# MARGINAL AND CONDITIONAL TREATMENT EFFECTS IN CLINICAL TRIALS


**Hong Tian PhD**

hong.tian@beigene.com

**on behalf of conditional vs marginal task force**

# Working Group

## Lead: Jiawei Wei (Novartis)

## Members:

Björn Bornkamp (Novartis), Ray Lin (Roche), Satrajit Roychoudhury (Pfizer), Hong Tian (BeiGene), Dong Xi (Gilead), **Jiajun Xu (J&J)**, Xin Zhang (Pfizer), Ziqiang Zhao (Novartis)

## Objective:

We would like to bring the complex concept and methods about conditional and marginal treatment effect into a simplified and interpretable way. Potential topics including adjusted or unadjusted analysis; stratified vs unstratified hazard ratio; collapsibility and subgroup; p-values; etc. We will give clinically relevant opinions and recommendations based on our interpretation and illustrate the idea using some case studies.

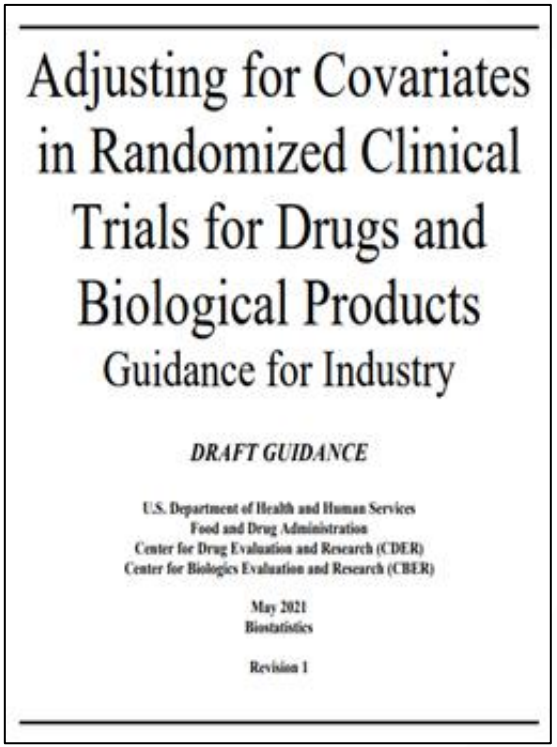

E9(R1) STATISTICAL PRINCIPLES FOR CLINICAL TRIALS: ADDENDUM: ESTIMANDS AND SENSITIVITY ANALYSIS IN CLINICAL TRIALS

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2021
ICH

Revision 1


Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

DRAFT GUIDANCE

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2021
Biostatistics

Revision 1

- ICH E9 encourages the identification of "covariates and factors expected to have an important influence on the primary variables"
- Adjusting for baseline covariates in statistical analysis of a randomized clinical trial can result in more efficient use of the data
  - Compensate for the chance imbalance
  - Improve efficiency in treatment effect estimation
  - Focus on prognostic baseline factors
- ICH E9(R1) requests a precise description of the treatment effect reflecting the clinical questions posed by the trial objective
  - Estimation for a treatment effect should align with the estimand

# Vocabulary

BeiGene

Estimand

**Marginal Treatment Effect**

- Average treatment effect had all patients taken treatment vs. had all patients taken control **in a pre-defined population**
- $Y(1)$ vs $Y(0)$

**Conditional Treatment Effect**

- Average treatment effect had all patients with covariates x taken treatment vs. had they all taken control
- $Y(1)|X = x$ vs $Y(0)|X = x$

Estimation

**Adjusted model**

- Regression model including treatment group + baseline factors

**Unadjusted model**

- Regression model including treatment group only

# Three Myths

- **MYTH # 1:** Marginal and conditional estimands can be compared directly

- **MYTH # 2:** Standardization methods using adjusted model can guarantee improvement in precision by addressing marginal estimand

- **MYTH # 3:** Conditional estimand by stratified analysis for time-to-event data may NOT be improved by adjusting additional prognostic covariates

# Myth # 1: Marginal and Conditional Estimands Can be Compared Directly

BeiGene

## Linear model

- Marginal estimand coincides with conditional estimand
- $Y = \beta_0 + \beta_1 Z + \beta_2 X$
- Efficiency gain by reducing residual variance if the covariates are prognostic

## Non-linear model

- Marginal estimand and conditional estimand differs for common efficacy measures, such as
  - $\text{logit}(\Pr(Y = 1)) = \beta_0 + \beta_1 Z + \beta_2 X$ for binary endpoint;
  - $\lambda(t) = \lambda_0(t) \exp(\beta_1 Z + \beta_2 X)$ for time-to-event endpoint
- Adjusted estimator associated with a larger variance and further away from the null

# Marginal Odds Ratio Differs from Conditional Odds Ratio Due to Non-Collapsibility

| | Percentage of target population | Success rate | | Odds ratio |
|---|---|---|---|---|
| | | New drug | Placebo | |
| Males | 50% | 80.0% | 33.3% | 8.0 |
| Females | 50% | 25.0% | 4.0% | 8.0 |
| Combined | 100% | 52.5% | 18.7% | 4.8 |

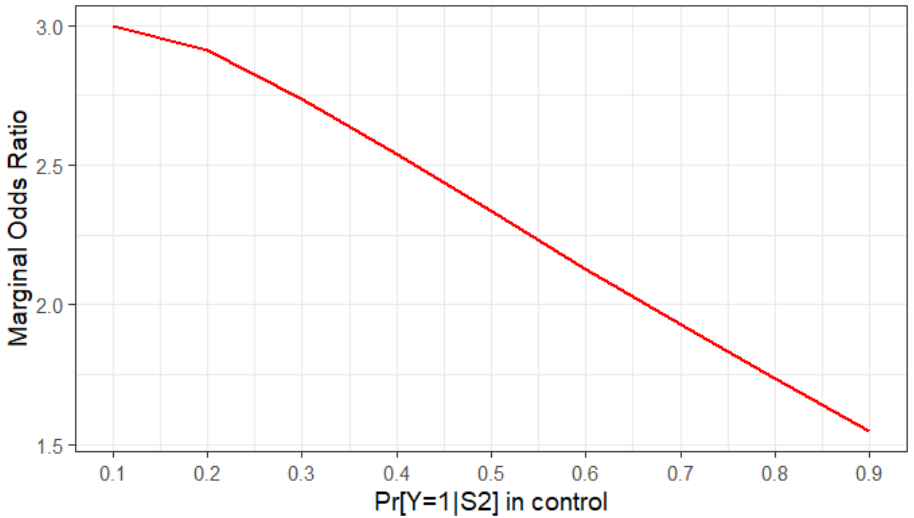
- Treatment effect in each subgroup defined by gender are identical, OR=8 (conditional)
- Treatment effect in the combined population is different, OR=4.8 (marginal)

FDA Guideline. (2021), Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance for Industry

# Marginal Odds Ratio Numerically Moves Away from the Conditional Odds Ratio As the Prognostic Effect Deepens

- Two subgroups ($s_1$ and $s_2$) with equal prevalence
- Odds ratio is constant within each subgroup (OR=3)
- In control arm
  - $\Pr(Y = 1|s_1) = 0.1$; $\Pr(Y = 1|s_2)$ varies in [0.1, 0.9]
- In treated arm
  - $\Pr(Y = 1|s_1)$ and $\Pr(Y = 1|s_2)$ are derived through the constant OR
- Marginal odds ratio is calculated through

$$\frac{\Pr(Y = 1|\text{treated}) / \Pr(Y = 0|\text{treated})}{\Pr(Y = 1|\text{control}) / \Pr(Y = 0|\text{control})}$$



# A Real Example with Feedback from FDA

- Primary estimand uses the marginal odds ratio $\frac{p_1}{1-p_1} / \frac{p_0}{1-p_0}$
- Primary analysis uses the logistic regression with baseline covariates
  - Regression coefficient as the estimate of the primary estimand
- FDA comments
  - Estimand uses the marginal odds ratio, but the logistic regression uses the conditional odds ratio, which does not align with the estimand
  - Clarify whether the population-level summaries are marginal odds ratios or conditional odds ratios, and provide adequate clinical justifications for these choices
  - For conditional odds ratios, the definitions of the estimands should include the variables (and their transformations) on which the odds ratios will condition

**The choice of estimand should inform the choice of analysis approach, not the reverse**

# MYTH # 2: Standardization Methods using Adjusted Model Can Guarantee Improvement in Precision Addressing Marginal Estimand

**Standardization** (estimating standardized outcome distributions using covariate specific estimates of the outcome distribution)

1. Model fitting: fit a regression model (e.g., GLM) considering treatment and pre-specified baseline covariates
2. Prediction: predict potential outcomes under treatment and under control for each patient
3. Average: average treatment effect on predicted outcome
4. Obtain a proper standard error (e.g.. Sandwich estimator, delta method or bootstrap)

# Simulation Studies to Understand Potential Benefit of Standardization

- Study I: Logistic regression – odds ratio
- Study II: Cox regression – hazard ratio

# Simulation

- **Data generation**
  - Generalized linear model on binary outcome:
    $\Pr(Y = 1|Z = z, X = x) = g^{-1}(\beta_0 + \beta_1 z + \beta_2 x)$
    - $Z$ is treatment; $X \sim N(0,1)$ is baseline covariate
    - $g^{-1}(\cdot)$ is the link function
  - Number of simulations = 1000
- **Target treatment effect**: marginal treatment effect
- **Analysis method**
  - Logistic regression without covariate adjustment
  - Logistic regression with covariate adjustment + standardization
  - The analysis models are misspecified
- Performance measure
  - Robustness of standardization: bias, SE
  - Efficiency of standardization: relative SE



13

# Robustness of Standardization

- Data generation $\mathrm{P}(\mathrm{Y}=1|Z=z,\ X=x)=\exp(-|-1.8+1.6z-0.6x|)$
- Summary measure: odds ratio (OR)

| Method | N=500 | N=1000 | N=2000 |
|---|---|---|---|
| | Bias(SE) | Bias(SE) | Bias(SE) |
| Unadjusted estimator | 0.32(1.84) | 0.20(1.22) | 0.06(0.83) |
| Standardization | 0.33(1.78) | 0.21(1.20) | 0.06(0.81) |

N is total sample size

- The standardized estimator is consistent when generalized linear model (with canonical link functions) is mis-specified in randomized trials



14

# Efficiency of Standardization

- Data generation
  - $\mathrm{P}(\mathrm{Y}=1|Z=z,\ X=x)=\boldsymbol{\Phi}(\mathbf{-1.8+1.6}\boldsymbol{z}+\boldsymbol{\beta_2}\boldsymbol{x})$,
  - $\beta_2 = 0, 0.5, 1, 1.5, 2, 4, 6, 8$
  - N=1000

| $\beta_2$ | **0** | **0.5** | **1** | **1.5** | **2** | **4** | **6** | **8** |
|---|---|---|---|---|---|---|---|---|
| Relative SE (standardization/unadjusted estimator) | 1.00 | 0.95 | 0.85 | 0.79 | 0.70 | 0.50 | 0.43 | 0.39 |

- Adjusting for prognostic covariates improves efficiency for estimating unconditional treatment effect



15

# Challenges on Time-to-event Outcome

- Proportional hazard assumption can only hold for either the analysis model for unconditional treatment effect or the analysis model for conditional treatment effect, and censoring distribution adds further complexity
- Application of standardization method to estimate a marginal causal hazard ratio is not straightforward, an analytical solution might be intractable (Daniel et al. 2020)
- Semiparametric estimators (Lu and Tsiatis 2008) are complex and its properties need further exploration
- One could consider alternative summary measures such as restricted mean survival time (RMST) or survival probability difference, however such experience in clinical trials is limited



16

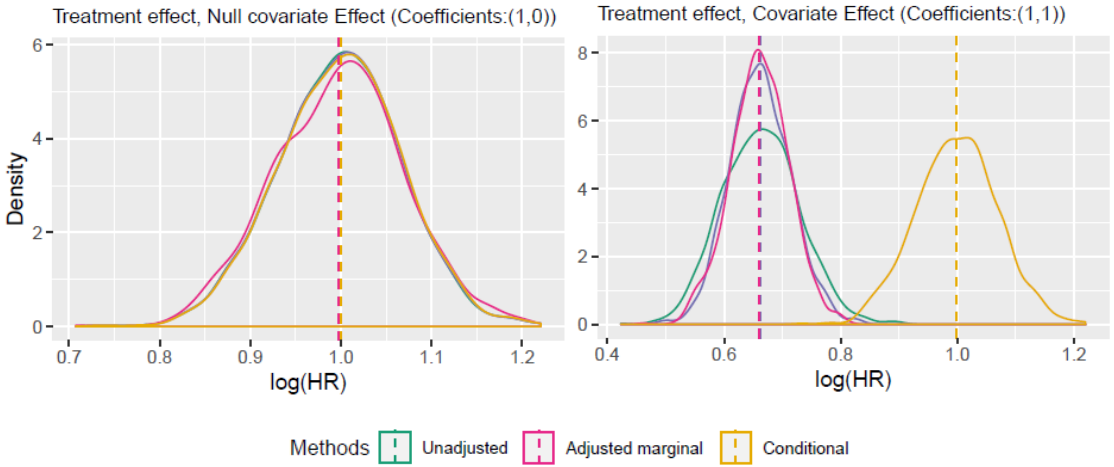
# Simulation Study by Daniel et al (2020)

***Data generation***

- $\lambda(t) = \gamma\lambda t^{\gamma-1}\exp(\theta \cdot z + \beta_x x)$;
- Uniform enrollment over 2 years
- Administrative censoring at 10 years

***Scenarios***

- $\theta = 1, \beta_x = 0$: treatment effect, null covariate effect
- $\theta = 1, \beta_x = 1$: treatment effect, covariate effect




$\lambda = 0.1, \gamma = 1.5$

- The simulation approach to obtain marginal treatment effect is computationally expensive. Need to carefully consider study time frame and censoring distribution
- Efficiency gain is not guaranteed for a given study

# Marginal Estimand Can be Helpful

- To enhance comparability - apple to apple comparisons across data sources
  - especially, when leveraging non-randomized data
- Estimating a marginal effect by standardization approach can in general lead to efficiency gain
  - The efficiency gain can be sizable when the covariate effect is strong
  - Enhance understanding of the strength of covariate effect before applying the approach with pre-specification
- Interpretation and numerical results is based on the pre-defined population of interest

# Myth # 3: Conditional Estimand can NOT be Improved by Adjusting Additional Prognostic Covariates

- Covariate adaptive randomization (e.g., stratified permuted block randomization) is commonly used
- For study with time-to-event endpoints, the primary analysis is often a stratified analysis following the stratified randomization
- Factors used in randomization is usually a subset of potential prognostic covariates
- Is there any room to improve for efficiency without losing robustness?

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

Original Article

# Robust tests for treatment effect in survival analysis under covariate-adaptive randomization

Ting Ye, Jun Shao

First published: 19 August 2020 | https://doi.org/10.1111/rssb.12392 | Citations: 5

# A Simulation to Evaluate the Efficiency of Covariates Adjusted Analysis in a Randomized Study

- **Scenario 1: stratification factor in randomization is subset of prognostic factors**
  - Data generation model: $\lambda(t) = \lambda_0 \exp(\theta \cdot z + \beta_x x_1 + \beta_x x_2^2)$
    - $\lambda_0 = \log(2)/12$; $z$: treatment indicator
    - Negative $\theta$ favors treatment
    - Prognostic factors: $x_1 = 0$ or 1 with $\Pr(x_1 = 1) = 0.5$; $x_2 \sim U(0,1)$
  - Randomization: stratified by $x_1$
- **Scenario 2: continuous prognostic factor is discretized for randomization**
  - Data generation model: same as scenario 1
  - Randomization: stratified by $x_1$ and discretized $x_2$ ($x_2 \leq 0.5$ and $x_2 > 0.5$)
- **Scenario 3: event time is not from the Cox type model**
  - Data generation model: $T = \exp(\theta \cdot z + \beta_x x_1 + \beta_x x_2^2) + \epsilon$
    - $z, x_1, x_2$ are same as scenario 1; $\epsilon \sim EXP(1)$
    - Positive $\theta$ favors treatment
  - Randomization: stratified by $x_1$ and discretized $x_2$ ($x_2 \leq 0.5$ and $x_2 > 0.5$)

# Testing Methods

BeiGene

## *Log-rank tests*

- Unstratified log-rank test
  - Type I error is conservative under stratified randomization
- Stratified log-rank test
  - Type I error is robust under stratified randomization
  - Only account for discrete covariates

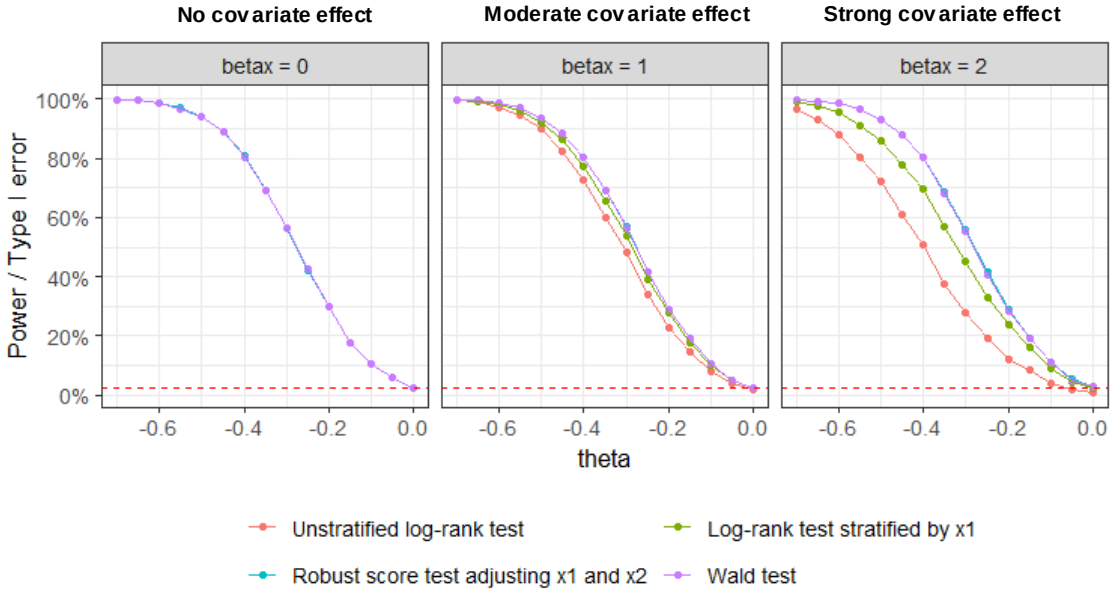
## *Tests based on Cox model :*

$$\lambda(t) = \lambda_0(t)\, exp(\alpha \cdot z + \beta_1 x_1 + \beta_2 x_2)$$

- Robust score test (Ye and Shao 2020)
  - Type I error is robust to stratified randomization and model misspecification
  - Inefficient if the analysis model is very different from the true model
- Wald test
  - Type I error is inflated if analysis model is wrong

Ye, T. & Shao J. (2020) Robust Tests for Treatment Effect in Survival Analysis under Covariate -Adaptive Randomization, J.R. Statist. Soc. B 82 (5) 1301-1323

# Scenario 1: Stratification Factor in Randomization is Subset of Prognostic Factors



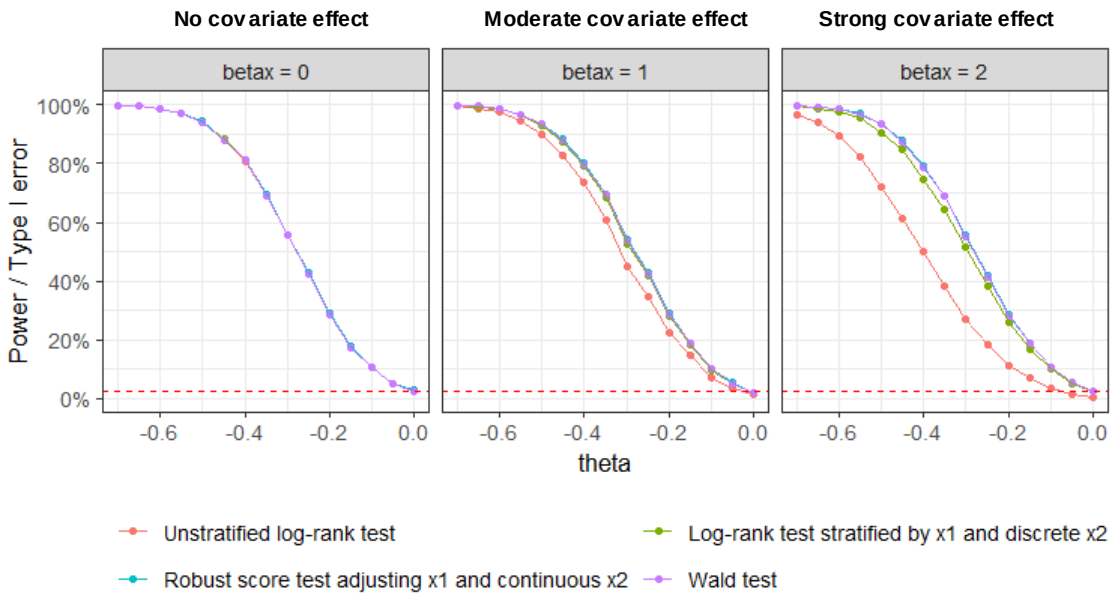

***Testing methods***

- Unstratified Log-rank test
- Log-rank test stratified by $x_1$
- Robust score test adjusting both $x_1$ and continuous $x_2$
- Wald test

***Observations***

- The unstratified Log-rank test is conservative under stratified randomization
- Power is enhanced by adjusting the covariate not considered in randomization (i.e., stratified log rank test and robust score test)
- Wald test performs well because the working model is close to true model

# Scenario 2: Continuous Prognostic Factor is Discretized for Randomization



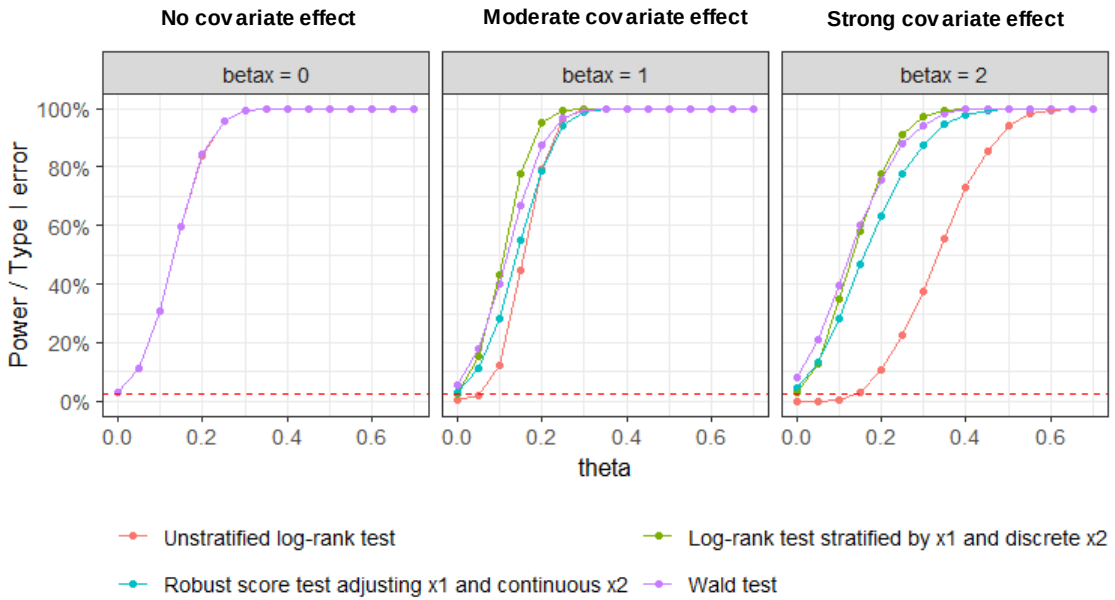

***Testing methods***

- Unstratified Log-rank test
- Log-rank test stratified by $x_1$ and discrete $x_2$
- Robust score test adjusting both $x_1$ and continuous $x_2$
- Wald test

***Observations***

- Adjusting for more prognostic factors in stratified log-rank test can enhance power

# Scenario 3: Event Time is not Simulated from a Cox Type Model




### *Testing methods*

- Unstratified Log-rank test
- Log-rank test stratified by $x_1$ and discrete $x_2$
- Robust score test adjusting $x_1$ and continuous $x_2$
- Wald test

### *Observations*

- Stratified log-rank test is superior to the robust score
- Type I error of Wald test is severely inflated since working model is very wrong

# What Works Well for the Time to Event Analysis

- Unstratified log-rank test tests a marginal treatment effect. It is conservative under stratified randomization
- For additional prognostic factors not part of stratification factors for randomization, including these variables into the analysis model may further enhance the study power
- Some covariates are continuous in nature, adjusting these covariates using their continuous scale may help to improve efficiency
- "All models are wrong", consider a robust approach to draw valid statistical inference
- Stratified log-rank test performs well in general

# Conclusions

- Estimation of treatment effect should align with the target estimand
- Conditional estimand and marginal estimand are both population level summary of treatment effect and should be clearly differentiated
- For binary outcome, standardization procedure is a robust approach to estimate the marginal treatment effect
  - Efficiency gain can be expected if the adjusted covariates have strong prognostic effect
- For time-to-event outcome, standardization procedure is tricky to implement, and complicated by censoring and time frame
- Stratified log-rank test performs well, robust score test approaches offer good promises

# References


- FDA Guideline (2021), Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance for Industry
- FDA Guideline (2021), E9 (R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials Guidance for Industry
- Daniel, R., Zhang, J., & Farewell, D. (2021). Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*, *63*, 528-557.
- Díaz, I., Colantuoni, E., Hanley, D. F., & Rosenblum, M. (2019). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. *Lifetime data analysis*, *25*(3), 439-468.
- Karrison, T., & Kocherginsky, M. (2018). Restricted mean survival time: Does covariate adjustment improve precision in randomized clinical trials?. *Clinical Trials*, *15*(2), 178-188.
- Lu, X., & Tsiatis, A. A. (2008). Improving the efficiency of the log-rank test using auxiliary covariates. *Biometrika*, *95*(3), 679-694.
- Rosenblum, M., & Van Der Laan, M. J. (2010). Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *The international journal of biostatistics*, *6*(1).
- Ye, T. & Shao J. (2020) Robust Tests for Treatment Effect in Survival Analysis under Covariate-Adaptive Randomization, J.R. Statist. Soc. B 82 (5) 1301-1323
- Trinquart, Ludovic, et al. "Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials." *Journal of Clinical Oncology* 34.15 (2016): 1813-1819.