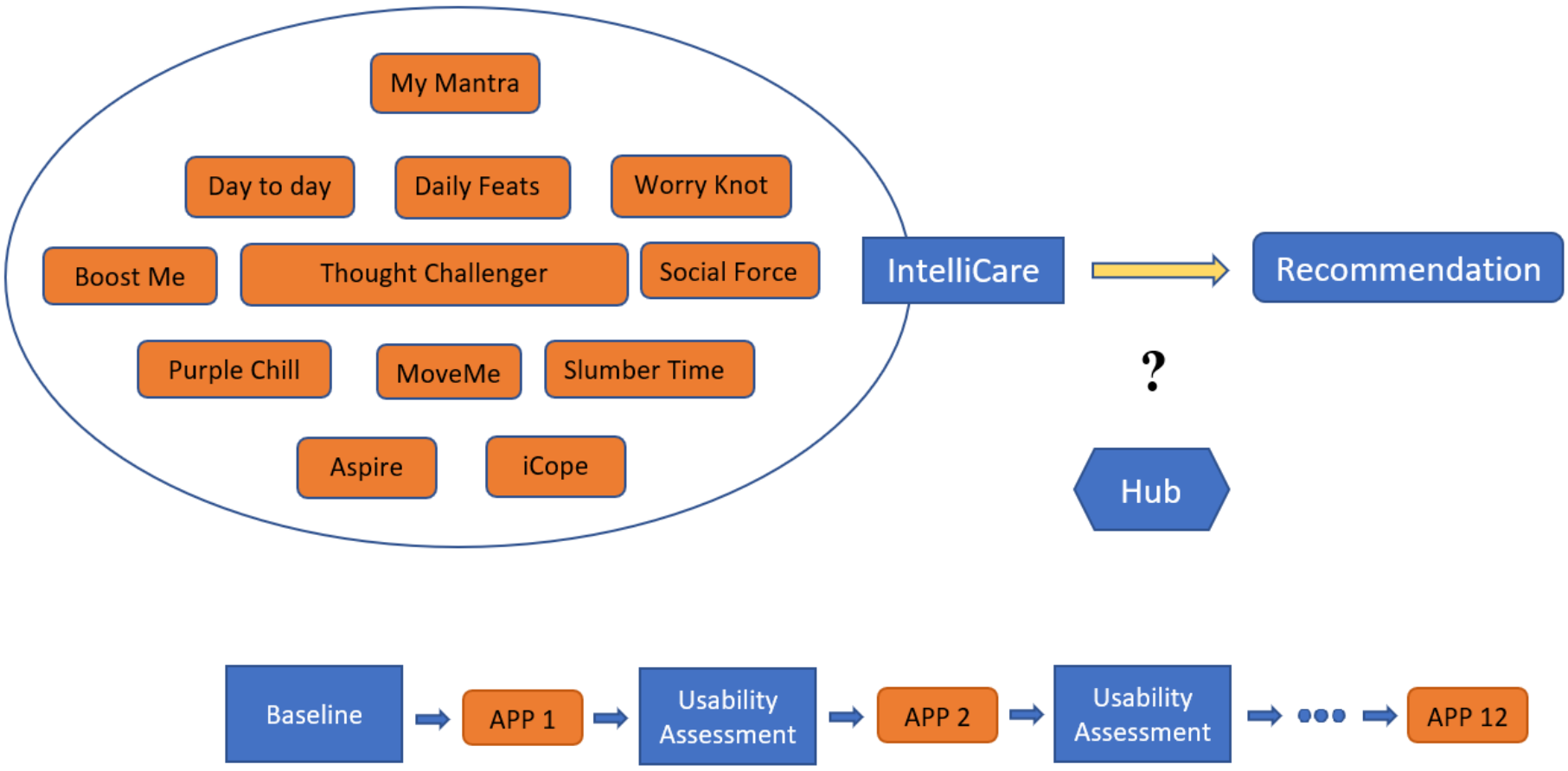
- CHOP: **C**yclophosphamide (750 mg/m^2), **H**ydroxydaunorubicin (50 mg/m^2), **O**ncovin (1.4 mg/m^2), **P**rednisone (40 mg/m^2)
- Response: complete/partial remission; Non-response: otherwise
- G-CSF: Granulocyte colony-stimulating factor

[1] Habermann et al. *J Clin Oncol*, 2006

Example 2: Depression Intervention

- Smartphone apps can be effective in managing depression
  - Pro1: more flexible than PC-based internet intervention
  - Pro2: collect individual information in real-time settings
  - Pro3: an active market provides more and more products
- Using smartphone apps for depression management has its own challenge
  - Simple apps are in favor
  - A wide variety of psychological strategies (e.g., goal setting, cognitive restructure, behavioral activation etc.) can be useful for patients with common mental health issues.
  - How to use the simple weapons to solve complicate problem?

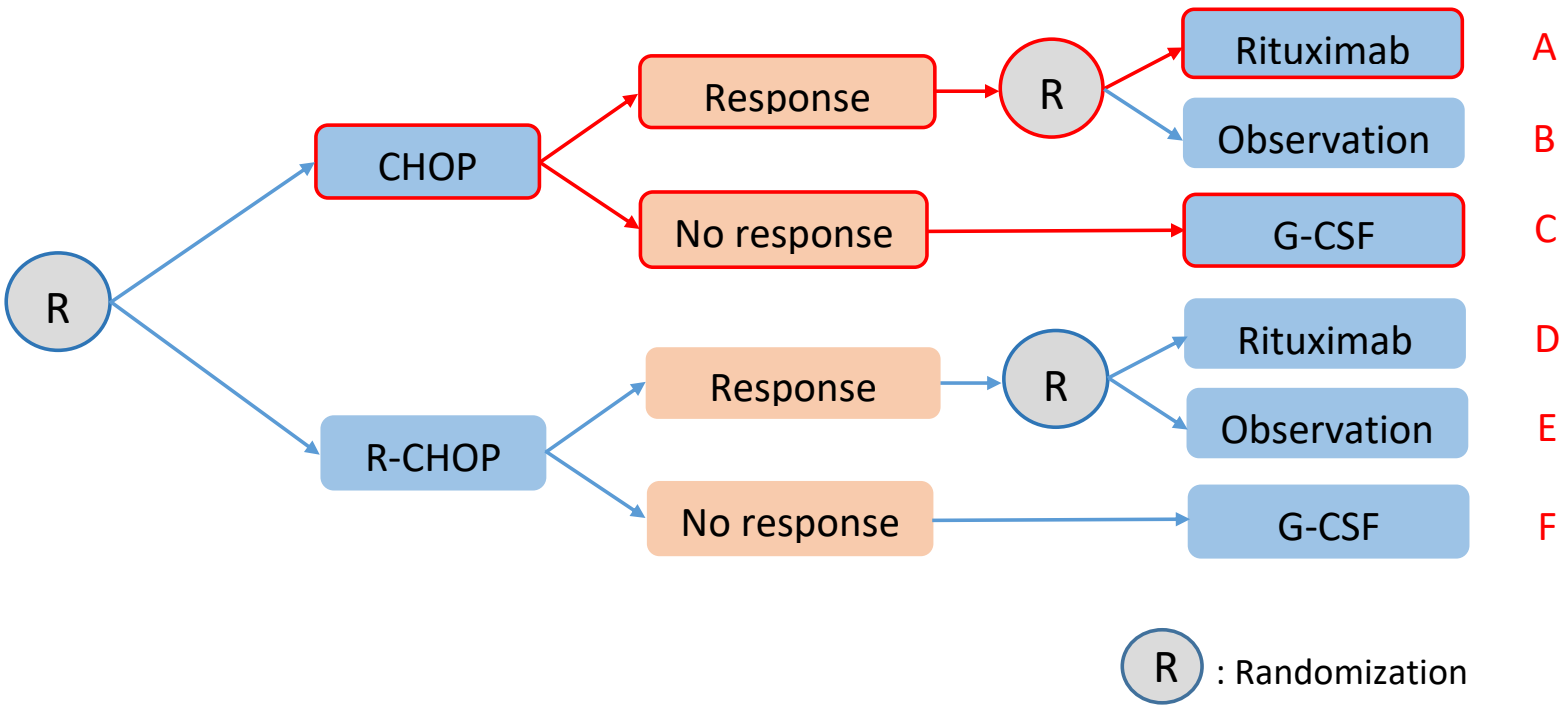
Example 2: IntelliCare (Cheung et al., 2018)



SMART Design

- Sequential multiple assignment randomized trial (SMART) is a clinical trial design that can be used to compare multiple adaptive interventions
  - SMART randomly assigns patients to **a collection of adaptive interventions** that may overlap in terms of treatment decisions
  - By virtue of randomization, it provides information for directly **comparing multiple adaptive interventions**

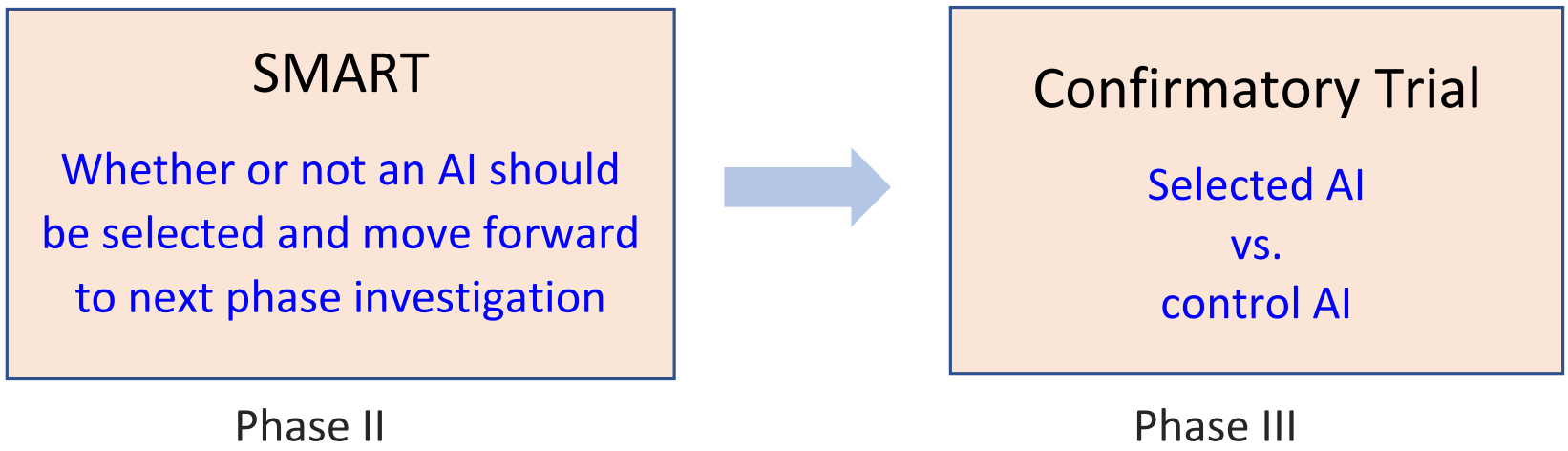
SMART Example: DLBCL Trial



There are 4 adaptive interventions embedded in this SMART, including

(A + C), (B + C), (D + F), (E + F)

Research Question



Statistical issue: multiple comparisons of adaptive interventions (AI)

Pairwise Comparison

- Pairwise comparison is commonly used for selecting adaptive interventions (AIs) embedded in a SMART
  - Conduct a series of pairwise tests based on the estimated values of all the AIs
    - e.g. Inverse probability weighted estimation (IPWE) [2]
  - Entail multiple comparisons
    - e.g. Bonferroni adjustment
- A SMART typically consists of numerous AIs
  - Inferential procedures based on pairwise comparisons of all interventions may suffer **substantial loss in power** after accounting for multiplicity

[2] Murphy, *Stat Med*, 2005

Gate-Keeping Approach

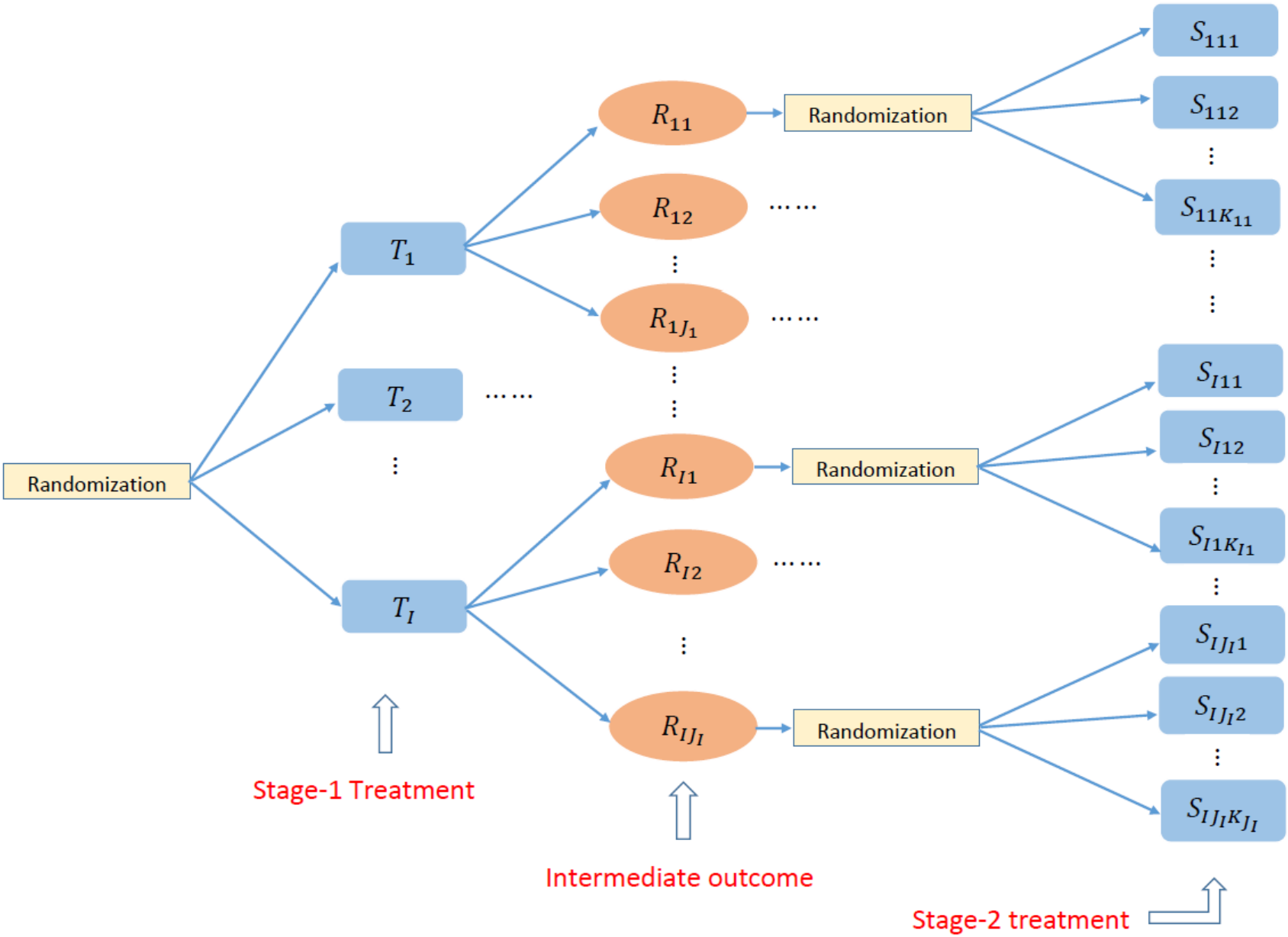
- Gate-keeping approach to avoid exhaustive search
  - Step 1: **conduct a gate-keeping test** with the null hypothesis that all the adaptive interventions have the same value
  - Step 2: **select the best adaptive intervention** only after the null hypothesis of the gate-keeping test is rejected
- Existing gate-keeping test for SMART
  - Wald test based on inverse probability weighted estimation (IPWE)[3]
    - can be used to control familywise type I error rate
    - less efficient (e.g., unbalanced randomization)
    - asymptotic properties of the test statistics has not been thoroughly studied
    - performance of selecting adaptive intervention remains unclear

[3] Ogabagabar, Karp and Wahed, *Stat Med*, 2016

The Proposed Gate-keeping Test

- We proposed a **likelihood-based** gate-keeping test for selecting the optimal adaptive intervention
  - derived the asymptotic distribution of test statistics and studied its properties
  - derived a formal sample size calculation formula
  - studied the finite sample performance of using the proposed gate-keeping approach for selecting the optimal adaptive intervention under different SMART settings, and compare it with the two existing methods

A General Two-stage SMART Design



Notation and Model

- Data for a patient can be summarized as (U, X, V, Y)
  - U: the stage-1 treatment
  - X: the intermediate response
  - V: the stage-2 treatment
  - Y: the final primary outcome
- Assuming (U, X, V, Y) to be i.i.d. distributed with

$$\Pr(U = T_i) = \pi_i,\ i = 1, \ldots, I,$$

$$\Pr(X = R_{ij} | U = T_i) = p_{ij},\ j = 1, \ldots, J_i, i = 1, \ldots, I,$$

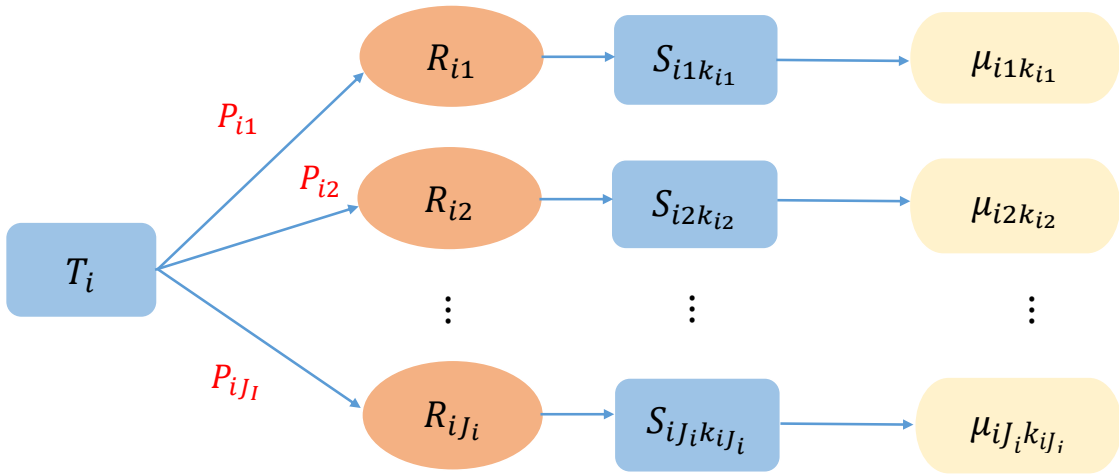
$$\Pr(V = S_{ijk} | U = T_i, X = R_{ij}) = \pi_{ijk},\ k = 1, \ldots, K_{ij}, j = 1, \ldots, J_i, i = 1, \ldots, I,$$

$$Y|(U = T_i, X = R_{ij}, V = S_{ijk}) \sim f\left(y|\mu_{ijk}, \tau_{ijk}\right),$$

$$\text{e.g. } Y|(U = T_i, X = R_{ij}, V = S_{ijk}) \overset{iid}{\sim} N\left(y|\mu_{ijk}, \sigma^2_{ijk}\right)$$

Estimand of Interest: Intervention Value

- Let $d_{i;k_{i1},\ldots,k_{iJ_i}} = (T_i; S_{i1k_{i1}}, \ldots, S_{iJ_ik_{iJ_i}})$ be an AI



- The value of an AI is defined as

$$\theta_{i;k_{i1},\ldots,k_{iJ_i}} = \sum_{j=1}^{J_i} p_{ij}\mu_{ijk_{ij}}.$$

MLE of AI Value

- MLE of an AI value

$$\hat{\theta}_{i;k_{i1},\ldots,k_{iJ_i}} = \sum_{j=1}^{J_i} \hat{p}_{ij}\hat{\mu}_{ijk_{ij}},$$

where $\hat{p}_{ij}$ and $\hat{\mu}_{ijk_{ij}}$ are obtained from maximizing the joint distribution of (U, X, V, Y).

Asymptotic Distribution

- Let $\boldsymbol{\Theta} = (\theta_1, \ldots, \theta_G)^T$ be all the adaptive intervention values embedded in a SMART. As $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Sigma}_I \end{pmatrix},$$

and

$$\text{rank}\left(\boldsymbol{\Sigma}_i\right) = \sum_{j=1}^{J_i} K_{ij} - J_i + 1. \tag{1}$$

Likelihood-based Wald Test

- Hypothesis:

$$H_0 : \theta_1 = \ldots = \theta_G \text{ vs. } H_1 : \theta\text{'s are not all equal.}$$

- Wald test:

$$Q = n(C\hat{\boldsymbol{\Theta}})^T (C\hat{\boldsymbol{\Sigma}}C^T)^- (C\hat{\boldsymbol{\Theta}}), \tag{2}$$

where $Q \xrightarrow{d} \chi^2_\nu$ under H_0 and $Q \xrightarrow{d} \chi^2_\nu(\lambda^*)$ under H_1.

- Degrees of freedom

$$\nu = \sum_{i=1}^{I} \sum_{j=1}^{J_i} K_{ij} - \sum_{i=1}^{I} J_i + I - 1. \tag{3}$$

Sample Size Calculation

- Conduct sample size calculation based on Wald test
  - Step 1: specify the study design

$$\{T_i, R_{ij}, S_{ijk}, \pi_i, \pi_{ijk}\}$$

  - Step 2: obtain λ^* by solving

$$\chi^2_{\nu,1-\beta}(\lambda^*) = \chi^2_{\nu,\alpha}(0) \tag{4}$$

  - Step 3: calculate the effect size

$$\Delta = (C\boldsymbol{\Theta}^*)^T (C\boldsymbol{\Sigma}^* C^T)^- (C\boldsymbol{\Theta}^*) \tag{5}$$

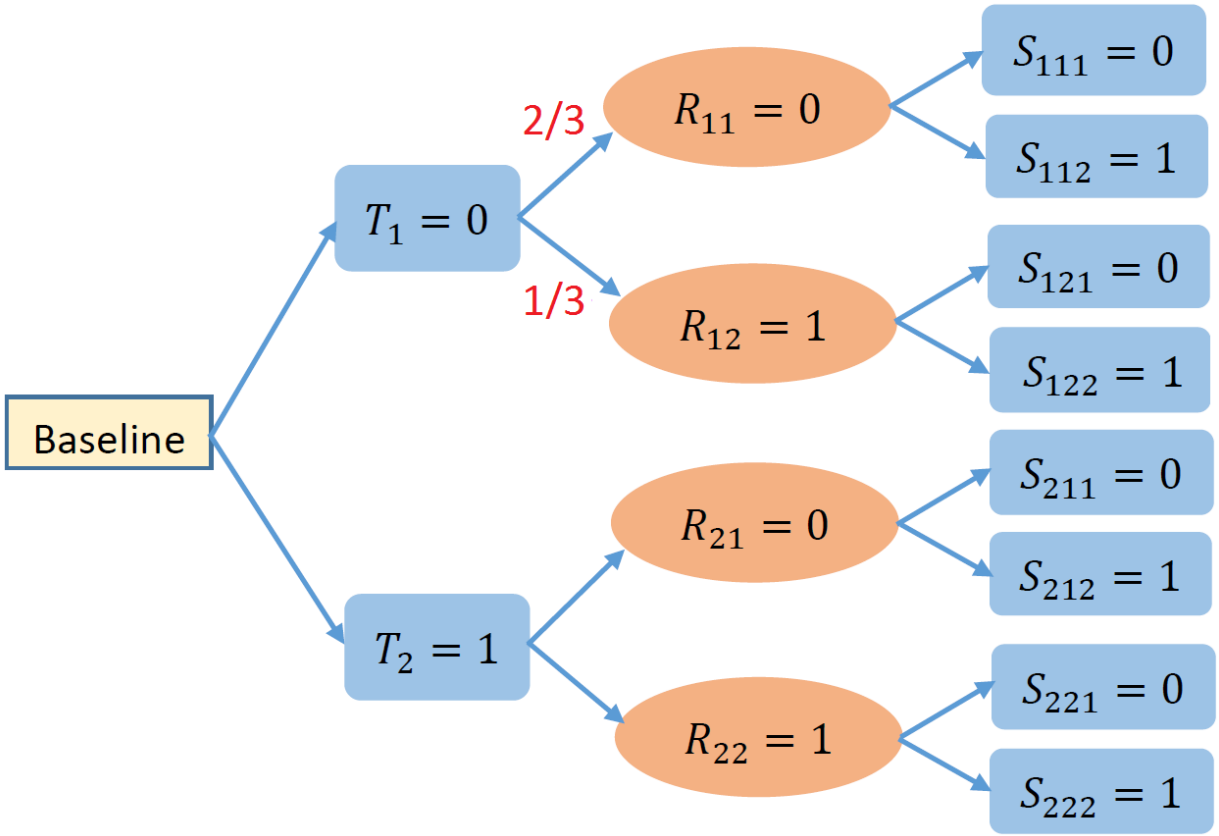
  - Step 4: total sample size of SMART

$$n = \frac{\lambda^*}{\Delta}. \tag{6}$$

Simulation Study

- Compare the type I error controls and the statistical powers
  1. Likelihood-based Wald test
  2. IPWE-based Wald test
  3. Pairwise tests with Bonferroni adjustment
     - Claim overall significance if any pairwise test is significant
- Assess the selection properties of the proposed gate-keeping approach
  - Step 1: conduct likelihood-based Wald test. If not significant, stop and claim no overall difference, otherwise, go to step 2
  - Step 2: select the best AI with the greatest observed value

Simulation: Design Structure



Simulation: Randomization

- Balanced randomization
  - $\Pr(U = 1) = 0.5$
  - $\Pr(V = 1|U, X) = 0.5$
- Unbalanced randomization
  - $\Pr(U = 1) = 0.7$
  - $\Pr(V = 1|U, X) = 0.7$
- Randomized play-the-winner
  - $\Pr(U = 1) = 0.5$
  - $\Pr(V = U|U, X = 1) = 0.7$
  - $\Pr(V = U|U, X = 0) = 0.3$

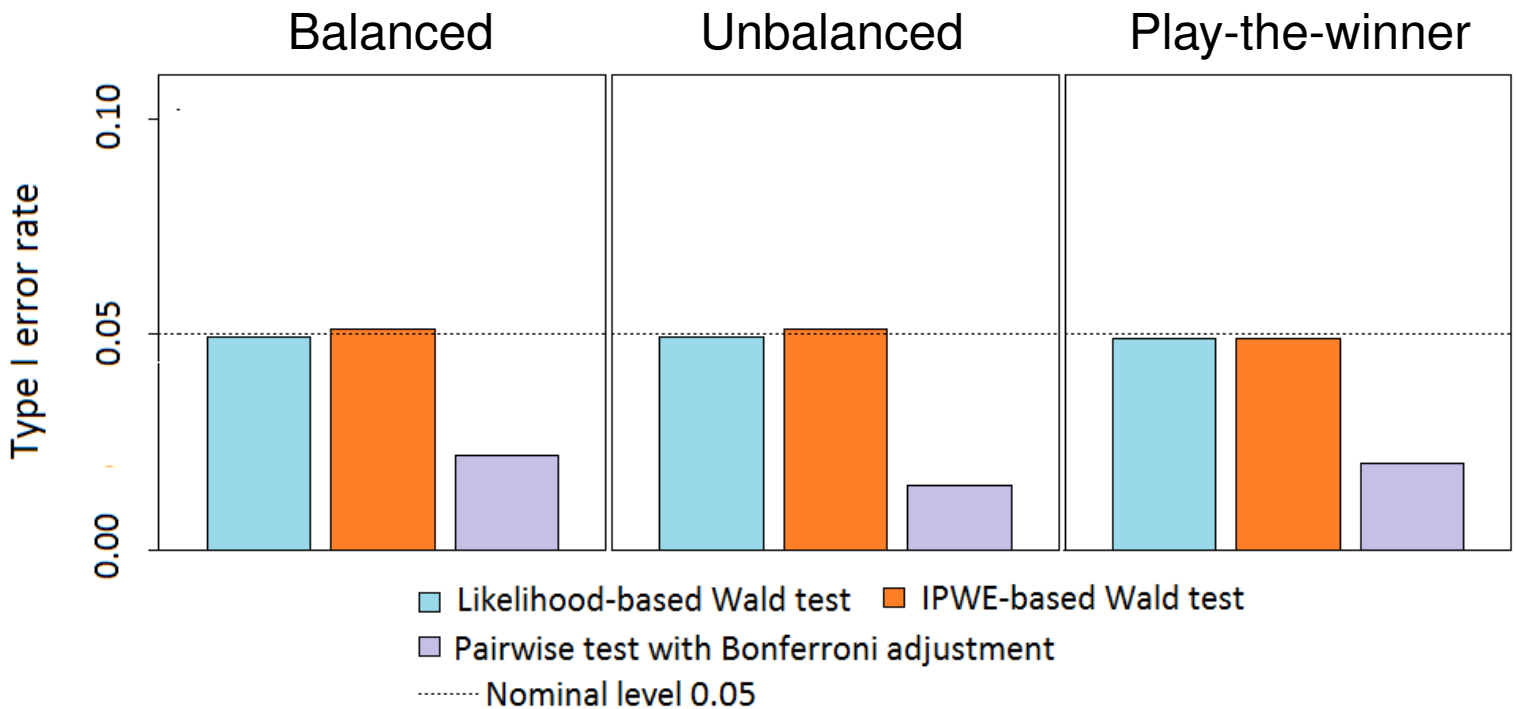
Simulation: Data Generation

- 5000 simulation replicates
- Sample size: $n = 200$
- Effect size: $\Delta = 0.10$
- Data generation
  - Intermediate response rates:
    - $\Pr(X = 1|U) = 1/3$ for $U = 0, 1$
  - Primary outcome

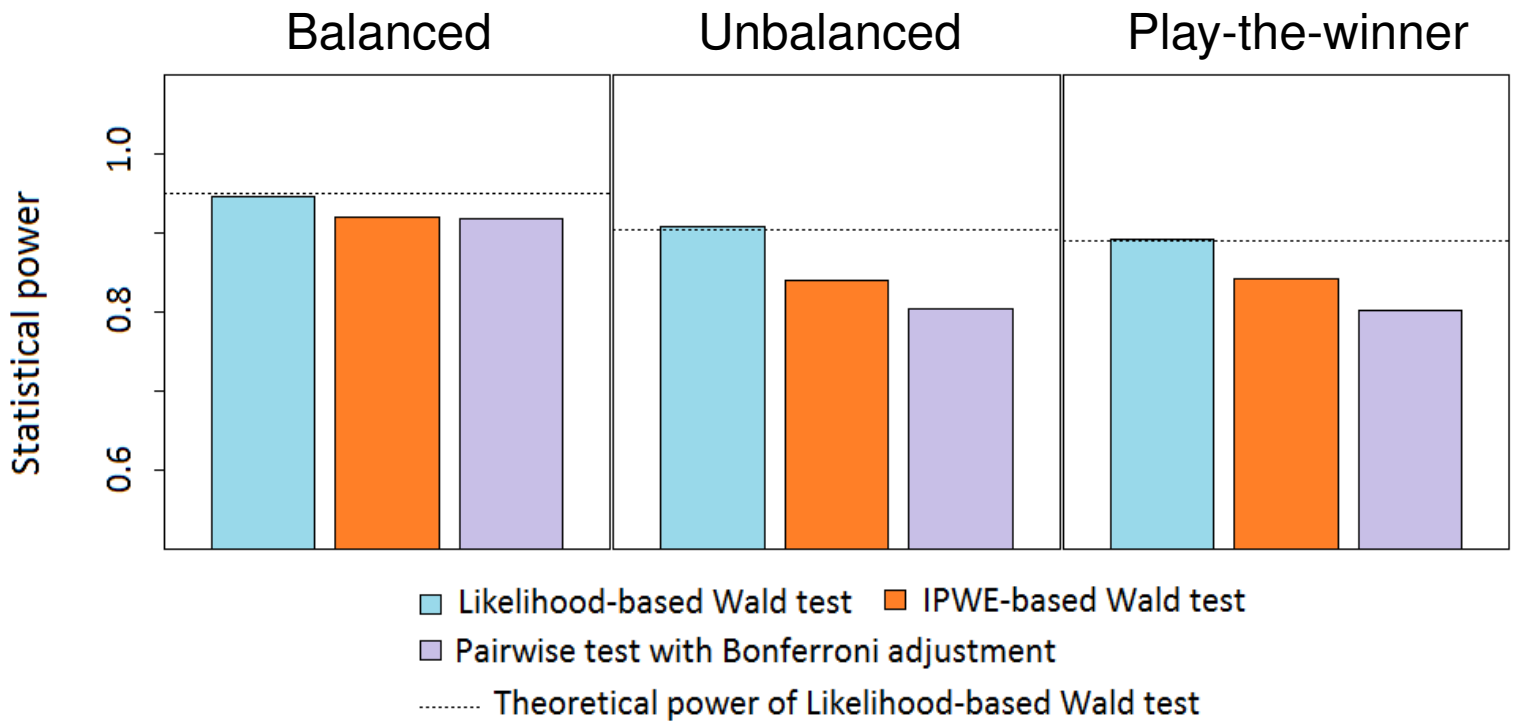
$$Y_{ijk} \overset{iid}{\sim} N(\mu_{ijk}, \sigma^2_{ijk})$$

    - $\sigma^2_{ijk} = 10^2$
    - $\mu_{ijk} = \beta_0 + \beta_1 T_i + \beta_2 R_{ij} + \beta_3 S_{ijk} + \beta_4 T_i R_{ij} + \beta_5 T_i S_{ijk} + \beta_6 R_{ij} S_{ijk} + \beta_7 T_i R_{ij} S_{ijk}$

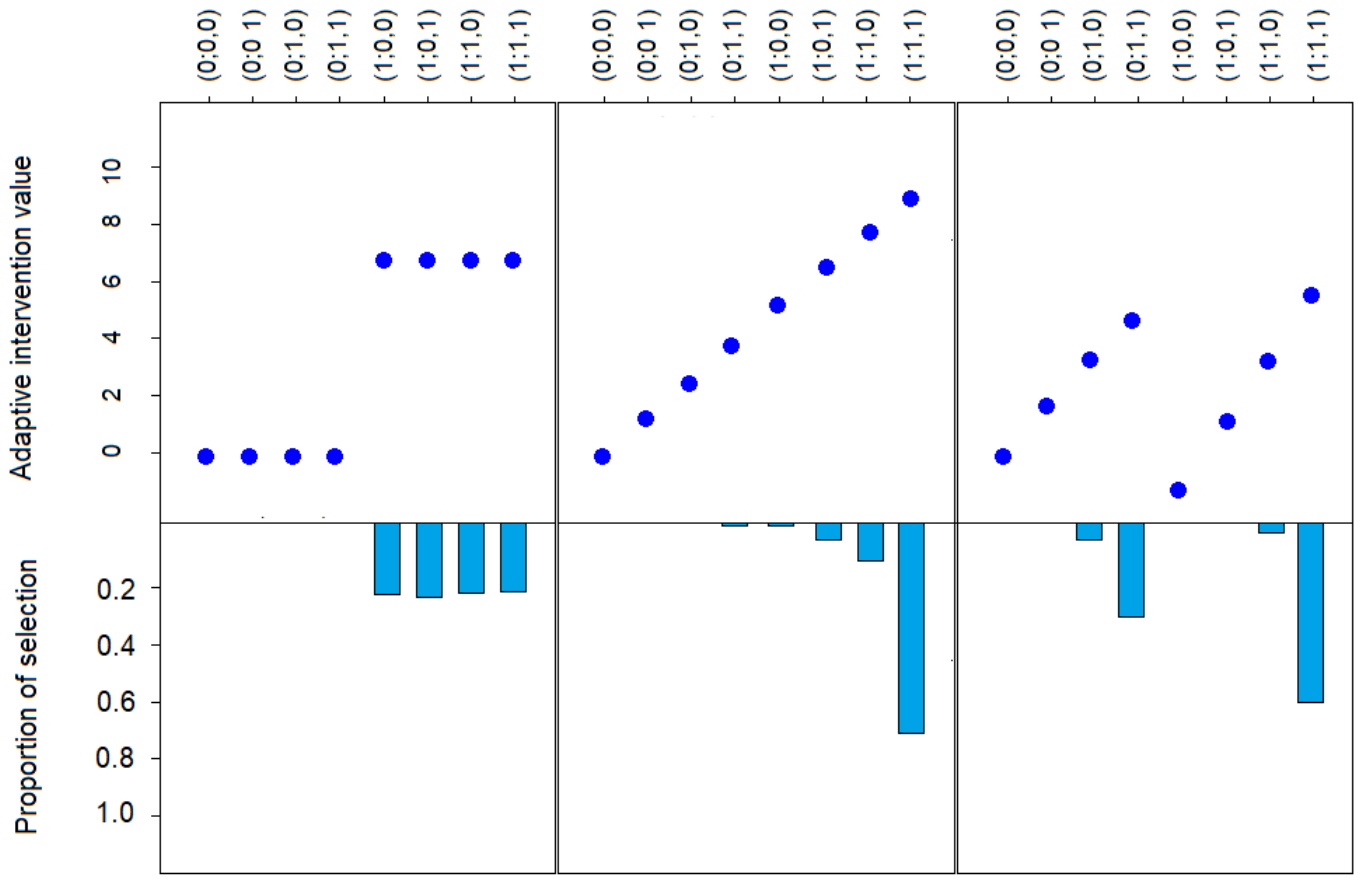
Results: Type I Error Rates (n=200)



Results: Statistical Powers (n=200)



Selection Properties: Results (n=200)



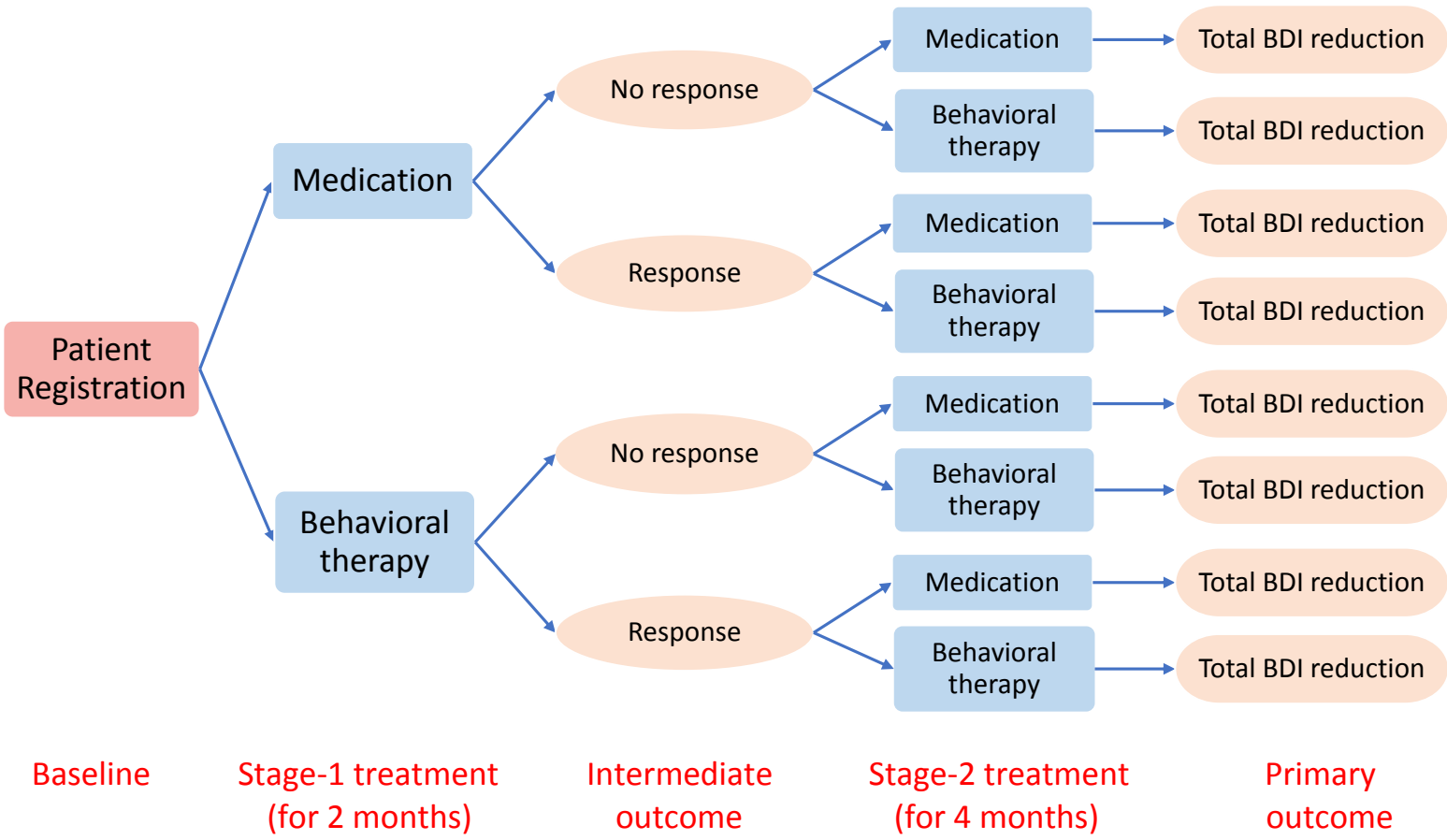
Note: under balanced randomization with effect size $\Delta = 0.10$

Real Data Analysis: CODIACS

- Comparison of Depression Interventions after Acute Coronary Syndromes (CODIACS)[4]
  - The study aimed to assess quality of depression care
  - The primary endpoint was reduction in the level of depression measured by Beck Depression Inventory (BDI)
  - More BDI reduction indicates a better treatment effect

[4] Cheung *et al. Biometrics*, 2015

CODIACS: Design Structure



CODIACS: Selecting the Best Strategy

- The objective was to select the optimal adaptive intervention to move forward to the next phase investigation

g	MLE	p-value
1	6.27	0.135
2	3.33	0.049
3	10.69	0.434
4	7.76	0.210
5	15.45	-
6	9.46	0.320
7	14.23	0.201
8	8.24	0.236

- Gate-keeping approach
  - Step 1: Likelihood-based Wald test ($p < 0.0001$)
  - Step 2: select $g = 5$ as the optimal
- Pairwise comparisons with Bonferroni's adjustment
  - Require a $p < 0.0018$ to achieve overall significance at 5%
  - No overall significance

Conclusion

- In this study
  - we explored the asymptotic distribution of the test statistics and its properties of the proposed likelihood-based Wald test
  - we derived a sample size calculation formula
- Pro: Likelihood-based Wald test
  - outperform the other two existing methods with better power
  - it can applied to varying design structures, randomization schemes and types of outcomes belonging to exponential family
- Con: The performance of likelihood-based test relies on the distributional assumption

References

Habermann T. M. Weller E. A., Morrison V. A., Gascoyne R. D., et al. Horning S. J. (2006), “Rituximab-CHOP versus CHOP along with maintenance Rituximab in older patients with diffuse large B-Cell Lymphoma,” *Journal of Clinical Oncology*, 24, 3121-3127.

Murphy S. (2005). An experimental design for the development of adaptive treatment strategy *Statistics in Medicine*, **24**: 1455-1481.

Ogbagaber, S. B., Karp, J., and Wahed, A. S. (2016). Design of sequentially randomized trials for testing adaptive treatment strategies. *Statistics in Medicine*, **35**: 840-854.

Cheung Y.K., Chakraborty B. and Davidson K.W. (2014). Sequential multiple adaptive randomized trial (SMART) with adaptive design for quality improvement in depression treatment program *Biometrics*, **71**: 450-459.

Cheung Y.K., Ling W., Karr C.J., Weingardt K., Schuller S.M., and Mohr D.C. (2018). Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. *Journal of the American medical Informatics Association*, **25**: 955-962.

Thank You !

Email: xiaobo.zhong@mountsinai.org