



IMS Health & Quintiles are now



How causal inference can fit the needs of a clinical trial (well kind of)

Principal stratum approach

Michael O'Kelly, Senior Statistical Director, IQVIA
and Bohdana Ratitch, Lilly

Won't randomisation allow a causal estimand?

- Doesn't randomisation provide “a ‘reasoned basis’ of testing the null hypothesis of no effect without resort to distributional assumptions such as normality” (Fisher)?
- Intercurrent events (ICEs) result in changes to treatment.
 - Trial can morph into no more than a survey whose only inference from randomisation is confined to the mere act of assigning a plan of treatment, a survey whose inference about the treatment regimen itself loses much of its credibility because those ICEs constitute non-randomized changes and distortions of the regimen to be tested.
- => idea of causal inference
 - E.g., a “principal stratum” of interest (obscure term 😊)
- Example of use of the causal-inference framework for an outcome censored by death; and other events and states
- Limitations to the methodologies.

What is a “principal stratum”?

- A principal stratum is the subset of the broader initially randomized population that **would have** a specific status with regard to an intercurrent event (ICE, ICE would/would not occur)...
 - **if they were assigned to a treatment arm** (possibly different from the arm to which they happened to be actually randomized).
- Are completers a principal stratum of “always completers”, for example? All have the status that no ICE occurred to prompt withdrawal, after all.
- Perhaps not, because a subject completing study under the actually assigned treatment A
 - ...might not (be able to) complete the study under treatment B.

What is a “principal stratum”?

- Why worry about what **might have happened if...**?
- Example
 - Treatment A has no side effects; treatment B has side-effects that are associated with, say, 50% withdrawal after a month.
 - Completers: we compare the treatment A subjects vs. the subset of treatment B subjects who are healthy enough and strong enough (or have genetic makeup such that they are able to) to complete the treatment.
 - If true efficacy for treatments A and B is the same in the general target population, which treatment will look better in a completers analysis?

 - A principal stratum approach would
 - › attempt to identify subjects that would complete the study, irrespective of assigned treatment;
 - › i.e. here, maybe end up with a comparison with a cohort of strong, relatively healthier subjects who can tolerate both treatments,
 - ...provides a “level playing field” for the comparison of treatment A vs. treatment B.

Causal framework and potential outcomes (PO)

- Potential outcomes:
 - $Y_i(z), z = 0,1$, is potential outcome that would be observed if a randomly selected patient i received treatment $Z = z$, even if contrary to the fact.
 - Here $Z=0$ means control treatment (e.g. placebo) and $Z=1$ means the experimental treatment
- Potential outcomes are connected with observed outcomes Y (“stable unit treatment value assumption”, SUTVA):
 - $Y_i = Y_i(1)Z_i + Y_i(0)(1 - Z_i)$
- Causal estimands can be defined as expectation over individual treatment contrasts:
 - $\delta = E(Y_i(1) - Y_i(0))$
- In a Randomized Clinical Trial (RCT), by design $Y(0)$ and $Y(1)$ are *independent* of randomized treatment assignment, Z , which allows us to express the causal estimand via observable outcomes Y :

$$\begin{aligned} E(Y(1) - Y(0)) &= E(Y(1)) - E(Y(0)) = \\ E(Y(1)|Z = 1) - E(Y(0)|Z = 0) &= \\ E(Y|Z = 1) - E(Y|Z = 0) \end{aligned}$$

Causal framework and potential outcomes (PO) – Cont.

- Incorporating post-randomization events in POs:
 - $Y_i(z, s)$ is a potential outcome that would be observed if an arbitrary patient i were assigned treatment $Z = z$ and a post-randomization event $S = s$
 - Two types of post-randomization events $S(z)$ that would be observed in a randomly selected patient on treatment $Z = z$ that may need to be accounted for in causal inference:
 - › Secondary outcomes (AE, relapse, death)
 - › Changes in treatment regimen, e.g., rescue, discontinuation from assigned treatment, etc.
- Varying degrees of “counterfactualness”:
 - If S is an outcome, such as AE or death, can we consider an outcome $Y_i(1,1)$, i.e. assuming $S = 1$ given that patient actually had no event, $S = 0$?
 - › Some researchers do not allow *manipulating* with intercurrent/post-randomization events that is, considering all possible hypothetical combinations: $Y_i(1,0), Y_i(1,1), Y_i(0,1), Y_i(0,0)$
 - › Some only focus on “de-facto” outcomes: $Y_i(1, S_i(1))$ and $Y_i(0, S_i(0))$

Example 1: undesirable post-randomization event

- Example: trial in multiple sclerosis
 - Primary outcome measure: confirmed disability progression
 - Post-randomization outcome: $S(z)$ = relapse
- Principal strata defined by a post-randomization outcome:

	$S(1) = 0$	$S(1) = 1$
$S(0) = 0$	Immune (I)	Harmed (H)
$S(0) = 1$	Benefiters (B)	Doomed (D)

PS of interest: Subjects that would not experience relapse regardless of treatment assignment

Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by post-randomization events

Baldur P. Magnusson¹, Heinz Schmidli¹, Nicolas Rouyrre¹, and Daniel O. Scharfstein²

¹Biostatistics & Pharmacometrics, Novartis Pharma AG, Basel, Switzerland

²Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD USA

2018-09-12

Example 2: death that censors the primary outcome measure

- Outcomes planned to be assessed at specific time points, but not ascertainable for all subjects.
 - E.g., quality-of-life assessment cannot be obtained after death
- Analysis based on those who *actually* survived on their assigned treatment (survivors analysis) is not suitable: comparability of treatments per randomization is broken.

	$S(1) = 0$	$S(1) = 1$
$S(0) = 0$	Doomed	Experimental-only survivors
$S(0) = 1$	Control-only survivors	Always survivors

- To restore comparability need to consider only subjects who *would survive on either treatment*:
principal stratum of “always survivors”
- Survivor Average Causal Effect (SACE) (Zhang & Rubin, 2003).
- Since patients are only assigned to one treatment group in a parallel design trial, membership in the “always survivors” stratum needs to be determined via statistical modeling.
- Approach involves important assumptions, including no unmeasured confounders, predictability of survival from baseline characteristics.

Example 3: rescue treatment interferes with regimen of interest

- Suppose estimand is *treatment effect ignoring rescue*
- Outcomes not relevant to the estimand if rescue affects outcome.
- Analysis based on those who did not *actually* take rescue treatment (analysis of subjects who did not take rescue) is not suitable: comparability of treatments per randomization is broken.

	$S(1) = 0$	$S(1) = 1$
$S(0) = 0$	Would always take rescue	Rescue on experimental trt only
$S(0) = 1$	Rescue in control group only	Never take rescue

- To restore comparability need to consider only subjects who *would not take rescue on either treatment*:
principal stratum of “never take rescue”
- Since patients are only assigned to one treatment group in a parallel design trial, membership in the “never take rescue” stratum needs to be determined via statistical modeling.
- Approach involves important assumptions, including no unmeasured confounders, predictability of taking of rescue medication from baseline characteristics.

Stating estimands with ICH E9(R1) strategies using POs

- *Treatment policy*, effect of initial randomization to treatment (sometimes referred to as ITT)
 - $E[Y(1, S(1)) - Y(0, S(0))] = E[Y(1, S(0)) - Y(0, S(1))] \dots = E[Y(1) - Y(0)]$
- *Principal stratification*. E.g. treatment effect in “always compliers”: a subset of patients who would comply regardless of whether assigned to treated or control
 - $E[Y(1) - Y(0) | S(1) = 1, S(0) = 0]$

Stating estimands with ICH E9(R1) strategies using POs

- *While on treatment*, E.g. an estimand based on a dynamic treatment regime, where $\tilde{T}(z)$ is the time on treatment z
 - $E[Y_{\tilde{T}(1)}(1) - Y_{\tilde{T}(0)}(0)]$
- *Hypothetical example 1*: average treatment effect if everyone were treated as planned, even if contrary to the fact
 - $E[Y(1,1) - Y(0,0)]$, or defining $S(z)$ as completing treatment z per protocol, $E[Y(1, S(1)) - Y(0, S(0))]$
- *Hypothetical example 2*: model post-treatment-discontinuation outcomes based on control group outcomes
 - Let \tilde{T}_+ indicate time after end of treatment to end of scheduled follow-up; let T_{EOS} = time at end of scheduled follow-up; define $\tilde{Y}(0) = Y(0,0)$; $\tilde{Y}(1) = Y_{\tilde{T}(1)}(1) + Y_{\tilde{T}_+(1)}(0)$; then $E[\tilde{Y}(1) - \tilde{Y}(0)]$

Estimators for estimands with **principal stratification** strategy

Some practical methods for SACE in the current literature

- Rubin's approach using Multiple Imputation (MI) for survival (Rubin, 1998; 2006)
 - MI takes account of the uncertainty of the model of survival.
- Analysis weighted by probability of survival (Hayden et al., 2005)
 - Method reminiscent of IPW / selection modelling.
- Crude estimate with “investigator-defined” sensitivity parameter (Chiba & VanderWeele, 2011)
 - Actual survivors estimate + strong assumptions + sensitivity parameter set by investigator “to what is plausible”.
- Crude estimate with lower and upper bounds (Colantuoni et al., 2018)
 - Reminiscent of trimmed mean with additional assumptions.

Example implementation of the principal stratum approach, due to Rubin (2006)

Statistical Science
2006, Vol. 21, No. 3, 299–309
DOI: 10.1214/088342306000000114
© Institute of Mathematical Statistics, 2006

Causal Inference Through Potential Outcomes and Principal Stratification: Application to Studies with “Censoring” Due to Death¹

Donald B. Rubin

Abstract. Causal inference is best understood using potential outcomes. This use is particularly important in more complex settings, that is, observational studies or randomized experiments with complications such as non-

Estimators for estimands with **principal stratification** strategy

SACE: Treatment effect in subjects who would survive to time K under either treatment

- Rubin's approach using Multiple Imputation (MI) for survival (Rubin, 1998; 2006)

- Estimate an MI model for survival within each treatment group based on baseline covariates.
- Predict survival of each subject on the opposite treatment. Create M predictions, as in MI.
- In each of M datasets, select subjects who are observed to survive to time T on their assigned treatment and predicted to survive to time X on the opposite treatment.
- Estimate treatment difference in each of M datasets and combine results using Rubin's rules.

- 😊 Provides an estimate of treatment effect at time K on functional outcome alone.
- 😊 MI takes account of the uncertainty of the model of survival.
- 😊 If treatment does not affect survival at time K , SACE can be estimated from actual survivors.
- 😞 Conclusions do not apply to all randomized subjects; difficult to determine in clinical practice who belongs to the stratum.
- 😞 Strong assumption of predictable survival. Death would need to be at least moderately prevalent.
- 😞 Difficulty choosing predictors of survival.

Multiple imputation-based strategy with logistic regression

- Fit a Bayesian logistic regression to observed data (S, Z, X) , where S is the column-vector for symptoms, $Z=\{0,1\}$ is the data column with treatment assignment at randomization and X is n by p -dimensional matrix of baseline covariates; we will denote X_0 as n_0 by p covariate matrix for control and X_1 the n_1 by p matrix for the experimental arm ($n_0 + n_1 = n$).
- For each treatment group z , Model $\tilde{p}(z)$ via $Logit[P(S = 1|X = x_z)] = a_0 + a_2^T x_z$ [z subscripts omitted from the a]
- Make M draws from a posterior distribution of parameters $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2$; For each draw, m
 - compute posterior probability of $S = 1$ for subjects in treatment group z using model for $\tilde{p}(1 - z)$ but given covariates X_z . So, for patients from control and experimental arm with covariate vectors x_{0i} and x_{1j} , the posterior probabilities will be $\tilde{p}(1, x_{0i})$ and $\tilde{p}(0, x_{1j})$, respectively.
 - Impute missing $S(0), S(1)$ by generating Bernoulli random variates from $\tilde{p}(1, x_{0i})$ and $\tilde{p}(0, x_{1j})$, respectively.
 - Based on observed and imputed $S(0), S(1)$, form a subset based on desired principal strata, e.g., $I = (S(0) = 0 \ \& \ S(1) = 0)$
 - Estimate treatment effect within subset I from the current draw $\hat{\delta}_m(I)$
- Compute MI point estimates $\hat{\delta}(I) = M^{-1} \sum_{m=1}^M \hat{\delta}_m(I)$
- Standard errors can be computed by bootstrap or Rubin's rule.

Estimators for estimands with **principal stratification** strategy

SACE: Treatment effect in subjects who would survive to time K under either treatment

- Rubin's approach using Multiple Imputation (MI) for survival (Rubin, 1998; 2006)
 - Difference in means weighted by the estimated probabilities of survival to time K on the opposite treatment.
 - Survival probabilities can be estimated using, e.g., a logistic regression model.
 - Method reminiscent of IPW / selection modelling.
- Analysis weighted by probability of survival (Hayden et al., 2005)

Principal strata risks or disadvantages

- Can we trust the model (e.g. can we trust the model for survival in the example?)
 - i.e. assumes “no unobserved/unmodelled confounders”.
- Could we even provide evidence to support the model’s credibility?

Counterfactuals

- See debate by Dawid, Rubin and others “Causal inference without counterfactuals” in JASA (2000), 95, 407-448
- Objection to counterfactuals cites Popper’s requirement: “How can a proposition be assessed if it’s not falsifiable?”
- The requirement of “no unmeasured confounders”, which is by definition impossible to show has been achieved, will continue to scare all contemplating the use of models for causal inference
 - However, there are reasonable checks for the model used to identify the principal stratum
 - › how well does the model fit the subjects from whom it was estimated (e.g. using a “model building” vs. a “training” data set)?

References

- Dawid A et al. (2000) Causal inference without counterfactuals (with discussion), *Journal of the American Statistical Association* 95 (450) 407-448
- Lipkovich I, Ratitch B, O'Kelly M (2016) Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints, *Pharmaceutical Statistics* 15:216-229

Extra slides

Estimators for estimands with **principal stratification** strategy

SACE: Treatment effect in subjects who would survive to time K under either treatment

- Rubin's approach using Multiple Imputation (MI) for survival (Rubin, 1998; 2006)
- Analysis weighted by probability of survival (Hayden et al., 2005)
- **Crude estimate with sensitivity parameter (Chiba & VanderWeele, 2011)**

- **Treatment effect estimate** = [treatment difference in actual survivors + α]
 - **Sensitivity parameter α** needs to be set by investigator "to what is plausible".

Similar 😊 and ☹️ as before + additional strong assumptions:

- **Monotonicity:** survival probabilities under control treatment are not greater than survival probabilities under the experimental treatment *for all individuals*.
 - Assumes no heterogeneity of treatment effect on survival.
 - When this assumption does not hold, 3 sensitivity parameters need to be specified.
- **$\alpha \leq 0$**
 - The population of "control survivors" is healthier than "experimental survivors"
 - Assuming that "control survivors" would have better outcomes under experimental treatment as well, assumption $\alpha \leq 0$ will be satisfied. Akin to assuming that experimental treatment does not have unpleasant side effects affecting functional outcomes.

Estimators for estimands with **principal stratification** strategy

SACE: Treatment effect in subjects who would survive to time K under either treatment

- Rubin's approach using Multiple Imputation (MI) for survival (Rubin, 1998; 2006)
 - Analysis weighted by probability of survival (Hayden et al., 2005)
 - Crude estimate with sensitivity parameter (Chiba & VanderWeele, 2011)
 - **Crude estimate with lower and upper bounds (Colantuoni et al., 2018)**
- **Estimate in the control group:** can be obtained from observed control survivors
 - Under monotonicity assumption, "control survivors" are also "experimental survivors".
 - **Estimate in experimental group:**
 - Observed experimental survivors are a mix of "always survivors" and "experimental-only survivors". Assume the former have better outcomes than the latter.
 - **Lower bound:** estimated from all observed experimental survivors
 - **Upper bound:** estimated from $n \times p$ best values, where proportion p is the proportion of always survivors (estimated from observed control survivors)

Reminiscent of trimmed mean with additional assumptions.