Assessing Causal Effects in the Presence of Treatment Switching through Principal Stratification

Alessandra Mattei
alessandra.mattei@unifi.it
University of Florence

Joint work with
Fabrizia Mealli (University of Florence) and Peng Ding (UC Berkeley)

BIRS Workshop
Emerging Challenges for Statistics and Data Sciences:
Complex Data with Missingness, Measurement Errors, and High Dimensionality
May 22 - 27, 2022

Randomized Clinical Trials with Treatment Switching

- Treatment switching is an intercurrent event commonly occurring in clinical trials designed to assess the effect of a treatment on the incidence of a disease
- Often switching is a prognosis-related choice
- ICH E9(R1) addendum provides guidelines on estimands and sensitivity analyses in clinical trials with treatment switching *(ICH, 2019)*
- There are various types of switching possibilities:
  - ✓ Control subjects may be allowed to start taking the active treatment
  - ✓ Treated subjects may be allowed to stop taking the active treatment
  - ✓ Subjects are allowed to start a non-trial treatment
- Focus on clinical trials with one-sided switching behavior

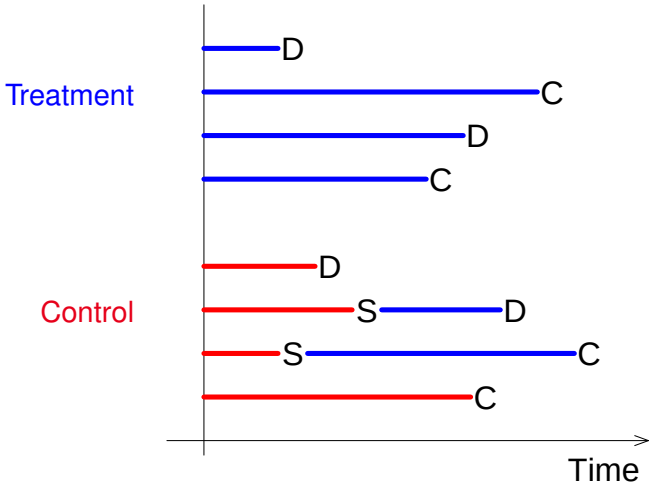
Motivating Study: Concorde Clinical Trial

(Concorde Coordinating Committee, 1994)

- Randomized controlled clinical trial involving patients with asymptomatic HIV infection
- Treatment variable: Immediate versus deferred treatment with zidovudine
  - ✓ In the control arm, treatment with zidovudine was deferred until the onset of symptoms of HIV/AIDS
- Outcome: Time-to-disease progression (time to ARC or AIDS) or death
- Some patients in the deferred arm switched to the active treatment starting zidovudine before the onset of symptoms of HIV/AIDS on the basis of low CD4 cell counts and other evidences of disease progression
- Synthetic data-set closely mimicking the Concorde trial *(White et al., 2002)*
  - ✓ The synthetic Concorde data do not include any pre-treatment variable
  - ✓ $N = 1\,000$ patients: $N/2 = 500$ patients are randomly assigned to immediate treatment with zidovudine; and $N/2 = 500$ patients are randomly assigned to deferred treatment with zidovudine

Data Structure

Randomized to:



Observed Synthetic Concorde Data

- Treatment actually assigned

  $Z_i = 1$ (Immediate zidovudine) and $Z_i = 0$ (Deferred zidovudine)

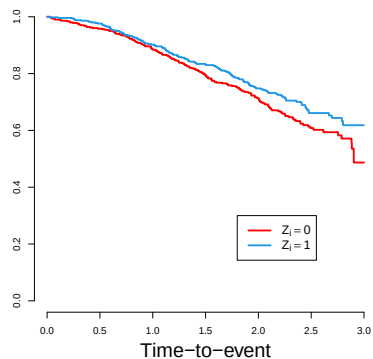
- Let Y_i^{obs} and S_i^{obs} denote the survival time and the switching time under the actual treatment assigned without censoring
- The survival time and the switching time are subject to censoring
  - ✓ The trial lasted 3 years, with staggered entry over the first 1.5 years
- Censoring time: $C_i \in [1.5, 3]$
- Observed survival time: $\tilde{Y}_i^{\text{obs}} = \min\{Y_i^{\text{obs}}, C_i\}$
- Observed switching status
  - ✓ For a patient i with $Z_i = 1$, $\tilde{S}_i^{\text{obs}} = S_i^{\text{obs}} = \overline{\mathbb{S}}$, where $\overline{\mathbb{S}}$ is a non-real value
  - ✓ For a patient i with $Z_i = 0$:

$$\tilde{S}_i^{\text{obs}} = \begin{cases} S_i^{\text{obs}} & \text{if } S_i^{\text{obs}} \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} \leq C_i \\ C_i & \text{if } (S_i^{\text{obs}} \in \mathbb{R}_+ \text{ and } S_i^{\text{obs}} > C_i) \text{ or } S_i^{\text{obs}} = \overline{\mathbb{S}} \end{cases}$$

Synthetic Concorde Data: Descriptive Statistics

Variable	Mean		
	All	$Z_i = 0$	$Z_i = 1$
Sample size	(1000)	(500)	(500)
Treatment assignment (Z_i)	0.5	0	1
$\mathbb{I}\{\tilde{S}_i^{\text{obs}} = C_i^{\text{obs}}\}$	–	0.62	–
Switching time $(\tilde{S}_i^{\text{obs}})$	–	1.55	–
$\mathbb{I}\{\tilde{Y}_i^{\text{obs}} = C_i^{\text{obs}}\}$	0.69	0.66	0.71
Survival time $(\tilde{Y}_i^{\text{obs}})$	1.93	1.89	1.97

Survival functions by assignment Z_i:
kaplan-Meier estimates



Existing Approaches to Treatment Switching

- Intention-to-Treat analysis
- Hypothetical strategy: Reconstructing the outcome a unit would have had if s/he had not switched *(ICH, 2019)*
  - ✓ Naive approaches: Censoring at switch (as-treated analysis); Excluding switchers (per-protocol analysis); Treatment as a time-varying covariate *(See Morden et al., 2011 for a review)*
  - ✓ More sophisticated approaches: Rank-preserving structural failure time model *(Robins and Tsiatis 1991; Robins 1994; White et al. 1997, 1999; White 2006)*; Inverse-probability of censoring weighting *(Robins and Finkelstein, 2000)* and Marginal structural models *(Hernán et al., 2000)*
- Time-varying treatment approach: Clinical trials with treatment switching as longitudinal causal studies with a time-varying treatment *(Petersen et al., 2014)*

Our Contribution

- We propose to re-define the problem of treatment switching using principal stratification *(Frangakis and Rubin 2002)*
  - ✓ The principal stratification approach is recognized in the ICH E9(R1) addendum as a strategy to deal with intercurrent events
- Causal estimands: principal causal effects for patients belonging to subpopulations defined by the switching behavior under the control treatment
  - ✓ Allow switching time to be nonignorable and to characterize treatment effect heterogeneity w.r.t. switching time
- Treatment switching can be viewed as a general form of noncompliance
  - ✓ Non-switchers are a specific type of compliers, because they will be exposed to treatment and control according to the protocol
- We use a Bayesian approach for inference, which allows us to properly take into account that
  - ✓ switching happens in continuous time generating a continuum of principal strata;
  - ✓ switching time is not defined for units who never switch in a particular study; and
  - ✓ both survival time and switching time are subject to censoring

Treatment Switching with Censoring: Potential Outcomes

- Patients: $i = 1, \ldots, N$
- Binary treatment: $z \in \{0, 1\} = \{\text{Control Treatment}, \text{Active Treatment}\}$
- The Stable Unit Treatment Value Assumption (SUTVA) is assumed
- $Y_i(z)$ = Survival time given assignment to treatment z, $z = 0, 1$
  - ✓ $Y_i(z)$ is a positive real number and may be right censored
- $C_i(z)$ = Censoring time given assignment to treatment z, $z = 0, 1$
  - ✓ Assumption: For $i = 1, \ldots, N$, $C_i(0) = C_i(1) = C_i$
- $S_i(z) =$ Switching status given assignment to treatment z, $z = 0, 1$
  - ✓ $S_i(1) = \overline{\mathbb{S}}$ and $S_i(0) \in \mathbb{R}_+ \cup \{\overline{\mathbb{S}}\}$
  - ✓ $S_i(0)$ might be right censored with censoring time C_i
- Natural constraint: $S_i(0) \leq Y_i(0)$, the switching time is censored by death with censoring event defined by $Y_i(0)$

Principal Stratification w.r.t. Switching Behavior

- The switching behavior is defined by $S_i(0) \in \mathbb{R}_+ \cup \{\overline{\mathbb{S}}\}$
- Basic principal strata
  - ✓ Non-switchers = $\{i : S_i(0) = \overline{\mathbb{S}}\}$: Units who would not switch to the active treatment if assigned to control no matter how long the follow-up is
  - ✓ Switchers = $\{i : S_i(0) = s, s \in \mathbb{R}_+\}$: Units who would switch to the active treatment if assigned to control at a given time point $s \in \mathbb{R}_+$
- All switchers = $\cup_{s \in \mathbb{R}_+} \{i : S_i(0) = s\}$

Treatment Switching with Censoring: Principal Causal Effects

- Average principal causal effects

$$ACE(s) = \mathbb{E}\left[Y_i(1) \mid S_i(0) = s\right] - \mathbb{E}\left[Y_i(0) \mid S_i(0) = s\right], \quad (s \in \{\overline{\mathbb{S}}\} \cup \mathbb{R}_+)$$

- Distributional principal causal effects for non-switchers

$$DCE(y \mid \overline{\mathbb{S}}) = P\left\{Y_i(1) > y \mid S_i(0) = \overline{\mathbb{S}}\right\} - P\left\{Y_i(0) > y \mid S_i(0) = \overline{\mathbb{S}}\right\}, \qquad (y \in \mathbb{R}_+)$$

- Conditional distributional principal causal effects for switchers

$$\begin{aligned} cDCE(y \mid s) &= P\left\{Y_i(1) > y \mid Y_i(1) \geq S_i(0), S_i(0) = s\right\} - P\left\{Y_i(0) > y \mid Y_i(1) \geq S_i(0), S_i(0) = s\right\} \\ &= P\left\{Y_i(1) > y \mid Y_i(1) \geq s, S_i(0) = s\right\} - P\left\{Y_i(0) > y \mid Y_i(1) \geq s, S_i(0) = s\right\}, \end{aligned}$$

$(y, s \in \mathbb{R}_+)$

- If $Y_i(1) \geq Y_i(0)$, then $Y_i(1) \geq S_i(0)$ and

$$cDCE(y \mid s) = DCE(y \mid s) \equiv P\left\{Y_i(1) > y \mid S_i(0) = s\right\} - P\left\{Y_i(0) > y \mid S_i(0) = s\right\} \quad (y, s \in \mathbb{R}_+)$$

with $cDCE(y \mid s) = DCE(y \mid s) = 0$ for $y \leq s$

Observed Data Pattern and Possible Latent Principal Strata

Z_i	$\tilde{S}_i^{\text{obs}}$	$\tilde{Y}_i^{\text{obs}}$	Principal strata	Principal stratum label
0	C_i	$Y_i^{\text{obs}} \in [0, C_i)$	$\{i : S_i(0) = \overline{\mathbb{S}}\}$	Non-switchers
0	$S_i^{\text{obs}} \leq C_i$	$Y_i^{\text{obs}} \in [S_i^{\text{obs}}, C_i]$	$\{i : S_i(0) = S_i^{\text{obs}}\}$	Switchers at time S_i^{obs}
0	$S_i^{\text{obs}} \leq C_i$	C_i	$\{i : S_i(0) = S_i^{\text{obs}}\}$	Switchers at time S_i^{obs}
0	C_i	C_i	$\{i : S_i(0) = \overline{\mathbb{S}}\}$ or $\{i : S_i(0) = s \in (C_i, +\infty)\}$	Non-switchers or Switchers at some time $s > C_i$
1	$\overline{\mathbb{S}}$	$Y_i^{\text{obs}} \in [0, C_i]$	$\{i : S_i(0) = \overline{\mathbb{S}}$ or $S_i(0) \in \mathbb{R}_+\}$	Non-switchers or Switchers
1	$\overline{\mathbb{S}}$	C_i	$\{i : S_i(0) = \overline{\mathbb{S}}$ or $S_i(0) \in \mathbb{R}_+\}$	Non-switchers or Switchers

Identification Issues under Randomization

- X_i: Vector of pre-treatment covariates
- Completely Randomized Experiment

$$P\{Z_i \mid S_i(0), Y_i(0), Y_i(1), C_i, X_i\} = P\{Z_i\}$$

- Ignorability of the Censoring Mechanism

$$P\{C_i \mid S_i(0), Y_i(0), Y_i(1), X_i\} = P\{C_i\}$$

- Randomization and ignorability of the censoring mechanism help inference, but the identification of average and distributional principal causal effects requires further structural and/or distributional assumptions

Bayesian Approach to Inference

- The Bayesian approach does not require full identification
  - ✓ “Weak identifiability” of partially identified parameters
- The Bayesian approach allows us to deal with all complications – missing data, truncation by death, censoring – simultaneously in a natural way
- In Bayesian analysis inferences are directly interpretable in probabilistic terms

Bayesian Principal Stratification

Under exchangeability, randomization, and ignorability of censoring:

$$
\begin{aligned}
& P\{\mathbf{C}, \boldsymbol{S}(0), \boldsymbol{Y}(0), \boldsymbol{Y}(1), \boldsymbol{X}\} \\
&= \int \prod_{i=1}^{n} P\{C_i, S_i(0), Y_i(0), Y_i(1), X_i \mid \boldsymbol{\theta}\}\, P(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \\
&= \int \prod_{i=1}^{n} P\{X_i \mid \boldsymbol{\theta}\}\, P\{C_i \mid X_i; \boldsymbol{\theta}\}\, P\{S_i(0) \mid C_i, X_i; \boldsymbol{\theta}\} \times \\
&\qquad P\{Y_i(0) \mid S_i(0), C_i, X_i; \boldsymbol{\theta}\}\, P\{Y_i(1) \mid Y_i(0), S_i(0), C_i, X_i; \boldsymbol{\theta}\}\, P(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \\
&\propto \int \prod_{i=1}^{n} P\{S_i(0) \mid X_i; \boldsymbol{\theta}\}\, P\{Y_i(0) \mid S_i(0), X_i; \boldsymbol{\theta}\}\, P\{Y_i(1) \mid Y_i(0), S_i(0), X_i; \boldsymbol{\theta}\}\, P(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}
\end{aligned}
$$

Bayesian Approach to Inference: Parametric Assumptions

- Sub-model for the Switching Behavior: A two-part model

$$\pi(X_i) = \mathbb{E}[\mathbb{I}\{S_i(0) = \overline{\mathbb{S}} \mid X_i\}] = P\left\{S_i(0) = \overline{\mathbb{S}} \mid X_i\right\} = \frac{\exp(\eta_0 + X_i'\boldsymbol{\eta})}{1 + \exp(\eta_0 + X_i'\boldsymbol{\eta})} \qquad \eta_0 \in \mathbb{R}, \boldsymbol{\eta} \in \mathbb{R}^K,$$

and

$$(S_i(0) \mid S_i(0) \in \mathbb{R}_+, X_i) \sim \text{Weibull}\left(\alpha_S, \beta_S + X_i'\boldsymbol{\eta}_S\right), \quad \alpha_S > 0, \beta_S \in \mathbb{R}, \boldsymbol{\eta}_S \in \mathbb{R}^K$$

- Sub-models for $Y_i(0)|S_i(0), X_i$,

$$\left(Y_i(0) \mid S_i(0) = \overline{\mathbb{S}}, X_i\right) \sim \text{Weibull}\left(\bar{\alpha}_Y, \bar{\beta}_Y + X_i'\bar{\boldsymbol{\eta}}_Y\right),$$
$$(Y_i(0) \mid S_i(0) \in \mathbb{R}_+, X_i) \sim S_i(0) + \text{Weibull}\left(\alpha_Y, \beta_Y + \lambda_0 \log(S_i(0)) + X_i'\boldsymbol{\eta}_Y\right),$$

with $\bar{\alpha}_Y > 0, \bar{\beta}_Y \in \mathbb{R}, \bar{\boldsymbol{\eta}}_Y \in \mathbb{R}^K$ and $\alpha_Y > 0, \beta_Y, \lambda_0 \in \mathbb{R}, \boldsymbol{\eta}_Y \in \mathbb{R}^K$

- Sub-models for $Y_i(1)|S_i(0), Y_i(0), X_i$,

$$\left(Y_i(1) \mid S_i(0) = \overline{\mathbb{S}}, Y_i(0), X_i\right) \sim \kappa Y_i(0) + \text{Weibull}\left(\bar{\nu}_Y, \bar{\gamma}_Y + X_i'\bar{\boldsymbol{\zeta}}\right),$$
$$(Y_i(1) \mid S_i(0) \in \mathbb{R}_+, Y_i(0), X_i) \sim \kappa Y_i(0) + \text{Weibull}\left(\nu_Y, \gamma_Y + \lambda_1 \log(S_i(0)) + X_i'\boldsymbol{\zeta}\right),$$

with $\kappa \in [0, 1]$, $\bar{\nu}_Y > 0, \bar{\gamma}_Y \in \mathbb{R}, \bar{\boldsymbol{\zeta}} \in \mathbb{R}^K$ and $\nu_Y > 0, \gamma_Y, \lambda_1 \in \mathbb{R}, \boldsymbol{\zeta} \in \mathbb{R}^K$

Identification of Some Model Parameters

Dependence between $Y_i(1)$ and $Y_i(0)$

- The parameter κ characterizes the dependence between $Y_i(1)$ and $Y_i(0)$ given $S_i(0)$ and X_i
  - ✓ If $\kappa = 0$ then $Y_i(1) \perp Y_i(0) \mid S_i(0), X_i$ and If $\kappa = 1$ then $Y_i(1) \geq Y_i(0)$
- The parameter κ can be viewed as a sensitivity parameter

Association between $Y_i(1)$ and $S_i(0)$

- The parameter λ_1 describes the association between $Y_i(1)$ and $S_i(0)$ given $Y_i(0)$ and X_i for switchers
  - ✓ Because $S_i(0)$ is never observed for treated units, the observed data provide no information about the association between $Y_i(1)$ and $S_i(0)$ given $Y_i(0)$ and X_i
- Parametric assumption: $\lambda_0 = \lambda_1 \equiv \lambda$
  - ✓ Because $S_i(0)$ and $Y_i(0)$ are jointly observed for some control patients, we have some information on λ

Sensitivity Checks

- The parameters λ and κ are not identifiable nonparametrically
- Under our parametric assumptions, λ and κ enter the observed data likelihood, and thus enter the Bayesian posterior inference
- Sensitivity analysis with respect to the prior specification for λ
- Sensitivity analysis by varying κ within the range $[0, 1]$

Prior Distribution and Posterior Distribution

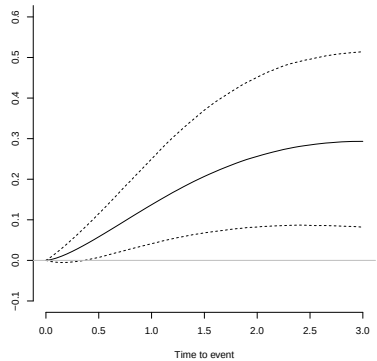
- We assume that the parameters are a priori independent
- Prior distribution
  - ✓ Normal prior distributions (mean $= 0$ and SD $= 100$) for the parameters of the logistic regression model for the probability of being a non-switcher, and for the intercept and the slope parameters of the Weibull distributions
  - ✓ Gamma prior distributions with parameters 1 and $10\,000$ for the shape parameters of the Weibull distributions
  - ✓ Normal and uniform prior distributions for the parameter λ
  - ✓ Dirac delta priors for κ concentrated at a pre-fixed value $\kappa_0 \in [0, 1]$
- Posterior distribution: MCMC Algorithm with Data Augmentation

Synthetic Concorde Clinical Data: Bayesian Principal Stratification Analysis ($\kappa = 0$)

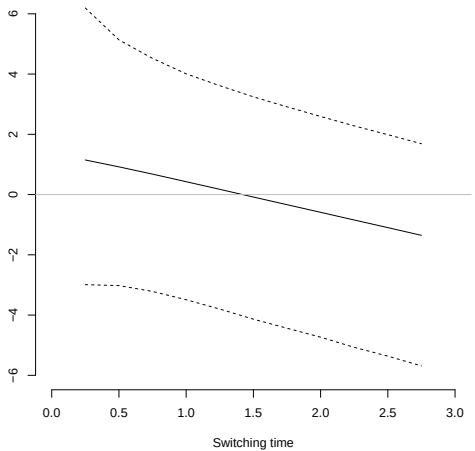
Posterior medians and 95% posterior credible intervals for principal causal effects for non-switchers

Estimand	Median	95% PCI 0.025	0.975
$\mathbb{E}[Y_i(0) \mid S_i(0) = \overline{\mathbb{S}}]$	2.02	1.44	2.97
$\mathbb{E}[Y_i(1) \mid S_i(0) = \overline{\mathbb{S}}]$	3.85	2.41	6.94
$ACE(\overline{\mathbb{S}})$	1.78	0.39	4.78

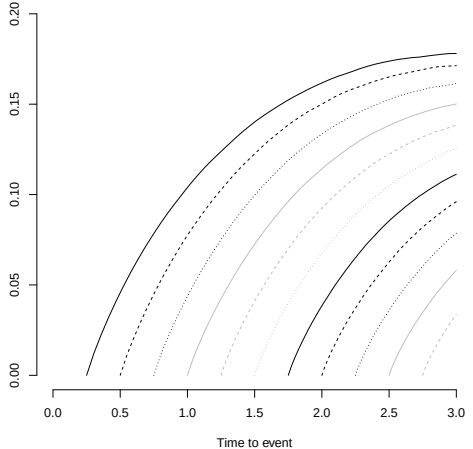
$DCE(y \mid \overline{\mathbb{S}})$



Posterior medians and 95% PCI of $ACE(s)$, $s \in \mathbb{R}_+$



Posterior medians of $cDCE(y \mid s)$ for $s = 0.25, 0.50, \ldots, 2.50, 2.75$



Sensitivity Analyses and Bayesian Posterior Predictive P-Values

- Sensitivity Analyses: Inference appears to be robust w.r.t. the prior specification for λ; some sensitivity w.r.t. the value of κ

- Bayesian PPPVs

Variable	Deviance	Signal	Noise	Signal to noise
Survival time	0.810			
Non-Switchers		0.333	0.542	0.329
Switchers		0.429	0.725	0.372
Switching time	0.478	0.398	0.336	0.568

$PPPV$ for BIC : 0.553

Discussion

- Clinical trials with treatment discontinuation
- The Role of the Pre-treatment Covariates
  - ✓ Conditioning on covariates makes structural and parametric assumptions more credible
  - ✓ Covariates usually lead to more precise inferences
  - ✓ In the principal stratification analysis, relevant information could also be obtained looking at the distribution of baseline characteristics within each principal stratum
- Extention: Treatment switching with non-ignorable censoring

References

Concorde Coordinating Committee. (1994). Concorde: MRC / ANRS randomised double blind controlled trial of immediate and deferred zidovudine in symptom-free hiv infection. Lancet 343, 871-881.

Frangakis C.E., Rubin D.B. (2002). Principal stratification in causal inference. Biometrics 58, 191-199.

Hernan M.A., Brumback B., Robins J.M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. Epidemiology 1, 561-570.

ICH. (2019). Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, E9(R1).

Morden J.P., Lambert P.C., Latimer N., Abrams K.R., Wailoo A.J. (2011). Assessing methods for dealing with treatment switching in randomised controlled trials: a simulation study. Med Res Methodol 1(4), Article number 4.

Robins J.M., Tsiatis A.A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun. Stat. -T&M 20(8), 2609-2631.

Robins J.M., Finkelstein D.M. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. Biometrics 56(3), 779-788.

White I.R., Walker S., Babiker A.G., Darbyshire J.H. (1997). Impact of treatment changes on the interpretation of the concorde trial. AIDS 11(8), 999-1006.

White I.R., Babiker A.G., Walker S., Darbyshire J.H. (1999). Randomization-based methods for correcting for treatment changes: examples from the concorde trial. Stat Med 18(19), 2617-2634.

White I.R., Walker S., Babiker A. (2002). strbee: Randomization-based efficacy estimator. Stata J 2(2), 140-150.