

Performance of Principal Stratification Method Adjusting for Treatment Noncompliance in Two Arms of a Randomized Trial

Lang'o Odondi¹

¹ ConDuCT Hub for Trials Methodology Research, School of Social and Community Medicine, University of Bristol, 39 Whatley Road, Bristol, UK

Correspondence: Lang'o Odondi, ConDuCT Hub for Trials Methodology Research, School of Social and Community Medicine, University of Bristol, 39 Whatley Road, Bristol BS8 2PS, UK. E-mail: Lango.Odondi@bristol.ac.uk

Received: September 11, 2013 Accepted: October 4, 2013 Online Published: October 22, 2013

doi:10.5539/ijsp.v2n4p93 URL: <http://dx.doi.org/10.5539/ijsp.v2n4p93>

Abstract

The method of principal stratification is a unifying framework for modelling cause and effect which is applicable to adjusting for treatment noncompliance in multiple arms of a trial. Baseline covariates which predict compliance with treatment are useful in addressing parameter identification problem associated with principal stratification. Roy, Hogan and Marcus (RHM) (2008) proposed a principal stratification framework in which they used baseline covariates to adjust for imperfect compliance in both arms of a two-active treatments trial. Key to the application of this method is a defining but untestable distributional assumption whose robustness is unknown. The present work uses statistically designed simulation studies in the framework of a clinical trial comparing two active treatments as applied to survival data under both homogeneous and heterogeneous treatment effect assumptions to evaluate the performance of the RHM method in terms of bias and 95% credible intervals. We first apply the standard proportional hazard model to obtain the ITT estimate and evaluate resulting bias if viewed as estimating a causal hazard ratio. We then compare the method's performance in terms of stratum-specific causal relative risk for different specifications of a user-defined spectrum parameter. The results showed no effect of the spectrum parameter on the ITT estimates. The RHM method performed poorly by producing significantly biased efficacy estimates in all strata with wider corresponding 95% credible intervals under heterogeneous treatment effect assumption. The resulting efficacy estimates varied a lot depending on the value of the unknown (user-defined) spectrum parameter.

Keywords: bias, causal relative risk, compliance, principal stratification, simulations

1. Introduction

Estimating causal effects is a primary objective in most medical studies which compare two or more interventions. This task may be achieved in randomized clinical trials under perfect compliance with treatment assignment. But the common phenomenon of noncompliance to treatment assignment which often manifests itself as treatment disruption, cessation, switches or patient withdrawal from the study complicates estimation of causal effects. The intention-to-treat (ITT) is considered the benchmark for estimating treatment efficacy under perfect compliance with allocation owing to the fact that it preserves the treatment groups' threshold comparability by contrasting treatment groups as assigned. However, when there is imperfect compliance with allocation to intervention, the ITT produces biased efficacy estimate when the effects of treatment non-compliers mixes with the effects of compliers (White & Pocock, 1996). While simple regression techniques adjusting for noncompliance may produce valid causal estimates under random noncompliance, the fact that noncompliance is often non-random in nature induces complication in making causal inference. In-treatment and as-protocol analyses are mostly used to augment ITT estimates while evaluating true treatment effects. However, these post-hoc analyses are devoid of the tenets of randomization and are likely to produce biased estimates due to selective choices arising from the underlying nature/pattern of arm-specific compliance (White, 2005; Little et al., 2009).

The problem of efficacy estimation is complicated more by the presence of noncompliance in two (or more) treatment arms, where ITT method produces biased estimates even under homogeneous (uniform) treatment effects assumption (Aalen, 1998; Baker & Kramer, 2005). The resulting identification problem due to noncompliance in multiple arms presents a challenge to adjust for imperfect compliance in such trials (Brittain & Lin, 2005; Chiba,

2009). As an example, a double-blind two-armed clinical trial comparing two active-ingredient treatments (say A and B) may be plagued by two levels of noncompliance in the form of simple noncompliance in both arms and additional arm-specific differential noncompliance due to possible breached/imperfect blinding or side effects from the treatment. While pairwise efficacy comparisons are suboptimal in the presence of multiple treatments, joint analysis may be more useful in providing additional analytical insights (Cheng & Small, 2006).

Frangakis and Rubin (2002) developed the principal stratification (PS) as a unifying framework of causal modelling which permits adjustment based on posttreatment variables (e.g. noncompliance status) to produce causal effect estimates which are properly defined. The PS is a robust framework which has been applied under different mathematical platforms to adjust for simple or partial form of noncompliance in multiple treatment arms (Roy et al., 2008; Long et al., 2010). Covariates which predict adherence with treatment allocation is crucial to solving parameter identification problems for the PS method. Roy, Hogan, & Marcus (RHM) (2008) introduced a PS method for adjusting imperfect compliance in two-active treatment trials using covariates recorded at baseline and which predict treatment compliance to mitigate such identification problem. Under all-or-nothing compliance assumption, the method used a set of arm-specific predictors of compliance to produce two arm-specific prediction models which are then combined into a causal model using a user-defined spectrum/sensitivity parameter (a function of arm-specific compliances and the correlation between compliances with treatment) to provide principal effects for each stratum.

The merits of comprehensive model selection have been demonstrated to be transferable to principal stratification for causal inference (Odoni & McNamee, 2013). But a key requirement in applying the RHM method is the plausibility of the distributional assumption positing conditional prediction, i.e. counterfactual response is assumed statistically ignorable (independent) of the set of selected baseline variables which predict treatment compliance for a given compliance type/stratum and treatment allocation. To assess the robustness of this defining (yet untestable) assumption, the present work uses statistical simulation studies to evaluate the performance of RHM method in terms of bias and 95% credible intervals in the platform of a clinical trial designed to compare two active treatments as applied to time-to-event data under both homogenous and heterogeneous treatment effect assumptions.

The rest of the paper is organized as follows: Section 2 describes the general notation and the relevant causal modeling assumptions implicit in the application of the RHM method. Section 3 presents a detailed description of the simulation design (aims and set-up). Section 4 outlines the methods of analysis (ITT, compliance prediction and the principal stratification method by RHM). Section 5 provide results of the simulation analysis under both homogeneous and heterogeneous treatment effect assumptions. In Section 6 we present a discussion of the simulation results.

2. Notation and Assumptions

Let us consider two generic treatments: an active control A and a new treatment B . In a clinical trial setup consisting of two arms, let $W \in \{0, 1\}$ denote a randomization indicator where $W = 1$ indicate randomization to the new treatment B and $W = 0$ indicates randomization to the active control A . We define the response as $Y \in \{0, 1\}$ (e.g. myocardial reinfarction or death). We let $C \in \{0, 1\}$ denote all-or-nothing compliance with assigned treatment. Under the potential outcome framework, every patient has two counterfactual compliance status C_0 and C_1 (comply with treatment A and B respectively) and two counterfactual responses Y_0 and Y_1 (response under A and B intervention respectively). However, the respective compliance and responses observed are represented by $C = WC_1 + (1 - W)C_0$ and $Y = WY_1 + (1 - W)Y_0$. By assumption, each patient belongs to one of four mutually exclusive (basic) principal strata which are defined by distinctive combinations of (C_0, C_1) where the principal strata comprise the set $\mathcal{S} = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. Of principal interest for causality is the common distributions $[(Y_0, Y_1)|S = s] \forall s \in \mathcal{S}$ which provides stratum-specific treatment effects in terms of causal relative risks.

Data analysis under the PS framework uses demographic (and environmental etc) variables X recorded at baseline together with the standard assumptions (i)-(v) for causal modelling (Angrist et al., 1996; Jin & Rubin, 2008) and the additional conditional (distributional) prediction assumption (vi) proposed by Roy et al. (2008):

- (i) Randomization: $W \perp \{Y_0, Y_1, C_0, C_1, X\}$ which posits statistical independence between treatment allocation and potential outcomes, potential treatment received and baseline covariates, i.e. W is randomly assigned.
- (ii) No interference (interaction) between treatment units, i.e. the stable unit-treatment value assumption (SUTVA).
- (iii) The exclusion restriction: $\Pr(Y_1|C_W, X) = \Pr(Y_0|C_W, X)$, i.e. no direct effect of treatment allocation on response except through treatment actually received.

- (iv) Monotonicity: $\Pr(C_1 = 1|C_0 = 1, X) \geq \Pr(C_1 = 1|C_0 = 0, X)$, i.e. there is no access to treatment for subjects randomized to the control arm of the trial.
- (v) Restricted access to treatment, i.e. no subject switches treatment.
- (vi) Selective prediction: $Y \perp X|S, W$, i.e. potential response is ignorable of the selected baseline variables which predict propensity to comply with treatment for a selected compliance type (principal stratum) and allocation.

The version of monotonicity assumption as applied in this framework ensures the type of compliance is observable for $W \neq C$, i.e. assuming no treatment defiers helps provide stable bounds of causal estimates (Taylor & Zhou, 2009). The RHM method additionally assumes admissibility of the ‘compound’ monotonicity assumption which posits no difference in the pattern/nature of compliance in both arms of intervention. This assumption will be reflected in our simulations by specifying a positive correlation (spectrum parameter ϕ) between C_0 and C_1 . The selective prediction assumption (vi) is key to identification of parameters for the RHM method. Although the assumption is untestable, a simulation study evaluating the performance of the method may provide an indication of its robustness.

3. Simulations Design

3.1 Aims of the Simulations

We use statistically designed simulation studies in the framework of a randomized controlled trial to compare two active treatments in terms of survival to evaluate possible bias due to noncompliance in the two treatment arms. First we evaluate the effect on the intention-to-treat (ITT) hazard ratio due to allocation to treatment B relative to treatment A where there may be noncompliance in either arm. As a check on the simulations, we evaluate the ITT effects for each stratum. Next we apply RHM method for survival data whose analysis requires specification of a spectrum parameter ϕ (positive and user-defined) chosen as a function of arm-specific compliances and the correlation between compliances with treatment, i.e. parameter ϕ is not estimated from data. With two factors separately assumed predictive of compliance, we first construct arm-specific prediction models of compliance using logistic models from which we estimate the probabilities of compliance with treatment in each arm. We use ϕ to combine the two arm-specific compliance models into one causal model which then provide stratum-specific treatment effects in terms of causal relative risks estimated from the means of posterior median relative risks of experiencing event within each subgroup: (i) risk arising from compliance with treatment B relative to compliance with treatment A , (ii) risk arising from compliance with treatment A only relative to baseline risk, and (iii) risk arising from compliance with treatment B only relative to baseline risk. We use Bayesian methods to estimate mean of the posterior median relative risks and their respective mean 95% credible intervals of experiencing event in three different strata defined by their corresponding compliance types (compliance with A and B , A only and B only) while assuming non-random compliance under both homogeneous and heterogeneous treatment effects assumptions. We use death as the generic outcome of interest.

3.2 Simulations Set-Up

The simulation study mimics a two-armed randomized trial with active treatments A and B lasting 24 months. There were 2000 replications for each scenario to ensure coverage lies within two standard errors of the nominal 95% coverage probability. Each simulation assumed a sample size of 1000 with equal probability of being randomly assigned to either treatment arm (to mimic Esprit data (Cherry et al., 2002; Odoni & McNamee, 2013)).

Each subject had three potential hazard rates: λ_{0i} , λ_{Ai} and λ_{Bi} corresponding to baseline risk under no treatment and under treatment A and B respectively. The effects of both treatments are assumed better than no treatment at all in all cases. The time-invariant hazard rates $\{\lambda_{0i}\}$ were generated from Gamma distribution with shape and scale parameters 2 and 0.006 respectively so as to have mean 0.012 and variance 7.2×10^{-5} . Each stratum assumed constant risk of death over time for both treatments A and B . The simulation model considered events in each month separately. For a given month the probability of dying if a specific treatment is taken in any stratum were taken as equal to $1 - \exp(-\lambda_{Ai})$ and $1 - \exp(-\lambda_{Bi})$ for treatment A and B respectively. Random numbers from the uniform distribution were used to decide which subjects actually died from either treatment arm. Time to death was taken as the end of each month: the minimum time is 1 month for those who died in the first month while the maximum time is taken as 24. Subjects were allocated to treatment arms at random and risks chosen according to arm and potential compliance type. We assume no switching of subjects between the treatment arms.

We considered all-or-nothing compliance to allocation for both treatments A and B up to 24 months. Compliance with treatment allocation is assumed to be predictable from two binary (0/1) baseline covariates. To mimic the Esprit data, we first specified two binary covariates to represent smoking status and diabetes risks. Each subject

belonged to one of four complier type (principal stratum): type 3 ($S = (1, 1)$) represent potential compliers to either treatment, type 0 ($S = (0, 0)$) represent people who would comply with neither treatment, types 1 ($S = (1, 0)$) and 2 ($S = (0, 1)$) represent compliers to treatment A only and B only respectively. The compliance types were determined independently by a subject's associated risk factors X and her baseline risk of death. We set the actual prevalence rates of history of smoking status and risk of diabetes at 25% and 60% respectively.

Next we describe the relationship between probabilities of compliance and the pair of covariates predicting compliance in terms of odds ratio. To link the risk factors and compliance, for each treatment arm, we specified three sets of statistics: (a) the probability of compliance to treatment allocation in the absence of both risk factors; this was set as 0.55 for treatment A and 0.30 for treatment B, (b) a compliance odds ratio for smoking: 2 for treatment A and 5 for treatment B and (c) a compliance odds ratio for diabetes: 4 for treatment A and 3 for treatment B. The joint effect of both factors on compliance was assumed to be multiplicative on the odds ratio scale. This is the same as using a logistic model with no interaction term to obtain actual compliance probabilities for individual cells. These assumptions imply that the probabilities of compliance given a set of covariates X , $\mu_A(x)$ and $\mu_B(x)$, say for groups A and B are such that $\mu_A(x) > \mu_B(x)$ as specified in the RHM model.

Finally we link the compliance type with the baseline hazard. We assume that potential compliance to treatment A and to B are positively correlated. Following Roy et al. (2008), we introduce non-random compliance in each stratum in the form of a spectrum parameter ϕ which was chosen as a positive function of arm-specific compliances and a correlation ρ between compliances to treatment. We compared results for different values of $\phi = 0, 0.2, 0.5, 0.8$. To allocate subject to compliance types, we specifically assigned highest ranked values of baseline risk λ_{0i} to represent those subjects who would comply with neither of the treatment allocations (type 0) while the lowest ranked values of λ_{0i} are assigned to compliers of either treatment (type 3). From the remaining middle set, we assign subjects at random to either compliance to treatment A only (type 1) or treatment B only (type 2) according to their respective weighted proportions as set for simulation model. We then worked out probability that compliance to A is i and compliance to B is j , given x_1 and x_2 , i.e. $\mu_{ij}(x_1, x_2)$ for a given value of ϕ . We use random numbers and the multinomial probabilities $\mu_{ij}(x_1, x_2)$ to determine the actual number of compliers for each of the four compliance types in a given simulation.

The simulations setup considered both homogeneous and heterogeneous treatment effects. An homogenous treatment effect corresponded to scenario when potential treatment effects were assumed same for all principal strata. The heterogeneous treatment effects assumption was reflected by setting the potential treatment effects among non-compliers with treatment A and B to be relatively smaller than potential effects among compliers for a specific stratum. For the homogeneous case $\lambda_{Bi} = [\exp(\psi)]\lambda_{Ai}$, where $\exp(\psi)$ is the true causal hazard ratio (THR), which was set at 0.667 for each stratum. For the heterogeneous case, the potential treatment effects among non-compliers to treatment A and B were set to be smaller than potential effects among compliers. Specifically we set the causal hazard ratio at 0.667, 0.750, 0.778 and 0.800 for stratum 3, 2, 1 and 0 respectively, i.e. we set best benefit from treatment B relative to A for patients of type 3 (1, 1), with the hazard ratio the same as in the homogenous case ($\text{THR}_{(1,1)} = 0.667$). The hazard rates for non-compliers among type 2 (0, 1) patients were set to be relatively lower ($\lambda_{Ai} = \frac{2}{3}\lambda_{0i}$) compared to hazard rates for non-compliers among type 3 patients. Conversely the hazard rates for non-compliers among type 1 (1, 0) patients was set to be relatively higher ($\lambda_{Bi} = \frac{7}{12}\lambda_{0i}$) compared to hazard rates for those classified to belong to type 3 (see Table 1). We used the ratio $\frac{\lambda_{Bi}}{\lambda_{Ai}}$ to obtain causal effects of treatment B relative to A for the subgroup who would comply with either treatment.

Table 1. Stratum-specific hazard rates (among patient compliers)

Type (stratum)	Homogeneous effects				Heterogeneous effects			
	$\bar{\lambda}_A$	$\bar{\lambda}_B$	THR	† TRR	$\bar{\lambda}_A$	$\bar{\lambda}_B$	THR	† TRR
3 (1,1)	0.009	0.006	0.667	0.729	0.009	0.006	0.667	0.729
2 (0,1)	0.009	0.006	0.667	0.594	0.008	0.006	0.750	0.594
1 (1,0)	0.009	0.006	0.667	0.815	0.009	0.007	0.778	0.815
0 (0,0)	0.009	0.006	0.667	1.000	0.010	0.008	0.800	1.000

† TRR calculated according to Equation (1).

Using the RHM model, we obtain the true (causal) relative risk (TRR) calculated as the ratio of average risk estimates of experiencing event in each treatment arm within a stratum. Specifically we use moment generating function results of Gamma distribution, i.e. $\lambda_i \sim \text{Gamma}(\alpha, \beta)$:

$$\text{TRR} = \frac{[1 - (1 + 24\bar{\lambda}_B)^{-\beta}]}{[1 - (1 + 24\bar{\lambda}_A)^{-\beta}]}, \quad (1)$$

so that for $\beta = 2$, $\text{TRR}_{(1,1)} = 0.729$, $\text{TRR}_{(0,1)} = 0.594$ and $\text{TRR}_{(1,0)} = 0.815$ for stratum $S = 3, 2$ and 1 respectively as shown in Table 1 (for $\lambda_{0i} = 0.012$, $\lambda_{Ai} = 0.009$, $\lambda_{Bi} = 0.006$).

4. Methods of Analysis

4.1 Intention-to-Treat

We begin the analysis with checks for the parameters as setup in the simulations. Next we obtain the ITT estimate by applying the Cox proportional hazards model ignoring treatment compliance in order to evaluate its bias (if any) for estimation of ψ . Specifically we evaluate the hazard ratio of death due to allocation to treatment B relative to treatment A for both homogeneous and heterogeneous treatment effects cases using the Cox proportional hazards model

$$h(t|W_i) = h_0(t) \exp[\psi W_i] : \quad W = \begin{cases} 1; & \text{if treatment B} \\ 0; & \text{if treatment A,} \end{cases} \quad (2)$$

where $h(t|W_i)$ denotes the hazard rate for experiencing event at time t given exposure, $h_0(t)$ is the baseline hazard for a subject allocated to treatment A and the ITT is estimated by $\exp(\hat{\psi})$.

All ITT results were assumed to provide an estimate $\hat{\psi}$ of ψ , the log causal hazard ratio in the simulation model; then the mean of the estimators, $\bar{\hat{\psi}}$, and their corresponding root mean squared errors (RMSE) are calculated: $\text{RMSE}(\bar{\hat{\psi}}) = \sqrt{[\bar{\hat{\psi}} - \psi]^2 + \text{var}(\hat{\psi})}$. In the table we show mean effect on the HR scale calculated as $\exp(\bar{\hat{\psi}})$. We use a one-sided t-test with $\alpha = 0.05$ to test for bias with t-statistic $\frac{\bar{\hat{\psi}} - \psi}{s/\sqrt{2000}}$, where s is the standard deviation of $\{\hat{\psi}_i\}$. Assuming that $s = 0.50$ or less, the simulation study was large enough to give 90% power to detect a bias of 0.01 or more on the ψ scale (i.e. $\bar{\hat{\psi}} - \psi$) for any statistical method. A non-significant test was taken as evidence of no important bias.

4.2 Predicting Compliance With Treatment Allocation

For a specified set of baseline predictors of compliance X , we can use logistic models to separately model the arm-specific likelihood to comply with treatment allocation:

$$\text{logit} [\mu_j(x)] = \left(\sum_{i=0}^n \gamma_{ji} x_i \right), \quad j = 0, 1, \quad (3)$$

where $\mu_j(x)$ provides the propensity (probability) to comply with allocation to intervention j (A/B) for selected baseline variables X . To mimic the Esprit study in our simulations, the two covariates represent smoking status and history of diabetes. The arm-specific probability of fidelity with treatment allocation may then be estimated by

$$\hat{\mu}_j(x) = \left[1 + \exp \left(- \sum_{i=0}^n \hat{\gamma}_{ji} x_i \right) \right]^{-1}, \quad j = 0, 1, \quad (4)$$

where γ is log of the odds ratio estimates of compliance with treatment.

Crucial to applying the RHM method is nature/form of correlation between the two compliance behaviours. By RHM's formulation, we define a positive spectrum parameter ϕ to capture the correlation ρ among compliances with treatment allocation (0/1) such that if $\hat{\mu}_A(x) > \hat{\mu}_B(x)$ then

$$\phi = \rho \sqrt{\frac{\hat{\mu}_A(x)[1 - \hat{\mu}_B(x)]}{\hat{\mu}_B(x)[1 - \hat{\mu}_A(x)]}}. \quad (5)$$

4.3 Causal Inference: Causal Relative Risk

The cause-effect inference of principal significance can be obtained from the common distributions $[(Y_0, Y_1)|S = s]$. By reparameterizing the causal model in terms of π and $\theta = f(\gamma, \phi)$, respectively representing the probability

of experiencing event and the logarithm of the odds ratio of complying with treatment for a specified spectrum parameter value, Roy et al. (2008) proved the likelihood for observable data obtainable as

$$L(\pi, \theta | Y, C, W, X) = \sum_{s=0}^3 [\pi_W^{S=s}]^Y [1 - \pi_W^{S=s}]^{1-Y} \Pr(S = s | X, \theta) G(s, C, W), \quad (6)$$

where $\pi_W^{S=s}$ is the probability of observed event $Y = 1$, given $S = s$ and arm of allocation W , and

$$G(s, C, W) = I(s=0)\{1 - C\} + I(s=1)\{C(1 - W) + (1 - C)W\} \\ + I(s=2)\{CW + (1 - C)(1 - W)\} + I(s=3)C.$$

By the exclusion restriction assumption, the hazard of suffering phenomenon of interest/outcome is assumed statistically ignorable of the assignment arm for the subset of patients not likely to comply with either treatment assignment, i.e. $\pi_1^{s=(0,0)} = \pi_0^{s=(0,0)}$. Using logistic models, the resulting likelihoods provide 7 parameters captured by π . Analysis using the RHM method then produce the causal relative risk calculated as risk ratios of experiencing event in each stratum s :

$$\tau_{11} = \frac{\pi_1^{s=(1,1)}}{\pi_0^{s=(1,1)}} = \frac{\hat{\pi}_2}{\hat{\pi}_6}, \quad \tau_{01} = \frac{\pi_1^{s=(0,1)}}{\pi_0^{s=(0,1)}} = \frac{\hat{\pi}_1}{\hat{\pi}_7}, \quad \tau_{10} = \frac{\pi_1^{s=(1,0)}}{\pi_0^{s=(1,0)}} = \frac{\hat{\pi}_4}{\hat{\pi}_5}. \quad (7)$$

Solution to (7) provide stratum-specific causal relative risk τ_{ij} calculated as ratio of risks of relevant mean posterior medians: (i) τ_{11} : causal relative risk of experiencing phenomenon for those complying with treatment B relative to treatment A in the subset of patients likely to comply with either one or the other treatment assignment, i.e. $S = (1, 1)$, (ii) τ_{01} : causal relative risk of experiencing phenomenon for those complying with treatment B only among the subset of patients likely to comply if allocated to it relative to baseline risk, i.e. $S = (0, 1)$, and (iii) τ_{10} : causal relative risk of experiencing event for those complying with treatment A only among the subset of patients likely to comply if allocated to it relative to baseline risk, i.e. $S = (1, 0)$.

We used Bayesian methods with suitable priors to estimate the parameters given by Equation (7). Specifically we used uniform $(0, 1)$ as priors for the probabilities of risks event in each stratum given the arm of allocation and compare results for different specified spectrum parameter values $\phi = 0, 0.2, 0.5$ and 0.8 . The use diffuse priors such as $\pi \sim U(0, 1)$ in our analyses may be considered plausible given the fact that a regular data from a trial is likely to monopolize corresponding priors and also that randomized clinical trials are often mainly constructed to provide definite evidence (Heitjan et al., 1991). We considered both homogeneous and heterogeneous treatment effect assumptions and ran three chains: chain one had null starting values while chains two and three had the arithmetic mean and median respectively from an initial trial run. To assess convergence, we conducted simulation for 1.1×10^4 iterations for every individual chain while excluding the first 1,000 for burn-in.

The causal relative risk estimates for each stratum τ_{ij} are calculated as ratio of probabilities of event among potential compliers to treatment B relative to A for each stratum given the arm of allocation. To evaluate their performance, we use the corresponding standard deviation (SD) of the median of the estimators, $\tilde{\tau}$, to calculate $RMSE(\tilde{\tau}) = \sqrt{[\tilde{\tau} - \tau]^2 + \text{var}(\tilde{\tau})}$ and used a one-sided t-test with $\alpha = 0.05$ to test for bias with t-statistic $\frac{\tilde{\tau} - \tau}{SD / \sqrt{30,000}}$, where SD is the standard deviation of $\{\tau_{ij}\}$. Assuming that $SD = 2$ or less, the simulation study was large enough to give 90% power to detect a bias of 0.01 or more on the τ scale for any statistical method. Also a non-significant test was taken as evidence of no important bias.

When applying the RHM method for survival data, we use relative risks to approximate hazard ratios. This may be justifiable for our simulation given that under short follow-up time and small event rates conditions, relative risk has been shown to be an algebraic approximation of hazard ratio, i.e. $\exp(\hat{\psi}) \cong \tilde{\tau}$ (Symons & Moore, 2002).

5. Results

5.1 Checking the Simulations

We obtained odds ratio estimates by fitting a logistic model to each simulation. For a moderate value of the spectrum parameter $\phi = 0.5$, the mean compliance odds ratios for smoking status were 2.015 and 5.105 for treatment A and B respectively, and the compliance odds ratios for risk to diabetes were 4.041 and 3.025 respectively for treatment A and B. In general, these results and simulation results for other values of ϕ were in agreement with the odds ratios pre-specified in the simulation design.

Table 2. Estimates of mean compliance proportion per stratum for different ϕ values

Stratum (type)	Homogeneous hazard rates				Heterogeneous hazard rates			
	ϕ				ϕ			
	0	0.2	0.5	0.8	0	0.2	0.5	0.8
$\tilde{\mu}_A$	0.744	0.745	0.747	0.749	0.745	0.746	0.746	0.746
$\tilde{\mu}_B$	0.542	0.543	0.543	0.544	0.541	0.542	0.542	0.541
$\tilde{\mu}_{11}$ (3)	0.426	0.449	0.484	0.518	0.425	0.449	0.483	0.518
$\tilde{\mu}_{01}$ (2)	0.115	0.092	0.057	0.023	0.116	0.092	0.058	0.023
$\tilde{\mu}_{10}$ (1)	0.319	0.296	0.261	0.227	0.320	0.297	0.262	0.228
$\tilde{\mu}_{00}$ (0)	0.139	0.162	0.197	0.231	0.139	0.162	0.197	0.231

$$\tilde{\mu}_A = \tilde{\mu}_{11} + \tilde{\mu}_{10} \text{ and } \tilde{\mu}_B = \tilde{\mu}_{11} + \tilde{\mu}_{01}.$$

Table 2 shows the mean (overall) proportion of compliance per stratum at different values of the spectrum parameter ϕ for both homogeneous and heterogeneous hazard rates. On average, the compliance proportion results were similar to the pre-specified probabilities in the simulation design, for example while for our set-up $\mu_{11} = 0.483$, the mean for the simulations was $\hat{\mu}_{11} = 0.478$. The mean proportions of compliance per stratum were similar for both homogeneous and heterogeneous cases: the mean proportion of compliance to treatment A was higher compared to mean compliance to treatment B (e.g. $\tilde{\mu}_A = 75\%$ and $\tilde{\mu}_B = 54\%$ when $\phi = 0.5$). As per our setup, the simulations ensured that potential compliers to either treatment (type 3) were the most frequent type while potential compliers to treatment B only (type 2) would be the least frequent for all chosen values of ϕ . Overall, the mean proportion of compliance to either treatment (type 3) and neither treatment (type 0) dominated (increased) as the sensitivity parameter ϕ increased. On the other hand, the mean compliance proportion reduced as ϕ values increased among those patients likely to comply with one treatment only (type 1 and 2). We note the small proportion of potential compliers to treatment B only (type 2) which approached total noncompliance as ϕ gets close to 1 (perfect correlation). In general, all the proportions of compliance were comparable to the expected weighted compliances proportions of the preset values. Overall, the general trend of proportion of compliance was the same under both homogeneous and heterogeneous treatment effect assumptions.

5.2 Effect Due to Intention-to-Treat

Table 3 provide the overall ITT estimates for both homogeneous and heterogeneous cases. The ITT hazard ratio 0.675 for the homogeneous treatment effects case model suggested that overall, the risk of event would reduce by 32% for those randomized to treatment B compared to those randomized to treatment A. The resulting small bias (0.008) for the ITT estimate was however statistically significant. On the other hand under the heterogeneous treatment effects assumption, the ITT hazard ratio 0.762 indicated an overall reduction of risk of event death by 24% for those randomized to treatment B compared to treatment A. The resulting bias (0.031) here for the heterogeneous ITT estimate was also statistically significant. We observe a bias-precision tradeoff where as expected the bias due to homogeneous hazard was negligible (relatively smaller) compared to bias from using heterogeneous hazard rates but the later had relatively smaller standard error compared to the former. However, in general we note that a study population is more likely to be heterogeneous than homogeneous.

Table 3. ITT estimates (hazard ratio) for homogeneous and heterogeneous hazard rates

	\dagger THR	$\exp(\hat{\psi})$	$SE(\hat{\psi}_i)$	$RMSE(\hat{\psi}_i)$	p-value
Homogeneous	0.667	0.675	0.161	0.162	0.001
Heterogeneous	0.731 \ddagger	0.762	0.155	0.160	< 0.001

\dagger True hazard ratio, Table 1 (\ddagger weighted using proportions in Table 2).

Table 4 shows performance of the ITT estimates for each stratum in terms of hazard ratio under both homogeneous and heterogeneous treatment effect assumptions for various specifications of the spectrum parameter ϕ . All selected values of ϕ produced essentially unbiased hazard ratio estimate of efficacy for patients complying with treatment B compared to compliers with treatment A among those patients likely to comply with either one treatment or the other ($S = (1, 1)$). This may be discernable from usual expectation of high compliance rates with treatment for this

subgroup which is likely to reveal true effects of both treatments. For a chosen value of ϕ in this stratum we also note similarity in the standard errors for the corresponding causal hazard ratio estimates under both homogeneous and heterogeneous treatment effect assumptions.

Table 4. Performance of ITT effects (hazard ratio) in each stratum

ϕ	Homogeneous hazard rates				Heterogeneous hazard rates			
	$\exp(\hat{\psi})$	$SE(\hat{\psi}_i)$	$RMSE(\hat{\psi}_i)$	p-value	$\exp(\hat{\psi})$	$SE(\hat{\psi}_i)$	$RMSE(\hat{\psi}_i)$	p-value
	<u>$\dagger THR_{(1,1)} = 0.667$</u>				<u>$THR_{(1,1)} = 0.667$</u>			
0	0.660	0.356	0.356	0.212	0.660	0.356	0.356	0.214
0.2	0.660	0.339	0.340	0.142	0.660	0.339	0.339	0.144
0.5	0.660	0.323	0.323	0.143	0.660	0.323	0.323	0.143
0.8	0.661	0.303	0.303	0.212	0.661	0.303	0.303	0.212
	<u>$THR_{(0,1)} = 0.667$</u>				<u>$THR_{(0,1)} = 0.750$</u>			
0	0.654	0.493	0.493	0.083	0.740	0.505	0.505	0.241
0.2	0.625	1.314	1.315	0.026	0.708	1.323	1.325	0.052
0.5	0.341	5.336	5.377	< 0.001	0.408	5.544	5.577	< 0.001
0.8	0.018	17.742	18.105	< 0.001	0.041	18.446	18.675	< 0.001
	<u>$THR_{(1,0)} = 0.667$</u>				<u>$THR_{(1,0)} = 0.778$</u>			
0	0.668	0.274	0.274	0.820	0.779	0.263	0.2636	0.851
0.2	0.670	0.286	0.286	0.472	0.782	0.275	0.275	0.450
0.5	0.667	0.307	0.307	0.981	0.778	0.296	0.296	0.954
0.8	0.664	0.329	0.329	0.537	0.777	0.317	0.317	0.811

\dagger True hazard ratio (see Table 1).

From Table 4 we observed inconsistency in results for the subgroup of patients who would comply with only one of either treatments. Under both homogeneous and heterogeneous treatment effect assumptions, we obtained unbiased ITT estimates for those patients likely to comply with treatment A only for all specified values of the spectrum parameter ϕ . These ITT estimates were, on average, invariant to the values of ϕ for this stratum. On the other hand, for those patients likely to comply with treatment B only the ITT estimates were significantly biased at higher values ($\phi = 0.5$ and 0.8) of the spectrum parameters under both homogeneous and heterogeneous treatment effect assumptions. We also observed relatively large standard errors corresponding to the estimates which increased with increase in spectrum parameter for this subgroup. The large standard errors (standard errors dominate corresponding RMSE) may be a manifestation of sparseness due to near ‘total’ noncompliance as ϕ approaches 1 (perfect correlation).

5.3 Performance of the RHM Method

Table 5 provides results from simulation evaluating the performance of the RHM method. The results show a comparison of stratum-specific causal relative risk estimates (calculated from posterior median relative risks) for different specifications of spectrum parameter ϕ values under both homogeneous and heterogeneous treatment effect assumptions. The resulting biases in the estimates for causal relative risks were all statistically significant in all strata at all values of ϕ under both homogeneous and heterogeneous treatment effect assumptions. We observe that only the causal relative risk estimate of efficacy among patients likely to comply with treatment B only compared to baseline was unbiased at the highest selected value of sensitivity parameter ($\phi = 0.8$) under homogeneous

treatment effect assumption. However, the corresponding standard error for this estimate was relatively very large compared to others in the same stratum. This may be considered a non representative (isolated) result given that compliance in the two arms is unlikely to be (nearly) perfectly correlated.

Table 5. Performance of the RHM method in terms of causal relative risk per stratum

ϕ	Homogeneous hazard rates			Heterogeneous hazard rates		
	$\tilde{\tau}$ (RR)	SE($\hat{\tau}_{ij}$)	RMSE($\hat{\tau}_{ij}$)	$\tilde{\tau}$ (RR)	SE($\hat{\tau}_{ij}$)	RMSE($\hat{\tau}_{ij}$)
	\ddagger TRR _(1,1) = 0.729 $\left(\frac{\bar{\lambda}_B}{\bar{\lambda}_A}\right)$			TRR _(1,1) = 0.729 $\left(\frac{\bar{\lambda}_B}{\bar{\lambda}_A}\right)$		
0	0.688	0.153	0.158	0.837	0.211	0.237
0.2	0.676	0.146	0.155	0.804	0.186	0.201
0.5	0.653	0.128	0.149	0.764	0.151	0.155
0.8	0.710	0.114	0.116	0.797	0.176	0.189
	TRR _(0,1) = 0.594 $\left(\frac{\bar{\lambda}_B}{\lambda_0}\right)$			TRR _(0,1) = 0.594 $\left(\frac{\bar{\lambda}_B}{\lambda_0}\right)$		
0	0.476	0.291	0.314	0.530	0.302	0.309
0.2	0.579	0.290	0.290	0.485	0.305	0.329
0.5	0.641	0.292	0.296	0.529	0.319	0.326
0.8	0.593 [§]	0.311	0.311	0.573	0.803	0.803
	TRR _(1,0) = 0.815 $\left(\frac{\bar{\lambda}_A}{\lambda_0}\right)$			TRR _(1,0) = 0.815 $\left(\frac{\bar{\lambda}_A}{\lambda_0}\right)$		
0	0.995	0.091	0.202	0.913	0.256	0.274
0.2	0.989	0.129	0.217	0.841	0.285	0.286
0.5	0.874	0.246	0.253	0.553	2.300	2.315
0.8	0.567	0.428	0.495	0.713	0.720	0.727

[‡]True relative risk (see Equation 1 and Table 1): $\lambda_{0i} = 0.012$; $\lambda_{Ai} = 0.009$; $\lambda_{Bi} = 0.006$; [§]unbiased.

For the homogeneous case, the causal relative risk estimate of efficacy for those complying with treatment B as contrasted to A in the subset of those patients likely to comply with either of the two treatments (type 3) consistently produced smaller biases for all ϕ values considered. Specifically the resulting bias in the causal relative risk estimate was smallest (-0.019) for higher values of $\phi = 0.8$. Also at a moderate spectrum value ($\phi = 0.5$) for the heterogeneous case, the causal relative risk of efficacy for patients complying with treatment B compared to treatment A in the subset of potential compliers with either treatment produced small bias (0.035), although the bias was statistically significant. In general compared to homogeneous case, we observe substantial increase in standard error corresponding to causal risk ratio estimates for the subgroup which would comply with either treatment under heterogeneous treatment effect assumption.

Results comparing efficacy estimates among patients likely to comply with one particular treatment only were less biased for the subgroup who would comply with treatment B only relative to those who would comply with treatment A only. But we also observe a bias-variance tradeoff in which the corresponding standard errors for the causal relative risk estimates of efficacy for patients likely to comply only with treatment B were relatively larger compared to those for compliance with treatment A only for all values of ϕ considered. Overall, causal relative risk estimates of efficacy among patients likely to comply only with treatment B compared to baseline risks were less biased (although statistically significant) under homogeneous treatment effect assumption than under heterogeneous case.

We note the fact that results presented in Table 5 were obtained after using the same value of spectrum parameter ϕ

for both data generation and analysis. An implication here is the likelihood of the simulations to give an optimistic view of the RHM method, i.e. the method's defining (distributional) selective compliance assumption introduced as ϕ may be over represented by first specifying a ϕ value to link the two marginal compliance model for each treatment arm during data generation and using the same value again for compliance models for each stratum in analysis.

Table 6. Comparing stratum-specific causal relative risks (mean 95% CI) for different values of ϕ ($\pi_Z^S = \Pr[Y = 1|S, Z] \equiv$ probability of death in a stratum given treatment allocation): (a) Homogeneous hazard rates (b) Heterogeneous hazard rates

ϕ	Comply with A and B			Comply with B only			Comply with A only		
	$\ddagger\hat{\pi}_2$	τ_{11}	$\hat{\pi}_6$	$\hat{\pi}_1$	τ_{01}	$\hat{\pi}_7$	$\hat{\pi}_4$	τ_{10}	$\hat{\pi}_5$
(a)	TRR _(1,1) = 0.729			TRR _(0,1) = 0.594			TRR _(1,0) = 0.815		
0	0.161	0.688 (0.420, 1.007)	0.238	0.114	0.476 (0.028, 1.010)	0.243	0.238	0.995 (0.808, 1.195)	0.240
0.2	0.158	0.676 (0.423, 0.993)	0.238	0.139	0.579 (0.039, 1.023)	0.243	0.233	0.989 (0.722, 1.258)	0.236
0.5	0.154	0.653 (0.429, 0.928)	0.240	0.153	0.641 (0.050, 1.055)	0.242	0.201	0.874 (0.260, 1.240)	0.235
0.8	0.168	0.710 (0.509, 0.956)	0.240	0.138	0.593 (0.038, 1.106)	0.237	0.123	0.567 (0.029, 1.311)	0.226
(b)	TRR _(1,1) = 0.729			TRR _(0,1) = 0.594			TRR _(1,0) = 0.815		
0	0.181	0.837 (0.531, 1.355)	0.221	0.125	0.530 (0.030, 1.040)	0.241	0.199	0.913 (0.433, 1.464)	0.221
0.2	0.178	0.804 (0.524, 1.241)	0.226	0.114	0.485 (0.023, 1.035)	0.240	0.185	0.841 (0.266, 1.393)	0.224
0.5	0.172	0.764 (0.528, 1.119)	0.230	0.122	0.529 (0.028, 1.095)	0.237	0.117	0.553 (0.030, 1.318)	0.221
0.8	0.174	0.797 (0.567, 1.250)	0.223	0.119	0.573 (0.029, 1.600)	0.220	0.146	0.713 (0.047, 1.695)	0.215

$$\ddagger\tau_{11} = \frac{\pi_1^{S=(1,1)}}{\pi_0^{S=(1,1)}} = \frac{\hat{\pi}_2}{\hat{\pi}_6}; \quad \tau_{01} = \frac{\pi_1^{S=(0,1)}}{\pi_0^{S=(0,1)}} = \frac{\hat{\pi}_1}{\hat{\pi}_7}; \quad \tau_{10} = \frac{\pi_1^{S=(1,0)}}{\pi_0^{S=(1,0)}} = \frac{\hat{\pi}_4}{\hat{\pi}_5} \text{ (see Equation 7).}$$

Table 6 shows results comparing causal relative risk estimates (calculated from mean posterior median relative risks) and corresponding mean 95% credible intervals for each stratum under both homogeneous and heterogeneous treatment effect assumptions for various specifications of spectrum parameter ϕ . A relative risk was calculated as proportion of the mean posterior median probability of suffering a phenomenon/event to that of survival within same stratum. Generally we observed lower median probabilities of experiencing event in those patients likely to comply only with treatment B compared to those who would comply only with treatment A under both homogeneous and heterogeneous treatment effect assumptions.

The risk of experiencing event was lowest among potential compliers with treatment B only ($\hat{\pi}_1$) for all selected values of ϕ but the risk increased with increase in ϕ . In contrast, risk of experiencing event for those patients likely to comply only with treatment A ($\hat{\pi}_4$) decreased as the specified values of ϕ increased. As we would expect, the probability of experiencing event among potential compliers to either treatment ($\hat{\pi}_2$) was intermediate between the

two risks of events for those patients likely to comply with only one particular treatment (A or B). On the whole, for potential compliers to either treatment we observe increased risks of death under heterogeneous treatment effect assumptions compared to the homogeneous treatment scenario.

At low or moderate values of ϕ under homogeneous treatment effect assumptions, the causal relative risk estimates of efficacy for patients complying with treatment B compared to A for the subgroup likely to comply with either treatment to which they were allocated were the least biased (although statistically significant). The resulting biases increased with increase in ϕ values. The corresponding 95% credible intervals for the causal relative risk estimates became narrower (smaller) as ϕ increased for this subgroup, indicating gain in precision. Moreover the causal relative risk estimates had relatively larger biases under heterogeneous treatment effect assumptions for all values of ϕ . In general, the mean 95% credible intervals for the causal relative risk estimates of efficacy for patients likely to comply with only one treatment (A or B) were generally wider compared to those for the subgroup likely to comply with either of the two treatments. These 95% credible intervals for causal relative risk estimates became wider with increase in ϕ values.

6. Discussion

The user-defined spectrum parameter ϕ had no effect on the ITT results and neither on the overall nor stratum-specific mean proportion of compliance for either treatment. This is not unusual and is expected since ITT estimation of efficacy ignores any form of (non-random) compliance information introduced by use of ϕ . While the principal effects of treatment for those patients likely to comply with their respective (either) treatment allocation were smaller than the ITT estimates under the assumption of homogeneous (constant) treatment effects, the effects were larger than ITT for the heterogeneous case. Overall, the principal effects of treatment for subset of patients complying only with treatment B allocation were smaller than ITT estimate under both homogeneous and heterogeneous treatment effect assumptions.

In general, causal relative risks estimating effects among potential compliers with treatment B compared to A in the subset of patients likely to comply with either of the two treatments produced the least bias (albeit statistically significant) compared to other strata. The corresponding 95% credible intervals for these estimates became narrower as the specified spectrum parameter ϕ values increased. In addition causal relative risk estimates of efficacy for patients likely to comply with one only treatment produced larger biases and corresponding wider 95% mean credible intervals which became even wider with increase in ϕ values. The proportion of potential compliers to treatment B only approached total noncompliance as ϕ approached perfect correlation. Such a phenomenon may be encountered in situations where a new treatment B produces unpleasant side effects likely to induce noncompliance among those randomized to it, i.e. resulting in dominance by the highly compliant type ($S = (1, 1)$) at the expense of those complying with treatment B only $S = (0, 1)$ since $\mu_B = \mu_{11} + \mu_{01}$.

Overall, the RHM method of principal stratification performed poorly and the causal relative risk estimates varied a lot depending on the (unknown) specified value of the spectrum parameter. Inference from the biased results of the simulation studies suggests that the RHM method may only be applicable when we have sufficient knowledge about the nature/pattern of compliance (e.g correlation) between the individual treatment arms. Given such knowledge, subgroup (stratum-specific) analyses may be useful towards understanding the nature of ITT bias by utilizing compliance information which would augment ITT results in efficacy estimation. Choosing non-compliers for a known inferior treatment from the tail of hazard rates' distribution may provide a practical and effective evaluation of principal effects, i.e. it may be considered more meaningful to associate noncompliance with a lower set of ranked baseline hazard rates and corresponding risk factors.

The restrictive nature of all-or-nothing compliance assumption may be a contributory factor in the method's poor performance. In general application of the RHM method may be limited to intermediates with fewer categories and extending the method to the more prevalent continuous compliance may suffer the problem of tractability (VanderWeele, 2011). Although Ma et al. (2011) extended the method to continuous compliance in which the joint distribution of the observed and latent/counterfactual compliance are specified by using copula to link the two arm-specific compliance distributions, we note that the underlying spectrum parameter remain unidentified. In addition to loss of information, principal stratification often coarsens data whose analysis is then likely to produce invalid or even contradicting estimates of causal effect for a variable used in stratification which is truly continuous but coarsened for analysis (Robins et al., 2007). In particular the exclusion restriction assumption may not be plausible if we condition on a coarsened simple surrogate version of the true compliance due to possible residual association. As a caveat, given that principal strata themselves are unidentified, any policy based on results from principal stratification method should be cautiously implemented because as Imai et al. (2011) points out there is

no statistical method with the ability to recover information that is not present in the observed data.

And finally the suitability of the applicability of the RHM method depends on admissibility of the strong but unverifiable distributional assumption positing selective compliance. It assumes that the response is statistically ignorable (independent) of the set of covariables recorded at baseline which predict propensity to comply with treatment given type of compliance type and treatment allocation. However, in practice risk factors which are strongly predictive of outcome are most likely to be related with the likelihood to comply with treatment allocation. In addition selection of suitable predictors of compliance is rarely a primary objective in trials and may only be feasible by exploiting data from pilot studies which ordinarily require more time and resources.

References

- Aalen, O. O. (1998). Frailty Models. In B. S. Everitt & G. Dunn (Eds.), *Statistical Analysis of Medical Data: New Developments* (pp. 59-72). London: Arnold.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444-455.
- Baker, S. G., & Kramer, B. S. (2005). Simple maximum likelihood estimates of efficacy in randomized trials and before-and-after studies, with implications for meta-analysis. *Statistical Methods in Medical Research*, *14*, 349-367.
- Brittain, E., & Lin D. (2005). A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Statistics in Medicine*, *24*, 1-10. <http://dx.doi.org/10.1002/sim.1934>
- Cheng, J., & Small, D. S. (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society, B68*(5), 815-836. <http://dx.doi.org/10.1111/j.1467-9868.2006.00568.x>
- Cherry, N., Gilmour, K., Hannaford, P., Heagerty, A., Khan, M. A., Kitchener, H., ... Buckley, H. (2002). Oestrogen therapy for prevention of reinfarction in postmenopausal women: A randomised placebo controlled trial. *Lancet*, *360*, 2001-2008.
- Chiba, Y. (2009). Bounds on causal effects in randomized trials with noncompliance under monotonicity assumptions about covariates. *Statistics in Medicine*, *28*, 3249-3259. <http://dx.doi.org/10.1002/sim.3724>
- Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, *58*, 21-29. <http://dx.doi.org/10.1111/j.0006-341X.2002.00021.x>
- Heitjtan, D. F., Moskowitz, A. J., & Whang, W. (1991). Bayesian estimation of costeffectiveness ratios from clinical trials. *Health Economics*, *8*, 191-201.
- Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational Studies. *American Political Science Review*, *105*(4), 765-789. <http://dx.doi.org/10.1017/S0003055411000414>
- Jin, H., & Rubin, D. B. (2008). Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*, *103*(481), 101-111. <http://dx.doi.org/10.1198/016214507000000347>
- Little, R. J. A., Long, Q., & Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics*, *65*, 640-649. <http://dx.doi.org/10.1111/j.1541-0420.2008.01066.x>
- Long, Q., Little, R. J. A., & Lin, X. (2010). Estimating causal effects in trials involving multitreatment arms subject to non-compliance: a Bayesian framework. *Journal of the Royal Statistical Society, C59*(3), 513-531. <http://dx.doi.org/10.1111/j.1467-9876.2009.00709.x>
- Ma, Y., Roy, J., & Marcus, B. (2011). Causal models for randomized trials with two active treatments and continuous compliance. *Statistics in Medicine*, *30*, 2349-2362. <http://dx.doi.org/10.1002/sim.4296>
- Odondi, L., & McNamee, R. (2013). Applying optimal model selection in principal stratification for causal inference. *Statistics in Medicine*, *32*, 1815-1828. <http://dx.doi.org/10.1002/sim.5649>
- Robins, J. M., Rotnitzky, A., & Vansteelandt, S. (2007). Discussions of "Principal stratification designs to estimate input data missing due to death". *Biometrics*, *63*(3), 650-653. http://dx.doi.org/10.1111/j.1541-0420.2007.00847_2.x

- Roy, J., Hogan, J. W., & Marcus, B. H. (2008). Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics*, *9*(2), 277-289.
- Symons, M. J., & Moore, D. T. (2002). Hazard rate ratio and prospective epidemiological studies. *Journal of Clinical Epidemiology*, *55*, 893-899.
- Taylor, L., & Zhou, X. H. (2009). Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics*, *65*, 88-95. <http://dx.doi.org/10.1111/j.1541-0420.2008.01023.x>
- VanderWeele, T. J. (2011). Principal stratification - uses and limitations. *The International Journal of Biostatistics*, *7*(1), 1-16. <http://dx.doi.org/10.2202/1557-4679.1329>
- White, I. R. (2005). Uses and limitations of randomization-based efficacy estimators. *Statistical Methods in Medical Research*, *14*, 327-347.
- White, I. R., & Pocock, S. J. (1996). Statistical reporting of clinical trials with individual changes from allocated treatment. *Statistics in Medicine*, *15*, 249-262.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).