Evaluating Causal Effects on Time-to-Event Outcomes in an RCT in Oncology with Treatment Discontinuation due to Adverse Events

Fabrizia Mealli

Fabrizia.Mealli@eui.eu

European University Institute

Joint work with Veronica Ballerini (HSPH), Alessandra Mattei (Unifi), Björn Borkamp (Novartis), Craig Wang (Novartis), Yufen Zhang (Novartis)

Workshop SISBAYES 2025

September 4-5, 2025, Padova, Italy

Randomized Clinical Trials with Treatment Discontinuation

- In randomized controlled trials (RCTs), patients sometimes discontinue study treatments prematurely due to reasons such as Adverse Events (AE)
- The addendum to the E9 guideline on 'Statistical principles in clinical trials', released by the International Council of Harmonization *(ICH, 2019)*, calls these events intercurrent events
- Tripartite estimand strategy *(Akacha et al., 2017)*:
  - ✓ Treatment effect for patients who adhere to the treatment for its intended duration
  - ✓ Proportion of patients who discontinue the treatment prematurely
  - ✓ Effect for patients who discontinue the treatment prematurely

- Rubin D.B. (1978) Bayesian Inference for Causal Effects: The Role of Randomization, *The Annals of Statistics*, 6 (1)
- Li F., Ding P., Mealli F. (2023). Bayesian causal inference: a critical review, *Philosophical Transactions A* 381(2247)

Motivating Study: A RCT in Oncology (Novartis Study)

- Randomized controlled clinical trial involving oncological patients
- Treatment variable: New treatment versus Standard of Care (SOC)
  - ✓ New treatment: New investigational drug + SOC
  - ✓ Standard / control treatment: SOC
- Outcome: Progression-free survival (Time from randomization until either disease progression or death)
- (One-sided) treatment discontinuation: Patients in the new treatment arm who incur AEs are allowed to discontinue the new investigational drug, but continue on SOC

✓ Treatment discontinuation can be viewed as a general form of noncompliance

- Treatment discontinuation is an intercurrent event because it occurs after treatment initiation, breaking initial randomization

Observed Data and Data Structure

- A three-dimensional vector of covariates: $X_i = (X_{i1}, X_{i2}, X_{i3})$ where X_{i1} is a risk score of progression; X_{i2} is a binary indicator for advanced metastatic status; and X_{i3} is binary indicator for high disease burden
- Treatment actually assigned: $Z_i = 1$ (Investigational drug + SOC) and $Z_i = 0$ (SOC)
- Let Y_i^{obs} and D_i^{obs} denote the survival time and the discontinuation time under the actual treatment assigned without censoring (in months)
- Duration of the study: 33 months with staggered patients' entry during the first 23 months
  - ✓ Censored progression-free survival and discontinuation time
- Censoring time: $C_i \in [10, 33]$

- Censored survival time:

$$\tilde{Y}_i^{\text{obs}} = \min\{Y_i^{\text{obs}}, C_i\}$$

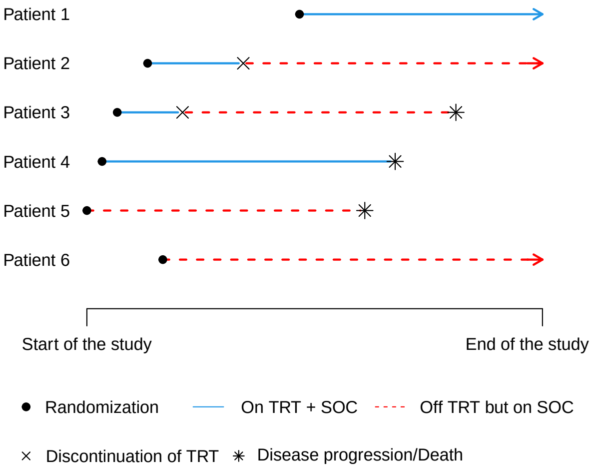
- Observed discontinuation time
  - ✓ For a patient i with $Z_i = 0$,

$$\tilde{D}_i^{\text{obs}} = D_i^{\text{obs}} = \overline{\mathbb{D}}_i$$

    where $\overline{\mathbb{D}}$ is a non-real value

  - ✓ For a patient i with $Z_i = 1$:

$$\tilde{D}_i^{\text{obs}} = \begin{cases} D_i^{\text{obs}} & \text{if } D_i^{\text{obs}} \in \mathbb{R}_+, D_i^{\text{obs}} \le C_i \\ C_i & \text{if } (D_i^{\text{obs}} \in \mathbb{R}_+, D_i^{\text{obs}} > C_i) \\ & \text{ or } D_i^{\text{obs}} = \overline{\mathbb{D}} \end{cases}$$



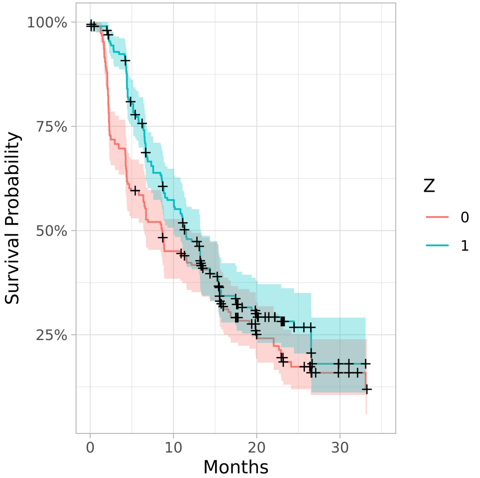
Synthetic Data: Descriptive Statistics

Sample size: $n = 389, n_1 = 200, n_0 = 189$

Variable	Mean	Mean (Proportion)	*SD*
Treatment assignment (Z_i)	0.51	$(200/389)$	–
$\mathbb{I}\{D_i^{\text{obs}} < C_i\}$	0.32	$(64/200)$	–
Discontinuation time $(\tilde{D}_i^{\text{obs}})^*$	4.85		6.73
$\mathbb{I}\{Y_i^{\text{obs}} < C_i\}$	0.71	$(278/389)$	–
Survival time $(\tilde{Y}_i^{\text{obs}})^*$	8.44		6.43
Covariates			
(Std) Risk score (X_{i1})	0.00		1.00
Metastatic status (X_{i2})	0.49	$(191/389)$	–
High disease burden (X_{i3})	0.32	$(124/389)$	–

*Means over patients who experience the event

Survival functions by assignment Z_i: kaplan-Meier estimates



Our Contribution

- We propose to re-define the problem of treatment discontinuation 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 discontinuation behavior under treatment
  - ✓ Allow discontinuation behavior to be nonignorable and to characterize treatment effect heterogeneity w.r.t. discontinuation behavior
- We use a Bayesian approach for inference, which allows us to properly take into account that
  - ✓ The discontinuation time is either not defined for patients who would never discontinue or continuous generating a continuum of principal strata; and
  - ✓ Both survival time and discontinuation time are subject to censoring

Treatment Discontinuation with Censoring: Potential Outcomes

- Patients: $i = 1, \ldots, n$
- Binary treatment: $z \in \{0, 1\} = \{\text{SOC}, \text{New drug + SOC}\}$
- The Stable Unit Treatment Value Assumption (SUTVA) is assumed
- $Y_i(z)$ = Survival time given assignment to treatment z, with $Y_i(z) \in \mathbb{R}_+$, $z = 0, 1$,
- $D_i(1) =$ Discontinuation time under the new treatment with $D_i(1) \in \mathbb{R}_+ \cup \{\overline{\mathbb{D}}\}$
- $D_i(1) \leq Y_i(1)$: the discontinuation time is censored by death with censoring event defined by $Y_i(1)$
- $C_i(z)$ = Censoring time given assignment to treatment z, $z = 0, 1$
  - ✓ Assumption: For $i = 1, \ldots, n$, $C_i(0) = C_i(1) = C_i$

Principal Stratification w.r.t. Discontinuation Behavior

- The discontinuation behavior is defined by $D_i(1) \in \mathbb{R}_+ \cup \{\overline{\mathbb{D}}\}$
- Basic principal strata
  - ✓ Never-discontinuing (ND) patients = $\{i : D_i(1) = \overline{\mathbb{D}}\}$: Patients who would not discontinue the new investigational drug if assigned to it no matter how long the follow-up is
  - ✓ Discontinuing (D) patients = $\{i : D_i(1) = d, d \in \mathbb{R}_+\}$: Patients who would discontinue the new investigational drug if assigned to it at a given time point $d \in \mathbb{R}_+$
- All D patients = $\cup_{d \in \mathbb{R}_+} \{i : D_i(1) = d\}$

Principal Causal Effects for Survival Outcomes

- Distributional principal causal effects for
  - ✓ ND patients:

$$DCE_{\mathrm{ND}}(y) = P\left\{Y_i(1) > y \mid D_i(1) = \overline{\mathbb{D}}\right\} - P\left\{Y_i(0) > y \mid D_i(1) = \overline{\mathbb{D}}\right\}, \qquad y \in \mathbb{R}_+$$

  - ✓ D patients:

$$DCE_{\mathrm{D}}(y \mid d) = P\left\{Y_i(1) > y \mid D_i(1) = d\right\} - P\left\{Y_i(0) > y \mid D_i(1) = d\right\}, \qquad y, d \in \mathbb{R}_+$$

- Restricted mean survival time principal causal effects

$$RMSTE_{\mathrm{ND}}(\tau) = \int_0^\tau DCE_{\mathrm{ND}}(y)\,\mathrm{d}y \qquad \textit{and} \qquad RMSTE_{\mathrm{D}}(\tau \mid d) = \int_0^\tau DCE_{\mathrm{D}}(y \mid d)\,\mathrm{d}y$$

for $\tau, d \in \mathbb{R}_+$

Principal Causal Effects and Overall Causal Effects

- Let $\pi_{\overline{\mathbb{D}}}$ be the probability that a patient is a ND patient
- Overall distributional causal effects

$$DCE(y) = P\{Y_i(1) > y\} - P\{Y_i(0) > y\} = \pi_{\overline{\mathbb{D}}} DCE_{\mathrm{ND}}(y) + (1 - \pi_{\overline{\mathbb{D}}}) DCE_{\mathrm{D}}(y) \quad y \in \mathbb{R}_+$$

where

$$DCE_{\mathrm{D}}(y) = \int_{\mathbb{R}_+} DCE_{\mathrm{D}}(y \mid d) f_{D(1)}(d)\, \mathrm{d}d$$

- Overall restricted mean survival time effects

$$RMSTE(\tau) = \int_0^{\tau} DCE(y)\, \mathrm{d}y = \pi_{\overline{\mathbb{D}}} RMSTE_{\mathrm{ND}}(\tau) + (1 - \pi_{\overline{\mathbb{D}}}) RMSTE_{\mathrm{D}}(\tau) \quad \tau \in \mathbb{R}_+$$

where

$$RMSTE_{\mathrm{D}}(\tau) = \int_0^{\tau} DCE_{\mathrm{D}}(y)\, \mathrm{d}y = \int_0^{\tau} \int_{\mathbb{R}_+} DCE_{\mathrm{D}}(y \mid d) f_{D(1)}(d)\, \mathrm{d}d\, \mathrm{d}y$$

- Remark: The overall causal effects are ITT effects

Observed Data Pattern and Possible Latent Principal Strata

- Both $Y_i(z)$ and $D_i(1)$ might be right censored with censoring time C_i
- Therefore, we observe

$$\tilde{Y}_i^{\text{obs}} = \min\{Y_i^{\text{obs}}, C_i\} = \min\{Z_i Y_i(1) + (1 - Z_i) Y_i(0), C_i\}$$

and

$$\tilde{D}_i^{\text{obs}} = \begin{cases} \min\{D_i(1), C_i\} & \text{if } D_i(1) \in \mathbb{R}_+ \text{ and } Z_i = 1 \\ C_i & \text{if } D_i(1) = \overline{\mathbb{D}} \text{ and } Z_i = 1 \\ \overline{\mathbb{D}} & \text{if } Z_i = 0 \end{cases}$$

Z_i	$\tilde{D}_i^{\text{obs}}$	$\tilde{Y}_i^{\text{obs}}$	Principal strata	Principal stratum label
1	C_i	$Y_i^{\text{obs}} \in [0, C_i)$	$\{i : D_i(1) = \overline{\mathbb{D}}\}$	ND patients
1	$D_i^{\text{obs}} \leq C_i$	$Y_i^{\text{obs}} \in [D_i^{\text{obs}}, C_i)$	$\{i : D_i(1) = D_i^{\text{obs}}\}$	D patients at time D_i^{obs}
1	$D_i^{\text{obs}} \leq C_i$	C_i	$\{i : D_i(1) = D_i^{\text{obs}}\}$	D patients at time D_i^{obs}
1	C_i	C_i	$\left\{i : D_i(1) = \overline{\mathbb{D}}\right\}$ or $\{i : D_i(1) = d \in (C_i, +\infty)\}$	ND patients or D patients at some time $d > C_i$
0	C_i	$Y_i^{\text{obs}} \in [0, C_i]$	$\left\{i : D_i(1) = \overline{\mathbb{D}} \text{ or } D_i(1) \in \mathbb{R}_+\right\}$	ND or D patients
0	C_i	C_i	$\left\{i : D_i(1) = \overline{\mathbb{D}} \text{ or } D_i(1) \in \mathbb{R}_+\right\}$	ND or D patients

Identification Issues under Randomization

- Completely Randomized Experiment

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

- Ignorability of the Censoring Mechanism

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

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

Bayesian Approach to Inference

- The Bayesian approach can handle weakly or 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

- Joint distribution of all observed and unobserved random variables,
$$(\boldsymbol{X}, \mathbf{C}, \boldsymbol{D}(1), \boldsymbol{Y}(0), \boldsymbol{Y}(1), \boldsymbol{Z}) = [X_i, C_i, D_i(1), Y_i(0), Y_i(1), Z_i]_{i=1}^n,$$
under exchangeability:
$$P\{\boldsymbol{X}, \mathbf{C}, \boldsymbol{D}(1), \boldsymbol{Y}(0), \boldsymbol{Y}(1), \boldsymbol{Z}\} = \int \prod_{i=1}^n P\{X_i, C_i, D_i(1), Y_i(0), Y_i(1), X_i, Z_i \mid \boldsymbol{\theta}\} P(\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}$$
$$= \int \prod_{i=1}^n P\{X_i \mid \boldsymbol{\theta}\} P\{D_i(1) \mid X_i; \boldsymbol{\theta}\} P\{Y_i(0), Y_i(1) \mid D_i(1), X_i; \boldsymbol{\theta}\}$$
$$P\{C_i \mid Y_i(1), Y_i(0), D_i(1), X_i; \boldsymbol{\theta}\} P\{Z_i \mid Y_i(1), Y_i(0), D_i(1), C_i, X_i; \boldsymbol{\theta}\} P(\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}$$
- Under randomization and ignorability of censoring, and conditioning on the empirical distribution of the covariates
$$P\{\boldsymbol{X}, \mathbf{C}, \boldsymbol{D}(1), \boldsymbol{Y}(0), \boldsymbol{Y}(1), \boldsymbol{Z}\} \propto \int \prod_{i=1}^n P\{D_i(1) \mid \boldsymbol{X}_i; \boldsymbol{\theta}\} P\{Y_i(1), Y_i(0) \mid D_i(1), \boldsymbol{X}_i; \boldsymbol{\theta}\} P(\boldsymbol{\theta}) \mathrm{d}\,\boldsymbol{\theta}$$
- Mixed causal effects *(Li, Ding, Mealli, 2023)*

Bayesian Approach to Inference: Parametric Assumptions

- We follow a similar modeling strategy as in *Mattei et al. (2025)*
- Sub-model for the Discontinuation Behavior: A two-part model with

$$\mathbb{I}_i^{\text{ND}} = \mathbb{I}\{D_i(1) = \overline{\mathbb{D}}\} \sim \text{Ber}(\pi(X_i)) \qquad \textit{with} \quad \pi(X_i) = \frac{\exp(\gamma_0 + X_i'\boldsymbol{\gamma})}{1 + \exp(\gamma_0 + X_i'\boldsymbol{\gamma})} \qquad \gamma_0 \in \mathbb{R}, \boldsymbol{\gamma} \in \mathbb{R}^K,$$

$$\left(D_i(1) \mid \mathbb{I}_i^{\text{ND}} = 0, X_i\right) \sim \text{Exp}\left(e^{\beta_D + X_i'\boldsymbol{\eta}_D}\right), \quad \beta_D \in \mathbb{R}, \boldsymbol{\eta}_D \in \mathbb{R}^K$$

- Assumption: $Y_i(1) \perp\!\!\!\perp Y_i(0) \mid D_i(1), X_i, \boldsymbol{\theta}$
- Sub-models for $Y_i(z) \mid D_i(1), X_i$ for *ND* patients, $z = 0, 1$:

$$\left(Y_i(0) \mid \mathbb{I}_i^{\text{ND}} = 1, X_i\right) \sim \text{Exp}\left(e^{\bar{\beta}_0 + X_i'\bar{\boldsymbol{\eta}}_0}\right), \quad \bar{\beta}_0 \in \mathbb{R}, \bar{\boldsymbol{\eta}}_0 \in \mathbb{R}^K$$

$$\left(Y_i(1) \mid \mathbb{I}_i^{\text{ND}} = 1, X_i\right) \sim \text{Exp}\left(e^{\bar{\beta}_1 + X_i'\bar{\boldsymbol{\eta}}_1}\right), \quad \bar{\beta}_1 \in \mathbb{R}, \bar{\boldsymbol{\eta}}_1 \in \mathbb{R}^K$$

- Sub-models for $Y_i(z) \mid D_i(1), X_i$ for D patients, $z = 0, 1$:

$$\left(Y_i(0) \mid \mathbb{I}_i^{\mathrm{ND}} = 0, D_i(1), X_i\right) \sim \mathrm{Exp}\left(e^{\beta_0 + X_i'\boldsymbol{\eta}_0 + \delta \log(D_i(1))}\right),$$
$$\left(Y_i(1) \mid \mathbb{I}_i^{\mathrm{ND}} = 0, D_i(1), X_i\right) \sim \mathrm{TExp}_{D_i(1)}\left(e^{\beta_1 + X_i'\boldsymbol{\eta}_1 + \delta \log(D_i(1))}\right),$$

  with $\beta_0, \delta \in \mathbb{R}, \boldsymbol{\eta}_0 \in \mathbb{R}^K$ and $\beta_1, \delta \in \mathbb{R}, \boldsymbol{\eta}_1 \in \mathbb{R}^K$, and $\mathrm{TExp}_{D_i(1)}$ stands for a left truncated Exponential distribution with truncation parameter $D_i(1)$

- For D patients, the parameter δ describes both the association between $Y_i(1)$ and $D_i(1)$ given X_i as well as the association between $Y_i(0)$ and $D_i(1)$ given X_i
  - ✓ Because $D_i(1)$ is never observed for control patients, the observed data provide no information about the association between $Y_i(0)$ and $D_i(1)$ given X_i
  - ✓ Because $D_i(1)$ and $Y_i(1)$ are jointly observed for some treated patients, we have some information on δ

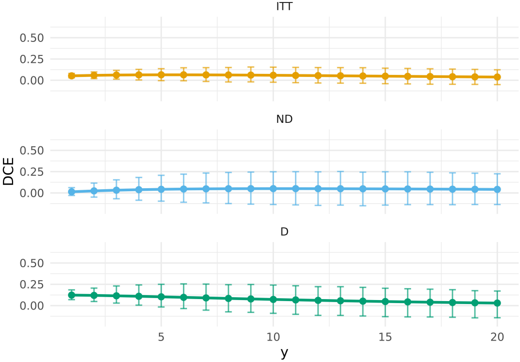
Prior Distribution and Posterior Distribution

- We assume that the parameters are a priori independent
- Prior distribution: Normal prior distributions for the parameters of the logistic regression model for the probability of being a ND patient, and for the intercept and the slope parameters of the Exponential distributions
- Posterior distribution: MCMC Algorithm with Data Augmentation
- Posterior predictive checks: We evaluate the credibility of our parametric assumptions with posterior predictive checks, computing Bayesian posterior predictive p-values Our PPPV suggest that our model adequately reproduce the features of the data reflected in the discrepancy measure

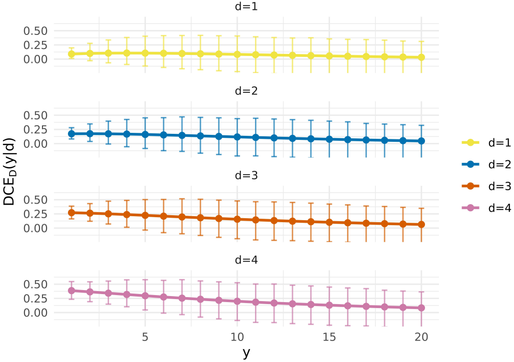
Results for the Synthetic Dataset

Estimand	Posterior Mean	95% HPD
π_{ND}	0.65	$[0.59; 0.70]$

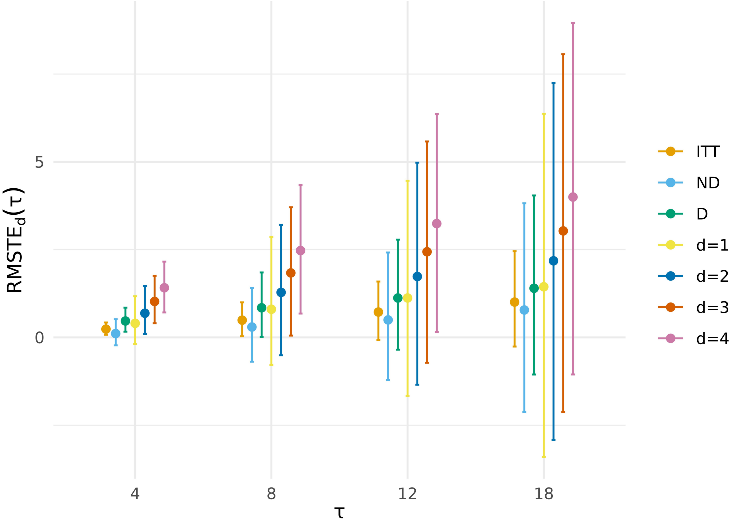
$DCE(y)$



$DCE_{\mathrm{D}}(y \mid d)$



Restricted mean survival time effects at $\tau = 4, 8, 12, 18$ months

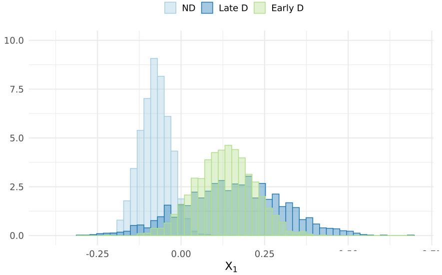


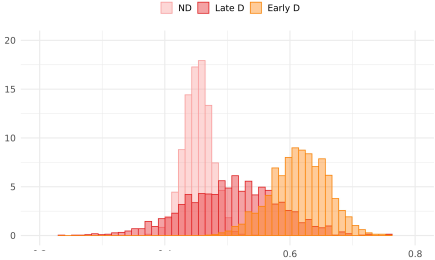
A Characterization of the Latent Principal Strata

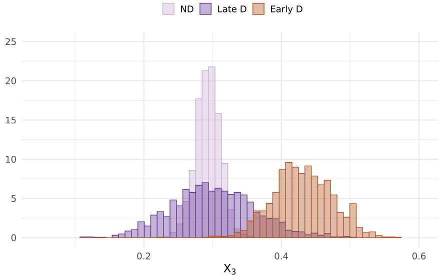
ND patients
$D_i(1) = \overline{\mathbb{D}}$

Patients who discontinue early
$D_i(1) <$ median$(D_i(1)) = 2.24$

Patients who discontinue late
$D_i(1) \geq$ median$(D_i(1)) = 2.24$







Discussion

- Principal stratification approach to clinical trials with one-sided 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
- The Bayesian approach to principal stratification
- Simulation study to investigate the performance of the Bayesian principal stratification approach in repeated sampling

References

Akacha M., Bretz F., Ruberg S. (2017). Estimands in clinical trials – broadening the perspective. Statistics in medicine 36(1), 5-19.

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).

Li F., Ding P., Mealli F. (2023). Bayesian causal inference: a critical review. Philosophical Transactions of the Royal Society A 381(2247), 20220153.

Mattei A., Ding P., Ballerini V., Mealli F. (2025). Assessing causal effects in the presence of treatment switching through principal stratification. Bayesian Analysis.