Improving sandwich variance estimation for marginal Cox analysis of cluster randomized trials

Xueqi Wang
Duke University School of Medicine

Joint work with:
Elizabeth L. Turner, Duke University
Fan Li, Yale University

Partially funded by NIH awards:
U01OD033247 & R01DC020026

Cluster Trials "Special Interest Group" Meeting
March 16, 2022

Duke Department of Biostatistics & Bioinformatics
Duke University School of Medicine

Duke GLOBAL HEALTH INSTITUTE

Introduction

Cluster randomized trials (CRTs)

- Unit of randomization: a cluster of individuals
- Commonly used in public health, education, and social policy

Introduction

Cluster randomized trials (CRTs)

- Unit of randomization: a cluster of individuals
- Commonly used in public health, education, and social policy

Challenge in CRTs: a limited number of clusters

- Logistical or resource constraints
- Systematic reviews
  - Fiero et al. (2016): from 86 CRTs published between $08/2013$ and $07/2014$, median number of clusters randomized was 24
  - Ivers et al. (2011): from 285 CRTs published between 2000 and 2008, median number of clusters randomized was 21
- Tend to inflate type I error rates for < 30 clusters (Murray et al., 2008)

Introduction

Generalized estimating equations (GEE) by Liang and Zeger (1986)

- Account for within-cluster correlations
- Population-averaged interpretation (Preisser et al., 2003)
- Robust sandwich variance estimator (ROB)
  - asymptotically valid inference
  - even when the correlation structure is not correctly specified

Introduction

Generalized estimating equations (GEE) by Liang and Zeger (1986)

- Account for within-cluster correlations
- Population-averaged interpretation (Preisser et al., 2003)
- Robust sandwich variance estimator (ROB)
  - asymptotically valid inference
  - even when the correlation structure is not correctly specified
- Limitation: ROB has negative finite-sample biases for < 30 clusters
- Bias-corrected sandwich variance estimators
  - Kauermann and Carroll (2001) (abbreviated as KC)
  - Mancl and DeRouen (2001) (abbreviated as MD)
  - Fay and Graubard (2001) (abbreviated as FG)
  - Morel et al. (2003) (abbreviated as MBN)

Introduction

Bias-corrected sandwich variance estimators

- Application to small CRTs
- Literature on comparing their finite-sample performances
  - in maintaining valid type I error rates
  - with a continuous, binary or count outcome

Introduction

Bias-corrected sandwich variance estimators

- Application to small CRTs
- Literature on comparing their finite-sample performances
  - in maintaining valid type I error rates
  - with a continuous, binary or count outcome
- Limited evaluations for censored time-to-event outcomes
  - Caille et al. (2021): from 186 CRTs from 2013 to 2018, time-to-event outcomes are not uncommon but appropriate statistical methods are infrequently used
  - Fay and Graubard (2001): the only study with a simulation evaluation

Introduction

Marginal Cox proportional hazards model (Wei et al., 1989; Lin, 1994)

- Clustered right-censored time-to-event data
- Hazard ratio as effect measure
- Assume an independence working correlation structure
- Robust sandwich variance estimator (Spiekerman and Lin, 1998)

Introduction

Marginal Cox proportional hazards model (Wei et al., 1989; Lin, 1994)

- Clustered right-censored time-to-event data
- Hazard ratio as effect measure
- Assume an independence working correlation structure
- Robust sandwich variance estimator (Spiekerman and Lin, 1998)

Our work

- Propose 9 bias-corrected sandwich variance estimators
  - for CRTs with time-to-event outcomes
  - under the marginal Cox model
- Provide practical recommendations
- Develop an R package **CoxBcv** accessible on CRAN

Section

Marginal Cox Proportional Hazards Model

Statistical model for a parallel-arm CRT

- n: number of clusters
- m_i: cluster size for cluster i ($i = 1, \ldots, n$)
- $\mathcal{O}_{ij} = (X_{ij}, \Delta_{ij}, \boldsymbol{Z}_{ij})$: observed data triplet
  - $X_{ij} = \min\{T_{ij}, C_{ij}\}$: observed event time, where
    T_{ij}: underlying failure time for the event of interest
    C_{ij}: censoring time
  - Δ_{ij}: event indicator; $\Delta_{ij} = 1$ if $X_{ij} = T_{ij}$ and $\Delta_{ij} = 0$ if $X_{ij} = C_{ij}$
  - $\boldsymbol{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijp})'$: a $p \times 1$ vector of baseline covariates

Marginal Cox Proportional Hazards Model

Marginal Cox proportional hazards model

$$\lambda_{ij}(t|\boldsymbol{Z}_{ij}) = \lambda_0(t) \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_{ij}\right) \tag{1}$$

- $\lambda_0(t)$: an unspecified baseline hazard function
- $\boldsymbol{\beta}$: a $p \times 1$ vector of regression parameters
- Estimate the population-averaged intervention effect
- Usually include only a cluster-level intervention indicator
  - $\boldsymbol{Z}_{ij}$: a scalar binary covariate
  - $\boldsymbol{\beta}$: the population-averaged hazard ratio

Marginal Cox Proportional Hazards Model

Estimate β in model (1) (Wei et al., 1989; Lin, 1994)

- Based on an independence working correlation structure
- An unbiased estimator $\widehat{\beta}$ solves the estimating equations

$$\boldsymbol{U}(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \Delta_{ij} \left\{ \boldsymbol{Z}_{ij} - \frac{\boldsymbol{S}^{(1)}(\beta; X_{ij})}{S^{(0)}(\beta; X_{ij})} \right\} = 0$$

  where

  - $\boldsymbol{S}^{(r)}(\beta; t) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} Y_{ij}(t) \exp\left(\beta' \boldsymbol{Z}_{ij}\right) \boldsymbol{Z}_{ij}^{\otimes r}$ for $r = 0, 1, 2$
  - $\boldsymbol{c}^{\otimes 0} = 1, \boldsymbol{c}^{\otimes 1} = \boldsymbol{c}, \boldsymbol{c}^{\otimes 2} = \boldsymbol{c}\boldsymbol{c}'$ for an arbitrary vector $\boldsymbol{c}$
  - $Y_{ij}(t) = I(X_{ij} \geq t)$: at-risk process

Marginal Cox Proportional Hazards Model

- $N_{ij}(t) = I(X_{ij} \leq t, \Delta_{ij} = 1)$: counting process for the failure time
- Breslow-type estimators
  - cumulative baseline hazard

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^t \frac{dN_{ij}(u)}{\sum_{k=1}^{n} \sum_{l=1}^{m_k} Y_{kl}(u) \exp\left(\boldsymbol{\beta}' \boldsymbol{Z}_{kl}\right)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^t \frac{dN_{ij}(u)}{S^{(0)}(\boldsymbol{\beta}; u)}$$

  - baseline hazard

$$\widehat{\lambda}_0(t)dt = \sum_{i=1}^{n} \sum_{j=1}^{m_i} S^{(0)}(\boldsymbol{\beta}; t)^{-1} dN_{ij}(t)$$

Marginal Cox Proportional Hazards Model

Robust sandwich variance estimator for $\widehat{\beta}$

- Extension from GEE with non-censored outcomes to the marginal Cox model (Wei et al., 1989; Spiekerman and Lin, 1998)
- Define the mean-zero martingale-score for each cluster

$$\boldsymbol{U}_i(\beta) = \sum_{j=1}^{m_i} \boldsymbol{U}_{ij}(\beta) = \sum_{j=1}^{m_i} \int_0^\infty \left\{ \boldsymbol{Z}_{ij} - \frac{\boldsymbol{S}^{(1)}(\beta; u)}{S^{(0)}(\beta; u)} \right\} dM_{ij}(u) \qquad (2)$$

where

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u)\lambda_0(u) \exp(\beta' \boldsymbol{Z}_{ij}) du$$

is the martingale

Marginal Cox Proportional Hazards Model

- Define $\boldsymbol{\Omega}_i(\boldsymbol{\beta}) = -\partial \boldsymbol{U}_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$
- Sandwich variance estimator

$$\widehat{\boldsymbol{V}}_s = \widehat{\boldsymbol{V}}_m \left(\sum_{i=1}^{n} \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \right) \widehat{\boldsymbol{V}}_m$$

where

- model-based variance estimator $\widehat{\boldsymbol{V}}_m = \left(\sum_{i=1}^{n} \widehat{\boldsymbol{\Omega}}_i\right)^{-1} =$ $\left(\sum_{i=1}^{n} \sum_{j=1}^{m_i} \int_0^{\infty} \left\{ \frac{\boldsymbol{S}^{(2)}(\widehat{\boldsymbol{\beta}};u)}{S^{(0)}(\widehat{\boldsymbol{\beta}};u)} - \frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}};u)\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}};u)'}{S^{(0)}(\widehat{\boldsymbol{\beta}};u)^2} \right\} dN_{ij}(u)\right)^{-1}$
- $\widehat{\boldsymbol{\Omega}}_i = \boldsymbol{\Omega}_i(\widehat{\boldsymbol{\beta}})$
- $\widehat{\boldsymbol{U}}_i = \boldsymbol{U}_i(\widehat{\boldsymbol{\beta}})$

Marginal Cox Proportional Hazards Model

Features of $\widehat{\boldsymbol{V}}_s$

- Unbiased in large samples regardless of the correct specification of the working independent correlation assumption
- Tend to underestimate the variance in small CRTs ($n < 30$)
  - inflated type I error rates
  - under-coverage
- Need small-sample bias corrections

Section

Bias Correction based on MR

Bias correction based on modification of the martingale residual (MR)

- Rewrite the martingale in Eq (2)

$$M_{ij}(t;\beta) = \widehat{M}_{ij}(t;\widehat{\beta}) - \left\{\widehat{M}_{ij}(t;\widehat{\beta}) - \widehat{M}_{ij}(t;\beta)\right\} - \left\{\widehat{M}_{ij}(t;\beta) - M_{ij}(t;\beta)\right\} \quad (3)$$

  - $\widehat{M}(t,\beta)$: baseline hazard estimated by the Breslow-type estimator
  - $\widehat{M}(t,\widehat{\beta})$: baseline hazard estimated by the Breslow-type estimator and β is estimated by $\widehat{\beta}$

Bias Correction based on MR

- Consider a first-order Taylor Series expansion to rewrite Eq (3)

$$M_{ij}(t;\beta) = \widehat{M}_{ij}(t;\widehat{\beta}) + \widehat{\boldsymbol{D}}'_{ij}(t;\beta)\widehat{\boldsymbol{V}}_m \sum_{k=1}^{n}\sum_{l=1}^{m_k} \boldsymbol{U}_{kl}(\beta) + \int_0^t Y_{ij}(u)\exp\left(\beta'\boldsymbol{Z}_{ij}\right)\frac{dM(u)}{S^{(0)}(\beta;u)}$$

where

- $M(t) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} M_{ij}(t)$
- $\boldsymbol{D}_{ij}$ is a gradient matrix

- Recall

$$\boldsymbol{U}_i(\beta) = \sum_{j=1}^{m_i} \boldsymbol{U}_{ij}(\beta) = \sum_{j=1}^{m_i} \int_0^\infty \left\{ \boldsymbol{Z}_{ij} - \frac{\boldsymbol{S}^{(1)}(\beta;u)}{S^{(0)}(\beta;u)} \right\} dM_{ij}(u)$$

Bias Correction based on MR

- Bias-corrected version of the estimated martingale-score $\widehat{\boldsymbol{U}}_i$

$$\widehat{\boldsymbol{U}}_i^{BC} = \left\{ \boldsymbol{I}_p + \sum_{j=1}^{m_i} \int_0^\infty \left\{ \boldsymbol{Z}_{ij} - \frac{\boldsymbol{S}^{(1)}(\widehat{\beta}; u)}{S^{(0)}(\widehat{\beta}; u)} \right\} d\widehat{\boldsymbol{D}}_{ij}'(u; \widehat{\beta}) \widehat{\boldsymbol{V}}_m \right\} \widehat{\boldsymbol{U}}_i$$
$$+ \sum_{j=1}^{m_i} \int_0^\infty \left\{ \boldsymbol{Z}_{ij} - \frac{\boldsymbol{S}^{(1)}(\widehat{\beta}; u)}{S^{(0)}(\widehat{\beta}; u)} \right\} Y_{ij}(u) \exp\left(\widehat{\beta}' \boldsymbol{Z}_{ij} \right)$$
$$\times S^{(0)}(\widehat{\beta}; u)^{-1} d\widehat{M}_{i\bullet}(u)$$

- MR bias-corrected sandwich variance estimator

$$\widehat{\boldsymbol{V}}_{MR} = \widehat{\boldsymbol{V}}_m \left\{ \sum_{i=1}^{n} \widehat{\boldsymbol{U}}_i^{BC} \left(\widehat{\boldsymbol{U}}_i^{BC} \right)' \right\} \widehat{\boldsymbol{V}}_m$$

Bias Corrections Based on Methods for GEE

Generalize multiplicative bias corrections developed for GEE

- Multiplicative bias corrections, following Wang et al. (2021): $\widehat{\boldsymbol{V}}_m \widehat{\boldsymbol{V}}_0 \widehat{\boldsymbol{V}}_m$ with

$$\widehat{\boldsymbol{V}}_0 = \sum_{i=1}^{n} \boldsymbol{C}_i \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \boldsymbol{C}_i' \tag{4}$$

  - $\boldsymbol{C}_i$: cluster-specific correction matrix for small CRTs
- Determine the form of $\boldsymbol{C}_i$
- Expand the estimating equations around $\widehat{\boldsymbol{\beta}}$:

$$\boldsymbol{U}_i \approx \widehat{\boldsymbol{U}}_i - \widehat{\boldsymbol{\Omega}}_i \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}} \right)$$

- Sum across all clusters and re-arranging terms:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx \widehat{\boldsymbol{V}}_m \left(\sum_{i=1}^{n} \boldsymbol{U}_i \right)$$

Bias Corrections Based on Methods for GEE

- Approximate the covariance of the estimated cluster-specific score:

$$E\left(\widehat{\boldsymbol{U}}_i\widehat{\boldsymbol{U}}_i'\right) \approx \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i\widehat{\boldsymbol{V}}_m\right)\boldsymbol{\Psi}_i\left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i\widehat{\boldsymbol{V}}_m\right)' + \widehat{\boldsymbol{\Omega}}_i\widehat{\boldsymbol{V}}_m\left(\sum_{j\neq i}\boldsymbol{\Psi}_j\right)\widehat{\boldsymbol{V}}_m'\widehat{\boldsymbol{\Omega}}_i' \quad (5)$$

  - $\boldsymbol{\Psi}_i = \text{Cov}(\boldsymbol{U}_i) = E\left(\boldsymbol{U}_i\boldsymbol{U}_i'\right)$: true covariance of $\boldsymbol{U}_i$

Bias Corrections Based on Methods for GEE

KC bias correction

- Assume $\boldsymbol{\Psi}_i \approx c \times \widehat{\boldsymbol{\Omega}}_i$ (Kauermann and Carroll, 2001)

$$E\left(\widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i'\right) \approx \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right) \boldsymbol{\Psi}_i \approx \boldsymbol{\Psi}_i \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)'$$

- Motivate $\boldsymbol{C}_i = \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1/2}$ in Eq (4)
- KC bias-corrected sandwich variance estimator

$$\widehat{\boldsymbol{V}}_{KC} = \widehat{\boldsymbol{V}}_m \left\{ \sum_{i=1}^{n} \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1/2} \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \left(\boldsymbol{I}_p - \widehat{\boldsymbol{V}}_m \widehat{\boldsymbol{\Omega}}_i\right)^{-1/2} \right\} \widehat{\boldsymbol{V}}_m$$

Bias Corrections Based on Methods for GEE

FG bias correction

- Analogous to Fay and Graubard (2001)
- $\boldsymbol{C}_i = \text{diag}\left\{\left(1 - \min\left\{r, [\widehat{\Omega}_i \widehat{\boldsymbol{V}}_m]_{jj}\right\}\right)^{-1/2}\right\}$ in Eq (4)
  - $r < 1$: a user-defined constant; usually set $r = 0.75$
- FG bias-corrected sandwich variance estimator

$$\widehat{\boldsymbol{V}}_{FG} = \widehat{\boldsymbol{V}}_m \left[\sum_{i=1}^{n} \text{diag}\left\{\left(1 - \min\left\{r, [\widehat{\Omega}_i \widehat{\boldsymbol{V}}_m]_{jj}\right\}\right)^{-1/2}\right\} \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \times \text{diag}\left\{\left(1 - \min\left\{r, [\widehat{\Omega}_i \widehat{\boldsymbol{V}}_m]_{jj}\right\}\right)^{-1/2}\right\}\right] \widehat{\boldsymbol{V}}_m$$

- Note: when $p = 1$, $\widehat{\boldsymbol{V}}_{FG} = \widehat{\boldsymbol{V}}_{KC}$ if $\widehat{\Omega}_i \widehat{\boldsymbol{V}}_m$ does not exceed r

Bias Corrections Based on Methods for GEE

MD bias correction

- Assume the last term of (5) is negligible (Mancl and DeRouen, 2001)

$$\boldsymbol{\Psi}_i \approx \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1} \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \left(\boldsymbol{I}_p - \widehat{\boldsymbol{V}}_m \widehat{\boldsymbol{\Omega}}_i\right)^{-1}$$

- Motivate $\boldsymbol{C}_i = \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1}$ in Eq (4)
- MD bias-corrected sandwich variance estimator

$$\widehat{\boldsymbol{V}}_{MD} = \widehat{\boldsymbol{V}}_m \left\{ \sum_{i=1}^{n} \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1} \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \left(\boldsymbol{I}_p - \widehat{\boldsymbol{V}}_m \widehat{\boldsymbol{\Omega}}_i\right)^{-1} \right\} \widehat{\boldsymbol{V}}_m$$

- Note: $\widehat{\boldsymbol{V}}_{MD}$ often leads to larger variance estimates than $\widehat{\boldsymbol{V}}_{KC}$

Bias Corrections Based on Methods for GEE

MBN bias correction

- An additive bias correction, analogous to Morel et al. (2003)
- MBN bias-corrected sandwich variance estimator

$$\widehat{\boldsymbol{V}}_{MBN} = \left(\frac{\sum_{i=1}^{n} m_i - 1}{\sum_{i=1}^{n} m_i - p} \times \frac{n}{n-1} \right) \widehat{\boldsymbol{V}}_s + \min\left(0.5, \frac{p}{n-p}\right) \widehat{\phi} \widehat{\boldsymbol{V}}_m \quad (6)$$

where

$$\widehat{\phi} = \max\left\{1, \left(\frac{\sum_{i=1}^{n} m_i - 1}{\sum_{i=1}^{n} m_i - p} \times \frac{n}{n-1} \right) \times \text{trace}\left[\widehat{\boldsymbol{V}}_m \left(\sum_{i=1}^{n} \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \right) \right] / p \right\}$$

- Advantage of the additive bias correction:
  it ensures a positive-definite covariance matrix (Morel et al., 2003)

Hybridizing Bias Corrections

Hybrid bias corrections

- Survival analysis concerns incompletely observed outcomes
- Bias correction may be insufficient when implemented alone

Hybridizing Bias Corrections

Hybrid bias corrections

- Survival analysis concerns incompletely observed outcomes
- Bias correction may be insufficient when implemented alone
- Hybridize the MR bias correction with either one of the multiplicative or additive bias-corrections: replace $\widehat{\boldsymbol{U}}_i$ with $\widehat{\boldsymbol{U}}_i^{BC}$ in $\widehat{\boldsymbol{V}}_{KC}$, $\widehat{\boldsymbol{V}}_{FG}$, $\widehat{\boldsymbol{V}}_{MD}$ and $\widehat{\boldsymbol{V}}_{MBN}$
- Hybrid bias-corrected sandwich variance estimators: $\widehat{\boldsymbol{V}}_{KCMR}$, $\widehat{\boldsymbol{V}}_{FGMR}$, $\widehat{\boldsymbol{V}}_{MDMR}$ and $\widehat{\boldsymbol{V}}_{MBNMR}$

Bias-Corrected Sandwich Variance Estimators

Table 1: A brief summary of different sandwich variance estimators.

Variance Estimator	Label	Formula	Feature
$\widehat{\boldsymbol{V}}_s$	ROB	$\widehat{\boldsymbol{V}}_m \left(\sum_{i=1}^n \boldsymbol{C}_i \widehat{\boldsymbol{U}}_i \widehat{\boldsymbol{U}}_i' \boldsymbol{C}_i'\right) \widehat{\boldsymbol{V}}_m$	$\boldsymbol{C}_i = \boldsymbol{I}_p$
$\widehat{\boldsymbol{V}}_{KC}$	KC		$\boldsymbol{C}_i = \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1/2}$
$\widehat{\boldsymbol{V}}_{FG}$	FG		$\boldsymbol{C}_i = \text{diag}\left\{\left(1 - \min\left\{r, [\widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m]_{jj}\right\}\right)^{-1/2}\right\}$
$\widehat{\boldsymbol{V}}_{MD}$	MD		$\boldsymbol{C}_i = \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1}$
$\widehat{\boldsymbol{V}}_{MBN}$	MBN	$c_1 \widehat{\boldsymbol{V}}_s + c_2 \widehat{\phi} \widehat{\boldsymbol{V}}_m$	c_1, c_2, $\widehat{\phi}$ defined in Eq (6)
$\widehat{\boldsymbol{V}}_{MR}$	MR	$\widehat{\boldsymbol{V}}_m \left(\sum_{i=1}^n \boldsymbol{C}_i \widehat{\boldsymbol{U}}_i^{BC} \left[\widehat{\boldsymbol{U}}_i^{BC}\right]' \boldsymbol{C}_i'\right) \widehat{\boldsymbol{V}}_m$	$\boldsymbol{C}_i = \boldsymbol{I}_p$
$\widehat{\boldsymbol{V}}_{KCMR}$	KCMR		$\boldsymbol{C}_i = \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1/2}$
$\widehat{\boldsymbol{V}}_{FGMR}$	FGMR		$\boldsymbol{C}_i = \text{diag}\left\{\left(1 - \min\left\{r, [\widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m]_{jj}\right\}\right)^{-1/2}\right\}$
$\widehat{\boldsymbol{V}}_{MDMR}$	MDMR		$\boldsymbol{C}_i = \left(\boldsymbol{I}_p - \widehat{\boldsymbol{\Omega}}_i \widehat{\boldsymbol{V}}_m\right)^{-1}$
$\widehat{\boldsymbol{V}}_{MBNMR}$	MBNMR	$c_1 \widehat{\boldsymbol{V}}_{MR} + c_2 \widehat{\phi} \widehat{\boldsymbol{V}}_m$	c_1, c_2, $\widehat{\phi}$ defined in Eq (6) with $\widehat{\boldsymbol{U}}_i$ replaced by $\widehat{\boldsymbol{U}}_i^{BC}$

Section

A Numerical Study

Study description

- Two-arm CRT with equal allocation
- Only 1 covariate in model (1):
  a binary cluster-level intervention indicator
  ($Z_i = 1$: intervention arm and $Z_i = 0$: control arm)

A Numerical Study

Study description

- Two-arm CRT with equal allocation
- Only 1 covariate in model (1):
  a binary cluster-level intervention indicator
  ($Z_i = 1$: intervention arm and $Z_i = 0$: control arm)
- Null hypothesis of no intervention effect $H_0 : \beta = 0$
- Two-sided Wald t-test with $n - 1$ degrees of freedom

A Numerical Study

Parameter specification, following Zhong and Cook (2015)

- Weibull distribution for cumulative baseline hazard: $\Lambda_0(t;\boldsymbol{\alpha}) = \int_0^t \lambda_0(s;\boldsymbol{\alpha})ds = (\lambda_0 t)^\kappa$ and $\boldsymbol{\alpha} = (\lambda_0, \kappa)'$
- Administrative censoring time $C^\dagger = 1$
- p_a: desired administrative censoring rate for the control group
- λ_0 solves $P(T_{ij} > C^\dagger | Z_i = 0) = p_a$
    - $p_a = 0.2$

A Numerical Study

Parameter specification, following Zhong and Cook (2015)

- Weibull distribution for cumulative baseline hazard: $\Lambda_0(t;\boldsymbol{\alpha}) = \int_0^t \lambda_0(s;\boldsymbol{\alpha})ds = (\lambda_0 t)^\kappa$ and $\boldsymbol{\alpha} = (\lambda_0, \kappa)'$
- Administrative censoring time $C^\dagger = 1$
- p_a: desired administrative censoring rate for the control group
- λ_0 solves $P(T_{ij} > C^\dagger | Z_i = 0) = p_a$
  - $p_a = 0.2$
- C_{ij}^*: random censoring time for individual j in cluster i
  - exponentially distributed with rate ρ
  - independent censoring within each cluster
- $C_{ij} = \min\{C_{ij}^*, C^\dagger\}$: true right-censoring time
- p_0: desired net censoring rate in the control arm
- ρ solves $P(T_{ij} > C_{ij} | Z_i = 0) = p_0$
  - $p_0 = 0.2$ or 0.5

A Numerical Study

Generate correlated failure times, from the Clayton copula (Clayton and Cuzick, 1985)

- $S_i(t_{i1}, \ldots, t_{im_i})$: joint survival distribution for m_i ($m_i \geq 2$) correlated observations $(T_{i1}, \ldots, T_{im_i})$ in a cluster
- $F_{ih}\left(t_{ih} \mid t_{i1}, \ldots, t_{i,h-1}\right)$: conditional cumulative distribution function for $T_{i1}, \ldots, T_{ih}$ $(h = 1, \ldots, m_i)$
  - $F_{ih}\left(t_{ih} \mid t_{i1}, \ldots, t_{i,h-1}\right) \sim$ Uniform$(0, 1)$
- Can generate m_i independent Uniform$(0, 1)$ variates
  - $u_{i1} = F_{i1}(t_1)$
  - $u_{ih} = F_{ih}\left(t_{ih} \mid t_{i1}, \ldots, t_{i,h-1}\right)$ for $h = 2, \ldots, m_i$
- Solve for t_{i1} and t_{ih} $(h = 2, \ldots, m_i)$

A Numerical Study

- Kendall's $\tau \in \{0.01, 0.05, 0.1, 0.25\}$
- Set $\beta = 0$ for assessing the empirical type I error rate
- Fix the nominal type I error rate at 5%

A Numerical Study

- Kendall's $\tau \in \{0.01, 0.05, 0.1, 0.25\}$
- Set $\beta = 0$ for assessing the empirical type I error rate
- Fix the nominal type I error rate at 5%
- $n \in \{6, 10, 20, 30\}$
- Generate m_i ($i = 1, \ldots, n$) from a gamma distribution with
  - mean equal to $\overline{m} \in \{20, 50, 100\}$
  - coefficient of variation (CV) ranging from 0 to 1 by increments of 0.1
  - m_i truncated at 2

A Numerical Study

- 5000 data replications for each scenario
- Fit the marginal Cox model for each replicate
- Considered 10 variance estimators for the intervention effect:
  - uncorrected robust sandwich variance estimator ROB
  - 9 bias-corrected sandwich variance estimators:
    MR, KC, FG, MD, MBN, KCMR, FGMR, MDMR, MBNMR

A Numerical Study

Results of interest

- Percent relative bias of the variance estimators:

$$\left\{\sum_{r=1}^{5000}(\widehat{V}_q)_r/5000 - \mathrm{Var}_{MC}(\widehat{\beta})\right\}/\mathrm{Var}_{MC}(\widehat{\beta}) \times 100$$

  - q: index of the evaluated variance estimator
  - $(\widehat{V}_q)_r$: $\widehat{V}_q$ from the rth simulated data replication
  - $\mathrm{Var}_{MC}(\widehat{\beta}) = \sum_{r=1}^{5000}(\widehat{\beta} - \sum_{r=1}^{5000}\widehat{\beta}/5000)^2/4999$

A Numerical Study

Results of interest

- Percent relative bias of the variance estimators:

$$\left\{\sum_{r=1}^{5000}(\widehat{V}_q)_r/5000 - \text{Var}_{MC}(\widehat{\beta})\right\}/\text{Var}_{MC}(\widehat{\beta}) \times 100$$

  - q: index of the evaluated variance estimator
  - $(\widehat{V}_q)_r$: $\widehat{V}_q$ from the rth simulated data replication
  - $\text{Var}_{MC}(\widehat{\beta}) = \sum_{r=1}^{5000}(\widehat{\beta} - \sum_{r=1}^{5000}\widehat{\beta}/5000)^2/4999$
- Empirical type I error rate under the null
  - Acceptable range of empirical type I error rates: $(4.4\%, 5.6\%)$

A Numerical Study

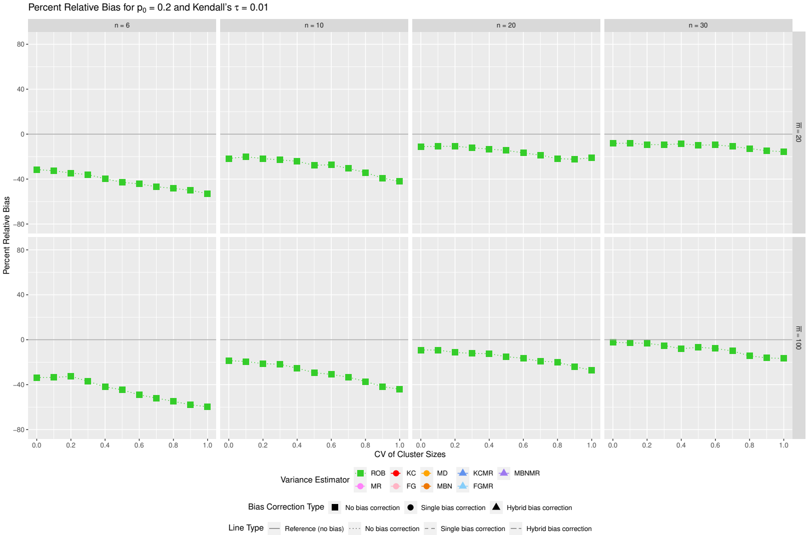

Figure 1: Percent relative biases of different variance estimators for $p_0 = 0.2$ and $\tau = 0.01$ under the marginal Cox model.

A Numerical Study

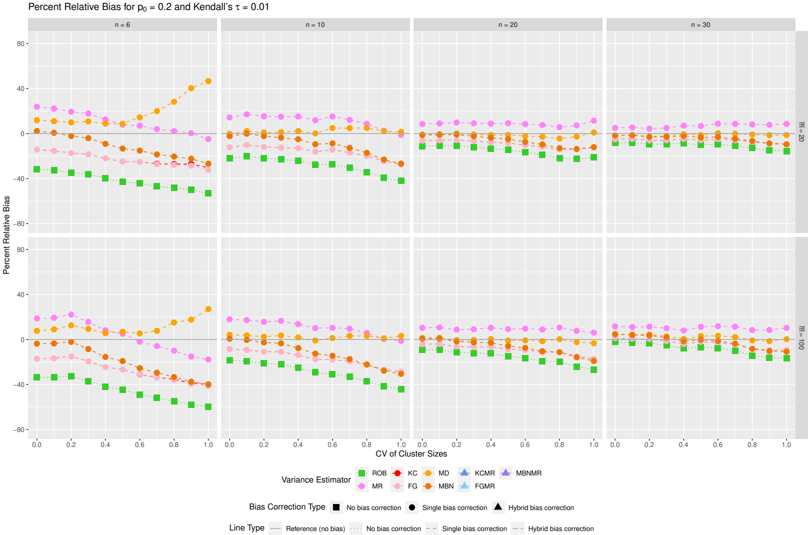

Figure 1: Percent relative biases of different variance estimators for $p_0 = 0.2$ and $\tau = 0.01$ under the marginal Cox model.

A Numerical Study

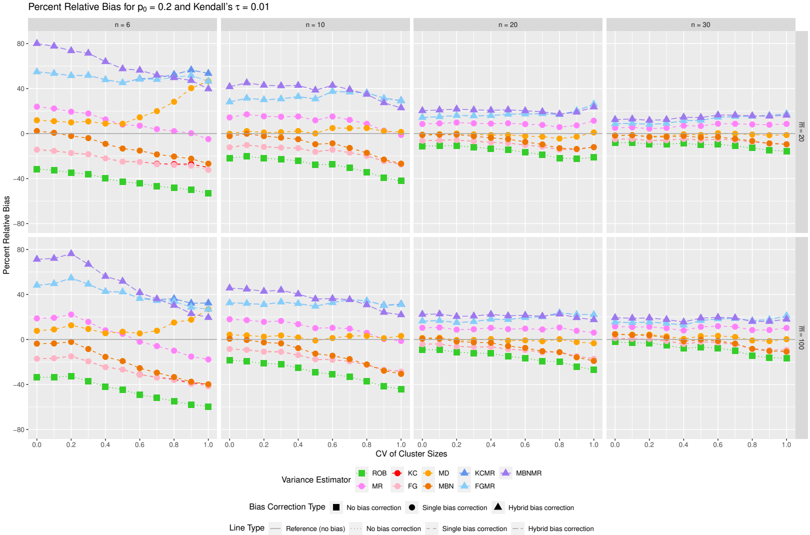

Figure 1: Percent relative biases of different variance estimators for $p_0 = 0.2$ and $\tau = 0.01$ under the marginal Cox model.

A Numerical Study

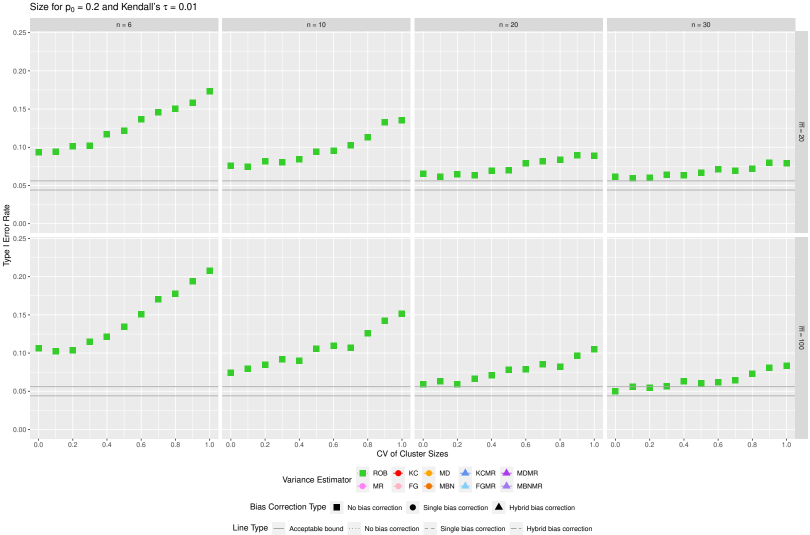

Figure 2: Empirical type I error rates of intervention effect tests for $p_0 = 0.2$ and $\tau = 0.01$ under the marginal Cox model, based on different variance estimators.

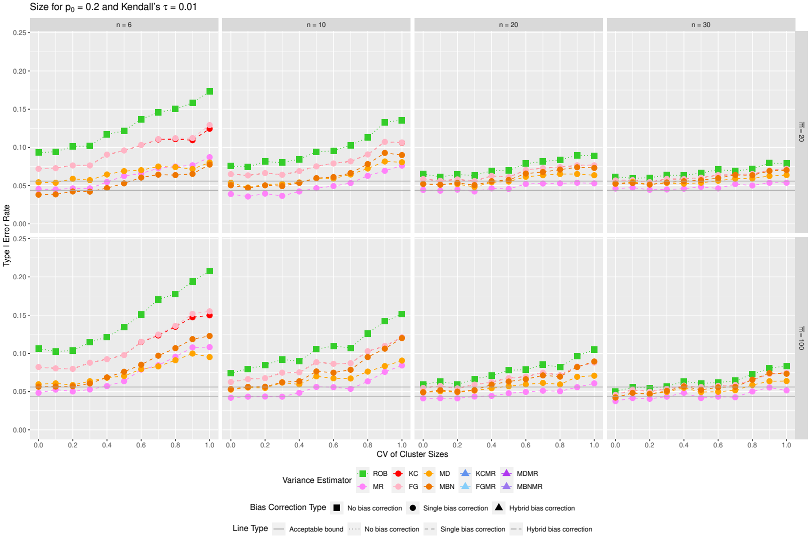

Figure 2: Empirical type I error rates of intervention effect tests for $p_0 = 0.2$ and $\tau = 0.01$ under the marginal Cox model, based on different variance estimators.

A Numerical Study

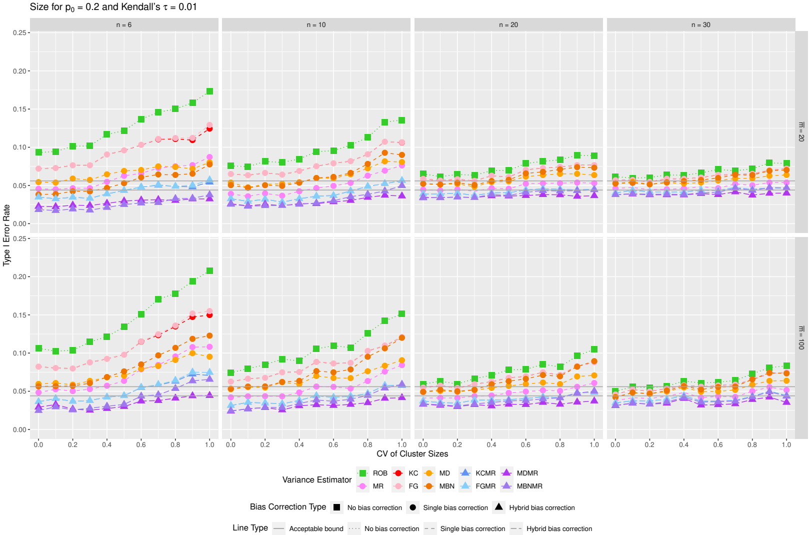

Figure 2: Empirical type I error rates of intervention effect tests for $p_0 = 0.2$ and $\tau = 0.01$ under the marginal Cox model, based on different variance estimators.

A Numerical Study

Summary

- When CV ≤ 0.4: MD performs best
- When CV ≥ 0.5: KCMR performs best, with exceptions under $n = 6$ and CV ≥ 0.8
  - seemingly discordant recommendation regarding positive bias and nominal test size
  - variability of KCMR when CV of cluster sizes increases (tendency to over reject) VS positive bias (tendency to be conservative)
- When $n = 6$ and CV ≥ 0.8: no bias corrections could lead to close to nominal test size

A Numerical Study

Practical recommendations

- Include at least 10 clusters in CRTs with time-to-event outcomes
- Use of MD bias-corrected sandwich variance estimator
  - robust to the moderate variation of cluster sizes (CV ≤ 0.4)
- Use of KCMR bias-corrected sandwich variance estimator
  - under larger variations of cluster sizes

Section

Application to the STOP-CRC CRT

STOP CRC trial (Coronado et al., 2018)

- Two-arm parallel CRT
- Compare 2 strategies: an EHR-embedded program & usual care
- Primary outcome: time to completion of colorectal cancer screening, administratively censored at 12 months

Application to the STOP-CRC CRT

STOP CRC trial (Coronado et al., 2018)

- Two-arm parallel CRT
- Compare 2 strategies: an EHR-embedded program & usual care
- Primary outcome: time to completion of colorectal cancer screening, administratively censored at 12 months
- Randomization level: health center clinic (cluster)
- Equal allocation to the two arms

Application to the STOP-CRC CRT

STOP CRC trial (Coronado et al., 2018)

- Two-arm parallel CRT
- Compare 2 strategies: an EHR-embedded program & usual care
- Primary outcome: time to completion of colorectal cancer screening, administratively censored at 12 months
- Randomization level: health center clinic (cluster)
- Equal allocation to the two arms
- Consider the subgroup of nonwhite participants
  - high CV of cluster sizes
  - application of bias-corrected sandwich variance estimator can lead to a different conclusion from the standard analysis

Application to the STOP-CRC CRT

Subgroup of nonwhite participants

- 4543 nonwhite participants in 26 clusters
  - 2513 (55.32%) females
  - mean (standard deviation [SD]) age: 58.72 (6.51) years
- Variable cluster sizes
  - range: $[8, 1054]$
  - mean (SD): 174.73 (246.95)
  - CV = 1.41
- Consider a proportional hazards model (1)
  - include only a binary cluster-level intervention indicator
  - compare ROB and bias-corrected sandwich variance estimators

Application to the STOP-CRC CRT

Table 2: Analysis results of the STOP CRC data (for the sub-population of nonwhite).

Variance Estimator	Log of Hazard Ratio[a] (95% CI[b])	Hazard Ratio (95% CI[b])	p-value
ROB	0.699 (0.151, 1.248)	2.012 (1.162, 3.483)	0.015
MR	0.699 (-0.011, 1.410)	2.012 (0.989, 4.094)	0.053
KC	0.699 (-0.018, 1.416)	2.012 (0.983, 4.120)	0.055
FG	0.699 (-0.018, 1.416)	2.012 (0.983, 4.120)	0.055
MD	0.699 (-0.274, 1.672)	2.012 (0.761, 5.323)	0.151
MBN	0.699 (0.129, 1.270)	2.012 (1.137, 3.560)	0.018
KCMR	0.699 (-0.249, 1.647)	2.012 (0.780, 5.191)	0.141
FGMR	0.699 (-0.249, 1.647)	2.012 (0.780, 5.191)	0.141
MDMR	0.699 (-0.605, 2.004)	2.012 (0.546, 7.417)	0.280
MBNMR	0.699 (-0.040, 1.438)	2.012 (0.961, 4.212)	0.063

[a] Estimate of β in model (1) with only one covariate of a binary cluster-level intervention indicator.

[b] CI: Confidence interval from the Wald t-test with $n-1$ degrees of freedom.

Section

Summary

Summary

- Propose 9 bias-corrected sandwich variance estimators for CRTs with time-to-event data analyzed through the marginal Cox model
- Conduct a comprehensive simulation study to evaluate proposed estimators
- Suggest that the Wald t-test with a bias-corrected sandwich variance estimator can maintain the nominal test size and generate reliable inferences for as few as 10 clusters
  - choice of bias-corrected sandwich variance estimators should take the variation of cluster sizes into account
- Develop an R package **CoxBcv** to implement proposed estimators

Summary

Future work

- Extend to non-independent working correlation structure
- Develop bias corrections for the correlation estimating equations, similar to the matrix-adjusted estimating equations proposed by Preisser et al. (2008)
- Generalize our recommendations under the covariate-adjusted analysis through the marginal Cox model in CRTs
- Extend to recurrent event data and multistate models

References I

Caille, A., Tavernier, E., Taljaard, M., and Desmée, S. (2021). Methodological review showed that time-to-event outcomes are often inadequately handled in cluster randomized trials. *Journal of Clinical Epidemiology* .

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society: Series A (General)* **148,** 82–108.

Coronado, G. D., Petrik, A. F., Vollmer, W. M., Taplin, S. H., Keast, E. M., Fields, S., and Green, B. B. (2018). Effectiveness of a mailed colorectal cancer screening outreach program in community health clinics: the stop crc cluster randomized clinical trial. *JAMA Internal Medicine* **178,** 1174–1181.

Fay, M. P. and Graubard, B. I. (2001). Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics* **57,** 1198–1206.

References II

Fiero, M. H., Huang, S., Oren, E., and Bell, M. L. (2016). Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. *Trials* **17,** 1–10.

Ivers, N., Taljaard, M., Dixon, S., Bennett, C., McRae, A., Taleban, J., Skea, Z., Brehaut, J., Boruch, R., Eccles, M., et al. (2011). Impact of consort extension for cluster randomised trials on quality of reporting and study methodology: review of random sample of 300 trials, 2000-8. *Bmj* **343,**.

Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* **96,** 1387–1396.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lin, D. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statistics in Medicine* **13,** 2233–2247.

References III

Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for gee with improved small-sample properties. *Biometrics* **57**, 126–134.

Morel, J. G., Bokossa, M., and Neerchal, N. K. (2003). Small sample correction for the variance of gee estimators. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* **45**, 395–409.

Murray, D. M., Pals, S. L., Blitstein, J. L., Alfano, C. M., and Lehman, J. (2008). Design and analysis of group-randomized trials in cancer: a review of current practices. *Journal of the National Cancer Institute* **100**, 483–491.

Preisser, J. S., Lu, B., and Qaqish, B. F. (2008). Finite sample adjustments in estimating equations and covariance estimators for intracluster correlations. *Statistics in Medicine* **27**, 5764–5785.

References IV

Preisser, J. S., Young, M. L., Zaccaro, D. J., and Wolfson, M. (2003). An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. *Statistics in Medicine* **22,** 1235–1254.

Spiekerman, C. F. and Lin, D. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93,** 1164–1175.

Wang, X., Turner, E. L., Preisser, J. S., and Li, F. (2021). Power considerations for generalized estimating equations analyses of four-level cluster randomized trials. *Biometrical Journal* .

Wei, L.-J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

References V

Zhong, Y. and Cook, R. J. (2015). Sample size and robust marginal methods for cluster-randomized trials with censored event times. *Statistics in Medicine* **34,** 901–923.

Thank You
Any questions?

Manuscript under review