American Journal of Epidemiology **Submitted Manuscript**

Title: A Bayesian approach for estimating the survivor average causal effect when outcomes are truncated by death in cluster-randomized trials

Authors: Guangyu Tong, Fan Li, Xinyuan Chen, Shashivadan P. Hirani, Stanton P. Newman, Wei Wang, and Michael O. Harhay

Correspondence Address: Correspondence to Dr. Guangyu Tong, Department of Biostatistics, Yale School of Public Health, Room 234, 135 College Street, New Haven, Connecticut 06510 (email: guangyu.tong@yale.edu)

Affiliations: Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, United States (Guangyu Tong, Fan Li); Center for Methods in Implementation and Prevention Science, Yale School of Public Health, New Haven, Connecticut, United States (Guangyu Tong , Fan Li); Department of Mathematics and Statistics, Mississippi State University, Mississippi State, Mississippi, United States (Xinyuan Chen); School of Health Sciences, City University London, London, United Kingdom (Shashivadan P. Hirani, Stanton P. Newman); Clinical Trials Methods and Outcomes Lab, Palliative and Advanced Illness Research (PAIR) Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States (Wei Wang, Michael O. Harhay); Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States (Wei Wang, Michael O. Harhay).

Funding: Research in this article was partially supported by the Patient-Centered Outcomes Research (PCORI, awards ME-2020C1-19220 to M.O.H. and ME-2020C3-21072 to F.L). M.O.H. is also funded by the United States National Institutes of Health (NIH), National Heart, Lung, and Blood Institute (grant number R00-HL141678). The WSD study was funded by the Policy Research Programme in the Department of Health of the United Kingdom.

Data Availability Statement: The WSD study is available upon reasonable request from Drs. Hirani and Newman. All code for the replication of the simulation study is available at https://github.com/ttyale/PSM_CRT.

Thanks: N/A

Conference Presentation: N/A

Disclaimer: 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, its Board of Governors or Methodology Committee, or the Department of Health of the United Kingdom.

Conflict of Interest: N/A

Running Head: Survivor average causal effect in cluster trials

Key words: Cluster-randomized trials, survivor average causal effect, counterfactual outcomes, always survivors, death truncation, quality of life, Bayesian estimation.

Abbreviations: SACE, survivor average causal effect

CRT, cluster-randomized trial

LMM, linear mixed model

MCMC, Monte Carlo Markov Chain

ABSTRACT

Many studies encounter clustering due to multicenter enrollment and non-mortality outcomes, such as quality-of-life, that are truncated due to death; i.e., missing not at random and nonignorable. Traditional missing data methods and target causal estimands are suboptimal for statistical inference in the presence of these combined issues, which are especially common in multicenter studies and cluster-randomized trials (CRTs) among the elderly or seriously ill. Using principal stratification, we developed a Bayesian estimator that jointly identifies the always-survivor principal stratum in a clustered/hierarchical data setting and estimates the average treatment effect among them (i.e., the *survivor average causal effect*, SACE). In simulations, we observed low bias and good coverage with our method. In a motivating CRT, the SACE and the estimate from complete case analysis differed in magnitude, but both were small, and neither was incompatible with a null effect. However, the SACE estimate has a clear causal interpretation. The option to assess the rigorously defined SACE estimand in studies with informative truncation and clustering can provide additional insight into an important subset of study participants. Based on the simulation study and CRT reanalysis, we provide practical recommendations for using the SACE in CRTs and code to support future research.

INTRODUCTION

Outcomes such as quality of life are frequently measured non-mortality outcomes used to assess general health, recovery, and the impact of medical interventions.[1–3] In studies with non-trivial mortality, such as among the elderly, or those with critical and serious illnesses, patient-centered outcome measures often cannot be captured for a sizeable proportion of participants who die during the study period.[4,5] Non-mortality outcomes missing due to death raises conceptual and empirical issues. First, death itself is an outcome of interest. Second, as the non-mortality outcome is empirically unmeasured (i.e., undefined) among those who die, imputation approaches may not appeal to certain stakeholders.[6–8] Relatedly, composite outcomes require that some subjective valuation be used concerning what value should be used for those who die.[9,10] Further, from a causal inference perspective, many strategies may not provide estimates with a clear causal interpretation under the counterfactual outcomes framework[11–13] because surviving study participants under treatment and control can be systematically different, which obscures the target estimand.

Adapted from Suzuki,[14] Figure 1 portrays what is often termed the "truncation-by-death" problem. In this setting, the survivor status (S) and quality of life outcome (Y) of an individual can be influenced by the treatment (D) as well as an unobserved variable (U). The unobserved variable (U) can also be potentially associated with the treatment (D), but would be less likely if the treatment (D) is assigned in a randomized trial setting. Under the counterfactual outcomes framework, there is considerable literature on estimating causal effects using principal stratification.[15–19] This framework has been applied to study non-mortality outcomes in individually randomized trials,[4,18] as well as to address protocol compliance, patient encouragement, and other problems in public health and social science settings (e.g., the effect of job training programs and behavioral health interventions).[16,17,20,21] Principal stratification seeks to identify strata of patients by their pre-exposure characteristics. The stratum of primary interest includes those who would always survive through the end of the trial period regardless of treatment assignment. The effect of an intervention in this stratum of "always-survivors" is termed the survivor average causal effect (SACE).

The SACE is a meaningful estimand because, without additional assumptions, only the always-survivors have both counterfactual outcomes well-defined. Thus, the SACE avoids survivor bias (i.e., observed and unobserved characteristics of survivors in treatment and control are likely different) and summarizes the treatment effect without needing additional assumptions on the counterfactual outcomes for the non-survivors.[5,10,22,23] However, the always-survivors target population is not directly observed; i.e., not all participants have a definitive strata membership, but principal strata must be identified to estimate the SACE. Bayesian inference is particularly attractive for this purpose as the posterior strata membership can be updated through a Monte Carlo Markov Chain (MCMC) algorithm, while posterior predictive distributions for the SACE can be conditional on the strata membership.[15,16]

Herein, we extend and use principal stratification to estimate the SACE using Bayesian inference in cluster-randomized trials (CRTs) with death truncation. In contrast to individual-level randomization, CRTs randomize groups of individuals to treatments.[24,25] Cluster-level randomization is used when the intervention is designed for a system-level improvement or when randomization is not feasible at the individual level (see Turner and colleauges[25] for a review). As a result, the outcome observations are often more similar within clusters than between clusters, causing a positive intracluster correlation that must be accounted for in the analysis stage to avoid inflated type I error. To estimate the SACE in CRTs, we developed a Bayesian approach that leverages the baseline covariates to predict the counterfactual survivor status and the counterfactual outcomes. As we explicate in the methods, our approach includes both a principal stratification model and an outcome model, for which we developed an iterative sampling algorithm to estimate the model parameters jointly, and hence the SACE in CRTs.

METHODS

Causal framework and assumptions

We consider the counterfactual outcome framework, where the causal effect is defined as the difference between the two counterfactual outcomes averaged across a common population.[12,13] We assume (1) the

Stable Unit Treatment Value Assumption (SUTVA) by which patients receive no different forms or versions of treatments and that no interference exists and (2) that the treatment is randomized at the cluster level and is independent of both the potential survivor status and counterfactual outcome of all individuals in each cluster. In a CRT with I clusters and n_i individuals in each cluster, denote the cluster-level treatment and control assignment as $D_i = 1,0$ respectively, where $i = 1, \dots, I$. For the j th individual ($j = 1, \dots, n_j$) in the i th cluster, we define $\{Y_{ij}(1), Y_{ij}(0)\}$ as the counterfactual outcomes for each individual under treatment and control. We are typically interested in the average causal effect, $\delta = E(Y_{ij}(1) - Y_{ij}(0))$ if the counterfactual outcomes are well defined for the entire trial population. However, when outcomes are truncated, the average causal effect can only be defined for a subset of patients. Estimating the average causal effect among a meaningful subgroup in this setting involves (1) using the covariates and survival status to identify the potentially unobserved always-survivor stratum, and (2), comparing counterfactual outcomes within the always-survivor stratum under treatment and control. We now describe models for each of these components.

Principal strata model for survivor status

We define S_{ij} as the observed survival status of a patient with $S_{ij} = 1$ indicating survival, and $S_{ij} = 0$ indicating death. Under the potential outcomes framework, the joint values of potential survival status produce four strata. These include:

1. *always-survivors* ($S_{ij}(1) = S_{ij}(0) = 1$): patients who will survive until the end of the study regardless of the treatment status.
2. *protected* ($S_{ij}(1) = 1, S_{ij}(0) = 0$): patients who will survive only under treatment.
3. *harmed* ($S_{ij}(1) = 0, S_{ij}(0) = 1$): patients who will survive only under control.
4. *never-survivors* ($S_{ij}(1) = S_{ij}(0) = 0$): patients who will not survive regardless of treatment.

We make an additional assumption of *monotonicity* such that the treatment does not lead to worse survival, and, thus, the harmed strata is assumed away.[17,26] Monotonicity is a plausible assumption in trials when interventions are carefully piloted for safety considerations. However, we acknowledge that an intervention can lead to worse survival in some patients, and in such instances, the monotonicity assumption will be questionable. In scenarios where the proportion of harmed strata is close to zero, modeling the harmed strata can also lead to computational challenges, in which case the monotonicity assumption represents a practical consideration for model fitting. In our motivating trial, we considered the harmed patients to be rare (if not non-existent) and undertook the analysis assuming monotonicity. Potential approaches to relaxing this assumption are discussed later. Under monotonicity, the principal strata membership is observed for survivors in the control group and the deceased in the treatment group, but is unknown for survivors in the treatment group or the deceased in the control group. Table 1 presents the strata membership attribution for each observed data group under monotonicity. Baseline covariates play a critical role in identifying the principal strata. Suppose $G_{ij} = \{00,10,11\}$ indicates principal strata membership, where $G_{ij} = 00$ for never-survivors, $G_{ij} = 10$ indicates protected individuals, and $G_{ij} = 11$ indicates always-survivors. Assuming X_{ij} as a covariate vector with both cluster-level and individual-level covariates and further including an intercept, the principal strata can be modeled using multinomial logistic regression with $G_{ij} = 11$ as the reference group,

$$P(G_{ij} = 00) = \frac{e^{\boldsymbol{X}_{ij}^T\boldsymbol{\beta}}}{1 + e^{\boldsymbol{X}_{ij}^T\boldsymbol{\beta}} + e^{\boldsymbol{X}_{ij}^T\boldsymbol{\gamma}}}, \quad P(G_{ij} = 10) = \frac{e^{\boldsymbol{X}_{ij}^T\boldsymbol{\gamma}}}{1 + e^{\boldsymbol{X}_{ij}^T\boldsymbol{\beta}} + e^{\boldsymbol{X}_{ij}^T\boldsymbol{\gamma}}},$$

$$P(G_{ij} = 11) = \frac{1}{1 + e^{\boldsymbol{X}_{ij}^T\boldsymbol{\beta}} + e^{\boldsymbol{X}_{ij}^T\boldsymbol{\gamma}}}$$

Here, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are p-dimensional regression coefficient vectors for never survivors versus always survivors and for the protected versus always survivors, respectively; each component of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is interpreted as the log odds ratio. To ensure numerical stability, we recommend choosing the (likely) largest strata as the reference category for the multinomial logistic model.

In addition, cluster-level random intercepts can be added when principal strata membership is believed to be correlated due to cluster randomization. Alternatively, a nested Probit model for the strata membership can be used,[16] however, the regression coefficients are more challenging to interpret. Thus, we did not pursue the nested Probit model further but derived its posteriors with a latent variable specification to support its use by interested readers (*Web Appendix 1*).

Models for potential outcomes

After defining the principal strata model, we specify counterfactual outcome models within each principal stratum. Specifically, only the always-survivors have well-defined counterfactual outcomes under both treatment and control ($Y_{ij}(0)$, $Y_{ij}(1)$) since they are not subject to truncation. Patients in protected strata have only one well-defined counterfactual outcome under treatment but their counterfactual outcome under the control condition is truncated by death ($Y_{ij}(0) = *$, $Y_{ij}(1)$). Among the always-survivors, each counterfactual outcome can be modeled as a function of covariates adjusting for clustering, whereas only one counterfactual outcome for each protected patient can be similarly modeled.

Now, let $\boldsymbol{\alpha}_1^{11}$, $\boldsymbol{\alpha}_0^{11}$, and $\boldsymbol{\alpha}_1^{10}$ be p-dimensional vectors of regression coefficients for covariates in these three groups: always-survivors in the treatment ($G_{ij} = 11$, $D_i = 1$), always-survivors in the control ($G_{ij} = 11$, $D_i = 0$), and protected individuals in the treatment ($G_{ij} = 10$, $D_i = 1$). Let Y_{ij} be the outcomes where $\{i, j \in S_{ij}(D) = 1\}$. Let η_i be the cluster-level random effects following a normal distribution with mean zero and variance of τ^2; let the residual errors follow a normal distribution with a mean equal to zero and variance of σ^2. The assumed linear mixed models (LMMs) for the counterfactual outcomes can then be summarized as in Table 2. The random-effects term η_i is required here to account for the intra-cluster correlation coefficient (ICC), $\rho = \frac{\tau^2}{\tau^2+\sigma^2}$, a quantity that is central to the design and analysis of CRTs, as ignoring the ICC in the outcome model leads to an inflated type I error rate.[27] With the outcome model specified for always-survivors, the SACE can be defined as,

$$\delta = E(Y_{ij}(1) - Y_{ij}(0)|G_{ij} = 11)$$

We leverage the mixture modeling assumptions and the monotonicity to jointly estimate the strata membership probability and the counterfactual outcome model parameters for each patient. This approach produces the treatment effect for each always survivor, which can be averaged over the always survivor subpopulation to identify the SACE. Beyond the mixture modeling approach, we acknowledge that other structural assumptions can be used to identify SACE even in the absence of monotonicity; see, for example, Hayden et al.[28] and Shepherd et al.[29]

Joint inference of the outcome model and the principal strata model

Posterior inference of the parameters in the principal strata model and the outcome model can be achieved through a MCMC algorithm. The algorithm is summarized in the *Appendix*, and detailed derivations for each step are in the *Web Appendix 1*. The algorithm uses Gibbs sampling steps to update outcome regression model parameters, where conjugate priors of normal and inverse gamma distributions are specified (details in *Web Appendix 1*). The algorithm further implements a Metropolis-Hastings step for the principal stratification model. To increase the convergence speed for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, we use a random-walk Metropolis algorithm[30] that draws proposals from multivariate *t* distributions, $t(s_\beta \boldsymbol{T}_\beta)$ and $t(s_\gamma \boldsymbol{T}_\gamma)$, that center at the values of the previous iteration. The parameters s_β and s_γ scale the covariance to achieve optimal acceptance rates, and both $\boldsymbol{T}_\beta$and $\boldsymbol{T}_\gamma$ is a $p \times p$-dimensional component-specific scale matrix. We use the adaptive proposal approach by Haario and colleagues[31] to tune $\boldsymbol{T}_\beta$ and $\boldsymbol{T}_\gamma$ by utilizing empirical covariance from an extended burn-in. As indicated by asterisks "*" in the *Appendix*, when the principal strata model also accounts for clustering (denoted as χ_i for the random intercept, where $\chi_i \sim N(0, \phi^2)$, it can be updated using the same approach as for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Simulation study

We conducted a simulation study to validate our algorithm. Specifically, we simulated a two-arm CRT with 1,500 individuals with varying cluster size m and number of clusters n as $(m, n) = \{(50,30), (25,60), (15,100)\}$. We simulated two continuous covariates following $X_{ij1} \sim N(0,4)$ and $X_{ij2} \sim Unif(-5,5)$, respectively. For the principal stratification model, we let $\boldsymbol{\beta} = \{-1,0.3,0.5\}$, and $\boldsymbol{\gamma} = \{-0.8,0.6,0.4\}$ so that the stratum proportions are 21.1% for never-survivors, 26.5% for protected, and 52.4% for always-survivors. We generated the potential outcomes following Table 2, where we set $\boldsymbol{\alpha}_1^{11} = \{1.5, 0.5, 0.8\}$,$\boldsymbol{\alpha}_0^{11} = \{0.2, 0.3, 0.6\}$, and $\boldsymbol{\alpha}_1^{10} = \{-1.5, 0.9, 0.5\}$. We set $\sigma^2 = 5$ and $\tau^2 = 1$ with an induced ICC of 0.167, which falls within the commonly reported range of 0-0.2.[32,33] For each simulated data set under each combination of (m, n), we implemented the Bayesian method with 10,000 MCMC iterations (with the first 2,500 iterations as burn-in). In addition, to evaluate the impact of varying cluster sizes, we considered scenarios with the mean cluster sizes and number of clusters as $(\bar{m}, n) = \{(50,30), (25,60), (15,100)\}$ and generated the data with a large coefficient of variability (defined as $\text{CV} = \bar{m}/\sqrt{var(m)}$) of 1.0. Additional simulations with smaller outcome ICCs (0.01 or 0.05), smaller numbers of clusters and/or cluster sizes, and when the principal strata models include random intercepts (with induced ICC of 0.05 or 0.10 on the latent response scale[34]) were also performed. We calculated the average posteriors means, relative bias, and coverage across 200 simulated data sets. All analyses were performed using R 4.0.1. R code for the simulation, including the data-generating process and the MCMC sampler, is available online.

Analysis of the motivating trial example

We applied our methodology to analyze the Whole Systems Demonstrator (WSD) Telecare Questionnaire Study.[35,36] The WSD study was a CRT randomized at the general practice (GP) level that evaluated the effect of telecare on the health-related quality of life and psychological well-being of 1,189 elderly recipients of social care in the United Kingdom over 12-months: 639 participants were randomized at the

cluster level into the telecare (TC) arm and 550 to usual care (UC). Recipients were additionally clustered within GPs across three English Local Authorities. The rationale of the TC intervention is not only its potential health benefit, but also its advantage in cost-effectiveness.[37] There were a total of 204 GPs, and the cluster size varied from 1 to 26. The TC arm installed electronic sensors in the home of recipients that provided safety monitoring (e.g., falls of recipients, fires at home). The UC arm did not receive TC. Our illustration focuses on the health-related quality of life measured by the EQ-5D-VAS index score, a self-rated scale (range: 0-100) on five domains of mobility, self-care, usual activities, pain and discomfort, and anxiety and depression, 12 months post-randomization.[38,39] Higher scores represent better overall quality of life. The sample size, cluster number, and results reported herein vary slightly from the original published analysis due to different analytic methods and outcome data.

We considered trial participants as non-survivors if they were deceased or had seriously deteriorated health such that a self-reported health outcome for them could not be measured or collected, and thus undefined. Recipients with seriously deteriorated health included those who were too ill, unable to continue due to dementia or deteriorated mental capacity, moved to long-term nursing care, residential care, sheltered housing, or family caregiver. For missing data in the baseline covariates and outcomes not due to death or seriously deteriorated health, we used multiple imputation[6,13] to impute a single dataset to fill those missing entries. Note that a fully Bayesian approach that incorporates the imputation in the proposed algorithm can also be implemented. Since the goal of our illustration is not focused on the missing covariate problem, we did not pursue this direction. The resulting dataset had 127 (10.7%) cases with truncated outcomes.

For both the principal stratification model and the potential outcome models, our baseline covariates included gender, age, ethnicity, participants' highest level of education, an indicator for only-adult household, number of comorbidities, impairment score, physical health score, mental health score, and EQ-5D-VAS index score.[35] We used LMMs for the outcome regression model and multinomial logistic

regression for the principal strata model, as previously specified. We did not pursue the more complex model that accounted for the random effects in the principal strata model because the average cluster size was too small, and a handful of clusters had a size of one, which would cause convergence issues (we discuss this further in the discussion). Two MCMC chains of 100,000 iterations were implemented where the first 25,000 iterations were set as burn-in. We started each chain using random initials. Model convergence and chain mixing were checked by traceplots. All analyses were performed using R 4.0.1. In addition, as a comparison model that might be frequently used in practice in the absence of our method, we estimated a LMM based on complete outcomes adjusting for the same baseline covariates.

RESULTS

Simulation study

Table 3 presents the simulation results of the key model parameters for the scenario of $(m, n) = (15,100)$. Relative bias and coverage were presented for $\boldsymbol{\alpha_1^{11}}$, $\boldsymbol{\alpha_0^{11}}$, the SACE, and the ICC. Our results show that the posterior means for most parameters are accurate with less than 10% bias and above 90% coverage. In particular, the SACE estimate is close to the truth (%Bias = -5.2%), and the coverage probability was 0.93. The results for the other two scenarios of $(m, n) = \{(50,30), (25,60)\}$ are similar where SACE estimates both have less than 5.0% percent bias and ≥ 0.95 coverage (see *Web Tables 1-3* for full results). Additional simulations with variable cluster sizes (*Web Tables 4-6*), principal strata model specification with random effects (*Web Tables 7-10*), a small number of clusters and/or cluster sizes (*Web Tables 11-12*) and variable outcome ICCs (*Web Tables 13-14*) showed similar performance.

Illustrative analysis

As noted, our approach allows for adjustment for covariates and clustering. As is common in CRTs, several prognostic variables were preselected for adjustment in the primary analysis; the variables we used in our analysis are summarized in Table 4. Our model identified 88.8% of recipients as always-survivors, 2.2% as protected, and 8.9% as never-survivors. The posterior mean of the ICC was 0.002 with

a credible interval (i.e., 2.5% and 97.5% of the posterior sample) of [0.000,0.020], suggesting a small intra-cluster correlation in EQ-5D-VAS.

Table 5 summarizes the analytical results of our analysis. The SACE point estimate for the effect of TC on the EQ-5D-VAS was -0.70, with a credible interval spanning potential effects that ranged from a decrease of -3.11 to an increase of 0.83. The estimated effect suggests that TC did not markedly improve the EQ-5D-VAS compared to UC among the principal strata of always-survivors.

Web Figure 1 depicts the average number of always-survivors, protected, and never-survivors by cluster based on the posterior sample of strata membership after burn-in. Many clusters have recipients possibly from all three strata. A notable advantage of our Bayesian approach is that the baseline characteristics of always-survivors can be obtained by averaging over the baseline covariates among always-survivors in the posterior sample of principal strata membership. This summary is provided in *Web Table 15*. In our illustration, the demographics, socioeconomic status, and baseline health status among the always survivors are similar to the overall population. This is unsurprising in this specific illustration as 88.8% of recipients were identified as always-survivors.

Finally, in Table 5, the LMM point estimate and 95% confidence interval based on the recipients with observed outcomes was -0.93 (95% confidence interval: -3.24 to 1.38). This result is in line (again due to overlap in this specific illustration, but this is not guaranteed), with the SACE. However, it is important to note that while similar in our illustration, the LMM estimate does not have a causal interpretation as the analytic sample for the LMM estimate included recipients deemed belonging to the protected stratum. Intuitively, it suggests that the protected recipients in the TC arm who were likely to die tended to be those with worse health outcomes in the treatment group.

DISCUSSION

The SACE is a well-defined causal estimand that describes the effect of an intervention among participants who would survive regardless of their randomized assignment in a trial. We used Bayesian principal stratification to estimate the SACE in a CRT where the hierarchical data structure due to clustered randomization is accounted for in the modeling and data analysis. Since strata membership is not fully identifiable for some participants, Bayesian estimation is a particularly attractive strategy to address uncertain strata membership of the counterfactual survivor status. Of note, our approach considers the model-based credible interval estimation under the Bayesian framework, and therefore differs from the usual cluster-robust variance approach under the frequentist framework.[40] The extent to which a cluster-robust variance approach applies to our Bayesian modeling framework merits additional research.

In our simulations, we observed low bias and good coverage of the true SACE parameter with our methods. In our analysis of the WSD Telecare trial, 88.8% of the participants were identified as always-survivors, and the SACE suggested no significant change in health-related quality of life measure at 12 months (-0.70, 95% credible interval: [-3.11, 0.83]). This point estimate was slightly smaller than that estimated from a complete case analysis using a naïve LMM model (-0.93, 95% confidence interval: [-3.24, 1.38]). We note that our result is only based on a one-time measure at 12 months, which is different from the original result published by Hirani and colleauges[35] where they utilized repeated outcome measures at different time points and considered different covariates. Adapting the principal stratification framework for CRTs with repeatedly measured outcomes requires additional methodological development. While our analysis of the WSD trial showed limited effect on the EQ-5D-VAS outcome, telecare may impact other physical health outcomes or have cost-effectiveness properties due to prevention or earlier intervention on health needs.

Practical recommendations

During the development and implementation of our methodology, we identified some analytic considerations that users may need to consider. First, the monotonicity assumption may be plausible for practice-level interventions like the one used in the WSD study setting, where installing electronic sensors in the TC arm was unlikely to harm participants. Before the implementation of many medical trials, interventions are evaluated in pilot studies with safety monitoring; thus, this assumption may often be reasonable. Relaxing the monotonicity assumption by adding the "harmed strata" (i.e., participants that die in treatment but survive in control) is possible. However, fitting the mixture model with an additional harmed strata is an added layer of computational considerations, and additional simulations are needed to fully understand the benefit of including this strata when the harmed population is relatively rare. Beyond the mixture modeling framework, other types of structural assumptions or sensitivity parameters are necessary to relax the monotonicity assumption,[28,29,41,42] and represent a fruitful direction for future investigations in the context of CRTs. Second, our simulation studies show that the use of non-informative priors can achieve adequate performance without sufficient knowledge of key model parameters from existing studies. However, Bayesian approaches have an inherent advantage of leveraging existing knowledge through informative priors on key parameters to sharpen the model performance. For example, Turner and colleauges[43,44] have demonstrated that compared to the non-informative priors, incorporating informative half-normal and beta priors on the outcome ICC parameter (based on published ICC estimates) can produce narrower credible intervals for the outcome ICC and variance components parameters. We anticipate this finding to be applicable for estimating the SACE. Third, while we have provided methods to account for clustering in the principal strata membership model, the model fitting can be substantially more challenging than its counterpart without clustering. In the analysis of the WSD trial, the principal strata membership model failed to converge due to several extremely small clusters. Thus, we did not consider clustering in the principal strata membership model. Therefore, in practice, specifying more complex principal strata membership models often require the absence of extremely small clusters. On the other hand, our additional simulation results (Web Tables 7-

10) have shown that when the ICC in the principal strata model is not exceeding 0.10, specifying the principal strata model without a random intercept can still achieve adequate performance characteristics and can be sufficient. However, we acknowledge that our assessment is limited to the data-generating process we considered, and a more systematic comparison between clustered and un-clustered principal strata models in CRTs is warranted. Lastly, our additional simulation (Web Tables 11-12) shows our method can be employed to estimate SACE in relatively small CRTs (e.g., 20 clusters with cluster size of 25). We caution against using our model in even smaller CRTs, as additional research to address small-sample challenges in these settings is needed.

The use of the SACE in CRTs (and individually-randomized trials) also requires some practical considerations from the design perspective as the always-survivors stratum is a subset of trial participants that cannot be identified until after randomization and trial completion. Though a more attractive target estimand, and thus alternative to composite outcomes, imputation, or other approaches to deal with truncated outcomes, there is always a threat of a loss of power with the SACE due to the inherently smaller sample size of this stratum. However, this concern may be partially offset when the treatment effect is likely to be larger among the always-survivors stratum than the overall population (i.e., average treatment effect).[45,46] Relatedly, without knowing the size or characteristics of the always-survivors stratum prior to a study, power calculations may be a challenge, particularly when sample sizes are constrained and cannot be increased. Thus, we believe we can offer three practical recommendations for using the SACE in a trial. First, the SACE may be most ideal as a pre-planned secondary analysis in trials with smaller available sample sizes, and only considered for the primary analysis in larger pragmatic trials or in trials where effect sizes and always-survivors rates can be anticipated with reasonable certainty to be in some range, and thus available sample sizes are adequate.[45] Second, as is recommended when working with other uncertain trial design elements, we recommend that Monte Carlo simulation studies be undertaken to assess statistical power.[47,48] Jo[45] provides statistical power guidance for trials with treatment noncompliance that is potentially relevant to those using the SACE. Closed-form sample size solutions

for noncompliance in CRTs could be extended to the SACE in future work. Third, when there is interest in conducting primary analysis to estimate the SACE, approaches for sample size re-estimation with pre-planned interim analysis[49] may be considered, but require future development.

REFERENCES

1. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med*. 1996;334(13):835-840.

2. Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*. New York, NY: Chapman and Hall; 2010.

3. Fayers PM, Hays R, Hays RD. *Assessing Quality of Life in Clinical Trials: Methods and Practice*. New York, NY: Oxford University Press; 2005.

4. 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.

5. Colantuoni E, Scharfstein DO, Wang C, et al. Statistical methods to compare functional outcomes in randomized controlled trials with high mortality. *BMJ*. 2018;360.

6. Tong G, Li F, Allen AS. Missing Data. In: Piantadosi S, Meinert C, eds. *Principles and Practice of Clinical Trials*. New York, NY: Springer, Cham; 2020:1-21.

7. Carreras G, Miccinesi G, Wilcock A, et al. Missing not at random in end of life care studies: multiple imputation and sensitivity analysis on data from the ACTION study. *BMC Med Res Methodol*. 2021;21(1):1-12.

8. Fielding S, Fayers PM, McDonald A, McPherson G, Campbell MK. Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data. *Health Qual Life Outcomes*. 2008;6(1):1-9.

9. Harhay MO, Ratcliffe SJ, Small DS, Suttner LH, Crowther MJ, Halpern SD. Measuring and analyzing length of stay in critical care trials. *Med Care*. 2019;57(9):e53.

10. Lin W, Halpern SD, Prasad Kerlin M, Small DS. A "placement of death" approach for studies of treatment effects on ICU length of stay. *Stat Methods Med Res*. 2017;26(1):292-311.

11. 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.

12. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688.

13. Rubin DB. Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In: *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Vol. 1. Alexandria VA: American Statistical Association; 1978: 20-34.

14. Suzuki E. Generalized Causal Measure The Beauty Lies in Its Generality: The authors respond. *Epidemiology*. 2015;26(4):490-495.

15. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Stat*. 1997; 25(1):305-327.

16. 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.

17. 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.

18. Yang F, Small DS. Using post-outcome measurement information in censoring-by-death problems. *J R Stat Soc Ser B Stat Methodol*. 2016;78(1):299-318.

19. McGuinness MB, Kasza J, Karahalios A, et al. A comparison of methods to estimate the survivor average causal effect in the presence of missing data: a simulation study. *BMC Med Res Methodol*. 2019;19(1):1-14.

20. 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.

21. Mattei A, Li F, Mealli F. Exploiting multiple outcomes in Bayesian principal stratification analysis with application to the evaluation of a job training program. *Ann Appl Stat*. 2013;7(4):2336-2360.

22. Wang C, Scharfstein DO, Colantuoni E, et al. Inference in randomized trials with death and missingness. *Biometrics*. 2017;73(2):431-440.

23. Rosenbaum PR. Comment: the place of death in the quality of life. *Stat Sci*. 2006;21(3):313-316.

24. Murray DM, Taljaard M, Turner EL, George SM. Essential Ingredients and Innovations in the Design and Analysis of Group-Randomized Trials. *Annu Rev Public Health*. 2020;41:1-19.

25. Turner EL, Li F, Gallis JA, et al. Review of recent methodological developments in group-randomized trials: part 1—design. *Am J Public Health*. 2017;107(6):907-915.

26. 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.

27. Tong G, Seal KH, Becker WC, et al. Impact of complex, partially nested clustering in a three-arm individually randomized group treatment trial: A case study with the wHOPE trial. *Clin Trials*. 2022;19(1):3-13.

28. Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*. 2005;61(1):305-310.

29. Shepherd BE, Redman MW, Ankerst DP. Does finasteride affect the severity of prostate cancer? A causal sensitivity analysis. *J Am Stat Assoc*. 2008;103(484):1392-1404.

30. Neelon B, Gelfand AE, Miranda ML. A multivariate spatial mixture model for areal data: examining regional differences in standardized test scores. *J R Stat Soc Ser C Appl Stat*. 2014;63(5):737-761.

31. Haario H, Saksman E, Tamminen J. Componentwise adaptation for high dimensional MCMC. *Comput Stat*. 2005;20(2):265-273.

32. Adams G, Gulliford MC, Ukoumunne OC, et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol*. 2004;57(8):785-794.

33. Eldridge SM, Ashby D, Feder GS, et al. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials*. 2004;1(1):80-90.

34. Hedeker D. A mixed-effects multinomial logistic regression model. *Stat Med*. 2003;22(9):1433-1446.

35. 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*. 2014;43(3):334-341.

36. Henderson C, Knapp M, Fernández JL, et al. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. *BMJ*. 2013;346.

37. Bower P, Cartwright M, Hirani SP, et al. A comprehensive evaluation of the impact of telemonitoring in patients with long-term conditions and social care needs: protocol for the whole systems demonstrator cluster randomised trial. *BMC Health Serv Res*. 2011;11(1):1-12.

38. Ware JE Jr, Kosinski M, Turner-Bowker DM, Sundaram M, Gandek B, Maruish ME. User's manual for the SF-12v2 Health Survey. 2nd edition. Lincoln, RI: *QualityMetric Inc*.; 2009.

39. Feng Y, Parkin D, Devlin NJ. Assessing the performance of the EQ-VAS in the NHS PROMs programme. *Qual Life Res*. 2014;23(3):977-989.

40. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.

41. 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.

42. Lou Y, Jones MP, Sun W. Estimation of causal effects in clinical endpoint bioequivalence studies in the presence of intercurrent events: noncompliance and missing data. *J Biopharm Stat*. 2019;29(1):151-173.

43. Turner RM, Omar RZ, Thompson SG. Constructing intervals for the intracluster correlation coefficient using Bayesian modelling, and application in cluster randomized trials. *Stat Med*. 2006;25(9):1443-1456.

44. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Stat Med*. 2001;20(3):453-472.

45. Jo B. Statistical power in randomized intervention studies with noncompliance. *Psychol Methods*. 2002;7(2):178.

46. Moerbeek M, Schie S van. What are the statistical implications of treatment non-compliance in cluster randomized trials: a simulation study. *Stat Med*. 2019;38(26):5071-5084.

47. Landau S, Stahl D. Sample size and power calculations for medical studies by simulation when closed form expressions are not available. *Stat Methods Med Res*. 2013;22(3):324-345.

48. Shi Y, Lee JH. Sample size calculations for group randomized trials with unequal group sizes through Monte Carlo simulations. *Stat Methods Med Res*. 2018;27(9):2569-2580.

49. van Schie S, Moerbeek M. Re-estimating sample size in cluster randomised trials with active recruitment within clusters. *Stat Med*. 2014;33(19):3253-3268.

Table 1. Data elements and principal strata membership based on observed survivor status and treatment status under the monotonicity assumption.

Observed group	Observed treatment status	Observed survival status	Observed outcome[a]	Unobserved outcome	(Possible) strata membership under monotonicity
$D_i = 1, S_{ij} = 1$	Yes	Yes	$Y_{ij} = Y_{ij}(1)$	$Y_{ij}(0)$	Always-survivor or Protected
$D_i = 1, S_{ij} = 0$	Yes	No	$Y_{ij} = Y_{ij}(1) = *$	$Y_{ij}(0)$	Never-survivor
$D_i = 0, S_{ij} = 1$	No	Yes	$Y_{ij} = Y_{ij}(0)$	$Y_{ij}(1)$	Always-survivor
$D_i = 0, S_{ij} = 0$	No	No	$Y_{ij} = Y_{ij}(0) = *$	$Y_{ij}(1)$	Protected or Never-survivor

[a]: '*' indicates truncation by death.

Table 2. Outcome model by principal strata and counterfactual of treatment status.

Principal Strata	Counterfactual in treatment ($D_i = 1$)	Counterfactual in Control ($D_i = 0$)
Always-survivors ($G_{ij} = 11$)	$Y_{ij}(1) = N\left(\boldsymbol{X}_{ij}^T \boldsymbol{\alpha}_1^{11} + \eta_i, \sigma^2\right)$	$Y_{ij}(0) = N(\boldsymbol{X}_{ij}^T \boldsymbol{\alpha}_0^{11} + \eta_i, \sigma^2)$
Protected ($G_{ij} = 10$)	$Y_{ij}(1) = N\left(\boldsymbol{X}_{ij}^T \boldsymbol{\alpha}_1^{10} + \eta_i, \sigma^2\right)$	
Never-survivors ($G_{ij} = 00$)		

Table 3. Bias in posterior means and coverage for $\boldsymbol{\alpha_1^{11}}$, $\boldsymbol{\alpha_0^{11}}$, SACE, and ICC for the scenario of $(m,n) = (15,100)$ over 200 MCMC sampler simulations each with 10,000 iterations and 2,500 burnin (see full results in *Web Appendix 2*).

Parameters &true values	Posterior mean	%Bias	Coverage
$\boldsymbol{\alpha_1^{11}} = \begin{pmatrix} 1.5 \\ 0.5 \\ 0.8 \end{pmatrix}$	1.28	-14.4	0.93
	0.46	-7.5	0.95
	0.77	-3.8	0.95
$\boldsymbol{\alpha_0^{11}} = \begin{pmatrix} -1.5 \\ 0.9 \\ 0.5 \end{pmatrix}$	-1.5	0.1	0.99
	0.9	-0.1	0.97
	0.5	-0.1	0.96
ICC = 0.17	0.17	0.0	0.96
SACE = 2.85	2.7	-5.2	0.93

Abbreviation: ICC, intracluster correlation coefficient; SACE: survivor average causal effect.

Table 4. Descriptive statistics for baseline covariates of 1,189 participants in the Whole Systems Demonstrator Telecare Questionnaire Study. Mean (SD) is presented for continuous variables and n (%) for categorical variables. Due to the use of a different imputation method, these numbers vary slightly from the original trial publication.

Covariates	Intervention		Control		Total	
	No.	%	No.	%	No.	%
Gender						
Female	205	37.3	219	34.3	424	35.7
Male	345	62.7	420	65.7	765	64.3
Age[a]	73.9 (14.3)		74.3 (13.6)		74.1 (13.9)	
Ethnicity						
White	485	88.2	568	88.9	1053	88.6
Non-white	65	11.8	71	11.1	136	11.4
Highest level of education						
No formal education	359	65.3	421	65.9	780	65.6
GCSE/ O'levels	92	16.7	132	20.7	224	18.8
A'levels/HNC	29	5.3	42	6.6	71	6.0
University level	26	4.7	16	2.5	42	3.5
Grad or professional	44	8.0	28	4.4	72	6.1
Only adult household						
Yes	286	52.0	344	53.8	630	53.0
No	264	48.0	295	46.2	559	47.0
Number of comorbidities[a]	1.1 (1.5)		1.1 (1.4)		1.1 (1.5)	
Impairment score [a]	27.7 (14.3)		28.6 (15.6)		28.2 (15.1)	
Physical health score[a]	28.3 (8.7)		27.9 (8.5)		28.1 (8.6)	
Mental health score[a]	33.0 (8.0)		33.1 (7.8)		33.1 (7.9)	
EQ-5D-VAS index score[a]	52.7 (22.0)		53.2 (22.0)		53.0 (22.0)	

Abbreviation: GCSE, general certificate of secondary education; HNC, higher national certificate.
[a] Values are expressed as mean (standard deviation).

Table 5. Results of the survivor average causal effect estimate and the proportion of recipients in each principal strata with the Bayesian joint modeling and linear mixed-effects model estimates.

Posteriors	Point estimate	95% Credible/Confidence interval
SACE	-0.70	-3.11, 0.83
$\bar{Y}(1)$	53.04	51.33, 54.74
$\bar{Y}(0)$	53.74	51.88, 55.03
Proportion of never-survivors	0.09	0.07, 0.11
Proportion of protected	0.02	0.00, 0.05
Proportion of always-survivors	0.89	0.88, 0.89
ICC	0.003	0.000, 0.020
Linear mixed-effects model	-0.93	-3.24, 1.38

Abbreviation: ICC, intracluster correlation coefficient; SACE: survivor average causal effect.

Appendix. Pseudo algorithm for the joint modeling of outcome and principal strata membership.

Pseudo-algorithm:

1. **Input:** Set multivariate normal priors for regression coefficients, $\boldsymbol{\alpha}_1^{11}$, $\boldsymbol{\alpha}_0^{11}$, $\boldsymbol{\alpha}_0^{10}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and inverse Gamma priors for σ^2, τ^2 (and ϕ^2).
2. Set random initials for all parameters.
3. **For S iterations, do:**
4. Sample $\boldsymbol{\alpha}_1^{11}$, $\boldsymbol{\alpha}_0^{11}$ and $\boldsymbol{\alpha}_0^{10}$ from multivariate normal posteriors.
5. For each cluster, sample η_i from a normal posterior.
6. Sample τ^2 from inverse Gamma posterior.
7. Sample σ^2 from inverse Gamma posterior.
8. Sample $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ jointly as follows:
    a. Draw candidates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from a proposal distribution (mv-t dist.).
    b. Compute rejection ratio, α based on likelihood and prior information.
    c. Sample a random number κ from a $Bernoulli(\alpha)$.
    d. If $\kappa = 1$ then accept the candidates; If $\kappa = 0$ then reject the candidates.
9. *For each cluster, sample χ_i using the rejection sampling approach.
10. *Sample ϕ^2 from inverse Gamma posterior.
11. Update the principal strata membership G_{ij} following the Bayes rule.
12. Estimate SACE among always-survivors ($G_{ij} = 11$).
13. **Output:** SACE

Figure 1. Conceptual diagram for the outcome truncated by death problem. (D: treatment; S: survivor status; Y: outcome; U: unobserved variables; solid lines with arrow: directed effect; dashed line: potential association). This figure is adapted from Reference 14, Suzuki E. Generalized Causal Measure The Beauty Lies in Its Generality: The authors respond. *Epidemiology*. 2015;26(4):490-495.