City Research Online

City, University of London Institutional Repository

**Citation:** Wang, W., Tong, G., Hirani, S. P., Newman, S. P., Halpern, S. D., Small, D. S., Li, F. & Harhay, M. O. (2023). A mixed model approach to estimate the survivor average causal effect in cluster-randomized trials. Staistics in Medicine, 43(1), pp. 16-33. doi: 10.1002/sim.9939

This is the published version of the paper.

This version of the publication may differ from the final published version.

**Permanent repository link:** https://openaccess.city.ac.uk/id/eprint/31803/

**Link to published version:** https://doi.org/10.1002/sim.9939

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

Check for updates

Received: 8 March 2022 | Revised: 5 September 2023 | Accepted: 12 October 2023

DOI: 10.1002/sim.9939

RESEARCH ARTICLE

Statistics in Medicine WILEY

# A mixed model approach to estimate the survivor average causal effect in cluster-randomized trials

**Wei Wang**[1] | **Guangyu Tong**[2,3,4] | **Shashivadan P. Hirani**[5] | **Stanton P. Newman**[5,6] | **Scott D. Halpern**[1,7] | **Dylan S. Small**[8] | **Fan Li**[3,4] | **Michael O. Harhay**[1,7]

[1]Clinical Trials Methods and Outcomes Lab, Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[2]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA

[3]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

[4]Center for Methods in Implementation and Prevention Science, Yale School of Public Health, New Haven, CT, USA

[5]School of Health Sciences, City University London, London, UK

[6]Division of Medicine, University College London, London, UK

[7]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[8]Department of Statistics and Data Science, The Wharton School, University of Pennsylvania, Philadelphia, PA, USA

**Correspondence**
Wei Wang, Clinical Trials Methods and Outcomes Lab, Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine, University of Pennsylvania, PA, USA.
Fan Li and Michael O. Harhay are co-senior authors.
Email: wwa@upenn.edu

**Funding information**
National Heart, Lung, and Blood Institute, Grant/Award Number: R01-HL168202; Patient-Centered Outcomes Research Institute, Grant/Award Numbers: ME-2020C1-19220, ME-2020C3-21072

In many medical studies, the outcome measure (such as quality of life, QOL) for some study participants becomes informatively truncated (censored, missing, or unobserved) due to death or other forms of dropout, creating a nonignorable missing data problem. In such cases, the use of a composite outcome or imputation methods that fill in unmeasurable QOL values for those who died rely on strong and untestable assumptions and may be conceptually unappealing to certain stakeholders when estimating a treatment effect. The survivor average causal effect (SACE) is an alternative causal estimand that surmounts some of these issues. While principal stratification has been applied to estimate the SACE in individually randomized trials, methods for estimating the SACE in cluster-randomized trials are currently limited. To address this gap, we develop a mixed model approach along with an expectation–maximization algorithm to estimate the SACE in cluster-randomized trials. We model the continuous outcome measure with a random intercept to account for intracluster correlations due to cluster-level randomization, and model the principal strata membership both with and without a random intercept. In simulations, we compare the performance of our approaches with an existing fixed-effects approach to illustrate the importance of accounting for clustering in cluster-randomized trials. The methodology is then illustrated using a cluster-randomized trial of telecare and assistive technology on health-related QOL in the elderly.

**KEYWORDS**
causal inference, cluster-randomized trials, linear mixed models, expectation–maximization, potential outcomes, principal stratification

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2023 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The informative "truncation" or "censoring" by death problem is a challenge in many studies. When death occurs, a non-mortality outcome cannot be measured or defined for those who die before the time of planned study measurement. Further, as the reason for the inability to measure an important study outcome is known and itself meaningful (ie, non-ignorable or missing not at random), this truncation is informative and can bias inference with statistical models that assume missing completely at random (MCAR) or missing at random (MAR).[1-4] A common example of this problem in the medical literature is missing QOL measures due to death. Researchers have a limited number of options in such settings.[3,4] Composite outcomes (eg, applying the worst possible QOL value to those who die) are a popular solution, but can be difficult to interpret, and similar summary values of a composite outcome in different studies may represent very different combinations of clinical outcomes.[5,6] Imputation strategies are another potential solution, but they also raise conceptual challenges as any imputed value for an informatively unmeasured outcome results in an explicit or implicit valuation of that health state to equal a value observed among those who did not die, which may not be appealing or logical to certain stakeholders. Other potential approaches, such as a survivors-only or complete case analysis, produce estimates that do not have a clear causal interpretation under the counterfactual outcome framework, that is, those who survive under treatment and control can be systematically different, and this obscures the target estimand.[7,8] In this article, we focus on cluster-randomized trials (CRTs) where randomization occurs at the cluster level, and, thus, participant observations are correlated, and we investigate the survivor average causal effect (SACE), defined using principal stratification, as a potential solution for the truncation-by-death problem.[1] Specifically for CRTs, we propose to model the non-mortality outcome using a finite mixture regression model with a random intercept, where the mixture weights are obtained through either fixed-effects (FE) or mixed-effects (ME) regression.

Under the potential outcomes framework, Frangakis and Rubin[9] first considered the principal stratification method to estimate the principal causal effects with independent and identically distributed data. The principal stratification approach has been used to address a wide range of similar problems, including truncation-by-death, noncompliance (eg, using the related complier average causal effect (CACE)[10]), and surrogate outcomes.[11] This framework (detailed more fully in Section 2) allows researchers to consider cohort members as belonging to one of four possible strata (Table 1). Under truncation-by-death, two strata include those who would or would not survive regardless of treatment: these are the *always survivors* and *never survivors* strata. The other two strata include those who would only survive under treatment or control, classified as the *protected* or the *harmed* (alternatively framed as the *compliant* or the *defiant*, respectively in the noncompliance literature). Because the pair of counterfactual non-mortality outcomes are only well-defined for those who would survive under both treatment and control, the SACE is defined as the effect of an intervention on the non-mortality outcome among the always survivors. In other words, the SACE as a causal estimand applies only to a specific subset of individuals rather than the original trial population.

Principal stratification and the SACE estimand have been used in a number of settings. For example, Zhang and Rubin used the SACE to examine the impact of an educational program on test scores when the scores were not possible to collect among those who dropped out of school before taking the test.[1] The SACE approach has also been used to gain insight into the effect of intensive care unit-based interventions for critically ill patients.[12-15] Bounds of the SACE providing the smallest and the largest possible values without modeling assumptions have also been derived.[1,16,17] In practice, however, the bounds may be too wide to be informative for deciding whether the effect is positive or negative,[18] and there is substantial interest in many medical studies to obtain a point estimate of the SACE. Zhang et al developed a likelihood-based method for SACE point estimation using the expectation–maximization (EM) algorithm,[19] and applied their method to estimate the effect of a job training program on wages in a setting where wages are only meaningfully defined for the employed individuals.[20] Since then, other methods for estimating the SACE have been developed and include the use of a sensitivity analysis procedure,[21] leveraging a pre-treatment covariate,[18] weighting and regression

**TABLE 1** The four principal strata of trial participants using principal stratification.

| | | |
|---|---|---|
| ss | Always survivors | $\{S_{ij}(1) = 1, S_{ij}(0) = 1\}$ |
| sn | Protected | $\{S_{ij}(1) = 1, S_{ij}(0) = 0\}$ |
| ns | Harmed | $\{S_{ij}(1) = 0, S_{ij}(0) = 1\}$ |
| nn | Never survivors | $\{S_{ij}(1) = 0, S_{ij}(0) = 0\}$ |

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

under a nonparametric structural equations model,[22] leveraging a substitution variable in the always survivors stratum,[23] and methods based on the principal scores.[24,25]

The aforementioned approaches focus on randomized trials or observational studies where treatment assignment is at the individual level and where data observations are typically independent (uncorrelated). However, many multicenter studies where the SACE would be useful must deal with clustering by study sites, and methods for estimating principal-strata-specific effects, such as the SACE, in such settings are currently limited with a few exceptions for estimating the average effect of treatment assignment for compliers (ie, CACE). Specifically, principal stratification has been applied in clustered encouragement designs where the treatment encouragement is assigned at the cluster level to address treatment non-compliance.[26,27] In CRTs, groups of individuals, instead of individuals, are randomized to different arms. Due to shared features among individuals within clusters, it is likely that the outcomes in the same cluster are more similar than those from different clusters.[28,29] The resulting similarity (or homogeneity) of patients (and thus their outcomes) within the same cluster then induces a positive intracluster correlation that must be accounted for in the analysis of the trial to ensure valid statistical inference. In CRTs with one-sided noncompliance, a multilevel mixture model has been developed to estimate the CACE.[30,31] Under one-sided noncompliance, individuals assigned to treatment could be a never-taker if they do not take the treatment, but individuals have no access to treatment if assigned to the control condition. This results in compliers and never-takers as two principal strata, and facilitates likelihood-based inference for the CACE based on multilevel models for both the outcome and compliance strata. Under one-sided noncompliance, the model fitting procedure based on the EM algorithm[19] and numerical integration has been implemented through the statistical modeling program Mplus.

In this article, we develop a mixed model estimators to support estimation of the SACE under a principal stratification framework adapted for CRTs. To explicitly account for possible intracluster correlations when estimating the SACE, we develop a clustered mixture model by introducing cluster-level random intercepts into the stratum-specific potential outcome models (Section 3.2). We adopt the multinomial logistic regression with and without random intercepts to model the membership probability in each principal stratum[20] and differentiate the latent strata in the trial population. In the outcome model, conditional on the random intercept, each individual's observed outcome in a treated cluster can come from the always survivor stratum or the protected stratum, and thus follows a mixture distribution.[20,32] Similarly, in a control cluster, a non-survivor whose outcome is not observed has a mixture representation (never survivors or the protected). Our formulation differs from methods developed for estimating the CACE in CRTs[30,31] in that we have considered three latent strata under monotonicity (instead of 2 strata under one-sided noncompliance) based on survival status and that the non-mortality outcomes of interest are not observable in certain strata. Our formulation is similar to Tong et al.,[33] who developed a Bayesian mixture model to estimate the SACE in CRTs. In contrast to Tong et al.,[33] we consider the frequentist paradigm and develop an EM approach for estimating the SACE in CRTs. Further comparison between the frequentist and the Bayesian approach in this context, however, will be left for future work.

The rest of this article is organized as follows. We define notation in Section 2, and introduce our models and the estimation procedures in Sections 3–5. In Section 6, we conduct a simulation study to compare our approaches of estimating the SACE with the approach of no random intercept and identify scenarios where it is necessary to account for clustering. We illustrate the proposed method to estimate SACE in a CRT evaluating a telehealth program in Section 7. Section 8 concludes.

## 2 | NOTATION AND SETUP

Here we review the potential outcomes framework with principal stratification to guide our work in the context of a parallel-arm CRT. We use the subscript $i$ to indicate the $i$th cluster, randomized to either treatment or control, and the subscript $j$ to indicate the $j$th individual in that cluster. Thereafter, we outline the assumptions used to support the estimation of the SACE in CRTs.

### 2.1 | Potential outcomes

In a CRT, individuals within the same cluster share the same treatment status. Let $D_i \in \{0, 1\}$ be the binary treatment assignment for the $i$th cluster, with $D_i = 1$ indicating treatment and $D_i = 0$ indicating control. Let $S_{ij}(D_i)$ be the potential survival status that would be observed under the treatment assignment $D_i$. Let $Y_{ij}(D_i)$ be the potential outcome that would

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

be observed under treatment condition $D_i$ at the end of the study. The outcome $Y_{ij}(D_i)$ is observed only for a survivor, that is, for whom $S_{ij}(D_i) = 1$ and therefore whose outcome measure is not truncated due to death or other forms of dropout.

## 2.2 | Principal stratification

Each individual has two potential survivor statuses under treatment and control. Based on them, under the principal stratification framework,[9] each individual can be classified into one of the four principal strata, defined by the joint potential values of the survival status (Table 1). Specifically, we let "s" stand for "survivor" and "n" stand for "non-survivor." We use "ss" to represent the always-survivors stratum (ie, $\{S_{ij}(1) = 1, S_{ij}(0) = 1\}$) where an individual would survive regardless of the treatment assignment. We let "sn" indicate the protected stratum (ie, $\{S_{ij}(1) = 1, S_{ij}(0) = 0\}$), where an individual would survive under treatment but would not survive under control. We use "ns" to represent the harmed stratum (ie, $\{S_{ij}(1) = 0, S_{ij}(0) = 1\}$), where individuals would not survive under treatment but would survive under control. Finally, we let "nn" indicate the never-survivors stratum (ie, $\{S_{ij}(1) = 0, S_{ij}(0) = 0\}$), where individuals would not survive in either treatment or control.

In practice, we cannot observe both $S_{ij}(1)$ and $S_{ij}(0)$ for any individual, that is, we observe only one of $S_{ij}(1)$ and $S_{ij}(0)$ depending on the assigned treatment status. However, we are able to track the principal stratum to some degree based on the observed treatment assignment and the observed survival status. In a treatment cluster with $D_i = 1$, a survivor is in either the "ss" stratum or the "sn" stratum, whereas a non-survivor is in either the "ns" or the "nn" stratum. In a control cluster with $D_i = 0$, a survivor is in the "ss" or the "ns" stratum while a non-survivor is in the "sn" or the "nn" stratum. In general, the potential outcomes $Y_{ij}(1)$ and $Y_{ij}(0)$ can only be well-defined if the patient survives until the time when the non-mortality outcome is measured. Thus, a common estimand of interest, the SACE, in the presence of truncation, can be defined in a CRT as a participant-level average of potential outcome contrasts among the always-survivors, or, equivalently,

$$\text{SACE} = \text{E}[Y_{ij}(1) - Y_{ij}(0) | S_{ij}(1) = S_{ij}(0) = 1]. \tag{1}$$

## 2.3 | Assumptions

We outline our key assumptions below.

> **Assumption 1.** (non-informative cluster size and cluster randomization) (a) Assume each cluster is drawn from a super-population of clusters, and the bounded cluster size $m_i$ is drawn from a common distribution of integers, $\mathcal{F}_m$. (b) For each individual in cluster $i$, the vector $\{S_{ij}(1), S_{ij}(0), S_{ij}(1)Y_{ij}(1), S_{ij}(0)Y_{ij}(0)\}$ follows a common marginal distribution $\mathcal{F}_{\{S(1),S(0),S(1)Y(1),S(0)Y(0)\}}$ that does not depend on the cluster size $m_i$. (c) The treatment is randomized at the cluster level and does not depend on any post-randomization potential outcomes, that is, $D_i \perp \{S_{ij}(1), S_{ij}(0), S_{ij}(1)Y_{ij}(1), S_{ij}(0)Y_{ij}(0); j = 1, \ldots, m_i\}$.

Assumptions 1(a) and 1(b) describe a super-population framework under which we are allowed to define the expectation of the SACE estimand (1). Specifically, Assumption 1(b) rules out informative cluster size[34] such that the distribution of potential survival statuses and potential outcome variables are independent of the cluster size.[35] It is Assumption 1(b) that permits us to define an individual-level causal estimand similar to the one typically defined for individually randomized trials. Assumption 1(c) holds as a result of the cluster-randomized design. While we assume away informative cluster size in our development, we will return to a discussion on this issue in Section 8. Of note, Assumption 1 is an adaption of an assumption introduced in Li et al.[36] for addressing post-randomization selection bias to accommodate the context of truncation-by-death in CRTs.

> **Assumption 2.** (Cluter-level SUTVA). Let $\mathbf{D}$ be the vector collecting the treatment assignment across all clusters. The cluster-level stable unit treatment value assumption (SUTVA) states that: (i) an individual's potential outcomes and survival status do not vary with treatment assigned to other clusters than the individual's own cluster; and (ii) for each cluster, there are no different versions of each treatment level. Mathematically, we have (i) $Y_{ij}(\mathbf{D}) \equiv Y_{ij}(D_i)$ and $S_{ij}(\mathbf{D}) \equiv S_{ij}(D_i)$; and (ii) If $D_i = D_i^*$, then $Y_{ij}(D_i) = Y_{ij}(D_i^*)$ and $S_{ij}(D_i) = S_{ij}(D_i^*)$.

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 2** Three key principal strata and a tabulation of the possibly observed non-mortality outcomes in each stratum.

| Principal stratum | | | $D_i = 1, S_{ij} = 1$ | $D_i = 0, S_{ij} = 1$ |
|---|---|---|---|---|
| ss | Always survivors | $\{S_{ij}(1) = 1, S_{ij}(0) = 1\}$ | $Y_{ij}(1)$ | $Y_{ij}(0)$ |
| sn | Protected | $\{S_{ij}(1) = 1, S_{ij}(0) = 0\}$ | $Y_{ij}(1)$ | * |
| nn | Never survivors | $\{S_{ij}(1) = 0, S_{ij}(0) = 0\}$ | * | * |

*Note*: Individuals in the "sn" or the "nn" stratum are alternatively framed as compliers or never-takers in the context of nontreatment compliance.

The cluster-level SUTVA does not exclude the possibility that $Y_{ij}$ can be influenced by the treatment received by other individuals in the same cluster.[27,37] However, because individuals in the same cluster receive the same treatment due to cluster randomization, the number of potential outcomes for each individual remains two, obviating the need to consider treatment interference within clusters. Therefore, the cluster-level SUTVA primarily assumes away interference between clusters. Finally, we assume monotonicity such that the cluster-level treatment does not lead to worse survival for each individual in the same cluster.

**Assumption 3.** (Monotonicity) $S_{ij}(1) \geq S_{ij}(0)$.

Under the monotonicity assumption, there are no harmed patients who survive under the control arm but would die otherwise. This is plausible in many CRTs with system-level interventions, which are likely beneficial, or at least not harmful, for patients. Each individual is therefore classified into one of the three principal strata, "ss", "sn", or "nn", prior to the treatment assignment. However, the "ss" stratum is the only group for which both $Y_{ij}(1)$ and $Y_{ij}(0)$ are well-defined due to the absence of death truncation. Inference regarding the causal effect of treatment on the outcome will then be drawn only for the always-survivors where the SACE is well-defined in equation (1). As a more concrete illustration, Table 2 shows the relationship between each principal stratum and the observed survival status; when the outcome is not defined for non-survivors, we use an asterisk, *.

Under Assumptions 1–3 alone, the SACE estimand (1) can only be partially or set identified. For example, one could follow the arguments in Zhang and Rubin[1] to develop large-sample bounds to interval identify the SACE estimand in a CRT. For point estimation, our approach follows Zhang et al.,[20] Jo et al.,[30,31] and Jo[38] and leverages multilevel parametric mixture models for principal stratification analyses. Under Assumptions 1–3, the multilevel parametric modeling assumptions are used to empirically identify the SACE estimand in CRTs. We assume the continuous outcomes follow a normal mixture model. Our model formulation does not assume principal ignorability,[24,25] because the design vector of measured covariates includes an intercept, and therefore, our models encode stratum-specific intercepts and coefficients. We formulate two sets of mixture models, one developed in Section 3 assuming a random intercept only in the potential outcome model, and the other developed in Section 5 assuming independent random intercepts in the principal strata model as well as the potential outcome model. Mathematical details of the models and the associated computation strategies are developed in Sections 3 and 5. We also discuss the caveats of this empirical identification strategy in Section 8.

## 3 | ESTIMATING THE SACE IN CLUSTER-RANDOMIZED TRIALS

In the following, we use subscripts "ss," "sn," or "nn" to denote the principal stratum-specific parameters or variables. Let elements of the random vector $\mathbf{z}_{ij} = (Z_{ss,ij}, Z_{sn,ij}, \text{ and } Z_{nn,ij})'$ be the multinomial indicator of the corresponding principal stratum. For example, $\mathbf{z}_{ij} = (1, 0, 0)'$ means the $(i, j)$th individual belongs to the "ss" stratum. Let $\mathbf{x}_{ij}$ be a column vector of measured covariates (including the intercept) of that individual.

### 3.1 | Principal strata membership model

Let $p_{ss,ij} = \mathrm{E}(Z_{ss,ij}|\mathbf{x}_{ij})$ be the probability that the the $(i, j)$th individual belongs to the "ss" stratum, and similarly we define $p_{sn,ij}$ and $p_{nn,ij}$. We consider a multinomial logistic membership model such that the reference "nn" stratum membership model is

$$p_{nn,ij} = \frac{1}{1 + \exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss}\} + \exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn}\}}, \tag{2}$$

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where $\boldsymbol{\alpha}_{ss}$ and $\boldsymbol{\alpha}_{sn}$ are the regression coefficient vectors. Under the multinomial logistic regression formulation, we further have $p_{ss,ij} = p_{nn,ij}\exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss}\}$, $p_{sn,ij} = p_{nn,ij}\exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn}\}$. For simplicity, we are not assuming the cluster-level random intercept in the principal strata model; an extension to allow for a cluster-level random intercept and hence the intra-cluster correlation (ICC) of the principal strata membership is developed in Section 5.

## 3.2 | Potential outcome models within principal strata

Let $\mathcal{N}$ denote a normal density function. In the $i$th treated cluster and under the treatment condition, we model the potential outcomes in the always-survivors stratum ("ss") with the random intercept $u_{1i}$ to account for the cluster-level random variation as

$$y_{ss,ij}(1) = \mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1} + u_{1i} + \epsilon_{ij},\ u_{1i} \sim \mathcal{N}(0,\tau^2),\ \epsilon_{ij} \sim \mathcal{N}(0,\sigma^2), \tag{3}$$

and the potential outcome in the protected stratum ("sn") as

$$y_{sn,ij}(1) = \mathbf{x}'_{ij}\boldsymbol{\beta}_{sn} + u_{1i} + \epsilon_{ij},\ u_{1i} \sim \mathcal{N}(0,\tau^2),\ \epsilon_{ij} \sim \mathcal{N}(0,\sigma^2). \tag{4}$$

We assume the observed outcome of a survivor $y_{ij}$ follows a mixture distribution with the probability $p_{ss,ij}$ from the model (3), and with the probability $p_{sn,ij}$ from the model (4). That is, conditional on $u_{1i}$, the density of a survivor's outcome in the $i$th treated cluster has the mixture model representation

$$f(y_{ij}|\mathbf{x}_{ij},u_{1i}) = \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1} + u_{1i},\sigma^2)p_{ss,ij} + \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{sn} + u_{1i},\sigma^2)p_{sn,ij}. \tag{5}$$

In the $i$th control cluster where $D_i = 0$, we model the potential outcome in the "ss" stratum with the random intercept $u_{0i}$ as follows:

$$y_{ss,ij}(0) = \mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,0} + u_{0i} + \epsilon_{ij},\ u_{0i} \sim \mathcal{N}(0,\tau^2),\ \epsilon_{ij} \sim \mathcal{N}(0,\sigma^2). \tag{6}$$

As noted above, by the monotonicity assumption, the outcome is not observed for a non-survivor who has the probability $p_{sn,ij}$ of belonging to the "sn" stratum, and the probability $p_{nn,ij}$ of belonging to the "nn" stratum. For identifiability purposes, we maintain the standard assumption that the random intercepts $u_{1i}$'s, and $u_{0i}$'s and the residual errors $\epsilon_{ij}$'s are all mutually independent for all $i$ and $j$. For simplicity, we assume the outcome models share the same variance parameters $\tau^2$ and $\sigma^2$. Under the above parameterization, the ICC among the always-survivors is given by $\tau^2/(\tau^2+\sigma^2)$. Previous simulations indicate that the restrictions imposed on variance parameters within the mixture model can influence the principal causal effect estimates.[38] When variance heterogeneity is a concern, our model can be extended to allow for stratum-specific or condition-specific variance parameters for the potential outcomes. We do not pursue this more complicated model formulation but discuss this extension as future work in Section 8.

## 3.3 | Likelihood function of the complete data

Table 3 shows the possible components of an individual's likelihood function. As a recap, the observed outcome density in the treatment group follows a mixture model as in (5). The observed data consist of the observed outcomes and the survival status of all the individuals. The missing data include the principal stratum indicator variables $\mathbf{z}_{ij}$'s and the random intercepts $u_{1i}$'s and $u_{0i}$'s. To arrive at the likelihood function, we now write $f$, which represents a generic density function. Suppose there are $m_{1,i}$ individuals in the $i$th treated cluster, and $m_{0,i}$ individuals in the $i$th control cluster. That is, $m_i = D_i m_{1,i} + (1-D_i)m_{0,i}$. Let $\mathbf{y}_i = (y_{i1}\ \cdots\ y_{i\ m_i})'$ be the observed outcome vector in each cluster. The likelihood of the complete data $(\mathbf{z}_{ij},\mathbf{y}_i,u_{1i})$, assuming the random intercepts are observed, in the $i$th treated cluster is

$$\prod_{j=1}^{m_{1,i}}\left[p_{ss,ij}^{Z_{ss,ij}}f_{ss,1}(y_{ij}|\mathbf{x}_{ij},u_{1i})^{Z_{ss,ij}}\right]\left[p_{sn,ij}^{Z_{sn,ij}}f_{sn}(y_{ij}|\mathbf{x}_{ij},u_{1i})^{Z_{sn,ij}}\right]\left[p_{nn,ij}^{Z_{nn,ij}}\right]f(u_{1i}),$$

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 3** Contribution of each individual's likelihood function given the survival status and the treatment assignment, and conditional on the random intercept.

| Observation | Principal stratum | | Treatment ($D_i = 1$) | Control ($D_i = 0$) |
|---|---|---|---|---|
| Survivor | ss | always-survivors | $p_{ss,ij}\mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1} + u_{1i}, \sigma^2)$ | $p_{ss,ij}\mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,0} + u_{0i}, \sigma^2)$ |
| | sn | protected | $p_{sn,ij}\mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{sn} + u_{1i}, \sigma^2)$ | * |
| Non-survivor | sn | protected | * | $p_{sn,ij}$ |
| | nn | never-survivors | $p_{nn,ij}$ | $p_{nn,ij}$ |

where $f_{ss,1}(y_{ij}|\mathbf{x}_{ij}, u_{1i}) = \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1} + u_{1i}, \sigma^2)$ and $f_{sn}(y_{ij}|\mathbf{x}_{ij}, u_{1i}) = \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{sn} + u_{1i}, \sigma^2)$ are from (3) and (4). Similarly, let $f_{ss,0}(y_{ij}|\mathbf{x}_{ij}, u_{0i}) = \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,0} + u_{0i}, \sigma^2)$ from (6), then in the $i$th control cluster, the likelihood of the complete data, according to Table 3, is

$$\prod_{j=1}^{m_{0,i}}\left[p_{ss,i}^{Z_{ss,ij}} f_{ss,0}(y_{ij}|\mathbf{x}_{ij}, u_{0i})^{Z_{ss,ij}}\right]\left[p_{sn,ij}^{Z_{sn,ij}}\right]\left[p_{nn,ij}^{Z_{nn,ij}}\right] f(u_{0i}).$$

## 4 | ESTIMATION VIA EXPECTATION–MAXIMIZATION

The parameters in our models include $\{\boldsymbol{\beta}_{ss,1}, \boldsymbol{\beta}_{sn}, \boldsymbol{\beta}_{ss,0}, \boldsymbol{\alpha}_{ss}, \boldsymbol{\alpha}_{sn}, \sigma^2, \tau^2\}$. To estimate these parameters, we developed an EM algorithm.[19] The details of the iterative estimation procedure are provided in the supplementary material (Section S1), and we only outline the main steps below. We obtain the complete data log-likelihood based on the calculation in Section 3.3. In the E-step, we calculate the conditional expectation of the complete data log-likelihood given the observed data and the estimates from the previous iteration. The calculation relies on the conditional mean of $Z_{ss,ij}$ and $Z_{sn,ij}$, and the conditional mean and variance of $u_{1i}$ and $u_{0i}$. As assumed, each of $Z_{ss,ij}$ and $Z_{sn,ij}$ is conditionally independent of $u_{i1}$ or $u_{i0}$.

In the M-step, we find estimates of $\boldsymbol{\beta}_{ss,1}$ and $\boldsymbol{\beta}_{sn}$ by maximizing the conditional complete data log-likelihood of the treated clusters, and similarly estimate $\boldsymbol{\beta}_{ss,0}$ based on the control clusters. We estimate the remaining parameters using data from both treated clusters and control clusters. After obtaining the parameter estimates, we also derive an estimator of the SACE and propose to use the cluster-bootstrap method to construct the confidence intervals.

### 4.1 | E-step

In the $i$th treated cluster, we can compute

$$\eta_{ss,ij} \equiv \mathrm{P}(Z_{ss,ij} = 1|S_{ij} = 1, \mathbf{x}_{ij}) = \frac{\mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1}, \sigma^2 + \tau^2)p_{ss,ij}}{\mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1}, \tau^2 + \sigma^2)p_{ss,ij} + \mathcal{N}(\mathbf{x}'_{ij}\boldsymbol{\beta}_{sn}, \tau^2 + \sigma^2)p_{sn,ij}},$$

and $\mathrm{P}(Z_{sn,ij} = 1|S_{ij} = 1, \mathbf{x}_{ij}) = 1 - \eta_{ss,ij}$. In the $i$th control cluster, we have

$$\eta_{sn,ij} \equiv \mathrm{P}(Z_{sn,ij} = 1|S_{ij} = 0, \mathbf{x}_{ij}) = \frac{p_{sn,ij}}{p_{sn,ij} + p_{nn,ij}}, \qquad \mathrm{P}(Z_{ss,ij} = 1|S_{ij} = 1, \mathbf{x}_{ij}) = 1.$$

To find the conditional mean and variance of $u_{1i}$, we perform a Monte Carlo E-step calculation.[39] Suppose there are $m_{1,i,1}$ survivors in the $i$th treated cluster. We find the joint density of $(u_{1i}, \mathbf{y}_i)$, and the marginal density of $\mathbf{y}_i$ respectively as follows:

$$f(u_{1i}, \mathbf{y}_i|\mathbf{x}_{ij}) = \left\{\prod_{j=1}^{m_{1,i,1}} f(y_{ij}|u_{1i}, \mathbf{x}_{ij})\right\} f(u_{1i}), \qquad f(\mathbf{y}_i|\mathbf{x}_{ij}) = \int f(u_{1i}, \mathbf{y}_i|\mathbf{x}_{ij})\mathrm{d}u_{1i}.$$

To reduce computational intensity, we randomly sample 100 $u_{1i}$'s from $N(0, \sigma^2)$ given the estimates from the previous iteration. Our experience based on simulations is that 100 samples provided adequate information about the underlying

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

normal distribution. Then we calculate $\mathrm{E}(u_{i1}|\mathbf{y}_i)$ and $\mathrm{Var}(u_{i1}|\mathbf{y}_i)$ numerically based on the posterior density of the random intercept, $f(u_{1i}|\mathbf{y}_i, \mathbf{x}_{ij}) = f(u_{1i}, \mathbf{y}_i|\mathbf{x}_{ij})/f(\mathbf{y}_i|\mathbf{x}_{ij})$. Specifically, under the linear mixed model setup, we obtain the conditional mean and variance of $u_{0i}$ in closed forms. Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{ij}, \ldots)$ be the design matrix of the $i$th cluster where the $j$th column is the vector of measured covariates of the $j$th individual. Let $\mathbf{1}_i$ be a column vector of 1's of appropriate order. Let $m_{0,i,1}$ be the number of survivors in the $i$th control cluster. We can arrive at

$$\mathrm{E}(u_{0i}|\mathbf{y}_i, \mathbf{X}_i) = \frac{\tau^2 \mathbf{1}_i'(\mathbf{y}_i - \mathbf{X}_i'\boldsymbol{\beta}_{ss,0})}{m_{0,i,1}\tau^2 + \sigma^2}, \qquad \mathrm{Var}(u_{0i}|\mathbf{y}_i, \mathbf{X}_i) = \frac{\tau^2\sigma^2}{m_{0,i,1}\tau^2 + \sigma^2}.$$

## 4.2 | M-step

Let $m_{1,1} = \sum_{i:D_i=1} m_{1,i,1}$ be the total number of survivors in the treatment clusters, and $m_{0,1} = \sum_{i:D_i=0} m_{0,i,1}$ be the total number of survivors in the control clusters. We obtain the estimates at the current iteration by the following the (generalized) least-squares solutions:

$$\hat{\boldsymbol{\beta}}_{ss,1} = \left\{ \sum_{i:D_i=1} \sum_{j\le m_{1,i,1}} \eta_{ss,ij}(\mathbf{x}_{ij}\mathbf{x}_{ij}') \right\}^{-1} \sum_{i:D_i=1} \sum_{j\le m_{1,i,1}} \eta_{ss,ij}\mathbf{x}_{ij}\left\{ y_{ij} - \mathrm{E}(u_{1i}|\mathbf{y}_i, \mathbf{X}_i) \right\},$$

$$\hat{\boldsymbol{\beta}}_{sn} = \left\{ \sum_{i:D_i=1} \sum_{j\le m_{1,i,1}} (1-\eta_{ss,ij})(\mathbf{x}_{ij}\mathbf{x}_{ij}') \right\}^{-1} \sum_{i:D_i=1} \sum_{j\le m_{1,i,1}} (1-\eta_{ss,ij})\mathbf{x}_{ij}\left\{ y_{ij} - \mathrm{E}(u_{1i}|\mathbf{y}_i, \mathbf{X}_i) \right\},$$

$$\hat{\boldsymbol{\beta}}_{ss,0} = \left\{ \sum_{i:D_i=0} \sum_{j\le m_{0,i,1}} (\mathbf{x}_{ij}\mathbf{x}_{ij}') \right\}^{-1} \sum_{i:D_i=0} \sum_{j\le m_{0,i,1}} \mathbf{x}_{ij}\left\{ y_{ij} - \mathrm{E}(u_{0i}|\mathbf{y}_i, \mathbf{X}_i) \right\}.$$

To update the variance components parameters, we let $y_{ss,ij,1} = y_{ij} - \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}_{ss,1} - \mathrm{E}(u_{1i}|\mathbf{y}_i, \mathbf{X}_i)$, $y_{sn,ij} = y_{ij} - \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}_{sn} - \mathrm{E}(u_{1i}|\mathbf{y}_i, \mathbf{X}_i)$, and $y_{ss,ij,0} = y_{ij} - \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}_{ss,0} - \mathrm{E}(u_{0i}|\mathbf{y}_i, \mathbf{X}_i)$, based on which we can derive

$$\hat{\sigma}^2 = \frac{\sum_{i:D_i=1}\sum_{j\le m_{1,i,1}} \left\{ \eta_{ss,ij}\, y_{ss,ij,1}^2 + (1-\eta_{ss,ij}) y_{sn,ij}^2 + \mathrm{Var}(u_{1i}|\mathbf{y}_i, \mathbf{X}_i) \right\}}{m_{1,1} + m_{0,1}} + \frac{\sum_{i:D_i=0}\sum_{j\le m_{0,i,1}} \left\{ y_{ss,ij,0}^2 + \mathrm{Var}(u_{0i}|\mathbf{y}_i, \mathbf{X}_i) \right\}}{m_{1,1} + m_{0,1}},$$

$$\hat{\tau}^2 = \frac{\sum_{i:D_i=1} \mathrm{E}(u_{1i}^2|\mathbf{y}_i, \mathbf{x}_{ij}) + \sum_{i:D_i=0} \mathrm{E}(u_{0i}^2|\mathbf{y}_i, \mathbf{X}_i)}{n_1 + n_0},$$

where $n_1$ is the number of treated clusters and $n_0$ is the number of control clusters.

Finally, we estimate the coefficient vectors $\boldsymbol{\alpha}_{ss}$ and $\boldsymbol{\alpha}_{sn}$ in the multinomial logistic regression using the Newton–Raphson algorithm. Based on the multinomial likelihood formulation, the first- and second-order derivatives of $\boldsymbol{\alpha}_{ss}$ are

$$\sum_{i:D_i=1} \sum_{j\le m_{1,i,1}} \eta_{ss,ij}\mathbf{x}_{ij}' + \sum_{i:D_i=0} \sum_{j\le m_{0,i,1}} \mathbf{x}_{ij}' - \sum_{i:D_i=1} \sum_{j\le m_{1,i}} \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})}{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{sn})}\mathbf{x}_{ij}'$$
$$- \sum_{i:D_i=0} \sum_{j\le m_{0,i}} \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})}{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{sn})}\mathbf{x}_{ij}',$$

and

$$- \sum_{i:D_i=1} \sum_{j\le m_{1,i}} \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})\left\{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{sn})\right\}}{\left\{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{sn})\right\}^2}\mathbf{x}_{ij}\mathbf{x}_{ij}' - \sum_{i:D_i=0} \sum_{j\le m_{0,i}} \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})\left\{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{sn})\right\}}{\left\{1+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{ss})+\exp(\mathbf{x}_{ij}'\boldsymbol{\alpha}_{sn})\right\}^2}\mathbf{x}_{ij}\mathbf{x}_{ij}'.$$

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Similarly, the first- and second-order derivatives of $\boldsymbol{\alpha}_{sn}$ are

$$\sum_{i:D_i=1}\sum_{j\le m_{1,i,1}}(1-\eta_{ss,ij})\mathrm{x}'_{ij}+\sum_{i:D_i=0}\sum_{j\le m_{0,i,0}}\eta_{sn,ij}\mathrm{x}'_{ij}$$
$$-\sum_{i:D_i=1}\sum_{j\le m_{1,i}}\frac{\exp(\mathrm{x}'_{ij}\boldsymbol{\alpha}_{sn})}{\left\{1+\exp(\mathrm{x}'_{ij}\boldsymbol{\alpha}_{ss})+\exp(\mathrm{x}'_{ij}\boldsymbol{\alpha}_{sn})\right\}}\mathrm{x}'_{ij}-\sum_{i:D_i=0}\sum_{j\le m_{0,i}}\frac{\exp(\mathrm{x}'_{ij}\boldsymbol{\alpha}_{sn})}{\left\{1+\exp(\mathrm{x}'_{ij}\boldsymbol{\alpha}_{ss})+\exp(\mathrm{x}'_{ij}\boldsymbol{\alpha}_{sn})\right\}}\mathrm{x}'_{ij},$$

and

$$-\sum_{i:D_i=1}\sum_{j\le m_{1,i}}\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn})\left\{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss})\right\}}{\left\{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss})+\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn})\right\}^2}\mathbf{x}_{ij}\mathbf{x}'_{ij}-\sum_{i:D_i=0}\sum_{j\le m_{0,i}}\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn})\left\{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss})\right\}}{\left\{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss})+\exp(\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn})\right\}^2}\mathbf{x}_{ij}\mathbf{x}'_{ij}.$$

These explicit forms of derivatives will be used in the iterative procedures required by the Newton–Raphson formula for updating the strata membership parameters.

### 4.3 | Estimating the SACE in cluster-randomized trials

Because the joint potential values of the survival status $\{S_{ij}(1), S_{ij}(0)\}$ are not simultaneously observable, it is not immediate how to derive the SACE from the parameter estimates under our modeling assumptions. To proceed, we base our estimate on an equivalent form of the SACE (a g-computation formula or model-based standardization formula) below. We summarize the result in a proposition, and the proof is provided in the supplementary material (Section S2.).

**Proposition 1.** (*A model-based g-computation formula*) *Under Assumptions 1–3 and assuming the multilevel mixture model, we have*

$$E[Y_{ij}(1)-Y_{ij}(0)|S_{ij}(1)=S_{ij}(0)=1]=\frac{\mathrm{E}\left[(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,1}+u_{1i})p_{ss,ij}\right]}{\mathrm{P}\left[S_{ij}(1)=S_{ij}(0)=1|D_i=1\right]}-\frac{\mathrm{E}\left[(\mathbf{x}'_{ij}\boldsymbol{\beta}_{ss,0}+u_{0i})p_{ss,ij}\right]}{\mathrm{P}\left[S_{ij}(1)=S_{ij}(0)=1|D_i=0\right]}.$$

After we obtain the parameter estimates, we get the estimated $\hat{p}_{ss,ij}$ from the principal strata model. In the $i$th treatment cluster, let $\hat{y}_{ij}=\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_{ss,1}+\hat{\mathrm{E}}(u_{1i}|\mathbf{y}_i,\mathbf{X}_i)$. In the $i$th control cluster, let $\hat{y}_{ij}=\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_{ss,0}+\hat{\mathrm{E}}(u_{0i}|\mathbf{y}_i,\mathbf{X}_i)$. Proposition 1 suggests that we can estimate the SACE using the sample analog

$$\widehat{\mathrm{SACE}}=\frac{\sum_{i:D_i=1}\sum_{j=1}^{m_{1,i}}\hat{p}_{ss,ij}\hat{y}_{ij}}{\sum_{i:D_i=1}\sum_{j=1}^{m_{1,i}}\hat{p}_{ss,ij}}-\frac{\sum_{i:D_i=0}\sum_{j=1}^{m_{0,i}}\hat{p}_{ss,ij}\hat{y}_{ij}}{\sum_{i:D_i=0}\sum_{j=1}^{m_{0,i}}\hat{p}_{ss,ij}}. \tag{7}$$

Finally, we obtain the confidence interval of the estimate using the bootstrap method of sampling the clusters with replacement.[40,41] We randomly sample the treated clusters and the control clusters separately. For each bootstrap sample, we fit the model and obtain the SACE estimate. The 95% confidence intervals are obtained from the 2.5% and the 97.5% quantiles of the bootstrap estimates.

## 5 | AN EXTENSION TO INCLUDE A CLUSTER-LEVEL RANDOM INTERCEPT IN THE STRATA MEMBERSHIP MODEL

In some cases, one may be interested in accounting for unmeasured cluster-level variation in the strata membership model. Specifically, we could include a random intercept $v_{1i}$ or $v_{0i}$ in the $i$th treated or the $i$th control cluster into the multinomial logistic membership model. In the $i$th treated cluster, the reference "nn" stratum membership model becomes

$$p_{nn,ij}=\frac{1}{1+\exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss}+v_{1i}\}+\exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn}+v_{1i}\}},$$

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and we further have $p_{ss,ij} = p_{nn,ij}\exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{ss} + v_{1i}\}$, and $p_{sn,ij} = p_{nn,ij}\exp\{\mathbf{x}'_{ij}\boldsymbol{\alpha}_{sn} + v_{1i}\}$. We assume the $v_{1i}$'s and $v_{0i}$'s independently follow the normal distribution $\mathcal{N}(0, \gamma^2)$. From Hedeker,[42] this model formulation induces an ICC (on the latent response scale) for the strata membership defined as $\gamma^2/(\gamma^2 + \pi^2/3)$. We also assume they are independent of the random intercepts in the outcome models,[31] and are independent of the residual errors there. Under this model specification, an individual's principal strata membership is assumed to be independent of the cluster-level random effect in the potential outcome model given the measured baseline covariates. This assumption will likely be more plausible when a rich set of risk factors are collected at baseline but will be violated when at least one cluster-level risk factor that happens to predict the strata membership is omitted from data collection. For example, under this last scenario, there can be a nonzero correlation between $v_{1i}$ and $u_{1i}$ given only the observed covariates and incorrectly assuming their independence may lead to biased parameter estimates.

With the extra random intercept in the strata model, the complete data of the $i$th treated clusters are $(\mathbf{z}_{ij}, \mathbf{y}_i, u_{1i}, v_{1i})$ whose likelihood is

$$\prod_{j=1}^{m_{1,i}}\left[p_{ss,ij}^{Z_{ss,ij}} f_{ss,1}(y_{ij}|\mathbf{x}_{ij}, u_{1i})^{Z_{ss,ij}}\right]\left[p_{sn,ij}^{Z_{sn,ij}} f_{sn}(y_{ij}|\mathbf{x}_{ij}, u_{1i})^{Z_{sn,ij}}\right]\left[p_{nn,ij}^{Z_{nn,ij}}\right] f(u_{1i}) f(v_{1i}).$$

Similarly, in the $i$th control cluster, the likelihood of the complete data is

$$\prod_{j=1}^{m_{0,i}}\left[p_{ss,i}^{Z_{ss,ij}} f_{ss,0}(y_{ij}|\mathbf{x}_{ij}, u_{0i})^{Z_{ss,ij}}\right]\left[p_{sn,ij}^{Z_{sn,ij}}\right]\left[p_{nn,ij}^{Z_{nn,ij}}\right] f(u_{0i}) f(v_{0i}).$$

Based on the complete data likelihood, we develop the EM algorithm to estimate the model parameters $\{\boldsymbol{\beta}_{ss,1}, \boldsymbol{\beta}_{sn}, \boldsymbol{\beta}_{ss,0}, \boldsymbol{\alpha}_{ss}, \boldsymbol{\alpha}_{sn}, \sigma^2, \tau^2, \gamma^2\}$. The iterative estimation procedure is similar to that in Section 4, and the detailed derivation is provided in the supplementary material (Section S3). Estimates of the parameters except for $\gamma^2$ have the analogous formulations as in the basic strata model without the random intercept. Importantly, the first- and second- order derivatives of $\boldsymbol{\alpha}_{ss}$ and $\boldsymbol{\alpha}_{sn}$ are now functions of the $v_{1i}$'s and $v_{0i}$'s. The E-step calculation also relies on the conditional mean of $v_{1i}$ and $v_{0i}$ given the observed data. We perform a Monte Carlo E-step calculation by random sampling 100 $v_{1i}$'s or $v_{0i}$'s from $N(0, \gamma^2)$ given the estimates from the previous iteration. Then we calculate the conditional expectation numerically based on the posterior density.

In addition to the observed outcome $\mathbf{y}_i$, the observed data also include the principal strata status $z_{nn,ij}$'s in a treated cluster or $z_{ss,ij}$'s in a control cluster. We let $\mathbf{z}_i^* = \{z_{nn,ij} : j = 1, \ldots, m_{1,i}\}$ in the $i$th treated cluster or $\mathbf{z}_i^* = \{z_{ss,ij} : j = 1, \ldots, m_{0,i}\}$ in the $i$th control cluster. An updated estimate of $\gamma^2$ is given by

$$\hat{\gamma}^2 = \frac{\sum_{i:D_i=1}\mathrm{E}(v_{1i}^2|\mathbf{z}_i^*, \mathbf{X}_i) + \sum_{i:D_i=0}\mathrm{E}(v_{0i}^2|\mathbf{z}_i^*, \mathbf{X}_i)}{n_1 + n_0}.$$

Once our model parameters are estimated (after convergence of the EM algorithm), the SACE estimate is computed by (7), where the estimate of the principal strata model $\hat{p}_{ss,ij}$ is replaced by the integrals $\int p_{ss,ij} f(v_{1i})\mathrm{d}v_{1i}$ or $\int p_{ss,ij} f(v_{0i})\mathrm{d}v_{0i}$ calculated numerically. We again use the bootstrap method, that is, sampling the clusters with replacement, to obtain the 95% confidence interval.

## 6 | A SIMULATION STUDY

We follow the ADEMP (aims, data-generating mechanisms, methods, estimands, performance measures) framework[43] to describe our simulation setups. The goal of our simulations is to examine the finite-sample performance of the proposed methods for estimating SACE in parallel-arm CRTs. We examine settings with the number of clusters $n_c = 30$ and 60 in each arm to capture medium and large numbers of clusters.[44,45] In each cluster, the number of individuals is simulated from a normal distribution after rounding to the nearest integer with mean $\overline{m} \in \{25, 50\}$, and standard deviation 3. The variance parameters satisfy $\tau^2 + \sigma^2 = 2$, and we set the ICC of the outcome models as $\rho = \tau^2/(\tau^2 + \sigma^2) = 0.01$, 0.05, and 0.1, corresponding to those commonly reported in the CRT literature.[46] These ICC values correspond to $\tau^2 = 0.02$, 0.1 and 0.2, and $\sigma^2 = 1.98$, 1.9 and 1.8. We simulate two covariates independently, one from a Bernoulli distribution with a probability of success as 0.5, and the other from a standard normal distribution. Values of the fixed effects, including the intercepts, are specified at $\boldsymbol{\beta}_{ss,1} = (-0.5, 1, 1.5)'$, $\boldsymbol{\beta}_{sn} = (-0.3, 0.8, 1.3)'$, and $\boldsymbol{\beta}_{ss,0} = (-0.2, 1, 1)'$.

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We simulate the principal strata membership variable $\mathbf{z}_{ij}$ from the multinomial logistic model introduced in Section 3.1 without the cluster-level random intercept, and then from the model in Section 5 with a cluster-level random intercept. In the latter case, we set the variances of the random intercept to be $\gamma^2 = 0.37$ and 0.8, which correspond to the ICC of the strata membership of 0.1 and 0.2 using the latent response definition.[42] In what follows, we write $\gamma^2 = 0$ to denote the principal strata simulation model with no random intercept. In each scenario, we specify two sets of coefficients in the principal strata models, that is, $\boldsymbol{\alpha}_{ss} = (1, 2, 1)'$ and $\boldsymbol{\alpha}_{sn} = (-0.5, -1.5, -1)'$ in setting A, and $\boldsymbol{\alpha}_{ss} = (1.6, 0.2, 0.1)'$ and $\boldsymbol{\alpha}_{sn} = (-0.1, -0.1, -0.2)'$ in setting B. We maintain the marginal principal strata proportion to be roughly the same in both settings (Section S4 of the supplementary material). In setting A, the odds ratios associated with the binary and continuous covariates are 7.39 and 2.72 in the "ss" stratum, and 0.22 and 0.37 in the "sn" stratum. In setting B, the odds ratios are 1.22 and 1.11 in the "ss" stratum, and 0.90 and 0.82 in the "sn" stratum. Of note, the covariates are strongly predictive of the principal strata membership (and hence survival status) in setting A, but are only weakly predictive of the principal strata membership in setting B.

Under each data-generating process, we compare the performance of the proposed ME approaches with the FE only outcome modeling approach to assess the necessity of accounting for clustering in the outcome model. The FE approach is an adaption from Zhang et al.[3,20] under the monotonicity and normality assumptions, where there is no cluster-level random intercept in either the strata-specific potential outcome model or principal strata model. Furthermore, we compare the methods in Section 3 and Section 5 to explore when it becomes necessary to adjust for clustering in the principal strata membership model beyond adjusting for clustering in the potential outcome models. For ease of reference, we abbreviate the approach in Section 3 with clustering in the outcome model only as the ME approach, and the approach in Section 5 with clustering in all models as the ME2 approach. Between the ME approach and the FE approach,[20] we retain the same principal strata membership model. The corresponding SACE estimate of the FE approach follows (7) with the random intercepts set to zero.

For all simulation scenarios, the true value of the SACE estimand is computed based on the following Monte Carlo procedure. Let $\mathbb{I}\{Z_{ss,ij} = 1\}$ be the indicator function which equals to 1 if the $(i, j)$th individual is in the "ss" stratum. Under each data-generating process where we know both potential outcomes, the SACE estimand can be obtained as follows:

$$\frac{\sum_{i:D_i=1}\sum_{j=1}^{m_{1,i,1}} y_{ij}(1)\mathbb{I}\{Z_{ss,ij}=1\}}{\sum_{i:D_i=1}\sum_{j=1}^{m_{1,i,1}}\mathbb{I}\{Z_{ss,ij}=1\}} - \frac{\sum_{i:D_i=0}\sum_{j=1}^{m_{0,i,1}} y_{ij}(0)\mathbb{I}\{Z_{ss,ij}=1\}}{\sum_{i:D_i=0}\sum_{j=1}^{m_{0,i,1}}\mathbb{I}\{Z_{ss,ij}=1\}}.$$

To ensure the stability of the computation for the true SACE estimand, we implement the above formula across all 200 simulated datasets, and then consider the average of the 200 SACE values to be the true value of the target estimand. Finally, for each data-generating process, due to extensive computation time, we carry statistical inference using 200 bootstrap samples in each of the generated 200 simulated datasets (for obtaining the confidence interval of the SACE). Specifically, since the FE approach assumes independence between individuals, we use individual-level bootstrap for estimating confidence intervals. Cluster-level bootstrap is used for inference with the proposed ME and ME2 approaches. To compare model performance, we examine the absolute value of the bias (Table S7 of the supplementary material), the mean-squared error (MSE) of the estimator (Table 4), as well as the coverage proportion of the bootstrap-based 95% confidence intervals (Table 5), relative to the true SACE we have obtained for each scenario.

Across the simulation scenarios, although the FE approach has a slightly lower bias overall, all of the approaches have a small bias of less than 2%. Importantly, the ME2 and ME approaches have smaller MSEs and close to the nominal 95% coverage. Interestingly, under the two levels of ICC for the principal strata membership variables we considered, the ME approach is generally robust even if the random intercepts are omitted from the principal strata model, but it is computationally more efficient and thus more convenient to implement. The coverage proportion of the FE approach deteriorates as the outcome ICC increases. Overall, this comparison demonstrates the necessity of accounting for outcome clustering in CRTs for more accurate inference with SACE, especially when the outcome ICC becomes larger. When the principal strata membership ICC does not exceed 0.2, our simulations demonstrate that inference about SACE is robust as long as the outcome ICC is already accounted for.

Finally, to provide additional intuitions on the performance of each method, we examine the estimation results in a single simulated dataset of 30 clusters, with the average cluster size being 25. The outcome and strata membership ICCs are $\rho = 0.1$ and $\gamma^2/(\gamma^2 + \pi^2/3) = 0$, respectively. The coefficients of the principal strata model are $\boldsymbol{\alpha}_{ss} = (1, 2, 1)'$ and $\boldsymbol{\alpha}_{sn} = (-0.5, -1.5, -1)'$. Table 6 shows the SACE estimates and the associated 95% confidence intervals. The target SACE estimand under this data-generating process is computed as −0.19. With that particular simulation dataset, the three methods ME2, ME, and FE produce the similar SACE point estimates. However, the confidence interval of the FE

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 4** Comparison of the MSE ($\times 10^{-2}$) of estimating SACE among the proposed ME approach with random intercepts in the principal strata models (ME2), the proposed ME approach and the FE approach.

| | | | | $\rho = 0.01$ | | | $\rho = 0.05$ | | | $\rho = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma^2$ | $\overline{m}$ | $n_c$ | ME2 | ME | FE | ME2 | ME | FE | ME2 | ME | FE |
| Setting A | 0 | 25 | 30 | 1.47 | 1.47 | 1.49 | 1.73 | 1.71 | 1.82 | 2.93 | 2.92 | 3.13 |
| | | | 60 | 0.70 | 0.70 | 0.71 | 0.90 | 0.91 | 0.95 | 1.26 | 1.27 | 1.38 |
| | | 50 | 30 | 0.95 | 0.96 | 0.99 | 1.19 | 1.19 | 1.24 | 2.09 | 2.11 | 2.23 |
| | | | 60 | 0.40 | 0.40 | 0.42 | 0.63 | 0.63 | 0.65 | 0.91 | 0.91 | 1.01 |
| | 0.37 | 25 | 30 | 1.65 | 1.66 | 1.69 | 1.71 | 1.72 | 1.76 | 2.69 | 2.66 | 2.86 |
| | | | 60 | 0.84 | 0.83 | 0.88 | 0.99 | 0.99 | 1.04 | 1.26 | 1.27 | 1.33 |
| | | 50 | 30 | 0.88 | 0.87 | 0.92 | 1.41 | 1.40 | 1.50 | 1.73 | 1.72 | 1.88 |
| | | | 60 | 0.37 | 0.37 | 0.40 | 0.63 | 0.64 | 0.68 | 1.00 | 1.00 | 1.05 |
| | 0.8 | 25 | 30 | 1.31 | 1.31 | 1.34 | 2.13 | 2.14 | 2.20 | 2.72 | 2.67 | 2.97 |
| | | | 60 | 0.72 | 0.73 | 0.75 | 0.95 | 0.95 | 1.03 | 1.04 | 1.02 | 1.15 |
| | | 50 | 30 | 0.75 | 0.74 | 0.77 | 1.18 | 1.18 | 1.30 | 1.94 | 1.96 | 2.14 |
| | | | 60 | 0.41 | 0.42 | 0.44 | 0.67 | 0.66 | 0.70 | 1.00 | 0.99 | 1.07 |
| Setting B | 0 | 25 | 30 | 1.60 | 1.63 | 1.66 | 1.75 | 1.73 | 1.90 | 2.58 | 2.60 | 2.52 |
| | | | 60 | 0.78 | 0.77 | 0.84 | 1.04 | 1.02 | 1.20 | 1.37 | 1.38 | 1.50 |
| | | 50 | 30 | 1.18 | 1.17 | 0.90 | 1.56 | 1.57 | 1.52 | 1.62 | 1.62 | 1.82 |
| | | | 60 | 0.41 | 0.41 | 0.45 | 0.64 | 0.63 | 0.78 | 0.83 | 0.82 | 0.90 |
| | 0.37 | 25 | 30 | 1.85 | 1.87 | 1.50 | 2.09 | 2.09 | 2.34 | 2.34 | 2.41 | 2.42 |
| | | | 60 | 0.81 | 0.80 | 0.91 | 1.07 | 1.06 | 1.15 | 1.37 | 1.38 | 1.48 |
| | | 50 | 30 | 0.95 | 0.96 | 0.97 | 1.77 | 1.77 | 1.78 | 2.10 | 2.07 | 2.18 |
| | | | 60 | 0.44 | 0.44 | 0.44 | 0.63 | 0.63 | 0.66 | 0.87 | 0.88 | 0.89 |
| | 0.8 | 25 | 30 | 1.79 | 1.79 | 1.77 | 1.82 | 1.81 | 2.08 | 2.23 | 2.26 | 2.52 |
| | | | 60 | 0.77 | 0.78 | 0.92 | 0.86 | 0.86 | 0.98 | 1.18 | 1.20 | 1.29 |
| | | 50 | 30 | 0.81 | 0.79 | 0.86 | 1.31 | 1.30 | 1.37 | 1.91 | 1.88 | 2.26 |
| | | | 60 | 0.41 | 0.42 | 0.44 | 0.60 | 0.58 | 0.63 | 0.86 | 0.85 | 0.86 |

approach fails to cover the true value. We also fit two conventional linear mixed effects model (LMM),[47] including only individuals with observed outcomes. Specifically, we adjust for the same set of covariates as main effects only (referred to as the mixed ANCOVA1 model). In addition, we fit the LMM by adjusting for the mean-centered covariates and all first-order treatment-by-covariate (mean-centered) interactions (referred to as the mixed ANCOVA2 model). In both models, we assess the treatment effect by examining the coefficient of the treatment variable. In contrast to the three SACE estimators (FE, ME, and ME2), the treatment effect estimates from the two LMM further deviate from the true SACE estimand, and the two 95% confidence intervals fail to contain the true SACE estimand either. This is not surprising because the LMM approach is based on the observed survivors only, and the treatment effect coefficient often does not bear a valid causal interpretation in the presence of truncation-by-death.

## 7 | APPLICATION TO THE WHOLE SYSTEMS DEMONSTRATOR (WSD) TELECARE QUESTIONNAIRE STUDY

To illustrate our methodology in a real-world trial setting, we reanalyzed the Whole Systems Demonstrator (WSD) Telecare Questionnaire Study, which is a pragmatic cluster-randomized WSD telecare trial conducted to compare participant-reported outcomes between home-based telecare with the usual care across three local authority sites in England.[48,49] In the study, participants registered within a particular general practice, and the general practice (cluster) is the

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 5** Comparison of the proportion of coverage (%) of SACE among the proposed ME approach with random intercepts in the principal strata models (ME2), the proposed ME approach, and the FE approach.

| | | | | $\rho = 0.01$ | | | $\rho = 0.05$ | | | $\rho = 0.1$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma^2$ | $\overline{m}$ | $n_c$ | ME2 | ME | FE | ME2 | ME | FE | ME2 | ME | FE |
| Setting A | 0 | 25 | 30 | 93.5 | 94.5 | 93.5 | 94.5 | 94.0 | 89.0 | 90.5 | 89.5 | 82.5 |
| | | | 60 | 92.5 | 94.0 | 93.0 | 94.5 | 93.0 | 91.0 | 94.5 | 94.5 | 80.5 |
| | | 50 | 30 | 91.5 | 92.5 | 90.5 | 94.0 | 94.5 | 86.0 | 92.5 | 91.5 | 75.0 |
| | | | 60 | 95.0 | 94.5 | 93.0 | 94.5 | 94.5 | 83.5 | 92.0 | 93.5 | 76.0 |
| | 0.37 | 25 | 30 | 93.5 | 93.0 | 92.0 | 94.5 | 95.5 | 92.0 | 94.0 | 94.0 | 84.0 |
| | | | 60 | 92.5 | 94.0 | 93.5 | 94.0 | 93.5 | 89.5 | 92.5 | 92.5 | 81.5 |
| | | 50 | 30 | 94.0 | 93.0 | 92.5 | 91.0 | 93.5 | 83.0 | 94.5 | 93.5 | 76.5 |
| | | | 60 | 94.5 | 93.0 | 92.5 | 94.5 | 95.5 | 84.0 | 95.5 | 93.5 | 68.0 |
| | 0.8 | 25 | 30 | 95.5 | 95.0 | 95.0 | 90.5 | 90.0 | 84.5 | 93.0 | 93.0 | 82.0 |
| | | | 60 | 95.5 | 96.0 | 94.5 | 93.5 | 95.5 | 91.0 | 95.5 | 97.0 | 88.5 |
| | | 50 | 30 | 96.0 | 96.0 | 95.0 | 94.5 | 95.5 | 85.5 | 95.0 | 95.5 | 75.0 |
| | | | 60 | 95.0 | 95.0 | 92.5 | 93.5 | 92.5 | 82.0 | 92.0 | 91.5 | 72.5 |
| Setting B | 0 | 25 | 30 | 94.5 | 94.5 | 94.5 | 97.0 | 97.0 | 95.0 | 95.5 | 95.0 | 86.5 |
| | | | 60 | 97.0 | 97.0 | 96.5 | 94.0 | 96.0 | 92.0 | 93.0 | 94.0 | 89.0 |
| | | 50 | 30 | 91.0 | 90.5 | 91.5 | 91.5 | 92.0 | 84.5 | 94.0 | 94.0 | 83.5 |
| | | | 60 | 94.0 | 92.5 | 92.5 | 94.0 | 94.5 | 87.0 | 93.0 | 93.5 | 83.5 |
| | 0.37 | 25 | 30 | 93.5 | 93.5 | 97.0 | 94.5 | 93.0 | 90.0 | 95.5 | 95.5 | 90.5 |
| | | | 60 | 94.5 | 94.5 | 95.5 | 95.0 | 93.0 | 93.0 | 93.5 | 92.0 | 87.5 |
| | | 50 | 30 | 93.0 | 92.5 | 94.0 | 90.0 | 90.0 | 83.5 | 90.0 | 91.0 | 79.5 |
| | | | 60 | 92.5 | 94.5 | 94.0 | 94.0 | 95.0 | 89.0 | 93.0 | 95.0 | 82.5 |
| | 0.8 | 25 | 30 | 94.0 | 94.0 | 96.5 | 95.5 | 95.0 | 90.5 | 94.0 | 94.0 | 89.5 |
| | | | 60 | 96.0 | 95.5 | 97.0 | 98.5 | 97.5 | 96.5 | 94.0 | 94.0 | 89.0 |
| | | 50 | 30 | 96.0 | 96.5 | 97.0 | 96.5 | 96.5 | 88.5 | 94.0 | 95.0 | 82.5 |
| | | | 60 | 96.5 | 98.0 | 97.0 | 95.0 | 95.5 | 91.0 | 94.5 | 93.5 | 85.5 |

*Note*: The number of clusters is $n_c = 30$ or 60; the average cluster size is $\overline{m} = 25$ or 50. The variance of the random intercept in the principal strata membership model is $\gamma^2 = 0$, 0.37, or 0.8, and the outcome ICC is $\rho = 0.01$, 0.05, or 0.1. Setting A: baseline covariates are strongly predictive of the principal strata membership; Setting B: baseline covariates are weakly predictive of the principal strata membership.

unit of randomization. There were 201 clusters, 100 in the telecare arm, and 101 in the usual care arm. The cluster sizes of the telecare group ranged from 1 to 20 with an average value of 5.24, and the cluster sizes in the usual care group ranged from 1 to 23 with an average value of 5.96. For missing data on the baseline covariates and outcomes not due to death, moving to residential or nursing care or seriously deteriorated health (considered as sources of truncation), we used multiple imputation to impute a single dataset to fill in missing entries. The imputation step followed from the illustrative analysis of the WSD telecare trial in Tong et al.,[33] who focused on the QOL outcome measured via the Euro-Qol Group's EQ-5D-VAS index, a self-rated scale (score range, 0–100) with 5 domains (mobility, self-care, usual activities, pain and discomfort, and anxiety and depression).

In this study, we examined another outcome: the physical quality of life outcome (PCS) of 1126 participants. The PCS was as assessed by the Short Form 12-item Survey (SF-12).[50] The outcome measures for the trial were collected at the baseline and also at 12 months, which was the final time of data collection. However, those outcomes were not measured if participants withdrew from the study. The main reasons for formal withdrawal included death, moving to residential or nursing care, and deterioration in condition (physical or mental capacity).[50] Among the 524 participants in the telecare arm, 45 (8.58%) of the outcomes were truncated due to the above reasons. In the usual care arm, 82 (13.62%)

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

**TABLE 6** An illustrative analysis of a single simulated dataset.

| Approach | Estimate | 95% Confidence interval | |
|---|---|---|---|
| ME2 | −0.46 | −0.74 | −0.16 |
| ME | −0.46 | −0.71 | −0.15 |
| FE | −0.46 | −0.68 | −0.26 |
| LMM (mixed ANCOVA1) | −0.52 | −0.78 | −0.27 |
| LMM (mixed ANCOVA2) | −0.51 | −0.66 | −0.35 |

*Note*: Simulation setting: the number of clusters is $n_c = 30$, and the average cluster size is $\overline{m} = 25$. The variance of the random intercept in the principal strata membership model is $\gamma^2 = 0$, and the outcome ICC is $\rho = 0.1$. The true SACE value is −0.19.

**TABLE 7** Estimation of the SACE and the proportion of a principal stratum by the proposed mixed-effects (ME2 and ME) approach and the fixed effects approach (FE) in the WSD telecare trial.

| | ME2 | ME | FE |
|---|---|---|---|
| SACE | 0.17 (−0.95,1.29) | 0.30 (−0.93,1.30) | 0.19 (−1.04, 1.24) |
| ss: always survivors | 0.84 (0.80,0.86) | 0.83 (0.79, 0.85) | 0.84 (0.80, 0.86) |
| sn: protected | 0.09 (0.05,0.13) | 0.06 (0.04, 0.11) | 0.07 (0.05, 0.12) |
| nn: never survivors | 0.08 (0.06,0.10) | 0.11 (0.08, 0.12) | 0.08 (0.06, 0.11) |

of the 602 outcomes were truncated due to the above reasons (in this illustrative example, we did not further differentiate these specific reasons and consider them collectively as truncation). Figure S1 in the supplementary material shows the outcome truncation proportion for each arm. The telecare arm had a slightly higher mean and standard deviation of the PCS outcome, 27.63 vs 27.21 and 8.50 vs 8.30. Figure S2 of the supplementary material shows the histograms and the QQ normal plots of the outcome for all participants and also by arms.

Each participant's demographic information was also recorded, including age, gender, ethnicity, number of comorbid conditions, level of education, and the Index of Multiple Deprivation score based on their postcode at the time of enrollment into the trial. We applied the proposed approaches (ME2 and ME) and the FE approach to analyze the PCS outcome adjusting for these baseline covariates and also the baseline measurement of PCS. The estimated outcome model ICC's were 0.02 and 0.01 from the ME2 and ME approaches, and the estimated ICC of the strata membership was 0.16 by the ME2 approach. Table 7 shows the SACE estimates together with their 95% confidence intervals based on 500 bootstrap replicates (obtained in a similar fashion as in our simulation study). The table also presents the estimated proportion of each principal stratum as the average of the individual probabilities for belonging to each stratum. Across the three methods, the estimated proportions are similar, and all confidence intervals include zero.

We further compared the three clustered SACE estimates with the two LMM approaches (mixed ANCOVA1 and mixed ANCOVA2) introduced in Section 6 which excludes the individuals with truncated quality-of-life outcomes. After excluding the truncated outcomes, the coefficient estimate from the mixed ANCOVA1 model was 0.06 with 95% confidence interval $(-0.81, 0.93)$. In addition, we obtained the main treatment effect coefficient as 0.08 with 95% confidence interval $(-0.80, 0.95)$ from the mixed ANCOVA2 model. It is important to note that, although the mixed ANCOVA models can provide consistent estimators of the average treatment effect in CRTs in the absence of informative truncation,[35] their target estimands are generally challenging to derive analytically under truncation-by-death (except that they are solutions to the observed-data score equations under the LMM likelihood formulation). In other words, the estimands under LMMs are ambiguous and difficult to interpret in the WSD telecare trial. In this data example, although the confidence intervals under the LMM similarly included zero, the magnitude of the point estimates were smaller than those offered by the mixture modeling approaches (whose estimands were the SACE). It is also worth noting that using a LMM excluded the 127 (11.28%) individuals from the analysis. In contrast, the ME, ME2, and FE methods leverage all data information to target the SACE, which bears a clear causal interpretation. For example, we interpreted the results under ME2, which account for clustering in both the potential outcome models and the principal strata membership model. That is, in the analysis of the WSD telecare trial, the average difference in physical quality of life measurement between the telecare arm and the usual care arm, among the participants whose outcome would not be truncated regardless of treatment assignment

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(the subset of always survivors), was estimated as 0.17. The associated 95% confidence interval was $(-0.95, 1.29)$ and included zero.

## 8 | SUMMARY AND DISCUSSION

In many trials, both individually and cluster randomized, participant outcomes will be unobserved due to truncation-by-death or other informative dropout reasons (eg, migration or leaving school). In such settings, the SACE estimand can provide valuable insights about treatment effects among a defined subset of individuals who would not experience truncation under either treatment or control. We develop a mixed-model approach to estimate the SACE in CRTs and explore the performance characteristics of this approach using simulations. It is worth noting that our methodology allows for the inclusion of covariates, which are commonly used to improve precision in CRTs, and can be easily extended to observational data settings when the cluster-level assignment is assumed to be ignorable. It is worth noting that we recommend adjusting for baseline covariates when estimating SACE, even in randomized trials, because randomization does not ensure covariate balance across principal strata, and failure to adjust for covariates can lead to a misspecified Gaussian mixture model, and thus biased SACE estimates. For point estimation, we describe a Monte Carlo EM algorithm and construct confidence intervals using cluster-level bootstrap. In our simulations under the CRT setting, we observe that confidence intervals derived from the proposed approach have closer to the nominal 95% coverage than the approach with fixed effects only, highlighting the importance of accounting for clustering in the potential outcome model. However, we also find that when the clustering in the potential outcomes is already accounted for, adjusting for clustering in the principal strata model does not lead to appreciable differences in the inference for SACE when the ICC of the strata membership variable is within the commonly reported range in CRTs (ie, less than 0.2). Although we have not identified issues on multiple solutions or non-convergence using our EM algorithm in our data application, in practice, one could consider using multiple sets of initial values to check the sensitivity of the EM estimates and avoid convergence to a local minima. Finally, while our development has centered around a continuous outcome, our mixed-model approach can be extended to study binary outcomes or count outcomes with modifications to the EM algorithm, and we plan to report those developments in a separate report.

The estimation of the SACE in CRTs with our method relies on several assumptions. Firstly, we have assumed monotonicity as a nonparametric (structural) assumption. This assumption, while seemingly plausible for the WSD trial studying cluster-level treatment with perceived benefit against no intervention, may be less reasonable for certain study designs, such as a comparative effectiveness trial comparing two active treatments. Under our likelihood-based inference strategy, relaxing the monotonicity assumption to include the harmed strata may lead to additional computational challenges and potential convergence issues of the EM algorithm (because the likelihood may become relatively flat with additional strata). Alternatively, nonparametric assumptions that replace the monotonicity assumption, such as the explainable nonrandom survival assumption studied in Hayden et al.,[14] may be worth further investigating for CRTs.

Secondly, we have relied on parametric assumptions for deriving a likelihood-based SACE estimator. In our setup, the likelihood assumption is necessary because Assumptions 1–3 can only partially identify or interval identify the SACE.[1] That is, the mixture model likelihood and the nonparametric assumptions are combined to identify the SACE in CRTs empirically. In the context of independent data and treatment noncompliance, Jo[38] has provided an insightful discussion on the mechanics of mixture models for point identification of principal causal effects (but only focused on the two-strata case). Jo[38] noted that the distance between principal strata (location parameters of each mixture component) could impact empirical identification of the principal causal effects under likelihood-based inference, and developed the moving exclusion restriction condition for diagnostics and sensitivity analyses. In general, when there is insufficient distance between the mixture components, the standard error and the width of the confidence interval for the SACE estimate may inflate, compromising the precision for inference. For Gaussian mixture modeling with independent data, Mercatanti et al.[51] developed analytical insights and empirical results to show that including a secondary outcome can effectively sharpen the inference on mixture model parameters for the primary outcome, and a greater improvement can be expected as the distance between mixture components increases. It is therefore reasonable to expect that, with a single outcome, violating the normality assumption can also lead to bias in the SACE estimates, but including a secondary outcome may reduce such bias. We anticipate that these general findings for mixture models (mostly developed for independent data) would be useful for our setting with correlated data, but a formal extension to accommodate multilevel data warrants future research.

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Thirdly, we have assumed an equal-variance constraint (across both arms and strata) in the potential outcome models as this offers a single, interpretable outcome ICC parameter that is often of interest in CRTs. While the equal-variance constraint is not a necessary condition for point identification, it can lead to biased inference for the SACE when the true outcome variances or ICCs differ by arm or strata.[38] We did not fully explore arm-specific and strata-specific variances and ICCs in this work, but such an extension is possible by modifying the steps of our EM algorithm. We leave a comprehensive evaluation of this more complicated model to future work.

Finally, we have assumed away informative cluster sizes (ie, the presence of a marginal association between cluster size and potential outcomes or their contrasts), which have been recently shown to have a critical impact on the clarity of target estimands in CRTs.[34,52] When informative cluster size is present, there are at least two versions of target estimands—one defined as a cluster-level average and one defined as an individual-level average. The implications of informative cluster sizes can be potentially more complex as not only the original cluster size can be informative, but the always-survivor strata size can also be informative. This complexity necessitates more precise definitions of target estimands to describe the SACE in CRTs, alternative identification conditions and possible modifications of the inferential procedure. A detailed treatment for estimating SACE in the presence of informative cluster sizes is beyond the scope of this article, and the impact of informative cluster sizes under truncation-by-death will be explored in our future work.

## AUTHOR CONTRIBUTIONS

Wei Wang, Fan Li, and Michael O. Harhay developed the idea for the manuscript. Wei Wang led the analysis. Wei Wang, Fan Li, and Michael O. Harhay led the writing of several iterations of the manuscript. All authors provided feedback on the manuscript structure and content and provided edits. All authors approved of the final submitted manuscript.

## ACKNOWLEDGMENTS

Research in this article was partially supported by the Patient-Centered Outcomes Research Institute (PCORI Awards ME-2020C1-19220 to Michael O. Harhay and ME-2020C3-21072 to Fan Li). Fan Li and Michael O. Harhay are funded by the United States National Institutes of Health (NIH), National Heart, Lung, and Blood Institute (grant number R01-HL168202). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the NIH or PCORI or its Board of Governors or Methodology Committee. The Whole Systems "Demonstrator" (WSD) telecare trial study was funded by the Policy Research Programme in the Department of Health, UK. The views expressed are not necessarily those of the Department. The WSD study was approved by Liverpool Research Ethics Committee (ref: 08/H1005/4).

## CONFLICT OF INTEREST STATEMENT

Michael O. Harhay has received consulting fees from Elsevier, the American Thoracic Society, and Unlearn.AI, all for work unrelated to the topics in this manuscript.

## DATA AVAILABILITY STATEMENT

R functions to implement our methods (ME2 and ME) and the FE method are available online at https://github.com/harhay-lab/SACE_PS_LMM.

## ORCID

*Wei Wang* https://orcid.org/0000-0002-0084-0814
*Guangyu Tong* https://orcid.org/0000-0002-7697-5029
*Dylan S. Small* https://orcid.org/0000-0003-4928-2646
*Fan Li* https://orcid.org/0000-0001-6183-1893
*Michael O. Harhay* https://orcid.org/0000-0002-0553-674X

## REFERENCES

1. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J Educ Behav Stat*. 2003;28(4):353-368.
2. Rubin DB. Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Stat Sci*. 2006;21(3):299-309.

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3. Colantuoni E, Scharfstein D, Wang C, et al. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. *BMJ*. 2018;360(j5748):1-6.
4. Harhay M, Ratcliffe S, Small D, Suttner L, Crowther M, Halpern S. Measuring and analyzing length of stay in critical care trials. *Med Care*. 2019;57(9):e53-e59.
5. Hernán MA, Schisterman EF, Hernández-Díaz S. Invited commentary: composite outcomes as an attempt to escape from selection bias and related paradoxes. *Amer J Epidemiol*. 2013;179(3):368-370.
6. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *J Am Med Assoc*. 2003;289(19):2554-2559.
7. Neyman J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J R Stat Soc*. 1934;97(4):558-625.
8. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688-701.
9. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002;58(1):21-29.
10. Little R, Rubin D. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Ann Rev Publ Health*. 2000;21:121-145.
11. VanderWeele TJ. Principal stratification – uses and limitations. *Int J Biostat*. 2011;7(1):28.
12. Yang F, Small DS. Using post-outcome measurement information in censoring-by-death problems. *J R Stat Soc Ser B*. 2016;78(1):299-318.
13. Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. *Am J Epidemiol*. 2011;173(7):745-751.
14. Hayden D, Pauler D, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*. 2005;61(1):305-310.
15. Chen X, Harhay MO, Tong G, Li F. A Bayesian machine learning approach for estimating heterogeneous survivor causal effects: applications to a critical care trial. arXiv:2204.06657 2022.
16. Imai K. Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Stat Probab Lett*. 2008;78(2):144-149.
17. Long DM, Hudgens MG. Sharpening bounds on principal effects with covariates. *Biometrics*. 2013;69(4):812-819.
18. Ding P, Geng Z, Yan W, Zhou XH. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *J Am Stat Assoc*. 2011;106(496):1578-1591.
19. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B*. 1977;39(1):1-22.
20. Zhang JL, Rubin DB, Mealli F. Likelihood-based analysis of causal effects of job-training programs using principal stratification. *J Am Stat Assoc*. 2009;104(485):166-176.
21. Egleston BL, Scharfstein DO, Freeman EE, West SK. Causal inference for non-mortality outcomes in the presence of death. *Biostatistics*. 2007;8(3):526-545.
22. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med*. 2014;33(21):3601-3628.
23. Wang L, Zhou XH, Richardson TS. Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*. 2017;104(3):597-612.
24. Jo B, Stuart E. On the use of propensity scores in principal causal effect estimation. *Stat Med*. 2009;28(23):2857-2875.
25. Ding P, Lu J. Principal stratification analysis using principal scores. *J R Stat Soc Ser B*. 2017;79(3):757-777.
26. Frangakis CE, Rubin DB, Zhou XH. Clustered encouragement designs with individual noncompliance: Bayesian inference with randomization, and application to advance directive forms. *Biostatistics*. 2002;3(2):147-164.
27. Forastiere L, Mealli F, VanderWeele TJ. Identification and estimation of causal mechanisms in clustered encouragement designs: disentangling bed nets using Bayesian principal stratification. *J Am Stat Assoc*. 2016;111(514):510-525.
28. Turner EL, Li F, Gallis JA, Prague M, Murray DM. Review of recent methodological developments in group-randomized trials: part 1 – design. *Am J Publ Health*. 2017;107(6):907-915.
29. Turner EL, Prague M, Gallis JA, Li F, Murray DM. Review of recent methodological developments in group-randomized trials: part 2 – analysis. *Am J Publ Health*. 2017;107(7):1078-1086.
30. Jo B, Asparouhov T, Muthén B. Intention-to-treat analysis in cluster randomized trials with noncompliance. *Stat Med*. 2008;27:5565-5577.
31. Jo B, Asparouhov T, Muthén B, Ialongo N, Brown C. Cluster randomized trials with treatment noncompliance. *Psychol Methods*. 2008;13(1):1-18.
32. McLachlan GJ, Lee SX, Rathnayake SI. Finite mixture models. *Ann Rev Stat Appl*. 2019;6(1):355-378.
33. Tong G, Li F, Chen X, et al. A Bayesian approach for estimating the survivor average causal effect when outcomes are truncated by death in cluster-randomized trials. *Am J Epidemiol*. 2023;192(6):1006-1015.
34. Kahan BC, Li F, Copas AJ, Harhay MO. Estimands in cluster-randomized trials: choosing analyses that answer the right question. *Int J Epidemiol*. 2023;52(1):107-118.
35. Wang B, Harhay MO, Small DS, Morris TP, Li F. On the mixed-model analysis of covariance in cluster-randomized trials. arXiv:2112.00832 2021.
36. Li F, Tian Z, Tian Z, Li F. A note on identification of causal effects in cluster randomized trials with post-randomization selection bias. *Commun Stat – Theory Methods*. 2022;1-13. doi:10.1080/03610926.2022.2116281
37. Forastiere L, Lattarulo P, Mariani M, Mealli F, Razzolini L. Exploring encouragement, treatment, and spillover effects using principal stratification, with application to a field experiment on teens' museum attendance. *J Bus Econ Stat*. 2019:1-15.
38. Jo B. Handling parametric assumptions in principal causal effect estimation using Gaussian mixtures. *Stat Med*. 2022;41:3039-3056.

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

39. Wei GCG, Tanner MA. A Monte Carlo implementation of the EM algorithm and the Poor Man's data augmentation algorithms. *J Am Stat Assoc*. 1990;85(411):699-704.
40. Field CA, Welsh AH. Bootstrapping clustered data. *J R Stat Soc Ser B*. 2007;69(3):369-390.
41. Huang FL. Using cluster bootstrapping to analyze nested data with a few clusters. *Educ Psychol Meas*. 2018;78(2):297-318.
42. Hedeker D. A mixed-effects multinomial logistic regression model. *Stat Med*. 2003;22(9):1433-1446.
43. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074-2102.
44. Ivers NM, Taljaard M, Dixon S, et al. 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*. 2011;343:1-14.
45. Kahan BC, Forbes G, Ali Y, et al. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. *Trials*. 2016;17(1):1-8.
46. Thompson DM, Fernald DH, Mold JW. Intraclass correlation coefficients typical of cluster-randomized studies: estimates from the Robert Wood Johnson Prescription for Health projects. *Ann Fam Med*. 2012;10(3):235-240.
47. Pinheiro J, Bates D, DebRoy S, Sarkar D. *nlme: Linear and Nonlinear Mixed effects Models. R package version 3.1-153*. Vienna, Austria: R Core Team; 2021, Vienna, Austria.
48. Hendy J, Chrysanthaki T, Barlow J, et al. An organisational analysis of the implementation of telecare and telehealth: the whole systems demonstrator. *BMC Health Serv Res*. 2012;12:403.
49. Hirani SP, Beynon M, Cartwright M, et al. The effect of telecare on the quality of life and psychological well-being of elderly recipients of social care over a 12-month period: the whole systems demonstrator cluster randomised trial. *Age Ageing*. 2013;43(3):334-341.
50. Ware J, Kosinski M, Turner-Bowker D, Gandeck B. *User's Manual for the SF-12v2 Health Survey (with a Supplement Documenting SF-12 Health Survey)*. Lincoln, RI: Quality Metric Incorporated; 2002.
51. Mercatanti A, Li F, Mealli F. Improving inference of Gaussian mixtures using auxiliary variables. *Stat Anal Data Mining: ASA Data Sci Jl*. 2015;8(1):34-48.
52. Wang X, Turner EL, Li F, et al. Two weights make a wrong: cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemp Clin Trials*. 2022;114:106702.


## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wang W, Tong G, Hirani SP, et al. A mixed model approach to estimate the survivor average causal effect in cluster-randomized trials. *Statistics in Medicine*. 2023;1-18. doi: 10.1002/sim.9939

10970258, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/sim.9939 by City, University Of London, Wiley Online Library on [04/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License