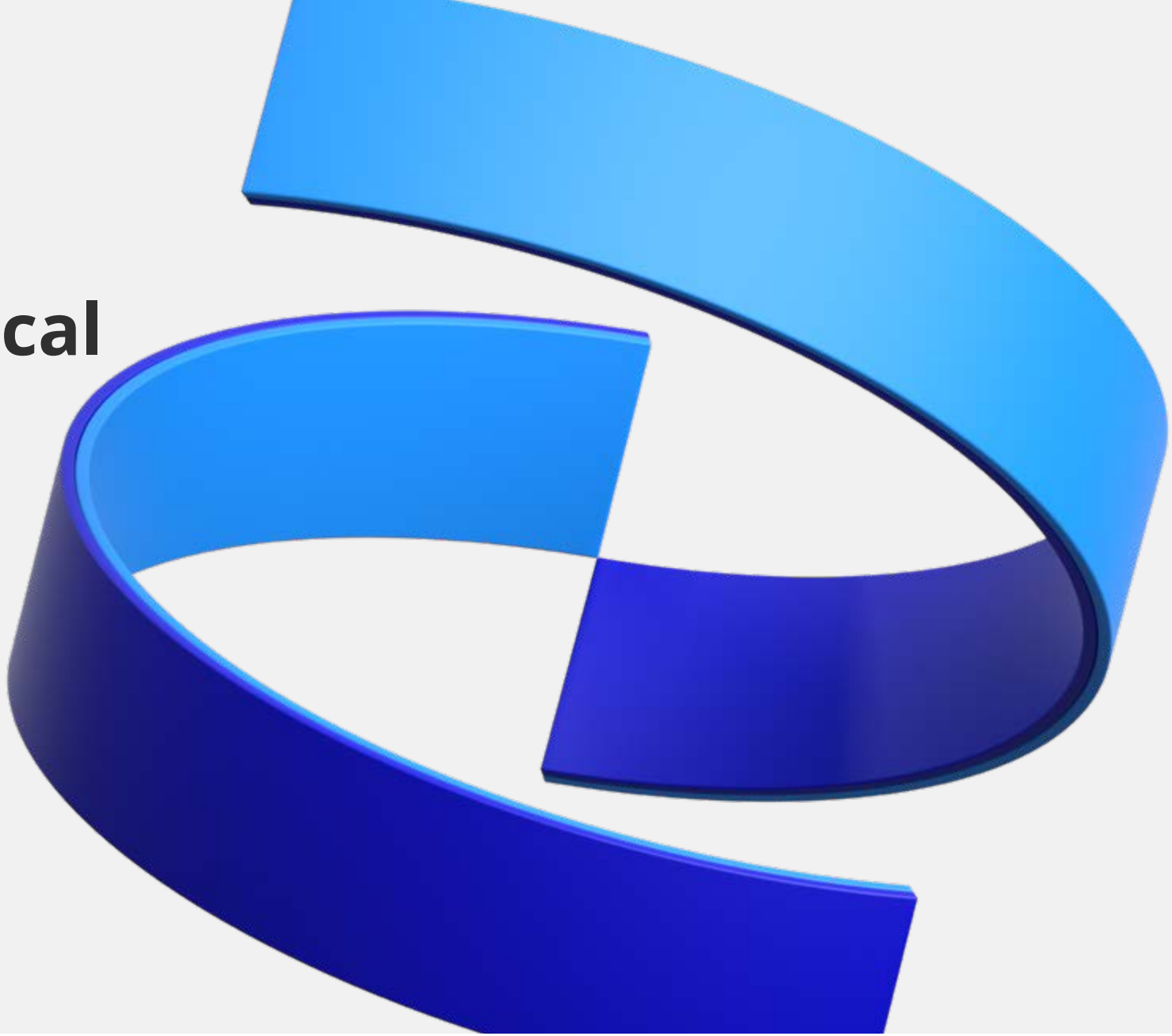
Discussion: Issues in Survival Analysis in Pharmaceutical Clinical Trials



Satrajit Roychoudhury
Statistical Research and Data Science
Pfizer Inc.

ASA Lifetime Data Science Conference
June 1st, 2023

Disclaimer

The views and opinions expressed herein are solely those of the presenter and are not necessarily those of Pfizer Inc. Any of these cannot and should not necessarily be construed to represent those of Pfizer Inc. or its affiliates.

Survival Analysis in Drug and Vaccine Development [Not an exhaustive List]

Confirmatory analysis

Oncology

Cardiovascular

Vaccine

Safety analysis

Supplementary/supportive analysis

Infectious disease

Hematology: ALS, Sickle cell disease

Neuroscience

Diabetes

Analysis Methodology

Based on the literature approximately 95% of the analysis includes

- Log-rank test
- Cox-PH
- Kaplan-Meier

Other variations include

- Composite strategy: Win ratio, joint rank analysis
- Competing-risk analysis and other multi-state models: used primary for exploratory setting
- Interval censored methods

Model based approaches beyond Cox-PH are still rare

Recent Discussions

- Violation of proportional hazard assumption
- Lack of causal interpretation of hazard ratio
- Appropriate censoring mechanism
- Handling “cured” population in analysis
- Consideration of multiple time to event outcome
- Treatment switching
- Patient focused summary measures: “Doctor, what are the chances I will do better on this new drug compared to no treatment?”

- Going over the 3 presentations
- Question and discussion

Talk 1: Efficiency of recurrent and time-to-first event methods in the presence of terminal events – Application to chronic heart failure trials

- Patrick Schlömer

Summary

- Discusses analysis strategy for handling co-primary endpoints in cardiovascular trial
- Substantial power again by using recurrent event methods in most cases compare to the traditional “time to first analysis” depending on drug discontinuation
- HHF+CVD seems to be more regulatory compliant due to strong control of type-I error
  - Shows power gain over the traditional analysis
  - Win-ratio as a summary measure in clinically interpretable
- Question: How to recommend appropriate analysis at the design stage?

Talk 2: Logic respecting efficacy measures in the presence of prognostic or predictive biomarker subgroups

-Liwei Wang

Summary

- This presentation emphasizes the issue with HR estimated by Cox PH as primary summary measure
- For general clarity and to avoid misunderstandings, associational concepts of dependence should clearly and formally be distinguished from causal measures of efficacy
- Authors proposed the use of logic respecting efficacy measure such as ratio of median as treatment effect summary
  - Additional structural assumption to ensure logical estimation
  - Proposed use of Subgroup Mixable Estimation (SME) based on δ-method to correctly analyze clinical trial results with prognostic subgroups
  - Used parametric model for the baseline

Few Thoughts…

- The non-collapsibility of HR and OR are mentioned in the recent FDA covariate adjustment guideline
  - However, the guidance is minimal for time to event data
- Difference for milestone time seems appealing to clinicians
  - Median may not be reached
- SME largely depends on delta-method
  - FDA rather recommended non-parametric bootstrap for SE calculation
  - This is problematic as common nonparametric bootstrap methods redraw the baseline covariates and thus do not estimate the correct standard error conditional
- Non-collapsibility vs confounding

Marginal Odds Ratio Differs from Conditional Odds Ratio

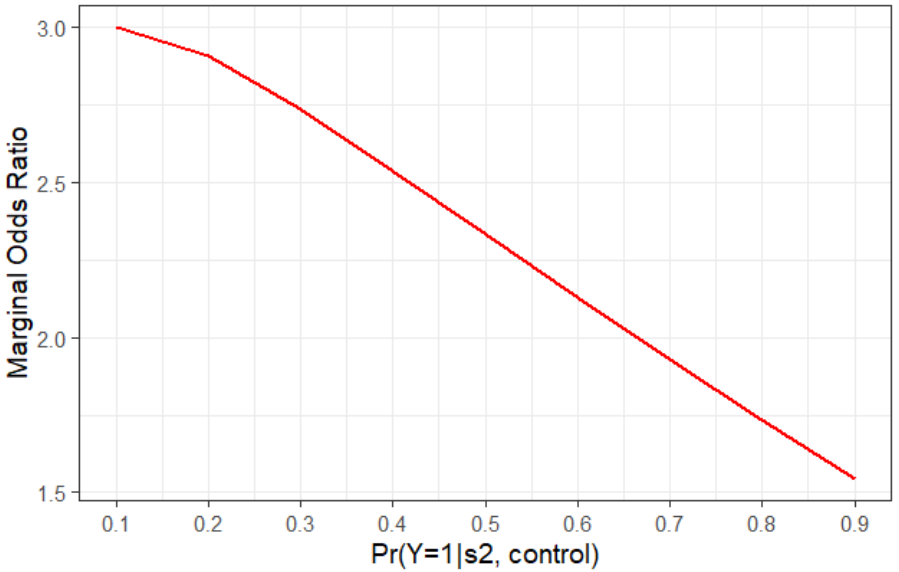
	Percentage of target population	Success rate		Odds ratio
		New drug	Placebo	
Males	50%	80.0%	33.3%	8.0
Females	50%	25.0%	4.0%	8.0
Combined	100%	52.5%	18.7%	4.8

- 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. (2023), Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products Guidance for Industry

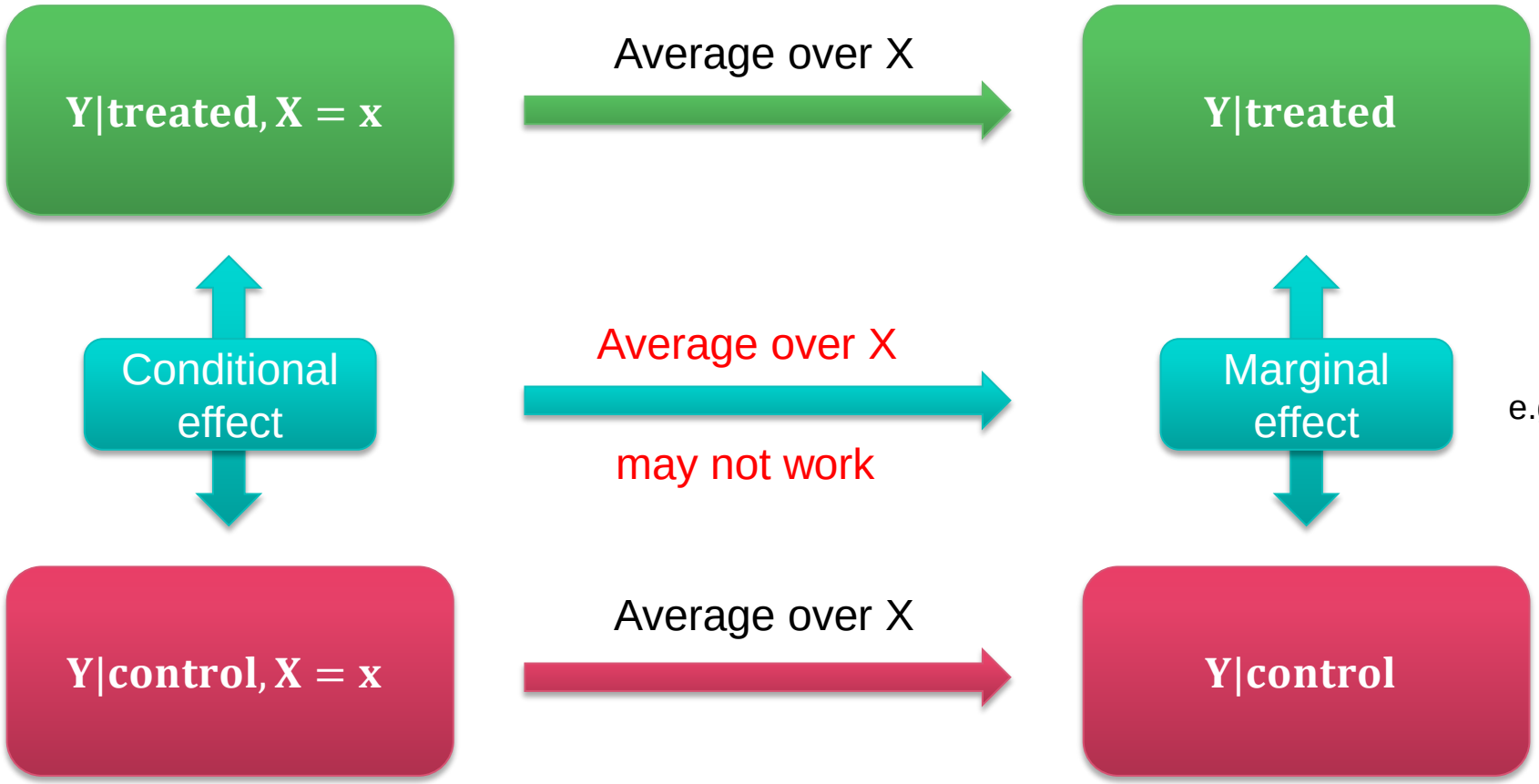
Marginal Odds Ratio Numerically Moves Away from the Conditional Odds Ratio As the Prognostic Effect Deepens

- 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.

Challenges on Time-to-event Outcome

- $\lambda(t) = \lambda_0(t)\exp(\theta z) \rightarrow$ unadjusted model → marginal estimand
- $\lambda(t) = \lambda_0(t)\exp(\theta z + \beta x) \rightarrow$ 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.

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

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 linea
the covariate-adaptive randomization method has bee
Often, practitioners will simply adopt a conventional
controversial since tests derived under simple rando
l 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 and
under covariate-adaptive randomization. We propose
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.

Talk 3: Balancing events, not patients, maximizes power in randomized survival studies

- Godwin Yung

Summary

- Looking into alternative ways to optimize the trial design
- Rubinstein's equation allows us to quickly and accurately estimate design parameters (e.g., power, trial duration, accrual rate)
- The proposal for unequal randomization is appealing as more patient gets the new treatment
  - Not sure always "optimal": against SOC, combination drug
  - Additional follow-up are often important for secondary endpoints or when PH assumption is not violated
- Theoretical query: It seems the Rubinstein's equation relies on exponential distribution. How much sensitive the results when such assumption violates (e.g., cure fraction)

Thank You