Causal inference and estimands in clinical trials

Ilya Lipkovich

Eli Lilly and Company

2022 Causal Machine Learning for Novel Settings Boot Camp,

Purdue Aug 20, 2022

Acknowledgements: Yongming Qu, Bohdana Ratitch, Craig Mallinckrodt

This talk is based on recent publications

Statistics in Biopharmaceutical Research

ISSN: (Print) 1946-6315 (Online) Journal homepage: https://www.tandfonline.com/loi/usbr20

Causal Inference and Estimands in Clinical Trials

Ilya Lipkovich, Bohdana Ratitch & Craig H. Mallinckrodt

To cite this article: Ilya Lipkovich, Bohdana Ratitch & Craig H. Mallinckrodt (2020) Causal Inference and Estimands in Clinical Trials, Statistics in Biopharmaceutical Research, 12:1, 54-67, DOI: 10.1080/19466315.2019.1697739

To link to this article: https://doi.org/10.1080/19466315.2019.1697739

Therapeutic Innovation & Regulatory Science (2021) 55:984–988
https://doi.org/10.1007/s43441-021-00297-6

DIA

ANALYTICAL REPORT

Implementation of ICH E9 (R1): A Few Points Learned During the COVID-19 Pandemic

Yongming Qu[1,2] · Ilya Lipkovich[1]

Received: 2 February 2021 / Accepted: 9 April 2021 / Published online: 13 May 2021

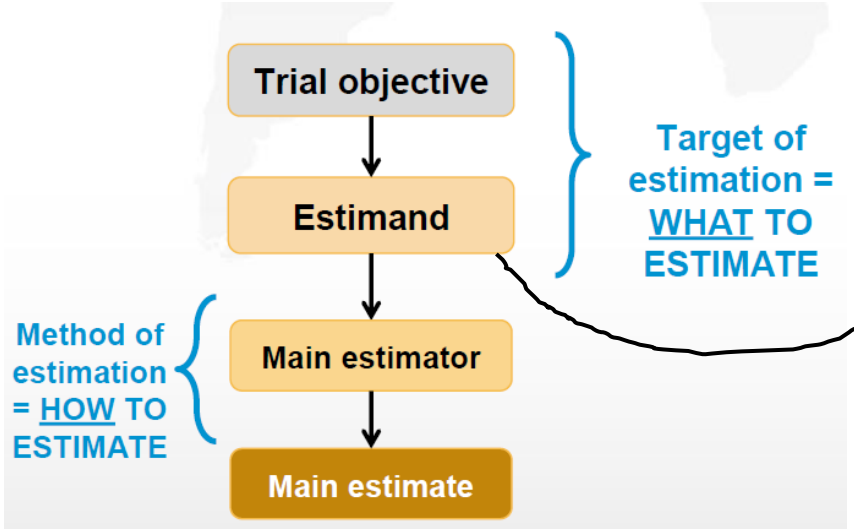
Historical perspective: from missing data to causal estimands

- Before National Research Council's (NRC) report on treatment of missing data (< 2010)
  - Discussions on likelihood based repeated measures (MMRM) vs. Last Observation Carried Forward (LOCF) – not always clear why LOCF is biased, as the target was not explicitly defined
  - Mechanism of missingness: MAR or MNAR?
  - Sensitivity analyses for departures from MAR: selection model, identifying influential patients
- From NRC's report to ICH E9 (R1): 2010-2017+
  - Put estimands first, missing data comes second
  - Distinguishing *Study vs Treatment* discontinuation: prevention of TD and encouraging data collection post TD
  - Sensitivity analysis: from selection model to more interpretable *Pattern-Mixture Models* (PMM)
    - not always clear whether we challenge the MAR assumption for primary estimand or propose a different estimand
- ICH E9 (R1) draft addendum (2017 to 2021)
  - EMA Step 5, 17 Feb 2020; FDA Guidance, May 2021
  - Strategies for defining *causal* estimands …

ICH E9 (R1) addendum

- Streamlines protocol development
  - The central role of estimands and strategies for dealing with *intercurrent events* (ICE)
  - ICE are defined as *events occurring after treatment initiation that affect either interpretation or existence of the measurements associated with clinical questions of interest*
- Emphasis on causal estimands (although not mentioning “potential outcomes”)
  - Treatment effects are quantified by “*how the outcome of treatment compares to what would have happened to the same subjects under different treatment conditions (e.g. had they not received the treatment or had they received a different treatment).*”

Estimand framework [ICH E9 (R1)]



- Treatment(s) of interest
- Population of interest
- Handling of relevant ICEs
- Outcome variable (endpoint) at patient level
- Population-level summary

ICEs, intercurrent events

Strategies to handle ICEs

- Treatment policy
- Hypothetical
- Composite variable
- While on treatment (WOT)
- Principal stratum (PS)

- ICH E9 (R1) provides a framework for defining estimand
- Key components to be considered
  - Treatment(s) of interest
  - Population of interest
  - Handling of relevant intercurrent events (ICEs)
  - Outcome variable (endpoint) at patient level
  - Population-level summary

ICEs, intercurrent events

Potential Outcomes (PO) framework

- For treatment $a = \{0,1\}$ define random variables $Y_i(a)$ as potential outcomes (PO) if treatment a is applied to patient $i = 1, .., n$ regardless of his/her actual treatment assignment, A_i
- In parallel randomized clinical trial, only one of the two potential outcomes can be observed.
- POs are linked with observables via consistency assumption implied by a more general SUTVA (Stable Unit Treatment Value Assumption)

$$Y_i = Y_i(0)(1 - A_i) + Y_i(1)\, A_i$$

Defining estimands in presence of ICE based on PO (Lipkovich et al., 2020)

- Y: outcome of interest
- S: stratum (subset) of the population
- A: treatment (0 = control; 1 = experimental treatment)
- $Y(a,b)$: the PO of Y for patients assigned to treatment a but actually taking b
- An example of ***causal*** estimand is the average treatment effect (ATE) for a subset S if patients would adhere to their assigned treatment

$$E[Y(1,1) - Y(0,0)|S]$$

- An example of ***non-causal*** estimand is the TE in completers $(S = 1)$ on each respective arm

$$E[Y(1,1)|S(1) = 1] - E[Y(0,0)|S(0) = 1]$$

- For the whole population (all randomized patients), we may remove S

$$E[Y(1,1) - Y(0,0)]$$

PO, potential outcome

A broader perspective

- We defined $Y(a, b)$ as a PO for assigned "a" and actual treatment "b".
- As we will see, actual treatment can be considered potential outcome on its own depending on intermediate outcomes following initial treatment , $b = B(a)$
- Note that by composition assumption for mediators (VanderWeele and Vansteelandt, 2009)

$$Y\big(a, B(a)\big) = Y(a)$$

- Therefore, $Y(a)$ may conceal any change of treatment occurring in a natural course of events following initial assignment, and if often left unspecified
- This often causes confusion when stating treatment policy strategy (see next)

Treatment policy (TP) strategy

- ICH E9 (R1) describes the TP strategy: "*the occurrence of the intercurrent event is considered irrelevant in defining the treatment effect of interest: the value for the variable of interest is used regardless of whether or not the intercurrent event occurs*."
- TP comes under different names and motivations
  - Intent to treat (ITT) – emphasizing that tested is the very fact of initial treatment assignment and ITT population
  - Treatment regimen – emphasizing that tested is *entire treatment strategy* that includes initial treatment and certain rules of treatment modification, whether pre-specified (e.g rescue) or spontaneous

Treatment policy (TP) strategy (cont.)

- It is often argued that TP combines the best of two worlds: RCT (randomization) and Real World (reflecting existing clinical practices, e.g. alternative/rescue medications)
- In fact, it may combine the worst of the two
  - The visit schedules, inclusion criteria, allowed rescue medication use, etc. make a clinical trial setting drastically different from those in real clinical practice making it hard to generalize
  - Ignoring changes of treatments makes it hard to attribute TP effect to any particular treatment
- To use TP strategy, it is recommended to clearly define the treatment regimen. For example,
  - *The treatment of interest is the randomized study medication with any additional rescue concomitant medications based on protocol-defined rescue criteria*
- TP is often defined in statements that can mean different things for different readers:
  - *The goal is evaluating difference in ….. irrespective of/regardless of premature discontinuation/change of treatment*

Treatment policy strategy (cont.)

- Often TP is denoted simply as $E\{Y_i(1) - Y_i(0)\}$, implicitly assuming composition assumption
- Let $A_i^* = \{A_i, g_i(Z_i(A_i))\}$ be the treatment regimen (policy) patient i takes
  - g_i maps intermediate outcomes Z_i to a treatment regimen (i.e., stopping study meds when having AE)
  - g_i generally is not precisely defined in the protocol (certain things may be left to physician's discretion)
- The estimand using this treatment policy strategy is defined by

$$E\left\{Y_i\left(1, g_i(Z_i(1))\right) - Y_i\left(0, g_i(Z_i(0))\right)\right\} \longrightarrow g_i(\cdot) \text{ with subject subscript}$$

- Estimand for the dynamic treatment regimen (DTR) (Murphy et al., 2001; Moodie et al., 2007)

$$E\left\{Y_i\left(1, g(Z_i(1))\right) - Y_i\left(0, g(Z_i(0))\right)\right\} \longrightarrow g(\cdot) \text{ without subscript } i$$

- The time-varying treatment regimen function g is defined clearly and *in the same way* for all patients

Composite strategies

- ICEs are used as part of the composite endpoint. It may be more straightforward to define the composite endpoint explicitly.
- For example,
  - In rheumatoid arthritis (RA), the binary variable of ACR20 is often used
  - Composite strategy may treat a patient with an ICE of using rescue medication as a non-responder
  - It is more appropriate to define the endpoint as a composite endpoint “achieving ACR20 at the end of study without using rescue medications”

Composite strategies: Binary outcome

- Assume a binary outcome Y
  - $Y = 1$ is clinical response, $Y = 0$ no response
  - $\Delta = 1$ discontinuation due to LOF
- Redefine potential outcome

$$\widetilde{Y}_i(a) = \begin{cases} Y_i(a), \Delta_i(a) = 0 \\ 0, \Delta_i(a) = 1 \end{cases}, a = 0,1$$

$$\delta_{CS} = E[\tilde{Y}(1) - \tilde{Y}(0)]$$

Composite strategies: Continuous outcome

- For continuous outcomes, use of win ratio to incorporate the ICE in the estimand
  - Define the composite endpoint as (assuming for Y, the smaller the better)

$$\tilde{Y}_i(a) = \begin{cases} Y_i(a), & \Delta_i(a) = 0 \\ \infty, & \Delta_i(a) = 1 \end{cases}, a = 0,1$$

  - Win probability: $\pi_w = \Pr\{\tilde{Y}_i(1) < \tilde{Y}_j(0)\}$ → Similar to Mann-Whitney test
  - Lose probability: $\pi_l = \Pr\{\tilde{Y}_i(1) > \tilde{Y}_j(0)\}$
  - Win ratio = π_w/π_l (estimand)
- We can introduce a hierarchy of ICEs
  - This allows us to compare any two subjects whether one is preferrable to the other

Hypothetical strategies

- Because causal estimands should be defined in terms of potential outcomes, most strategies for handling ICEs should be “hypothetical”)
- We introduce 3 different hypothetical strategies
  - *Controlled direct hypothetical* (CDH) strategy
  - *No treatment hypothetical* (NTH) strategy
  - *Partial treatment hypothetical* (PTH) strategy

Controlled direct hypothetical (CDH) strategy

- The PO of interest is the outcome if patients could complete the treatment even in the presence of ICEs
- The estimand is

$$E\{Y_i(1,1) - Y_i(0,0)\}$$

- "Controlled direct" following *controlled direct effect* (Pearl, 2009)
- This approach may be appropriate for
  - ICEs due to administrative reasons (e.g., ICEs related to COVID-19 controlled measures)
  - ICEs that do not represent the "normal" time (e.g., COVID-19 illness), because we would like to generalize to "normal" time
  - ICEs due to LoE (arguably an important benchmark for hypothetical efficacy)

LoE, lack of efficacy; PO, potential outcome

No treatment hypothetical (NTH) strategy

- The PO of interest is the outcome assuming patients with ICEs would have no benefit from the treatment (as if left untreated starting from randomization):

$$E\left[\left\{Y_i(1,-1)\Delta_i(1)+Y_i(1,1)\left(1-\Delta_i(1)\right)\right\}-\left\{Y_i(0,-1)\Delta_i(0)+Y_i(0,0)\left(1-\Delta_i(0)\right)\right\}\right]$$

  where "-1" in the second argument of $Y_i(\cdot,\cdot)$ indicates no treatment received and $\Delta_i(a)$ is the ICE indicator (0 for no ICE and 1 for ICE occurring).

- This approach may be appropriate for ICEs due to certain AEs (during "normal" times)

AE, adverse event; ICE, intercurrent events; PO, potential outcome

Partial treatment hypothetical (PTH) strategy

- The PO of interest is the outcome if the patient may benefit from (or be harmed by) the study medication until occurrence of the ICE and then stops taking the medication
- The estimand is defined as

$$E\left[\left\{Y_i(1, g_i(T_i(1)))\Delta_i(1) + Y_i(1,1)\left(1-\Delta_i(1)\right)\right\}\right.$$
$$\left.-\left\{Y_i(0, g_i(T_i(0)))\Delta_i(0) + Y_i(0,0)\left(1-\Delta_i(0)\right)\right\}\right]$$

  where $T_i(a)$ is the time to the ICE under treatment $a \in \{0,1\}$ and $g_i\left(T_i(\mathrm{a})\right)$ is the treatment regimen: taking treatment a until the occurrence of the ICE and then having no access to treatment until a specified assessment time

- This strategy may be suitable for handling ICEs due to AE under "normal circumstances" (not for AE related to the COVID-19 pandemic), especially for treatments with potential long-term or disease-modification effects

ICE, intercurrent events; PO, potential outcome

CHMP: Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus (draft)

- "… Specifically, since patients are not expected to benefit once treatment is discontinued (e.g. due to adverse events) the treatment effect should be estimated based on observed or modelled data reflecting adherence to treatment as observed in the clinical trial." **PTH or NTH strategy**
- "… Therefore, the treatment effect can be estimated under the assumption that rescue medication, or use of other medications that will influence HbA1c values, was not introduced (hypothetical scenario), provided that a reliable estimate of that effect can be obtained." **CDH strategy**

While-on-treatment strategy

- The while-on-treatment (WOT) strategy yields a direct effect of the initial treatment that by construction obviates the need to account for subsequent ICEs through defining the outcome up to the point where an ICE occurs
  - Define potential time on treatment until ICE or end pf study for patients randomized to control and active treatment, as $T(0)$, $T(1)$, respectively
  - Let potential outcome up to time u, be $Y_u(a), a = 0,1.$
- WOT estimand

$$\delta_{WOT} = E\left[Y_{T(1)}(1) - Y_{T(0)}(0)\right]$$

- It is sometimes tempting for sponsors to disguise the old LOCF as WOT
  - Typically examples of WOT are measures summarizes benefits over time (area under curve or average slope) up to change in treatment

Principal stratification (PS) strategy

- PS should be considered in the context of defining population of interest rather than a strategy for dealing with ICE
- In fact, PS is a hypothetical population that can be defined based on any post-randomization variable
  - e.g. a post-baseline biomarker $S = I(Z > c)$, early responder
- For example, if ICE is adherence $S = 1$ if treated and we are interested in the CDH (hypothetical) strategy, the estimand is

$$E\{Y(1,1) - Y(0,0)|S(1) = 1\}$$

- What if we are interested in TP strategy?

$$E\{Y(1) - Y(0)|S(1) = 1\}$$

Use a mix of strategies for handling ICEs in a study (Darken et al., 2020; Qu et al., 2020)

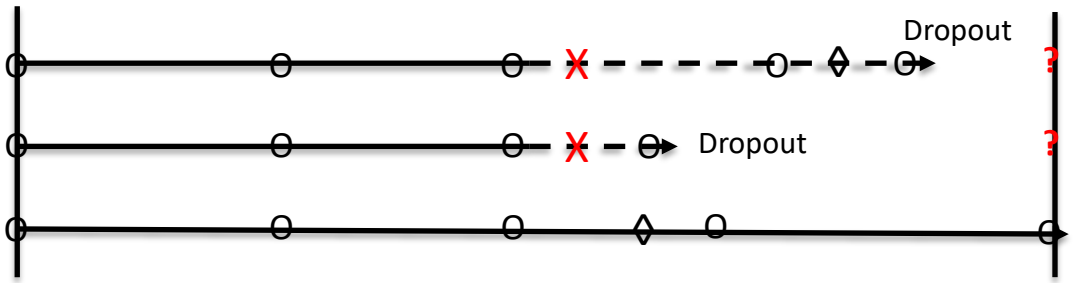
- One common drawback in most current clinical studies is that only ONE strategy is used to handle all ICEs
- Strategies for handling ICEs should be based on the underlying reasons
  - ICEs due to AE
    - AE at “normal time”
    - AE of COVID-19 illness
  - ICEs due to lack of efficacy (LoE)
    - Treatment discontinuation due to LoE
    - Use of rescue medication due to LoE
  - ICEs due to administrative reasons
    - Relocation, family situation changed, COVID-19 controlled measures, etc.

Missing values

- Missing values
  - As a result of handling ICEs with hypothetical strategies
  - *True* missing values caused by data not being collected
- Assumptions for missingness and methods for handling missing values should be based on the underlying reasons of ICEs or missingness
  - ICEs due to AE
    - AE at “normal circumstances”
    - AE of COVID-19 illness
  - ICEs due to LoE
  - ICEs due to administrative reasons
  - Not due to ICEs

AE, adverse event; ICEs, intercurrent events; LoE, lack of efficacy

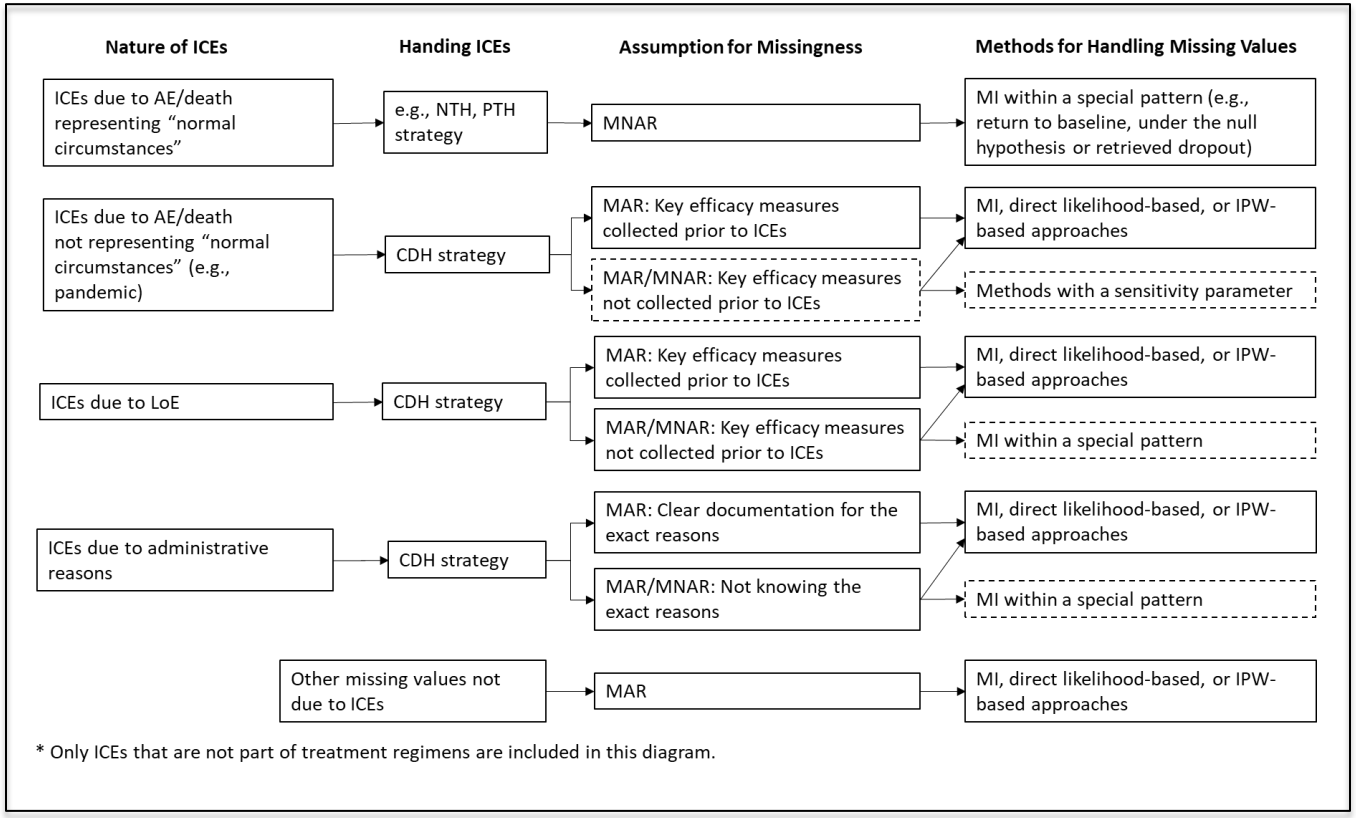
Understanding the potential outcome before imputing



O: measurement available
? : measurement missing
X: treatment discontinuation
◊: Use of rescue medication
Dropout: study discontinuation

- What is the potential outcome of interest at last scheduled visit?
  - Not taking study medication after X but using rescue medication
  - Not taking study medication after X and not using rescue medication (having no access to treatment)
  - Continuing to take study medication after X rather than using rescue medication

Handling ICEs and missing values according to the nature of ICE/missingness



AE, adverse event
CDH, controlled direct hypothetical
ICEs, intercurrent events
IPW, inverse probability weighting
LoE, lack of efficacy;
MAR, missing at random
MI, multiple imputation
MNAR, missing not at random
NTH, no treatment hypothetical
PTH, partial treatment hypothetical

Summary and Recommendations

- **Describing estimands**
  - Using PO language may help define and communicate estimands more succinctly. It also helps evaluate the plausibility of certain strategies for handling ICEs
- **Defining ICEs**
  - Prior to identifying possible ICEs, treatment regimens of interest need to be defined precisely
  - To be considered an ICE, this event should be a deviation from the treatment regimens of interest
- **Handling ICEs**
  - If intending to use a composite strategy to handle certain ICEs, these ICEs should be explicitly included in the composite endpoint
  - Hypothetical strategies should be predominately used to define causal estimands
  - Using a mix of strategies (rather than a single strategy) for handling ICEs is often clinically relevant.
- **Estimation**
  - Multiple imputations is a flexible tool allowing for implementing a mix of strategies for handling ICEs
  - Use the most ***plausible*** assumptions (not the most conservative assumptions)

References

Bergenstal, R. M., Lunt, H., Franek, E., Travert, F., Mou, J., Qu, Y., ... & Bastyr III, E. J. (2016). Randomized, double-blind clinical trial comparing basal insulin peglispro and insulin glargine, in combination with prandial insulin lispro, in patients with type 1 diabetes: IMAGINE 3. Diabetes, Obesity and Metabolism, 18(11), 1081-1088.

Darken, P., Nyberg, J., Ballal, S., & Wright, D. (2020). "The attributable estimand: A new approach to account for intercurrent events". Pharmaceutical Statistics 19(5): 626-635.

ICH E9 (R1) (2019). "ICH Harmonised Guideline Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1) ," International Conference on Harmonisation [online], Available at https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Hernán MA and Scharfstein D (2018) Cautions as regulators move to end exclusive reliance on Intention to Treat. *Annals of Internal Medicine*, 168, 515–516

Lipkovich, I., Ratitch, B., & Mallinckrodt, C. H. (2020). Causal inference and estimands in clinical trials. *Statistics in Biopharmaceutical Research*, *12*(1), 54-67.

Little, R. J. A. (1995), "Modeling the Drop-Out Mechanism in Repeated Measures Studies," Journal of the American Statistical Association, 90, 1112–1121.

Meyer, R. D., Ratitch, B., Wolbers, M., Marchenko, O., Quan, H., Li, D., ... & Englert, S. (2020). Statistical Issues and Recommendations for Clinical Trials Conducted During the COVID-19 Pandemic. Statistics in Biopharmaceutical Research, (just-accepted), 1-22.

Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9," Statistical Science, 5, 465–480.

Pearl, J. (2009), "Causal Inference in Statistics: An Overview," Statistics Surveys, 3, 96–146.

Qu, Y., Shurzinske, L., & Sethuraman, S. (2020). Defining estimands using a mix of strategies to handle intercurrent events in clinical trials. Pharmaceutical Statistics. DOI: 10.1002/PST.2078

Ratitch B, Bell J, Mallinckrodt C, Bartlett JW, Goel N, Molenberghs G, O'Kelly M, Singh P, and Lipkovich I (2019) Choosing estimands in clinical trials: putting the ICH E9(R1) into practice. *Therapeutic Innovation & Regulatory* Science 54(2):324-341

Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," Annals of Statistics, 6, 34–58.

Rubin, D. B. (1987), "Multiple Imputation for Nonresponse in Surveys", New York: Wiley.

Thank you!

Q & A