HealthCore

A Step-By-Step Guide to Causal Study Design

1 Research Question[1]

Association
Most biases disregarded by definition[2]

Causal Effect
Move to Step 2

Estimand

2 Which Kind of Effect?[3]

Intention-to-Treat (ITT)
≈ First-treatment-carried-forward (FTCF)

Per-protocol (PP) ≈ As-treated

3 Effect in *Whom*?

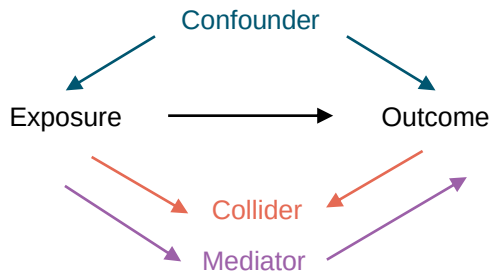
Target Population For Counterfactual Contrast

- Average treatment effect (ATE)
- ATE in the treated/untreated (ATT/ATU)
- Conditional ATE (subgroups)
- Individual TE (ITR)

4 Measure of Effect?

Difference or ratio?
Risk, rate, hazard, odds, cost…?

5 Create Directed Acyclic Graph (DAG)



6 Navigate the Land of Biases

- Measured confounding
- Unmeasured confounding
- Collider bias
- Selection bias
- Immortal time bias
- Protopathic bias (reverse causality)
- Healthy adherer effect
- Prevalent user bias
- Confounding by indication
- Effect modification ↔ Generalizability
- Dependent/informed censoring
- Misclassification

Etc.[4]

7 Explore the Land of Solutions

- "Target trial" thinking[5]
- New user design with active comparator
- Confounder adjustment[6]
  - *Time-invariant ("baseline")*
    - Matching or weighting
    - Best with propensity scores
  - *Time-varying*
    - Survival analysis with time-varying covariates
    - Mixed models, GEE
    - MSM with IPTW (if confounders are affected by prior treatment)
- Evaluate confounder balance
- Evaluate unmeasured confounding (e.g., via E-value or use of instrumental variables)
- IPCW to account for loss-to-follow-up/censoring

8 Conduct QC & Sensitivity Analyses

- Test if model assumptions are fulfilled
- Use different estimand or model
- Quantitative bias analysis[7]

Acronyms: GEE, generalized estimating equations; IPC/TW, inverse probability of censoring/treatment weighting; ITR, individual treatment response; MSM, marginal structural model; TE, treatment effect

Please refer to the glossary for detailed description of terms.

[1]Ensure that the exposure and outcome are well-defined based on literature and expert opinion

[2]More specifically, measures of association are not affected by issues such as confounding and selection bias because they do not intend to isolate and quantify a single causal pathway. However, information bias (e.g., variable misclassification) can negatively affect association estimates, and association estimates remain subject to random variability (and are hence reported with confidence intervals).

[3]"Trial ≈ real world data" parallels are inexact. PP especially can be expanded within the target trial framework.

[4]This list is not exhaustive; it focuses on frequently encountered biases

[5]To assess bias in a nonrandomized study, use of the ROBINS-I tool is recommended (Sterne 2016; http://www.bristol.ac.uk/population-health-sciences/centres/cresyda/barr/riskofbias/robins-i/resources/)

[6]Only a selection of the most popular approaches is presented here. Other methods exist; e.g., g-computation and g-estimation for both time-invariant and time-varying analysis; instrumental variables; and doubly-robust estimation methods. There are also program evaluation methods (e.g., difference-in-differences, regression discontinuities) that can be applied to pharmacoepidemiological questions.

Conventional outcome regression analysis is NOT recommended for causal estimation due to issues determining balance, correct model specification, and interpretability of effect estimates.

[7]Online tools include, among others, an E-value calculator for unmeasured confounding (https://www.evalue-calculator.com/) and the P95 outcome misclassification estimator (http://apps.p-95.com/ISPE/).

Five suggested articles for further reading

(additional articles can be found in the glossary reference section)

- Hernán MA. The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data*. Am J Public Health*. 2018;108(5):616-619.
- Franklin JM, Platt R, Dreyer NA, et al. When Can Nonrandomized Studies Support Valid Inference Regarding Effectiveness or Safety of New Medical Treatments? *Clin Pharmacol Ther*. 2021;10.1002/cpt.2255.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424.
- Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ*. 2017;359:j4587.
- Lash TL, Fox MP, MacLehose RF, Maldonado G, McCandless LC, Greenland S. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-1985.

HealthCore

Causal Study Design Glossary

Active comparator: Describes a study design where the drug of interest is compared with another drug commonly used for the same indication and the same stage of disease, as opposed to no treatment. This is the preferred approach in real-world studies, as it reduces confounding by measured and unmeasured factors.

References: Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials?. *Clinical Pharmacology & Therapeutics*. 2017 Dec;102(6):924-33.

Yoshida K, Solomon DH, Kim, SC. Active-comparator design and new-user design in observational studies. *Nat Rev Rheumatol*. 2015 Jul;11(7):437-41.

Association: Association, or correlation, is the statistical relationship between two variables. Association does not imply a causal relationship between the two variables.

References: https://www.cdc.gov/csels/dsepd/ss1978/glossary.html

Altman N, Krzywinski M. Points of Significance: Association, correlation and causation. *Nature Methods*. 2015 Oct 1;12(10).

As-treated analysis: An “as-treated” analysis is based on the treatment actually received (i.e. it accounts for treatment switching, discontinuation, etc.) and not on the original treatment assignment (i.e. the treatment on index date). An "as-treated" analysis of RCTs becomes similar to an analysis of observational data in the sense that additional adjustment for confounding, informed censoring, etc. become necessary.

References: Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55.

Smith VA, Coffman CJ, Hudgens MG. Interpreting the Results of Intention-to-Treat, Per-Protocol, and As-Treated Analyses of Clinical Trials. *JAMA*. 2021;326(5):433–434.

Average treatment effect (ATE): ATE describes the average over the entire population of the individual treatment effects. For example, what is the expected impact of *everyone* in the airport eating a chicken sandwich versus *everyone* not eating one? This terminology (as well as ATT and ATU, defined below) is related to the potential outcomes framework. Choice of ATE/ATT/ATU is one component of the estimand. In general, the ATE is a weighted average of the ATT and ATU, with weights equal to the relative sample sizes.

References: Whitney Newey, course materials for 14.386 New Econometric Methods, Spring 2007. MITOPENCOURSEWARE (OCW) (http://ocw.mit.edu), Massachusetts Institute of Technology.

Little RJ, Lewis RJ. Estimands, Estimators, and Estimates. *JAMA*. 2021;326(10):967–968.

Faries D, Zhang X, Kadziola Z, et al. 2020. Real World Health Care Data Analysis: Causal Methods and Implementation Using SAS®. Cary, NC: SAS Institute Inc. Chapter 2.

Average treatment effect on the treated (ATT): ATT describes the average treatment effect in the subpopulation of treated people. For example, what is the expected impact among those in the airport eating a chicken sandwich who actually did eat one? In an RCT with perfect adherence, the ATT is identical to the ATE. Note that ATT weights are also referred to as “standardized morbidity or mortality ratio weights” (SMRW) and “weighting by the odds” in the literature.

Reference: Whitney Newey, course materials for 14.386 New Econometric Methods, Spring 2007. MITOPENCOURSEWARE (OCW) (http://ocw.mit.edu), Massachusetts Institute of Technology.

Rubin DB. Causal Inference Using Potential Outcome’s: Design, Modeling, Decisions. *J Am Stat Assoc*; Mar 2005; 100, 469.

Faries D, Zhang X, Kadziola Z, et al. 2020. Real World Health Care Data Analysis: Causal Methods and Implementation Using SAS®. Cary, NC: SAS Institute Inc. Chapter 2.

Average treatment effect on the untreated (ATU): ATU describes the average of the individual treatment effects across the untreated subgroup of the full sample/population. The ATU is of interest whenever we want to estimate how an intervention would affect a group of individuals who have not received it.

References: Wang A, Nianogo RA, Arah OA. G-computation of average treatment effects on the treated and the untreated. *BMC Med Res Methodol*. 2017;17(1):3.

Garrido MM, Dowd B, Hebert PL, Maciejewski ML. Understanding Treatment Effect Terminology in Pain and Symptom Management Research. *J Pain Symptom Manage*. 2016;52(3):446-452.

Balance assessment: Balance assessment is the assessment of covariate balance between the treatment and comparator arms prior to outcomes analysis. It is usually part of an iterative approach (e.g. propensity score model fitting). Several metrics can be used to assess covariate balance, such as standardized mean differences (SMD) or variance ratios.

Reference: Ali MS, Prieto-Alhambra D, Lopes LC et al. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. *Frontiers in Pharmacology*. 2019:973.

Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15. Chapter 5.

Case-control: Rather than identifying exposed and unexposed people and then following them over time, a case-control study identifies people with the outcome (cases) and people without the outcome (controls), and calculates the odds ratio comparing exposed to unexposed. It is important to select the controls from the same source population as the cases to avoid bias. This study is an efficient approach to dealing with rare outcomes.

References: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances

Murphy, SW. *Longitudinal Studies 1: Determination of Risk in Clinical Epidemiology: Practice and Methods 2nd Ed.* Edited by Parfrey PS & Barret BJ. Humana Press, 2015; New York. Pp 59-62.

Causation: In health services research, causation means that a factor has a causal impact on a health outcome. Several causal inference methods have been developed such as Rubin's causal model and the directed acyclic graph.

References: Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*. 1986 Sep 1;15(3):413-9.

Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15. Chapter 2.

Channeling bias: "Channeling is a form of allocation bias, where drugs with similar therapeutic indications are prescribed to groups of patients with prognostic differences. Claimed advantages of a new drug may channel it to patients with special pre-existing morbidity, with the consequence that disease states can be incorrectly attributed to use of the drug." (Reference 1). Similarly, contraindications may channel some people away from a drug.

References: Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. *Stat Med*. 1991 Apr;10(4):577-81.doi: 10.1002/sim.4780100409. PMID: 2057656.

Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15. Chapter 4, page 44.

Cohort: "In a cohort study, the investigator identifies a population at risk for the outcome of interest, defines two or more groups of people (referred to as study cohorts) who are free of disease and differ according to their extent of exposure, and follows them over time to observe the occurrence of the disease in the exposed and unexposed cohorts. A cohort study may also include a single cohort that is heterogeneous with respect to exposure history, and occurrence of disease is measured and compared between exposure groups within the cohort."

> **Reference:** The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and guidances.

Collider and Collider Bias: "When an exposure and an outcome independently cause a third variable, that variable is termed a 'collider'. Colliders do not need to be adjusted for because they block the causal pathway (i.e., collide). Inappropriately controlling for a collider variable, by study design or statistical analysis, results in collider bias. Controlling for a collider can induce a distorted association between the exposure and outcome, when in fact none exists. This bias predominantly occurs in observational studies."

> **References:** https://catalogofbias.org/biases/collider-bias/
>
> Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf
>
> Pearce N, Lawlor DA. Causal inference-so much more than statistics [published correction appears in Int J Epidemiol. 2020 Feb 1;49(1):358]. *Int J Epidemiol.* 2016;45(6):1895-1903.

Confounding: "Confounding occurs when the estimate of measure of association is distorted by the presence of another risk factor. For a variable to be a confounder, it must be associated with both exposure and outcome, without being on the causal pathway between exposure and outcome."

> **References:** The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances
>
> https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704-ep713_confounding-em/bs704-ep713_confounding-em_print.html

Confounding adjustment: Any method that attempts to reduce or remove bias in the treatment effect estimate stemming from confounding by observable covariates, including matching, weighting, randomization, stratification, etc. A large literature exists on how to select confounders and the advantages and disadvantages of different methodologies for confounding adjustment.

> **References:** https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module11-Confounding-EMM/PH717-Module11-Confounding-EMM5.html
>
> Jager KJ, Zoccali C, Macleod A, Dekker FW. Confounding: what it is and how to deal with it. *Kidney Int*. 2008;73(3):256-260.

Confounding by indication: "Confounding by indication is when the outcome may in fact be caused by the indication for the treatment exposure. For example, the difference in outcomes between the exposed/treatment and non-exposed/control may be due to difference in underlying disease severity or underlying risk profiles. Confounding by severity is a type of confounding by indication".

> **Reference:** The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances.

Conditional ATE: The ATE in a subgroup of patients defined by observable covariates. For example, one can estimate a Conditional ATE (CATE) for the subgroup of female patients in a study, defined as the ATE in female patients.

References: https://egap.org/resource/10-things-to-know-about-heterogeneous-treatment-effects/

Abrevaya J, Hsu YC, Lieli RP. Estimating conditional average treatment effects. *Journal of Business & Economic Statistics.* 2015 Oct 2;33(4):485-505.

Conditional effect: "A conditional effect is the average effect, at the subject level, of moving a subject from untreated to treated (as in an ATT estimand). The regression coefficient for a treatment assignment indicator variable from a multivariable regression model is an estimate of a conditional or adjusted effect." (Reference 1). The effect is calculated for a one unit change in the exposure when all the other covariate values are held constant, in other words, conditional on covariates.

Reference: Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*. 2013;32(16):2837-2849.

Conditioning on the future: Conditioning on the future is when the study cohort is defined by future exposure information rather than baseline/current information. Conditioning on future exposure to define study cohorts can induce selection and immortal time biases.

Reference: Lund JL, Horváth-Puhó E, Komjáthiné Szépligeti S, et al. Conditioning on future exposure to define study cohorts can induce bias: the case of low-dose acetylsalicylic acid and risk of major bleeding. *Clin Epidemiol*. 2017 Nov 23;9:611-626.

Counterfactual: A counterfactual is a hypothetical alternate exposure scenario. For example, what would have happened to patient X had they taken drug B instead of drug A? This is related to the potential outcomes framework.

References: Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol.* 2002;31(2):422-429.

Höfler M. Causal inference based on counterfactuals. *BMC Medical Research Methodology*. 2005 Dec;5(1):1-2.

Direct acyclic graph (DAG): A causal path diagram showing at least two variables (exposure and outcome), connected by an arrow depicting the direction of the causal relationship. Variables are usually arranged in temporal or causal order (left to right) starting from the exposure and ending at an outcome. Variables are introduced between the causal path to showcase mediation, moderation, collision, or confounding.

Reference: http://www.dagitty.net/learn/graphs/index.html

Doubly robust estimation: A method of causal effect estimation which combines two approaches to estimating the causal effect of the exposure on an outcome: one regression model for outcomes with covariates (generating predicted outcomes), and one regression model for exposures (generating propensity scores), to get an estimator that is robust to misspecification of one of the two models. When these models are used independently, estimation of the causal effect is only unbiased if each model is correctly specified. However, using both outcome and exposure models together means that only one of the two must be correctly specified for an unbiased estimator, hence doubly robust.

References: Funk MJ, Westreich D, Wiesen C, et al. Doubly robust estimation of causal effects. *Am J Epidemiol*. 2011 Apr 1;173(7):761-7.

Liu L, Hudgens MG, Saul B, Clemens JD, Ali M, Emch ME. Doubly Robust Estimation in Observational Studies with Partial Interference. *Stat (Int Stat Inst)*. 2019;8(1):e214.

Effect modification: A third variable other than the exposure or the outcome that modifies the causal effect between them, based on the levels of this third variable. Effect modification can be additive or multiplicative. Also known as interaction or moderation.

Reference: Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics, 3rd Ed*. Jones and Bartlett Learning, 2014 pp 185-187.

Estimand: The causal effect of interest for a given study objective (distinct from an *estimator*, i.e. the specific statistical technique). The estimand takes into account the exposure, the population, the endpoint of interest, and handling of intercurrent events (i.e., occurring after index which may affect measurements or outcomes [like discontinuation, new treatment, or death]).

Reference: Lawrance R, Degtyarev E, Griffiths P, et al. What is an estimand & how does it relate to quantifying the effect of treatment on patient-reported quality of life outcomes in clinical trials?. *Journal of Patient-Reported Outcomes*. 2020 Dec;4(1):1-8. Available at https://doi.org/10.1186/s41687-020-00218-5.

Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15 Chapter 2, page 19.

Estimator: An estimator is a statistic used to estimate an unknown parameter. An estimator is a function of the data in a sample. Common estimators are the sample mean and sample variance which are used to estimate the unknown population mean and variance.

Reference: https://analytics-toolkit.com/glossary/estimator/.

E-value: A measure of the minimum strength of an association that an unmeasured confounder would need to have with both the exposure and the outcome to fully overcome the observed association between the exposure and the outcome, conditional on the covariates included in the model. E-values are measured on the risk-ratio scale. As the E-value increases, the strength of the unmeasured confounder would need to increase to overcome the exposure-outcome association. E-values are one example of quantitative bias analysis.

Reference: Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15 Chapter 13.

VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med. 2017;167(4):268-274.

Exposure misclassification: Occurs when a patient is mistakenly labeled as exposed when unexposed, or unexposed when exposed. It contributes to misclassification bias.

Reference: Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics, 3rd Ed*. Jones and Bartlett Learning, 2014 pp 185-187.

Exposures (more than two): In some cases, there may be a desire to compare more than two exposures for their causal effect on the outcome. In this case, propensity scores can be generated using the generalized propensity score method, where multinomial logistic models are performed for dummy variables representing each exposure to get propensity scores for each; the vector of these generalized propensity scores is used for weighting or matching.

Reference: Thoemmes F, Ong AD. A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*. 2016 Feb;4(1):40-59.

Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15 Chapter 10.

First treatment carried forward: An approach that ignores treatment discontinuation or switching. It is similar to the “intention-to-treat” analysis used in RCT. This approach avoids healthy adherer bias, at the cost of exposure misclassification.

Reference: Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep*. 2015;2(4):221-228.

Frequency matching: While individual matching selects an unexposed person with characteristics (or a propensity score) closest to that of an exposed person, frequency matching randomly selects one or more unexposed people who fall within specified ranges of each matching characteristic (or ranges of propensity score values; e.g., deciles, tertiles) for every exposed person included in the same range.

References: https://www.goldsteinepi.com/blog/matchinginacasecontrolstudyinpractice/index.html

Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *Journal of clinical epidemiology*. 2006 May 1;59(5):437-e1.

https://isctm.org/public_access/Feb2020/Posters/Turkoz-Poster.pdf

https://devline.medunigraz.at/mug_dev/wbAbs.getDocument?pThesisNr=44589&pAutorNr=&pOrgNr=1

Generalized estimating equations: A method of analyzing longitudinal or repeated measures (cluster) data using generalized linear models that accounts for correlation of repeated measures within a patient and handles non-normally distributed outcomes well.

Reference: Ballinger GA. Using Generalized Estimating Equations for Longitudinal Data Analysis. *Organizational Research Methods*. 2004;7(2):127-150.

G-computation / G-formula: (with or without a preceding "parametric"). Like IPTW, this approach can be used to estimate a Marginal Structural Model (MSM) parameter for time-varying as well as single timepoint exposures and confounders. Both IPTW and g-computation can be equivalent to standardization. G-computation can estimate ATE/ATT/ATU and consists of calculating the estimand for different scenarios (e.g., all patients treated, all patients untreated) and then calculating their ratio or difference. In this way, it explicitly employs the potential outcomes framework in causal inference. G-computation is a viable (but more complex) alternative to IPTW because g-computation produces more efficient (i.e., smaller standard errors) and more stable estimates in parametric settings and can better handle heterogeneity involving time-varying exposure and confounding.

Reference: Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol.* 2011;173(7):731-738.

G-estimation: G-estimation is a g-method, but it is unlike g-computation/g-formula in that it involves a Structural Nested Model (SNM) which is conceptually distinct from a Marginal Structural Model (MSM). Compared to IPTW and g-computation / g-formula, it is even more complex.

References: Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ*. 2017 Oct 16;359.

Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *International journal of epidemiology.* 2017 Apr 1;46(2):756-62.

G-methods: The 'g' in G-methods stands for "generalized." These methods include inverse probability weighted marginal structural models, g-estimation of a structural nested model, and the g-formula (or g-computation) approach. These methods do not "fix" the post-baseline variable values and thus avoid problems associated with "conditioning on the future." They can account for feedback between time-varying treatments and time-varying confounders.

References: https://towardsdatascience.com/causal-inference-in-data-science-g-estimation-of-structural-nested-models-d11b1e3c9360

Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *International journal of epidemiology.* 2017 Apr 1;46(2):756-62.

Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Hazard ratio: A measure of how often an event happens in one group compared to how often it happens in another group, over time. A hazard ratio of 1 means no difference in occurrence, while a hazard ratio greater than one means a higher occurrence of an event in the group of interest (numerator) relative to the comparator group.

Reference: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/hazard-ratio

Barraclough H, Simms L, Govindan R. Biostatistics primer: what a clinician ought to know: hazard ratios. *J Thorac Oncol*. 2011;6(6):978-982.

Healthy adherer effect/Healthy user bias: Users of medications or other treatments may be healthier than non-users due to factors other than the treatment (e.g., better exercise and diet).

Reference: Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. *J Gen Intern Med*. 2011;26(5):546-550.

Immortal time bias: "Immortal time" is time during which the outcome cannot occur. It is important to design studies with an eye for immortal time so that it can be avoided. This bias was first described in heart transplant research in the 1970s. Patients who received a transplant were those who *survived long enough* to receive it. This makes heart transplants look beneficial by design. Similarly, in prevalent user bias, prevalent users of a drug have already survived while using the drug for some period.

References: Gleiss A, Oberbauer R, Heinze G. An unjustified benefit: immortal time bias in the analysis of time-dependent events. *Transplant International.* 2018 Feb;31(2):125-30.

Prada-Ramallal G, Takkouche B, Figueiras A. Bias in pharmacoepidemiologic studies using secondary health care databases: a scoping review. *BMC medical research methodology*. 2019 Dec;19(1):1-4.

Suissa S, Immortal Time Bias in Pharmacoepidemiology, *Am J Epidemiol*, Volume 167, Issue 4, 15 February 2008, Pages 492–499.

Individual treatment effect (or individual ATE): Defined as the difference between outcomes if a given patient received treatment 1 vs. treatment 2. Since a patient can only receive a single treatment, this outcome difference cannot be calculated. Instead, researchers calculate an average treatment effect (ATE) using mean outcomes for patients receiving treatment 1 or treatment 2. Potential effect modification/ heterogeneous treatment effects imply that a given patient may not themselves experience the ATE. Recent years have seen the development of many methodologies to support personalized medicine by estimating quantities such as individual treatment effects or rules (ITR), typically via machine learning algorithms.

References: Meid AD, Ruff C, Wirbka L, et al. Using the Causal Inference Framework to Support Individualized Drug Treatment Decisions Based on Observational Healthcare Data. *Clin Epidemiol*. 2020;12:1223-1234.

Faries D, Zhang X, Kadziola Z, et al. Real world health care data analysis: causal methods and implementation using SAS. SAS Institute; 2020 Jan 15 Cary, NC: SAS Institute Inc. Chapter 15.

https://www.vanderschaar-lab.com/individualized-treatment-effect-inference/

Individual matching: While frequency matching randomly selects one or more unexposed people within ranges of each matching characteristic of interest (or propensity scores) for every exposed patient within the same range, individual matching selects one or more unexposed people with a combination of characteristics (or propensity score) closest to that of an exposed person. A common method for individual matching is "nearest neighbor" matching.

Reference: https://www.goldsteinepi.com/blog/matchinginacasecontrolstudyinpractice/index.html

Information bias: If the association between the treatment and outcome is strengthened or weakened as a result of an error in the measurement of a variable, the corresponding bias is called information bias. It

is also known as measurement bias or measurement error. Recall bias, protopathic bias and surveillance bias are all subtypes of information bias. Sometimes the errors in the variable may depend on the distribution of errors in another variable (known as differential misclassification).

References: Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC14.pdf

https://catalogofbias.org/biases/information-bias/

Informed/Informative censoring: Informed or informative censoring occurs if an individual is censored (lost to follow-up) due to events that are related to the exposure of interest, and the censoring carries information about what would have happened in the absence of censoring. Examples are death of the patient, patient experiencing drug toxicity or worsening of clinical condition, or a patient getting cured and dropping out of the study.

Reference: Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003;89(2):232-238.

Intention to treat (ITT): Intention-to-treat is a methodological approach used in RCTs. In an intention-to-treat approach, patients are analyzed according to their initial exposure group, regardless of subsequent changes to their exposure status during follow-up: patients are not censored if there are interruptions or switching in treatment. This is in contrast to a "per protocol" design. For observational studies, ITT is also known as “first treatment carried forward,” while the “per protocol” approach is often referred to an “as-treated" design.

Reference: https://www.pcori.org/assets/Standards-for-Causal-Inference-Methods-in-Analyses-of-Data-from-Observational-and-Experimental-Studies-in-Patient-Centered-Outcomes-Research.pdf

Interventional: An interventional study is a type of study that is typically done prospectively and is tailored to evaluate the direct impact of a treatment, preventive measure, care management program, etc. on a specific outcome (e.g., disease incidence). It stands in contrast to observational study designs, which are typically retrospective.

Reference: Thiese MS. Observational and interventional study design types; an overview. *Biochem Med (Zagreb)*. 2014;24(2):199-210.

Instrumental variable: An Instrumental Variable (IV) is used to control for confounding and measurement error in observational studies so that causal inferences can be made. It should fulfil three assumptions: (1) it should affect treatment or be associated with treatment by sharing a common cause; (2) it should be a factor that is as good as randomly assigned so that it is unrelated to patient characteristics, and (3) it should be related to the outcome only through its association with treatment. A common IV in the healthcare context is "distance to the nearest hospital" as a proxy for likelihood of receiving a particular intervention. Due to the heroic assumptions required for this design to work, the practical use of IV analysis is limited.

References: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances.

Soumerai SB, Koppel R. The Reliability of Instrumental Variables in Health Care Effectiveness Research: Less Is More. *Health Serv Res*. 2017;52(1):9-15.

https://effectivehealthcare.ahrq.gov/products/instrumental-variable-methods/research

Inverse probability of censoring weighting (IPCW): A method of adjusting analysis by using weights based on probability of being censored during study follow-up due to being exposed to a specific treatment, noncompliance with the initial treatment regimen, or the occurrence of a competing outcome. It

is used to determine what the endpoints of the artificially censored participants would have been had they never been exposed to the specific intervention, complied with the initial regimen, or not developed the competing outcome. This approach has been used to address bias due to noncompliance with fixed or dynamic treatment plans, as well as to estimate survival in the presence of competing risks.

References: https://ccss.stjude.org/content/dam/en_US/shared/ccss/documents/concept-prop/concept-prop-11-09.pdf

Howe CJ, Cole SR, Chmiel JS, Muñoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *Am J Epidemiol.* 2011;173(5):569-577.

Inverse probability of treatment weighting (IPTW): IPTW is a weighting method where weights are assigned to patients based on the inverse of their probability of receiving treatment, as estimated by the propensity score. Weighting subjects by the inverse probability of treatment received creates a synthetic sample in which treatment assignment is independent of measured baseline covariates. IPTW can reduce bias from measured confounding when estimating treatment effects (ATE, ATT, or ATU, depending on how the weights are constructed). Weighting replicates patients and induces a lack of independence in the data such that naïve standard error estimates will tend to underestimate the true variance, hence a robust variance estimator or a bootstrap are required.

References: Allan V, Ramagopalan SV, Mardekian J, et al. Propensity score matching and inverse probability of treatment weighting to address confounding by indication in comparative effectiveness research of oral anticoagulants. *J Comp Eff Res*. 2020;9(9):603-614.

Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679.

Marginal: This can simply mean not conditional on covariates (see “conditional effect”), but in the context of Marginal Structural Models (MSMs) it refers to estimation of the marginal distribution of the potential (counterfactual) outcomes. This is most clearly explained in the *Appendix* of Breskin (2018).

References: Breskin A, Cole SR, Westreich D. Exploring the Subtleties of Inverse Probability Weighting and Marginal Structural Models. *Epidemiology*. 2018;29(3):352-355.

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res*. 2011;46(3):399-424.

Marginal effect: A marginal effect is the average effect, at the population level, of moving an entire population from untreated to treated. It is the effect in the population when the treated and untreated are balanced in their covariate distribution (e.g., through standardization/weighting or matching). This is in contrast to a "conditional effect". (Note that "marginal effect" also has an entirely different meaning as the effect of a one-unit change in a covariate of a regression model when all other covariates are held constant.)

Reference: Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med.* 2013;32(16):2837-2849.

Marginal structural model (MSM): Marginal structural models (MSMs) are a class of models that were developed to account for time-varying confounders when examining the effect of a time-dependent exposure (e.g., treatment) on an outcome in the presence of censoring. IPTW is used to create a pseudo-population with balanced covariates in each time period, and IPCW to adjust for informed censoring. Here, "marginal" refers to the fact that the marginal distribution of the counterfactual random variables Y1 [outcome if treated] and Y0 [outcome if untreated], rather than the joint distribution, is modelled (see Appendix of Breskin (2018)). "Structural" refers to the final model, where the exposure-outcome relationship is estimated with the weights applied. (Note that MSM is also an acronym for “multistate modeling” in the decision-analytic literature.)

Reference: Breskin A, Cole SR, Westreich D. Exploring the Subtleties of Inverse Probability Weighting and Marginal Structural Models. *Epidemiology*. 2018;29(3):352-355

Matching: Matching is a commonly used approach to reduce confounding bias by ensuring that patients are compared only to other patients with similar values for particular confounders.

Reference: Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.

Measured confounding: Variables that are observed in the dataset used for analysis, and that influence both exposure and outcome, are measured confounders (e.g., baseline disease severity or prior medication use).

Reference: Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Mediator: A variable is a mediator when it is affected by the treatment or exposure, and also affects the outcome itself. When a mediator is hypothesized, the total effect can be broken down into direct (effect from exposure to outcome) and indirect effects (effect from exposure to outcome through the mediator). For example, cholesterol level is a mediator of the effect of statins on CV outcomes.

Reference: https://www.publichealth.columbia.edu/research/population-health-methods/causal-mediation

Misclassification: Misclassification occurs when individuals are incorrectly assigned to a different category of an exposure, outcome, or both instead of the category where they actually belong. This can lead to misinterpretations as any associations observed between the exposure and the outcomes of interest may be spurious. Misclassification can be of two types: differential and non-differential. Differential misclassification occurs when the probability of being misclassified differs between groups of exposure or outcome. Non-differential misclassification occurs when the probability of individuals being misclassified is equal across all groups of exposure or outcome.

Reference: https://catalogofbias.org/biases/misclassification-bias/

Missing data: ENCePP Guide (Page 55) defines missing data as "(…) the data value that is not stored for a variable in the observation of interest. Missing data are a common problem in all datasets and can have a significant effect on the conclusions that can be drawn from the data (…)" There are different types of mechanisms assumed to generate missing data patterns, and different methods to deal with them (e.g., complete case analysis or multiple imputation).

Reference: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances

Mixed models with repeated measures (MMRM): Mixed models with repeated measures are models explicitly accounting for fixed and random effects occurring due to the correlations between repeated measurements within each patient. This model avoids model misspecification and is unbiased for data missing completely at random or at random. MMRM are sometimes called hierarchical linear models, multilevel models, or mixed effect models.

References: Bell ML, Rabe BA. The mixed model for repeated measures for cluster randomized trials: a simulation study investigating bias and type I error with missing continuous data. *Trials*. 2020;21(1):148. Published 2020 Feb 7.

https://www.apa.org/science/about/psa/2017/01/multilevel-modelling

https://www.theanalysisfactor.com/multilevel-hierarchical-mixed-models-terminology/

Model assumptions: Model assumptions denote the large collection of explicitly stated (or implicitly premised) conventions, choices and other specifications on which any statistical model is based. For example, constant variance of terms, normally distributed error terms, independent observations.

Reference: Pearl J. An introduction to causal inference. *Int J Biostat*. 2010;6(2):7. Published 2010 Feb 26.

New user design: A "new user" design follows patients beginning at the initial time of exposure and enables researchers to establish clear temporality among baseline confounders, exposures, and outcomes.

Reference: https://www.pcori.org/assets/Standards-for-Causal-Inference-Methods-in-Analyses-of-Data-from-Observational-and-Experimental-Studies-in-Patient-Centered-Outcomes-Research.pdf

Observational: An observational study is a type of study where the investigator does not influence receipt of exposure. It can be used to assess causation in exposure-outcome relationships.

Reference: Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Outcome misclassification: Outcome misclassification occurs when patients are erroneously assigned to a different outcome category than the one to which they should be assigned. As an example, patients who do not have a family history of dementia may be tested less often than those who do. This may result in a greater proportion of missed dementia cases in populations without a family history of dementia than those with it. The influence of misclassification on the point estimate should be quantified or, if this is not possible, its impact on the interpretation of the results should be discussed.

Reference: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances

Overadjustment: Defined inconsistently in the literature and common use. Two related ideas:

1) Overadjustment bias: controlling for any variable that increases bias of a causal effect estimate between an exposure and an outcome (e.g., including a mediator in a regression model).
2) Unnecessary adjustment: Controlling for a variable in a way that does not create bias but lowers precision (e.g., a variable only related to the exposure but not the outcome).

These effects are relative to the causal effect of interest, relative to the method chosen to estimate that effect, and relative to the other variables that are controlled for.

References: Lu H, Cole SR, Platt RW, Schisterman EF. Revisiting Overadjustment Bias. *Epidemiology*. 2021;32(5):e22-e23.

VanderWeele TJ. On the relative nature of overadjustment and unnecessary adjustment. *Epidemiology*. 2009;20(4):496-499.

Per protocol: In a per protocol design, only patients who adhere to the study protocol (e.g., those who adhere to a particular intervention) are analyzed. Patients are censored if there are interruptions or switching in treatment in the follow-up period. When this approach is used for RCTs that were randomized for an ITT design, Murray et al. note that "Per-protocol effects estimated using this method are highly susceptible to the placebo paradox, also called the "healthy adherers" bias, where individuals who adhere to placebo appear to have better survival than those who don't." In other words, analyzing RCT data – and observational data – using a "per protocol" (or "as treated") design requires additional design considerations and statistical adjustments.

References: Shah PB. Intention-to-treat and per-protocol analysis. *CMAJ*. 2011;183(6):696.

Murray EJ, Caniglia EC, Petito LC. Causal survival analysis: A guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence. *Research Methods in Medicine & Health Sciences*. 2021;2(1):39-49.

Potential outcomes framework: Also known as the Neyman-Rubin Causal Model, or Counterfactual Framework. Potential outcomes are what would happen to a person had they been in the contrasting treatment group. Think, "Would I have still gotten sick if I had not eaten that chicken sandwich?" Y(0) is your outcome if you eat the sandwich and Y(1) is your outcome if you do. This framework is how we arrive at thinking about the "average effect", i.e. the ratio or difference between Y(0) and Y(1) at the population level. In other words, the risk ratio or difference comparing what would happen if all people had not eaten the chicken sandwich versus if all people did (this would be the ATE).

References: Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf (See Section 13.4 IP weighting or standardization?)

Prevalent user bias: Prevalent user bias occurs when the exposed group is comprised of individuals who have been exposed to the pharmaceutical of interest for some period of time before the study initiation. If the risk varies with time, the early period of pharmacotherapy can introduce substantial selection bias as groups of prevalent users may be enriched for individuals who are resilient to any influence that the drug exerts on disease risk. Also known as "depletion of susceptibles".

Reference: Ahern TP. Pharmacoepidemiology in pharmacogenetics. *Advances in Pharmacology*. 2018 Jan 1;83:109-30., Pharmacogenetics; Edited by Kim Brøsen.

https://www.sciencedirect.com/topics/medicine-and-dentistry/selection-bias

Probabilistic bias analysis: Probabilistic bias analysis is an analysis where bias parameter values are sampled from an investigator-assigned distribution and this process is repeated multiple times using Monte-Carlo techniques to generate a frequency distribution of bias-adjusted effect estimates. It is a method that helps with quantitative bias assessment.

Reference: Hunnicutt JN, Ulbricht CM, Chrysanthopoulou SA, Lapane KL. Probabilistic bias analysis in pharmacoepidemiology and comparative effectiveness research: a systematic review. *Pharmacoepidemiology and drug safety*. 2016 Dec;25(12):1343-53.**Propensity score:** According to ENCePP guide (page 51) a propensity score is "The conditional probability of exposure to a treatment given observed covariates." Conditional on the true propensity score and under certain assumptions, exposure status is independent of measured baseline covariates. Therefore, estimates of PS can be used to create cohorts balanced on observable baseline characteristics in order to replicate the balance achieved by a randomized study design.

Reference: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at: http://www.encepp.eu/standards_and_guidances

Protopathic bias (Reverse causality): "Protopathic bias arises when the initiation of a drug (exposure) occurs in response to a symptom of the (at this point undiagnosed) disease under study (outcome). For example, use of analgesics in response to pain caused by an undiagnosed tumor might lead to the erroneous conclusion that the analgesic caused the tumor."

Reference: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances Page 47/168.

Quantitative bias analysis: To estimate the potential magnitude and direction of biases arising from systematic errors affecting an estimate, and to quantify the uncertainty about these biases. Inclusive of conventional sensitivity analyses, E-value, etc.

Reference: Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis. Int J Epidemiol. 2014;43(6):1969-1985.

Risk ratio: Compares the risk of a health event (disease, injury, death, etc.) among one group, typically the exposed, with the risk among another group, typically the unexposed/control, through dividing the former by the latter. Also called relative risk.

Reference: https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section5.html

Risk difference: Compares the risk of a health event (disease, injury, death, etc.) among one group, typically the exposed, with the risk among another group, typically the unexposed/control, through subtracting the latter from the former.

Reference: https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_association/EP713_Association5.html

Regression: "A set of statistical processes for estimating the relationships between a dependent variable (outcome) and one or more independent variables ('predictors', 'covariates', 'explanatory variables')" (Wikipedia).

Reference: https://en.wikipedia.org/wiki/Regression_analysis

Retrospective study: An observational study that examines patients who already have a disease or condition at the time of the study. This is in contrast to prospective studies where patients are enrolled before they develop the disease or outcome in question.

Reference: https://www.statisticshowto.com/retrospective-study/

Selection bias: A possible bias that arises when study members are selected by a process that is affected by both the treatment and the outcome of interest (e.g., loss to follow-up, self-selection, missing data, healthy adherer/worker bias).

Reference: https://catalogofbias.org/biases/selection-bias/

Standardized morbidity (or mortality) ratio weights (SMRW): Used to estimate the ATT. They differ from IPTWs in that treated patients are given a weight of 1. Untreated (referent group) individuals get a weight of PS/(1-PS), where PS is the propensity score. While IPTWs standardize the results to the covariate distribution of the entire study population (i.e., treated and untreated together), SMRWs standardize the results to the covariate distribution of the treated population.

Reference: https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_standardizedrates/EP713_StandardizedRates_print.html

Stabilized weights: In a conventional IPTW approach, the *denominator* of the weight is used to reduce confounding bias. An additional *numerator* term can be added to reduce excess variation in the weights (such variation can lead to inflated sample sizes). For example, the denominator in simple IPTW is based on the propensity score (predicted treatment probability given observable covariates), while a stabilizing numerator would be based on just the probability of receiving treatment (i.e., the given proportions of treated and untreated patients in the full sample). Thus, IPTWs are stabilized by multiplying the weight by the marginal probability of receiving treatment.

References: https://www.mayo.edu/research/documents/biostat-84-pdf/doc-20024406

Hernán MA, Robins JM. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Standardization: “Standardization” is a class of methods within epidemiology that addresses confounding by weighting data to the distribution of confounder covariates in the population for whom knowledge of an effect is desired. It allows for the comparison of two populations with different

distributions of important confounders. For example, comparing rates of death or cancer between two different states. If one state tends to attract retirees, the death rate may appear higher. A 'fairer' comparison can be made by weighting the death frequencies by the proportion of people in each age category and comparing to some standard (e.g., the U.S. population, the other state, etc.). Standardization can be thought of as weighting and is similar to calculating a GPA in a report card. G-formula (g-computation) and IPTW are equivalent to standardization in their simplest implementations.

Reference: https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_standardizedrates/EP713_StandardizedRates_print.html

Stratification: To examine an association of interest at different levels of a potential confounding factor. Outcomes are compared within strata (subgroups) of "like" patients and then averaged across strata. As direct stratification becomes infeasible for more than a few strata at a time, other approaches (such as stratification using propensity scores) have been developed.

Reference: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/bs704_multivariable3.html

Hernán MA, & Robins J M. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Survival analysis: An analysis where the outcome is *time until an event occurs*. This outcome has special features (such as nonnegativity, skewness, and the issue of censoring) that require specific types of estimators (e.g., Kaplan-Meier curves). Time to death is a common event of interest (hence the term "survival analysis"), but other events can also be analyzed this way (and then "survival" refers to a patient not experiencing the event of interest).

Reference: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/BS704_Survival_print.html

Target trial: Creating a target trial is an approach to help avoid biases in observational research. In the first step, we "design a hypothetical ideal randomised trial that would answer the research question of interest. The target trial is described with regards to all design elements: the eligibility criteria, the treatment strategies, the assignment procedure, the follow-up, the outcome, the causal contrasts and the analysis plan. In the second step, we specify how to emulate the design elements of the target trial and what analytic approaches to take given the trade-offs in an observational setting." (Reference 1)

Reference: The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 5). EMA/95098/2010. Available at http://www.encepp.eu/standards_and_guidances Page 40/168

Hernán MA, & Robins J M. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Time lag bias: The rapid or delayed publication of research findings, depending on the nature and direction of the results. Typically, the results of "negative" studies (which fail to reject the null of no treatment effect) take substantially longer to publish than "positive" studies (which do reject the null).

Reference: https://handbook-5-1.cochrane.org/chapter_10/10_2_1_1_time_lag_bias.htm

Time window bias: Time-related biases in cohort studies can produce illusory "beneficial" effects of medications due entirely to an artifact of the analytic design. For example, time window bias "results from the use of time-windows of different lengths between cases and controls to define time-dependent exposures" (Suissa 2011). The "magnitude of the bias is proportional to the ratio of the unequal time-window lengths" (Suissa 2011).

References: Suissa S, Dell'aniello S, Vahey S, Renoux C. Time-window bias in case-control studies: statins and lung cancer. *Epidemiology*. 2011;22(2):228-231

Suissa S, Dell'Aniello S. Time-related biases in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2020;29(9):1101-1110.

Time-invariant confounding: Presence of common causes of both treatment and outcome that can reasonably be considered static during the period of analysis (for example age, sex, chronic comorbidities, socioeconomic status).

Reference: Hernán MA, & Robins J M. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Time-varying confounding: Presence of common causes of (time-varying) treatments and the outcome of interest that can change over time and affect the choice of the (time-varying) treatments (for example disease severity, side effects, biomarkers). A particular challenge is presented by treatment-confounder feedback, when a confounder affects the treatment, and the treatment affects the confounder (e.g., a biomarker that responses to treatment but also determines the next treatment).

Reference: Hernán MA, & Robins J M. Causal Inference: What if? December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Time-varying treatments: A study design choice whereby the timing dimension of clinical treatments is explicitly incorporated into the methodology and analysis. For example, allowing the initial treatment assignment to change on a fixed (e.g., quarterly) basis (in a retrospective study), or to change at regularly scheduled clinic visits (in a prospective study). As there are many possible treatment strategies in this context, there is no single definition of causal effect for time-varying treatments.

Reference: Hernán MA, Robins JM. Causal Inference: What if? (2020). December 31, 2020. Available as a free eBook at https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf

Total effects: Describes the entire effects of an exposure on an outcome: the sum of direct and indirect effects. For example, when calculating the total effect, the effect of exposure on a mediator which then itself affects the outcome is included. Controlling for mediators typically leads to a biased estimate of the total effect of the exposure; however, the *controlled direct effect* thereby obtained may also be of research interest.

Reference: https://www.publichealth.columbia.edu/research/population-health-methods/causal-mediation

Unmeasured confounding: Existence of one or more confounders that are not controlled for in the analysis, typically due to missing data. Presence of unmeasured confounding can usually not be excluded, because we are seldom certain that we know all the confounding domains. Approaches such as instrumental variables may to some extent address issues of unmeasured confounders. There are also tools to quantify how strong the association of an unmeasured confounder with the treatment and outcome would need to be in order to bias the measured treatment effect (such as the E-value).

References: Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?. *Clin Pharmacol Ther*. 2017;102(6):924-933.

https://training.cochrane.org/handbook/current/chapter-25#section-25-2

Variance ratio: Ratio of the variance of a variable measured in the treated subjects to the variance of the variable in untreated subjects. Can be used to assess balance in baseline characteristics across treatment and control cohorts. Acceptable range is 0.5 to 2.0. Can be calculated for each baseline covariate as well as for the propensity score and help inform specification of the adjustment model (e.g., propensity score logistic regression).

Reference: Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci*. 2010;25(1):1-21.

Weighting: Weighting is used to create a pseudo-population that is a better fit for the planned analysis (for example, weighting survey respondents to be more representative of the population they were sampled from). In the causal design framework discussed here, weighting aims to create populations of treated and untreated patients that are balanced with respect to baseline covariates. This is typically done by weighting with the inverse of the propensity score (IPTW).

Reference: Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015;34(28):3661-3679