JAMA Guide to Statistics and Methods

Estimands, Estimators, and Estimates

Roderick J. Little, PhD; Roger J. Lewis, MD, PhD

The primary goal of most randomized clinical trials (RCTs) is to draw conclusions about the effect of a treatment in a specific population of patients. The true effect of the intervention, termed the *estimand*, is estimated with the data acquired in the trial, subject to limitations associated with variations in adherence to treatment, patients being lost to follow-up, and data quality.

The choice of estimand and associated target population should reflect the goals of the trial, and can vary according to who designed or sponsored the study, who will use the results of the study, and the motivating scientific question. In the PIONEER 3 trial,[1] investigators compared 3 doses of oral semaglutide with sitagliptin, added to background therapy, in adults with type 2 diabetes. The primary end point was the change in glycated hemoglobin (HbA_{1c}). The trial design considered 2 estimands for summarizing treatment effect, termed the *treatment policy estimand* and the *trial product estimand*.

Explanation of the Concept

What Is an Estimand?

The true effect of the intervention is the estimand. The estimand is a target quantity (ie, what the study aspires to measure). It is a summary of patient outcomes, such as a difference in mean outcomes or a difference in mortality rates in the population, comparing patients who receive the investigational treatment with those who do not. Estimands can describe both therapeutic benefits and adverse effects; thus, more than 1 estimand may be needed to capture fully the results of a study.

Trial data provide only estimates of trial estimands because trial participants are sampled from the population and outcomes are not always observed for all randomized participants, and because there are practical limitations in clinical trial execution, such as participants' not following the prescribed protocol or not completing the study. An estimator, in contrast to an estimand, is a formula or algorithm used to estimate the target quantity from the clinical trial data, such as the difference in sample means between 2 treatment groups, or the Kaplan-Meier estimator of a survival curve. Statistical inference for an estimand requires a choice of the estimator and a measure of its precision. Typical methods of statistical inference are hypothesis tests, confidence intervals, and posterior credibility intervals in a bayesian analysis. The estimate is the numeric value obtained when the estimator is applied to the actual data from the trial.

Why Is the Choice of Estimand Important?

The optimal choices of estimands and associated target populations are determined by the goals of the trial. For example, the sponsor of a trial may be interested in a per-protocol estimand: the efficacy of a treatment in patients adherent to their assigned treatment. However, a payer or clinician may be interested in an intention-to-treat (ITT) estimand: the effectiveness of a treatment for all individuals assigned to it, irrespective of adherence. The different choices of estimands address fundamentally different questions.

Clinical trials seek to measure causal effects of treatments. The choice of estimands determines how these effects are measured, and a lack of clarity in this choice obscures the interpretation of trial results. Good estimands and estimators have the following features:

1. **The estimand compares outcomes that capture the main benefits and risks of treatments.** This consideration is particularly important when surrogate or near-term outcomes are used. For example, in attempts to characterize the utility of drugs for preventing sudden cardiac death, investigators compared suppression of ventricular ectopy on the electrocardiogram.[2] Drugs were approved based on randomized trials that showed improvements in this surrogate measure of drug activity, but the definitive CAST trial,[3] with an estimand measuring survival, demonstrated that the approved drugs actually reduced survival. This difference may have resulted in a substantial cost in lives.
2. **Estimators should summarize the causal effects of treatments in the sample of individuals in the study.** The *individual* causal effect of a treatment is the difference in outcome if the individual were assigned the treatment vs assigned the control.[4] This individual effect is usually unmeasurable because only 1 of the possible outcomes is observed; namely, the outcome for the therapy actually assigned. However, summary causal effects, such as the average, can be estimated for groups of individuals. Internal validity is the ability to estimate the summary causal effect of a treatment for the sample of individuals in the study. This requirement may not be met if, for example, the assumptions of the statistical analysis strategy are violated.
3. **Estimands should summarize the causal effects of treatments in the target population.** External validity refers to the ability to establish the average causal effect of a treatment for the target population, usually composed of patients who would receive the treatment in clinical practice. Given internal validity, the main threat to external validity is effect modification; that is, the degree to which treatment effects vary across individuals according to their characteristics. Substantial effect modification that reduces external validity occurs because the target population is often loosely defined and individuals in RCTs are usually not randomly sampled from it, but rather are self-selected volunteers. A common strategy to assess effect modification in trials is to compare average treatment effects across subgroups according to baseline characteristics, such as demographic variables or initial disease severity. Similarity of the estimated treatment effects across subgroups suggests that effect modification is small, increasing the evidence of external validity.
4. **The estimator should provide a valid and unbiased estimate of the study estimand.** For this to be true, study estimators must have good internal and external validity. RCTs are the criterion standard for generation of evidence because random assignment to treatments tends to eliminate bias from observed and unobserved confounders, and hence increase internal validity. Observational studies are often larger and less costly than RCTs, and may have better external validity than RCTs if potential confounders are measured and controlled. However, they are vulnerable to bias from unmeasured confounders. Because internal validity is a requirement for external validity, RCTs will continue to be a crucial source of evidence; however, there is substantial potential in creative combinations of evidence from RCTs and observational databases.[5]

Limitations of Alternative Choices of Estimands

Estimands are summaries and cannot capture all relevant features of treatments. The robustness of the findings in an RCT or observational study (ie, the degree to which causal effects of treatments have been established) depends on the statistical analysis, and the extent to which the analysis rests on untestable assumptions, such as absence of unobserved confounders or the assumption that missing data are missing at random.[6] This can vary greatly according to the choice of estimand. For example, per-protocol estimates in RCTs often rely more heavily on assumptions than ITT estimates because the exclusive consideration of participants who follow the treatment protocol may undo the balance created by randomization and result in marked differences between the treatment groups, confounding estimates of the treatment effect.

Substantial missing data are a threat to validity, and the amount of missing data can vary greatly, depending on the choice of estimand. For example, in the ATLAS ACS 2–TIMI 51 study,[7] a large clinical trial that assessed rivaroxaban for the treatment of acute coronary syndrome, estimating a modified ITT estimand involved far fewer missing data than estimating the strict ITT estimand; of 15 526 enrolled patients, 9.7% were missing the ITT outcome and 5.1% the modified ITT outcome. A 2010 report from the National Research Council[8] recommended that the amount of missing information should be a factor when alternative estimands are considered.

How Were Estimands Used in the PIONEER 3 Trial?

The PIONEER 3 RCT compared 3 doses of oral semaglutide with sitagliptin, added to background therapy, in adults with type 2 diabetes. The primary end point was change in HbA_{1c} from baseline to week 26.[1]

Two estimands were defined, an ITT treatment policy estimand and a per-protocol trial product estimand. Both of these estimands concern the difference in the mean end point between a treatment group and the comparator group, but for 2 different populations. The treatment policy population was an ITT population, defined as "all randomized patients regardless of trial product discontinuation or use of rescue medication." In contrast, the trial product population was a per-protocol population, described as "the treatment effect for all randomized patients under the assumption that all patients continued taking trial product for the entire planned duration of the trial and did not use rescue medication." An alternative per-protocol definition could have been the subpopulation of all individuals who would adhere to all of the compared treatments if assigned. The assumption that all patients would adhere to assigned treatment is avoided in this definition by restricting the estimand to the subpopulation consisting of study participants who were adherent to all the treatments.

How Do the Choices of Estimands Affect the Interpretation of the PIONEER Trial?

Estimating the trial product estimand defined in the PIONEER trial required predicting hypothetical outcomes for individuals who discontinued the assigned treatment, if they had continued with that treatment. Those hypothetical outcomes represent missing data. The repeated-measures statistical model used to predict the missing values involved untestable and perhaps unreliable assumptions. It is often preferable when possible to define estimands that involve only information arising while a trial participant is receiving the assigned treatment. These estimands can be called "on-treatment summaries,"[9] and the ICH E9 addendum[10] calls this approach the "while on treatment" strategy. In the PIONEER 3 study, a possible on-treatment summary estimand could have been defined as follows: "Of the time between baseline and week 26 when the participant was on the assigned treatment, the proportion of time when the diabetes was controlled," in which "controlled" is defined as below some threshold value of HbA_{1c}. Creative choices of estimand that limit the influence of missing data can improve the robustness of clinical trial findings.

ARTICLE INFORMATION

Author Affiliations: Department of Biostatistics, University of Michigan, Ann Arbor (Little); Department of Emergency Medicine, Harbor-UCLA Medical Center, Torrance, California (Lewis); Department of Emergency Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California (Lewis); Berry Consultants, Austin, Texas (Lewis).

Corresponding Author: Roderick J. Little, PhD, Department of Biostatistics, University of Michigan, M4071 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109 (rlittle@umich.edu).

Section Editor: Robert Golub, MD, Deputy Editor.

Published Online: August 23, 2021. doi:10.1001/jama.2021.2886

Conflict of Interest Disclosures: Dr Lewis reported being the senior medical scientist at Berry Consultants LLC, a statistical consulting firm that specializes in adaptive and platform clinical trial design. No other disclosures were reported.

REFERENCES

1. Rosenstock J, Allison D, Birkenfeld AL, et al; PIONEER 3 Investigators. Effect of additional oral semaglutide vs sitagliptin on glycated hemoglobin in adults with type 2 diabetes uncontrolled with metformin alone or with sulfonylurea: the PIONEER 3 randomized clinical trial. *JAMA*. 2019;321(15):1466-1480. doi:10.1001/jama.2019.2942

2. Rydén L, Arnman K, Conradson T-B, et al. Prophylaxis of ventricular tachyarrhythmias with intravenous and oral tocainide in patients with and recovering from acute myocardial infarction. *Am Heart J*. 1980;100(6 pt 2):1006-1012.

3. Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med*. 1989;321(6):406-412. doi:10.1056/NEJM198908103210629

4. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688-701. doi:10.1037/h0037350

5. Walicke P, Abosch A, Asher A, et al. Launching effectiveness research to guide practice in neurosurgery: a National Institute of Neurological Disorders and Stroke workshop report. *Neurosurgery*. 2017;80(4):505-514. doi:10.1093/neuros/nyw133

6. Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367(14):1355-1360.

7. Little RJ, Wang J, Sun X, et al. The treatment of missing data in a large cardiovascular clinical outcomes study. *Clin Trials*. 2016;13(3):344-351.

8. National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials: Panel on Handling Missing Data in Clinical Trials*. National Academy Press; 2010.

9. Little R, Kang S. Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Stat Med*. 2015;34(16):2381-2390.

10. Guidance Document E9(R1): statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. FDA. May 2021. Accessed July 15, 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e9r1-statistical-principles-clinical-trials-addendum-estimands-and-sensitivity-analysis-clinical