DIA 中国统计社区

Conditional and Marginal Treatment Effects in Clinical Trials

Jiajun Xu PhD

Statistics and Decision Science, Janssen R&D

on behalf on conditional vs marginal effects task force

主持人

韦加为 博士

诺华全球药物研发生物统计部门

生物统计总监

韦加为博士2011年加入诺华，现任高级方法和数据科学部的统计方法总监。主要从事内部生物统计方法顾问以及创新统计方法研究，主要研究领域包括估计目标，复发数据分析，多重检验，适应性设计等。被授予诺华leading scientist 荣誉.

加入诺华前在Texas A&M大学获得统计博士学位，并从事一年助理教授工作。是Statistics in Biopharmaceutical Research的副主编，同时在复旦大学担任兼职导师职位。韦加为博士参与了ICH M11方案模板中估计目标部分的撰写，同时也是国际肿瘤估计目标条件和边际治疗效应工作组成员。

演讲嘉宾

徐家俊 博士

强生全球统计模型与高级方法部

高级资深统计师

徐家俊博士于2016年加入强生，现任强生全球统计模型与高级方法部的高级资深统计师。主要支持从试验设计到项目递交阶段复杂与创新统计方法的探索与应用，是国际肿瘤估计目标工作组的重要成员，主要研究领域包括生存分析，估计目标，适应性设计，多重检验，贝叶斯统计等。徐博士同时也作为临床项目统计师，负责多个产品在实体瘤与疫苗领域的临床研究与注册。 徐家俊博士于2016年获得香港大学统计学博士学位。

讨论嘉宾

赵耐青 教授

缔脉公司首席科学家

复旦大学生物统计学教授、博导

- 专长复杂临床试验设计，临床试验中的统计学方法
- **国家**传染病医学中心（**复旦大学附属**华山医院）**双聘**专家
- **复旦大学学**报（**医学版**）**副主**编
- **国家**药监局药品审评中心外审专家和参与临床试验指导原则撰写
- **上海市区域**伦理委员会委员和**3家三**级医院伦理委员和**2家三**级医院临床试验学术会委员
- **中国**卫生信息学会卫生统计教学专业委员会副主任委员

讨论嘉宾

William Koh 博士

美国食品药品管理局(FDA)

高级数理统计师

William Koh现为美国食品药品管理局(FDA)的高级数理统计师。William在华盛顿大学获得生物统计学博士和硕士学位，他的论文研究主要关注对至事件发生终点的适应性临床试验设计与成组序贯设计的比较。William于2016年加入FDA生物计量学II部生物统计室，参与过肺、肾、心脏、及风湿疾病产品，和生物仿制药的NDA/BLA/IND评审。他目前的研究兴趣包括成组序贯分析，缺失数据处理，estimand，生存分析，层级建模，以及协变量调整。

讨论嘉宾

田红 博士

百济神州全球统计负责人

田红博士现任百济神州的全球统计负责人。田博士领导和管理百济神州的统计部门，包括实体瘤临床、血液学临床、上市后、以及科研和生产统计。

加入百济神州前，田博士在强生全球制药研发 (J&J) 拥有超过 15 年的临床开发经验，任肿瘤学领域早期开发统计负责人和 BCMA CAR-T 项目全球临床团队统计主管；在此之前，田博士还支持了跨治疗领域的各阶段的药物开发，并为肿瘤、心血管和代谢疾病以及传染病领域的多项创新药的全球注册上市做出了贡献，包括 DARZALEX、ERLEADA、XARELTO、INVOKANA 和 SIRTURO。田红博士获得哥伦比亚大学生物统计学博士学位。

讨论嘉宾

Ray Lin 博士

罗氏/基因泰克

高级资深统计师

Ray Lin 现为罗氏/基因泰克公司的高级资深统计师。Ray工作12年来一直支持肿瘤学的早期和晚期药物开发，对统计学方法应用于临床试验拥有极大的热情。Ray还活跃于各个生物统计社区：他是ASA旧金山湾区地方分会(SFASA)前主席，湾区生物制药统计研究会(BBSW)副主席，跨制药公司的非比率风险模型（NPH）工作组领导成员，以及Estimand条件和边际治疗效应研究组成员。Ray于2010年获得斯坦福大学统计学博士学位。

Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to DIA, its directors, officers, employees, volunteers, members, chapters, councils, Communities or affiliates, or any organization with which the presenter is employed or affiliated.

Conditional and Marginal Treatment Effects in Clinical Trials

Acknowledgement

Conditional vs Marginal Effects Working Group

Lead: Jiawei Wei (Norvatis)

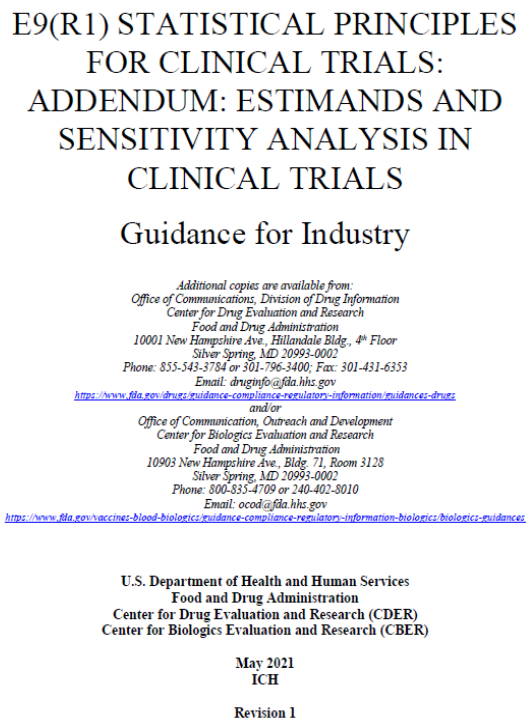
Members

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.

Motivation



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

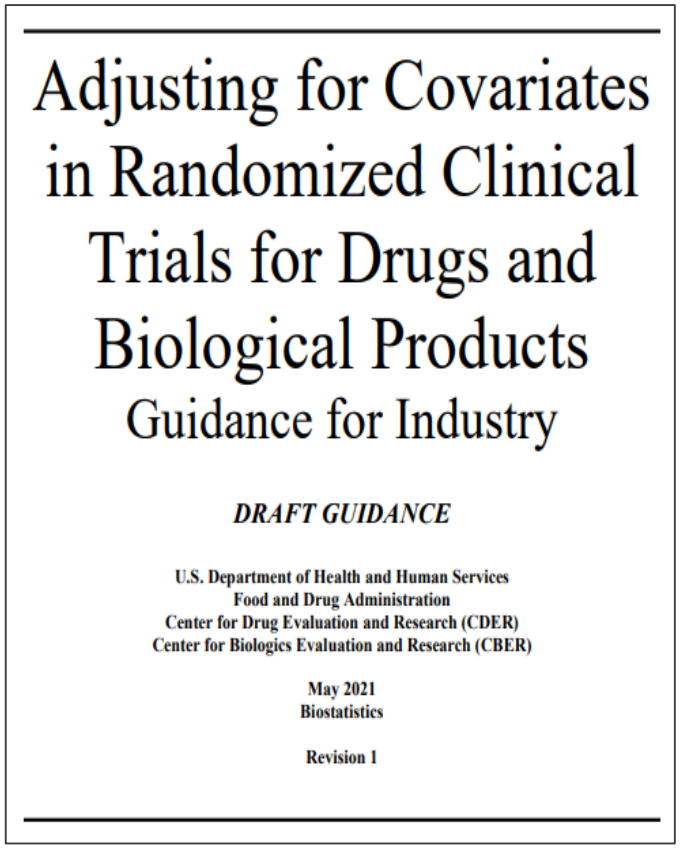
Guidance for Industry

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs
and/or
Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances

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
- 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

Adjusting for Covariates is a Common Practice

Continuous endpoint

$$Y = \beta_0 + \beta_1 Z + \beta_2 X$$

Adjustment for baseline covariates can

- Compensate chance imbalance between treatment groups
- Reduce the variability of the estimated treatment effects (narrower confidence interval, more powerful hypothesis testing)
- Still be valid for inference on the average treatment effect even when the regression model dose not fully capture the relationships between the outcome, treatment and covariates (Lin 2013)

Does these good properties also apply to

Binary endpoint

$$\text{logit}(\Pr(Y = 1)) = \beta_0 + \beta_1 Z + \beta_2 X$$

Time-to-event endpoint

$$\lambda(t) = \lambda_0(t) \exp(\beta_1 Z + \beta_2 X)$$

Note: Z for treatment; X for covariates

Source: Lin, W., 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*, *7*(1), pp.295-318.

Adjusted Model vs Unadjusted Model
Conditional Treatment Effect vs Marginal Treatment Effect

Unadjusted Model

$$Y = \beta_0 + \beta_1 Z$$
$$\text{logit}(\Pr(Y = 1)) = \beta_0 + \beta_1 Z$$
$$\lambda(t) = \lambda_0(t)\exp(\beta_1 Z)$$

- Average treatment effect had all patients taken treatment vs had they all taken control

Marginal treatment effect

Estimation

Estimand

Adjusted Model

$$Y = \beta_0 + \beta_1 Z + \beta_2 X$$
$$\text{logit}(\Pr(Y = 1)) = \beta_0 + \beta_1 Z + \beta_2 X$$
$$\lambda(t) = \lambda_0(t)\exp(\beta_1 Z + \beta_2 X)$$

- Average treatment effect had all patients with covariates X taken treatment vs had they all taken control
- Assuming constant treatment effect across subgroups defined by covariates

Conditional treatment effect

Linear Model

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

Non-linear Model

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

Marginal Mean Difference Equals to Conditional Mean Difference

	Percentage of target population	Mean of change		Mean difference Δ
		New drug	Placebo	
Males	50%	8	4	4
Females	50%	6	2	4
Combined	100%	7	3	4

- Treatment effect in each subgroup defined by gender are identical, $\Delta = 4$ (conditional)
- Treatment effect in the combined population is the same, $\Delta = 4$ (marginal)

Marginal Odds Ratio Differs from Conditional Odds Ratio

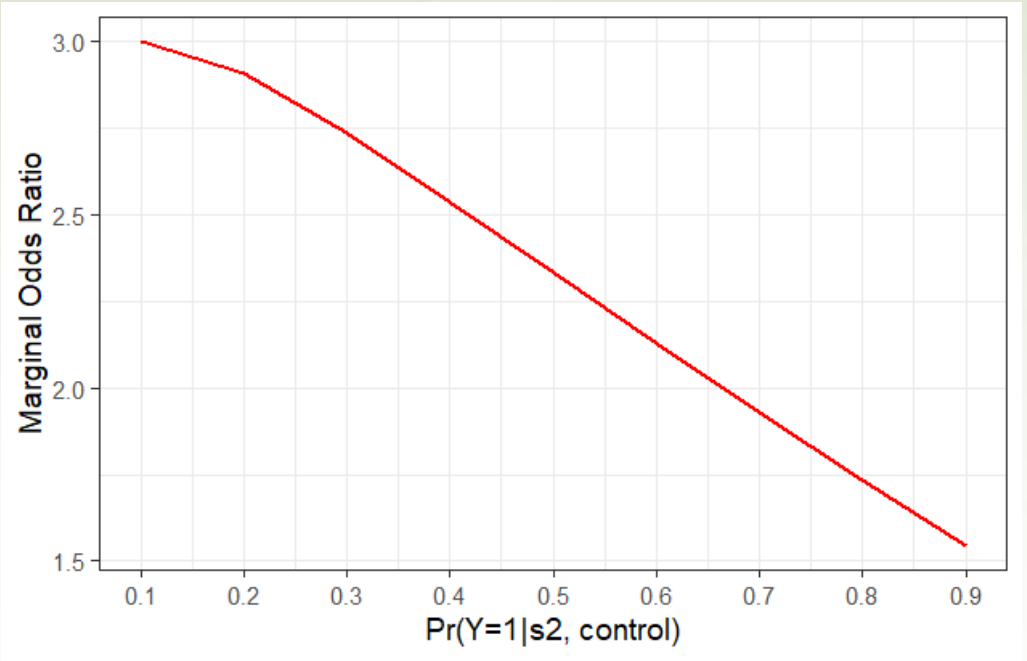
	Percentage of target population	Success rate		Odds ratio
		New drug	Placebo	
Males	50%	80.0%	33.3%	8.0
Females	50%	25.0%	4.0%	8.0
Combined	100%	52.5%	18.7%	4.8

- 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)

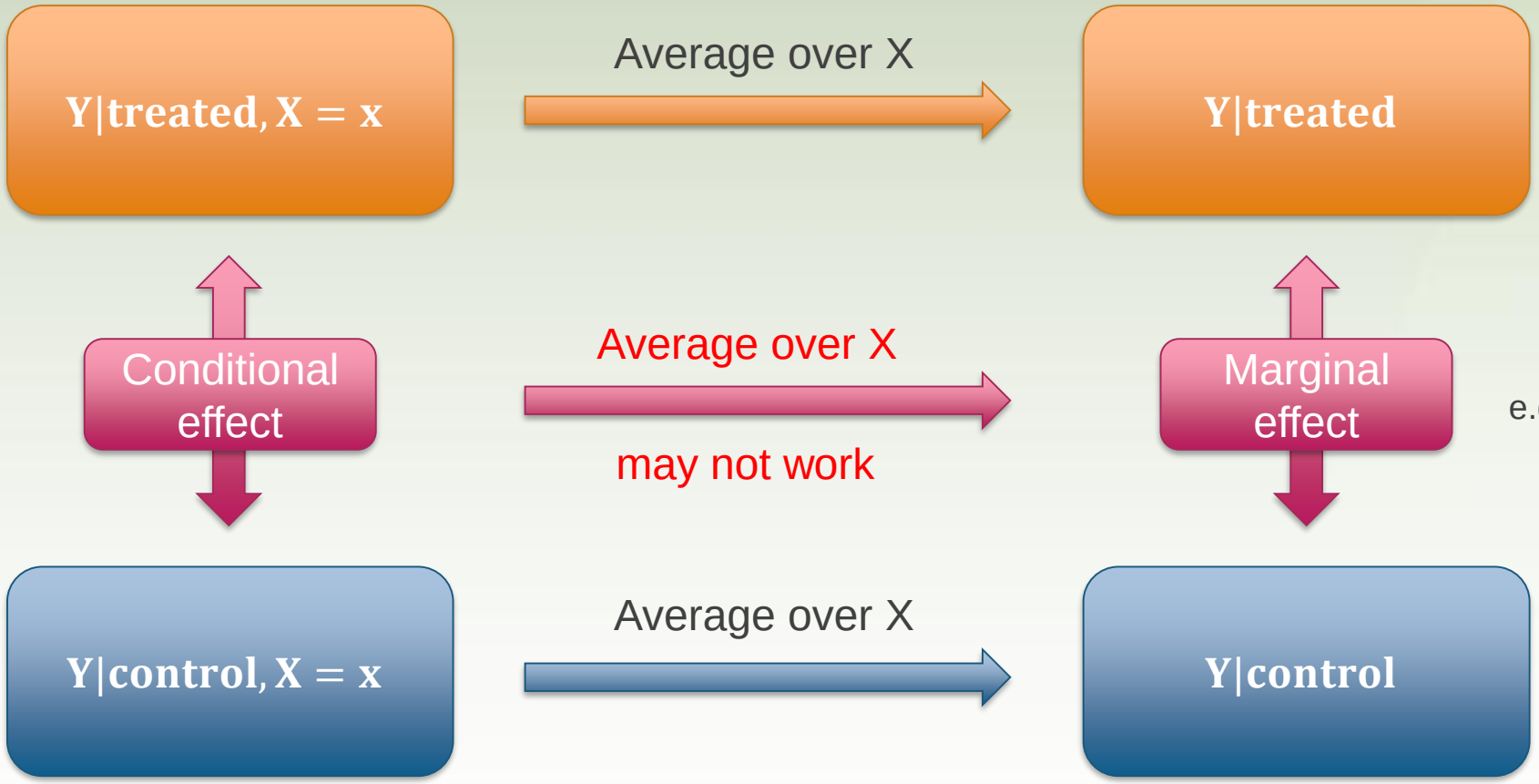
Source: FDA Guideline. (2021), Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance for Industry

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

- Two subgroups (s_1 and s_2) have equal prevalence
- Odds ratio is constant within each subgroup (OR=3)
- In control arm
  - $\Pr(Y = 1|s_1, \text{control}) = 0.1$
  - $\Pr(Y = 1|s_2, \text{control})$ varies in [0.1, 0.9]
- In treated arm
  - $\Pr(Y = 1|s_1, \text{treated})$ and $\Pr(Y = 1|s_2, \text{treated})$ can be derived through the constant OR
- **Marginal odds ratio** in the overall population is calculated through
  - $$\frac{\Pr(Y = 1|\text{treated})/\Pr(Y = 0|\text{treated})}{\Pr(Y = 1|\text{control})/\Pr(Y = 0|\text{control})}$$



Why Marginal Effect Differs From Conditional Effect



e.g., $E_X\left[\frac{E(Y|treated, X = x)}{E(Y|control, X = x)}\right] \neq \frac{E(Y|treated)}{E(Y|control)}$

For more comprehensive and formal explanation please refer to: 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.

A Real Example

- A randomized placebo-controlled trial to compare drug A to placebo for a disease
- Primary endpoint is a binary response variable
- Primary estimand uses the marginal odds ratio $\frac{p_1}{1-p_1} / \frac{p_0}{1-p_0}$
  - p_1 and p_0 are response rates in treatment and control arms, respectively
- Primary analysis uses the logistic regression including treatment and 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 estimand 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"*

Estimate Marginal Effect with Covariate Adjustment

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

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., delta method, bootstrap)

Simulation Studies to Understand the Potential Benefit of Standardization

- **Data generation**
  - Generalized linear model for 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
    - $g^{-1}()$ is the link function where we considered **log link** and **probit link**
  - Number of simulations: 1000
- **Target treatment effect**: marginal treatment effect
- **Analysis method**
  - Logistic regression without covariate adjustment
  - Logistic regression with covariate adjustment + standardization
- **Performance measure**
  - Robustness of standardization: bias, SE
  - Efficiency of standardization: relative SE

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=200	N=500	N=1000	N=2000
	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)
Unadjusted estimator	0.93(3.43)	0.32(1.84)	0.20(1.22)	0.06(0.83)
Standardization	0.95(3.31)	0.33(1.78)	0.21(1.20)	0.06(0.81)

- The standardized estimator is consistent when generalized linear model is **misspecified** in randomized trials (Rosenblum and Steingrimsson, 2018)

Efficiency of Standardization

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

β_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 marginal treatment effect

Marginal Estimand Can be Helpful

- 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 the covariate effect before applying the approach with pre-specification
- To enhance comparability – apple to apple comparisons across data sources
- Interpretation and numerical results is based on the pre-defined population of interest

Challenges on Time-to-event Outcome

- $\lambda(t) = \lambda_0(t)\exp(\theta z)$ → unadjusted model → marginal estimand
- $\lambda(t) = \lambda_0(t)\exp(\theta z + \beta x)$ → adjusted model → conditional estimand
- Proportional hazard assumption can only hold for at most one of the above models
- If the adjusted model is true, marginal hazard ratio in the overall population varies over time
  - $\theta \rightarrow \theta(t)$
  - The estimated HR under marginal model can be interpreted as average HR (Rauch et al 2018)
  - The censoring distribution also plays a role in the interpretation, which adds further complexity

Rauch, G., Brannath, W., Brückner, M. and Kieser, M., 2018. The Average Hazard Ratio–A Good Effect Measure for Time-to-event Endpoints when the Proportional Hazard Assumption is Violated?. *Methods of information in medicine*, *57*(03), pp.089-100.

A Simulation Approach to Estimate Marginal HR Through Covariates Adjusted Model (Daniel et al 2020)

Step A

- Model fitting: fit a Cox regression model considering treatment and pre-specified baseline covariates
- Prediction: predict the survival function under treatment and under control for each patient
- Average: average the survival function under treatment and under control:
- Simulation: simulate a set of event times using the average survival functions

Step B

- Same as part A but reverse the censoring indicator (being censored is the event of interest)

Step C

- The observed time is the minimum of the simulated event time (from Part A) and the censoring time (from Part B)
- Fit a Cox model on the simulated data using treatment as the only explanatory variable to estimate the marginal HR

Step D

- Repeat Step A to Step C to get the empirical distribution of the estimated marginal HR

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.

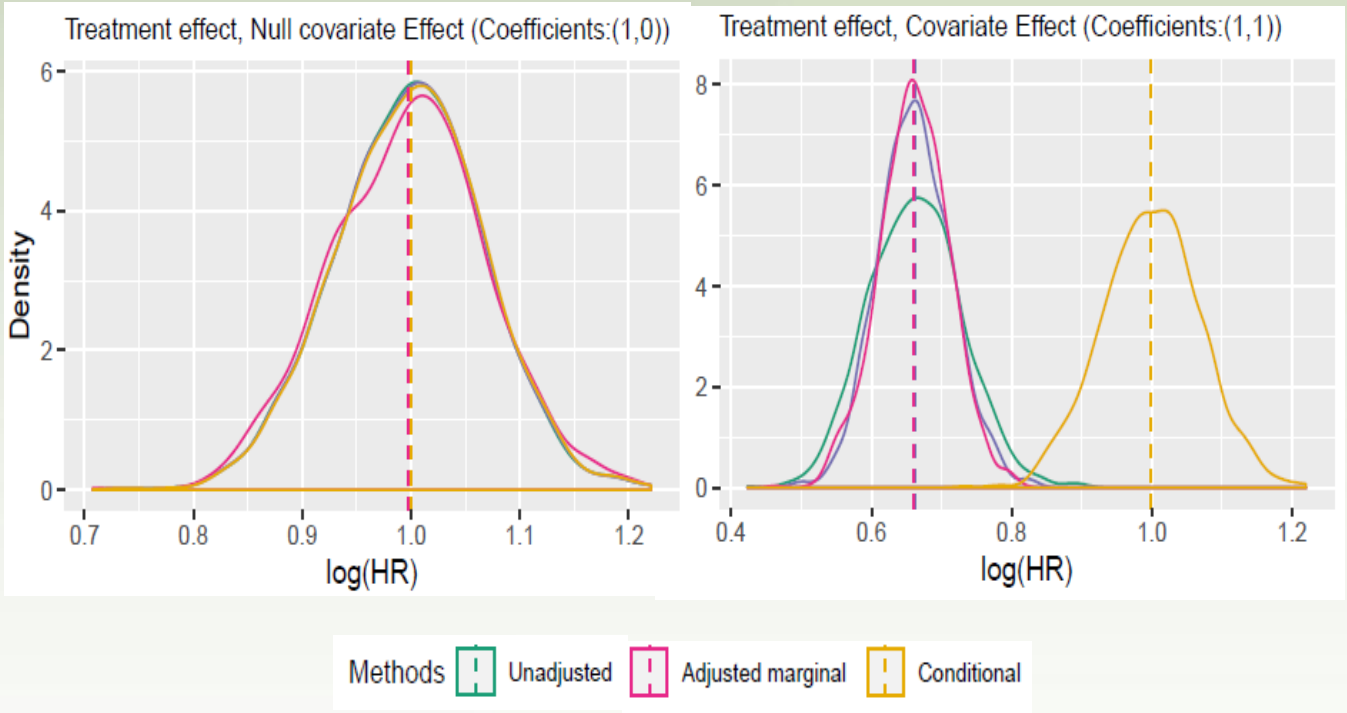
Simulation Study (Daniel et al 2020)

Data generation

- $\lambda(t) = \gamma\lambda t^{\gamma-1}\exp(\theta z + \beta_x x)$
  - $\lambda = 0.1, \gamma = 1.5$
- 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



- *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*

Current Practice for Study with Time-to-event Endpoint

- Covariate adaptive randomization is commonly used (e.g., stratified permuted block randomization)
  - Ensure prognostic factors are balanced between treatment groups
- Factors used in randomization is usually a subset of potential prognostic covariates
  - To avoid too many strata
- For study with time-to-event endpoints, the primary analysis is often a stratified analysis following the stratified randomization
  - Stratified analysis targets a conditional estimand
  - Unstratified analysis targets a marginal estimand. Conservative under stratified randomization
- If the ***conditional estimand*** is interested, is there any room to improve for efficiency without losing robustness?
  - Model misspecification is often concerned for conditional model

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2020)
82, *Part* 5, *pp.* 1301–1323

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

Ting Ye and Jun Shao

East China Normal University, Shanghai, People
University of Wisconsin—Madison, USA

[Received December 2018. Final revision July 202(

Summary. Covariate-adaptive randomization is popu
patients for balancing treatment assignments across
on the response. However, existing theory on tests
adaptive randomization is limited to tests under line
the covariate-adaptive randomization method has be
Often, practitioners will simply adopt a conventional
controversial since tests derived under simple rando
I error under other randomization schemes. We deriv
likelihood score function under covariate-adaptive ra
subject to possible model misspecification. Using thi
likelihood score test that is robust against model mis
is no longer robust but conservative under covariat
that the unstratified log-rank test is conservative an
under covariate-adaptive randomization. We propos
the partial likelihood score test, which leads to a scor
trary model misspecification under a large family of c
including simple randomization. Furthermore, we sho
test derived under a correctly specified model is more
of Pitman's asymptotic relative efficiency. Simulation
of various tests are presented under several popular

Keywords: Cox model; Log-rank test; Pitman's asym
against model misspecification; Stratified permuted b

Published in final edited form as:
J Am Stat Assoc. 2018 ; 113(524): 1784–1796. doi:10.1080/01621459.2017.1375934.

Inference under Covariate-Adaptive Randomization*

Federico A. Bugni,
Department of Economics, Duke University, federico.bugni@duke.edu

Ivan A. Canay, and
Department of Economics, Northwestern University, iacanay@northwestern.edu

Azeem M. Shaikh
Department of Economics, University of Chicago, amshaikh@uchicago.edu

Abstract

This paper studies inference for the average treatment effect in randomized controlled trials with covariate-adaptive randomization. Here, by covariate-adaptive randomization, we mean randomization schemes that first stratify according to baseline covariates and then assign treatment status so as to achieve "balance" within each stratum. Our main requirement is that the randomization scheme assigns treatment status within each stratum so that the fraction of units being assigned to treatment within each stratum has a well behaved distribution centered around a proportion π as the sample size tends to infinity. Such schemes include, for example, Efron's biased-coin design and stratified block randomization. When testing the null hypothesis that the average treatment effect equals a pre-specified value in such settings, we first show the usual two-sample t-test is conservative in the sense that it has limiting rejection probability under the null hypothesis no greater than and typically strictly less than the nominal level. We show, however, that a simple adjustment to the usual standard error of the two-sample t-test leads to a test that is exact in the sense that its limiting rejection probability under the null hypothesis equals the nominal level. Next, we consider the usual t-test (on the coefficient on treatment assignment) in a linear regression of outcomes on treatment assignment and indicators for each of the strata. We show that this test is exact for the important special case of randomization schemes with $\pi = \frac{1}{2}$, but is otherwise conservative. We again provide a simple adjustment to the standard errors that yields an exact test more generally. Finally, we study the behavior of a modified version of a permutation test, which we refer to as the covariate-adaptive permutation test, that only permutes treatment status for units within the same stratum. When applied to the usual two-sample t-statistic, we show that this test is exact for randomization schemes with $\pi = \frac{1}{2}$ and that additionally achieve what we refer to as "strong balance." For randomization schemes with $\pi \neq \frac{1}{2}$, this test may have limiting rejection probability under the null hypothesis strictly greater than the nominal level. When applied to a suitably adjusted version of the two-sample t-statistic, however, we show that this test is exact for all randomization schemes that achieve "strong balance," including those with $\pi \neq \frac{1}{2}$. A simulation study confirms the practical relevance of our theoretical results. We conclude with recommendations for empirical practice and an empirical illustration.

Simulation Studies 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: $\lambda(t) = \lambda_0 \exp(\theta z + \beta_x x_1 + \beta_x x_2^2)$
    - z is the treatment indicator, negative θ 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: 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 generated from the Cox type model**
  - Data generation: $T = \exp(\theta z + \beta_x x_1 + \beta_x x_2^2) + \varepsilon$
    - z, x_1, x_2 are the same as scenario 1; $\varepsilon \sim EXP(1)$
    - Positive θ favors treatment
  - Randomization: stratified by x_1 and discretized x_2 ($x_2 \leq 0.5$ and $x_2 > 0.5$)

Testing Methods

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

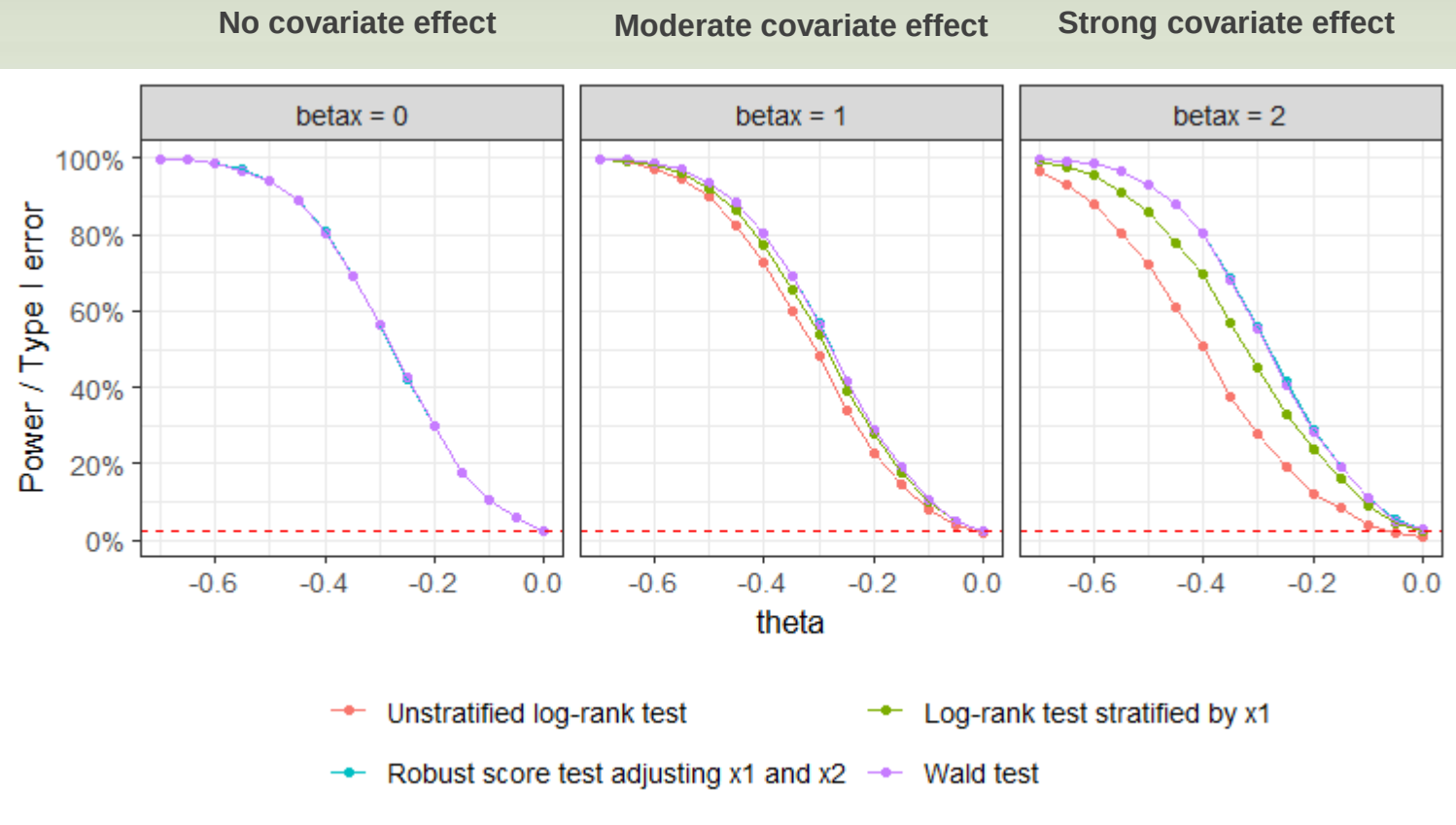
Tests based on Cox model

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

- 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

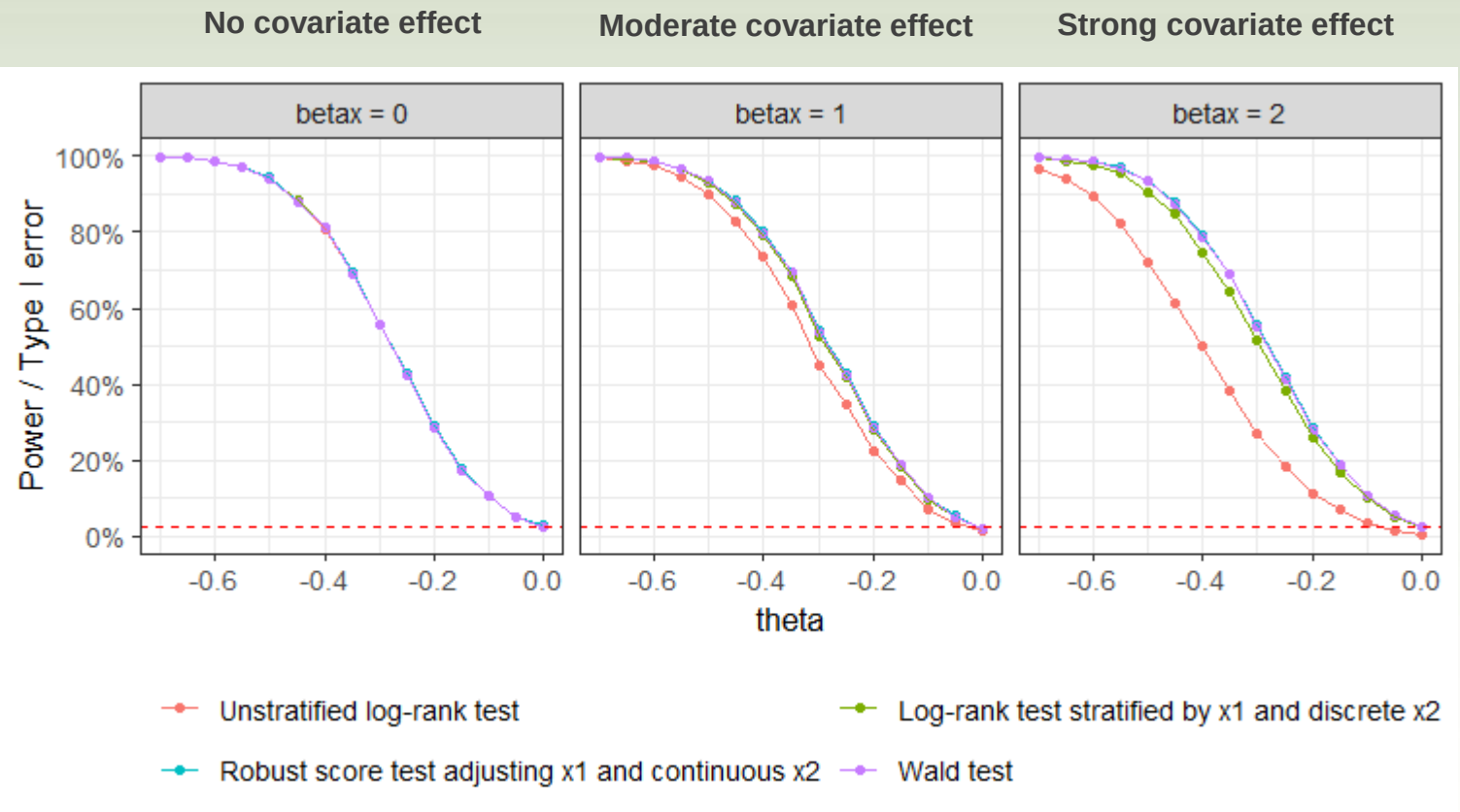
Ye, T. & Shao J. (2020) Robust Tests for Treatment Effect in Survival Analysis under Covariate-Adaptive Randomization, J.R. Statist. Soc. B 82 (5) 1301-1323

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



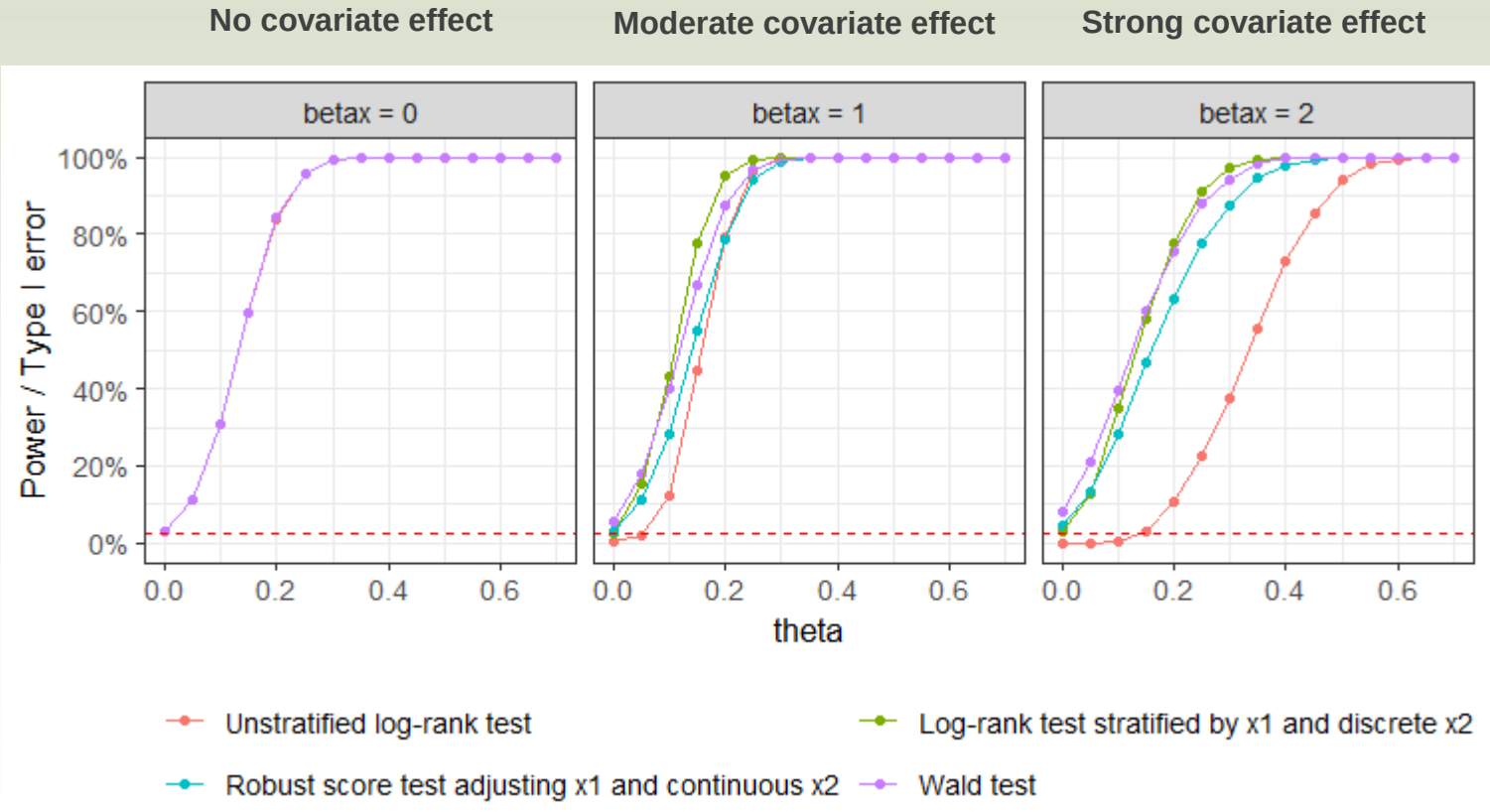
- When there is no covariate effect, all methods lead to same result
- The unstratified Log-rank test is conservative under stratified randomization
- Power is enhanced by adjusting the covariate not considered in randomization (i.e., robust score test)
- Wald test performs well because the working model is close to true model

Scenario 2: Continuous Prognostic Factor is Discretized for Randomization



- Power is enhanced by adjusting the original continuous covariate (i.e., robust score test)
- 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



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

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

Connection with Classic Theory

"**In proportional hazards model with omitted covariates, $\hat{\boldsymbol{\beta}}$ as an estimator of the regression parameter in the true model is asymptotically biased toward zero*"

– ***"Misspecified proportional hazard models"*** by C.A. Struthers and J.D. Kalbfleisch

Table 1. *Values of β^* when Z_2 is omitted from the model*

α_1	α_2	β^*	$\exp(\alpha_1)$	$\exp(\beta^*)$	α_1	α_2	β^*	$\exp(\alpha_1)$	$\exp(\beta^*)$
1·0	0·5	0·95	2·72	2·59	2·0	0·5	1·92	7·39	6·82
	1·0	0·85		2·34		1·0	1·73		5·64
	2·0	0·68		1·97		2·0	1·34		3·83
	3·0	0·59		1·80		3·0	1·10		3·00
1·5	0·5	1·43	4·48	4·18					
	1·0	1·28		3·60					
	2·0	1·01		2·75					
	3·0	0·86		2·36					

- True model: $\lambda(t) = \lambda_0(t)\exp(\alpha_1 z_1 + \alpha_2 z_2)$
- Analysis model: $\lambda(t) = \lambda_0(t)\exp(\beta^* z_1)$
  - The covariate z_2 is omitted

α_1 and β^* are targeting on different estimand!

* Rephrased from Struthers, C.A. and Kalbfleisch, J.D., 1986. Misspecified proportional hazard models. *Biometrika*, *73*(2), pp.363-369.

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

Reference

- 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
- Lin, W., 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*, *7*(1), pp.295-318.
- 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).
- Rauch, G., Brannath, W., Brückner, M. and Kieser, M., 2018. The Average Hazard Ratio–A Good Effect Measure for Time-to-event Endpoints when the Proportional Hazard Assumption is Violated?. *Methods of information in medicine*, *57*(03), pp.089-100.
- 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
- Struthers, C.A. and Kalbfleisch, J.D., 1986. Misspecified proportional hazard models. *Biometrika*, *73*(2), pp.363-369.
- Lin, D.Y. and Wei, L.J., 1989. The robust inference for the Cox proportional hazards model. *Journal of the American statistical Association*, *84*(408), pp.1074-1078.
- Bugni, F.A., Canay, I.A. and Shaikh, A.M., 2018. Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, *113*(524), pp.1784-1796.

DIA 中国统计社区

?

Panel Discussion – Question 1

1. **随机**临床试验中的协变量调整通常仅限于用于随机化的分层因素：
   - **如何**纳入其他预先指定的协变量进行分析（例如，二元终点/logistic**回**归）？
   - **如何正确的分析？**
   - 监管机构是否接受？

Panel Discussion – Question 2

2. 当前版本的FDA协变量调整指导原则讨论了无条件或边际治疗效应的协变量调整估计值。与二元终点相比，使用Cox回归分析至事件时间终点的边际风险比存在一些解释上的挑战，因为比例风险假设无法同时在条件模型和边际模型中成立，它最多只适用于其中一个。在这种情况下：
   - 假设条件比例风险模型时，如何正确解释边际风险比？
   - 大家对替代汇总指标（例如 milestone probability、RMST）有何看法?

Panel Discussion – Question 3

3. 边际和条件治疗效应之间的差异如何影响临床医生的解释？如何解释两者的差异？在实践中有什么想法或建议吗？