Covariate Adjustment in RCTs: **Practical Implementation and Outlook**

Marlena Bannick, PhD

Incoming Assistant Professor

Fred Hutchinson Cancer Center

EFSPI / PSI Webinar on Covariate Adjustment

30 June 2026

Acknowledgement: ASA Biopharmaceutical Section Covariate Adjustment Working Group Software Subteam

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

The RobinCar Family

Supported by the ASA Biopharmaceutical Section Covariate Adjustment Working Group

Statistical Method for Covariate Adjustment

Widely accepted and standard methods are directly implemented.

RobinCar

A "sandbox" for analyzing clinical trials using covariate adjustment.

- *Published in peer-reviewed journals with theoretical and empirical justification*
- *In-demand by clinical trials practitioners*

RobinCar2

Curated subset of methods with streamlined interface and GxP-ready package.

The RobinCar Family

Link to arXiv preprint with all of the methods here (+ additional ones)

	RobinCar	RobinCar2
Standardization using canonical generalized linear models	✓	✓
Standardization using SuperLearner (includes augmentation term for bias correction and cross-fitting)	✓	
Covariate-adjusted log-rank test and marginal hazard ratio (+ option for stratifying the test)	✓	✓

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Using Covariates in Design and Analysis

- Trials commonly use **covariates at the design stage** during randomization:
  - E.g., stratified permuted block randomization*, stratified biased coin, Pocock and Simon's minimization.
- This breaks the "independent and identically distributed" assumption that is commonly employed in the theoretical analysis of estimators.

  Bannick et al. (2025) *Biometrika;* Ye et al. (2024) *JASA:*

AIPW (or G-computation) estimators are still consistent for marginal estimand.

Variance estimators that are valid under simple randomization may not be valid under other randomization schemes.

* Stratified permuted block randomization is *not* the same thing as stratified randomization. Stratified randomization has the same asymptotic properties as simple randomization.

Using Covariates in Design and Analysis

- Ignoring the randomization scheme often leads to conservative inference.
- An important exception is using a canonical generalized linear interaction model with the randomization variables included as covariates.
- The asymptotic variance of the G-computation estimator based on the working model below is **invariant to the randomization scheme**.

$$E(Y|A=a,Z,X) = g^{-1}(\nu_a + \gamma_a Z + \beta_a X)$$

Canonical link function

Treatment assignment

Discrete covariates used in randomization

Other covariates

RobinCar provides variance estimators under several randomization schemes if you choose not to use this working model.

Using Covariates in Design and Analysis

- When doing a **log-rank test**, any variables used as the design stage need to either be accounted for via stratification (stratified log-rank test), or through covariate adjustment (covariate-adjusted log-rank test).
- The choice of approach should be driven by the *null hypothesis (and estimand) of interest*.

Unconditional Null
(corresponds to marginal HR)

$$H_0 : \lambda_1(t) = \lambda_0(t)$$

- Covariate-adjusted log-rank test
  (same null as log-rank test)

Conditional Null
(for all levels of Z)

$$H_0 : \lambda_1(t|Z) = \lambda_0(t|Z)$$

- Stratified log-rank test
- Covariate-adjusted stratified log-rank test

Ye et al. (2024) in *Biometrika*

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Dealing with Imperfect Data: *Missing Data*

- RobinCar and RobinCar2 require users to provide a **complete** dataset.
- **Missing covariate data** can be imputed and still lead to efficiency gains!*
- **Missing outcome data** is more complicated, but it needs to be dealt with regardless of whether covariate adjustment is used.

** See: Zhao and Ding (2022) in JASA; or Song, Hughes, and Ye (2024) in Biometrics.*

Dealing with Imperfect Data: *Finite Samples*

- Asymptotic variance is one thing – constructing a well-behaved variance estimator in finite samples is another!
- **Influence function-based variance estimators** require plug-in estimates for quantities relating to the working model $f_a(x) = E(Y|A = a, X = x)$
  - Different variance options for situations where
    - the outcome is zero-inflated, or
    - distribution of covariates differs substantially by treatment group
- **Sandwich-based variance estimators** have correction terms that may perform well in finite samples – but this is only a valid approach for ANCOVA (linear model, no interactions, linear contrast).

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Approaches for High-Dimensional Data

Approach 1: Use a "simple" model with lots of covariates.

Data on Trial Participants

Generalized Linear Model

$\widehat{f}_0(x), \widehat{f}_1(x)$

Predicted outcomes under control and experimental

Marginalize over covariate distribution in the trial to estimate treatment effect

$$\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{f}_1(X_i) - \widehat{f}_0(X_i)\right]$$

Approaches for High-Dimensional Data

Approach 1: Use a “simple” model with lots of covariates.

Approach 2: Use an ML or AI model with lots of covariates.

Data on Trial Participants

Marginalize over covariate distribution in the trial to estimate treatment effect

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{f}_1(X_i) - \hat{f}_0(X_i)\right]$$

+ augmentation term for consistency

$$\hat{f}_0(x), \hat{f}_1(x)$$

Predicted outcomes under control and experimental

Approaches for High-Dimensional Data

Approach 1: Use a "simple" model with lots of covariates.

Approach 2: Use an ML or AI model with lots of covariates.

Schuler et al. (2021) *Int. J Biostat.*

Approach 3: Use an ANCOVA model with one pre-trained covariate ("super covariate").

Data on External Controls

Data on Trial Participants

Conditional mean outcome under control, based on external data

$\widehat{f}_0^{\mathrm{EXT}}(x)$ → $\widehat{f}_0^{\mathrm{EXT}}(X_i)$ → ANCOVA with this as a *single* covariate

All three of these methods are supported in RobinCar (you must provide the super covariate).

Approach 2 *requires cross-fitting to make sure you're not overfitting to the trial data.*

Approach 3 *only captures the covariate association with outcomes under control conditions (not experimental treatment).*

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Outline

Practical implementation of covariate adjustment methodology

The RobinCar Family

Using Covariates in Design and Analysis

Dealing with Imperfect data

Approaches for High-Dimensional Data

Time-to-Event Estimands

Time-to-Event Estimands

- With time-to-event data, there could be several estimands of interest:
  - Comparing two full survival curves visually.
  - A comparison of some summary metric of the curves: E.g., difference in restricted mean survival time, difference in survival probability at a landmark time.
  - Marginal hazard ratio.
- Shao, Zhang, and Ye (2026) in *SBR*: estimate treatment propensity using logistic model, and inverse-weight the event and at-risk processes in the typical Kaplan-Meier estimator form.

$$\widehat{S}_a(t) = \prod_{t_j \le t} \left(1 - \frac{d_a(t_j)}{r_a(t_j)}\right)$$

Kaplan-Meier Estimator

$$\widehat{S}^w_a(t) = \prod_{t_j \le t} \left(1 - \frac{d^w_a(t_j)}{r^w_a(t_j)}\right)$$

Weighted Kaplan-Meier Estimator

Thank you for listening!

RobinCar Collaborators

Yanyao Yi	*Eli Lilly*
Yuhan Qian	*UW Biostatistics*
Daniel Sabanes Bove	*Novartis / RCONIS*
Faith Bian	*Eli Lilly*
Dong Xi	*Gilead*
Gregory Chen	*Merck*
Liming Li	*Roche / AstraZeneca*