Development of a Web-based Tool for Comparative Effectiveness Research Using Observational Data

Yi Zhang[1], Mae Thamer[2], Onkar Kshirsagar[3]

[1,2,3] Medical Technology and Practice Patterns Institute, 5272 River Road, Suite #500, Bethesda, MD 20816

Abstract

Core patient-centered outcomes research (PCOR) questions are causal: should I do A or B? To conduct comparative effectiveness (CER) research for causal questions using observational data, two key steps are required: 1. Formulating a well-defined causal question that is relevant to patients and useful for decision making; 2. Providing a valid answer using the best available data and CI analysis methods. Getting both steps right is crucial: a poorly formulated causal question may not lead to an actionable answer even when using high-quality data and cutting edge methodology; an incomplete data or inappropriate analytical approach may lead to a biased answer even for a well defined causal question. Step 1 (formulating the causal question) requires input from PCOR stakeholders including clinicians and patients. Step 2 (providing a valid answer) requires input from researchers with expertise in CER and statistical methods. In practice, there will typically be an iterative process to complete these steps and thus sound PCOR necessitates the close collaboration of stakeholders and researchers throughout the entire research process. In this article, we describe an effort to develop a web-based tool CERBOT (Comparative Effectiveness Research Based on Observational Data to Emulate a Target Trial**)** that aims to 1) provide guidance and support for CER based on 'real-world' observational data, and 2) facilitates close collaboration and communication between researchers and stakeholders

Key Words: PCOR, CER, Causal inference

1. The Problem

Improving the validity of observational studies using causal inference methods is one primary objective of Patient-Centered Outcomes Research (PCOR). When one seeks to understand the extent to which a given therapy, intervention, or strategy affects a particular patient outcome, causal inference (CI) becomes necessary. For example, one of the 4 core PCOR questions raised by PCOR Institute (PCORI), "What are my options and what are the potential benefits and harms of those options?" is a typical causal question. While randomized clinical trials (RCTs) remain the gold standard for answering such questions that compare the effects of interventions on patient outcomes, they often are not ethically, practically, or economically feasible. In these situations, causal inference must be based instead on observational data to provide timely answers. Indeed, observational methods are being increasingly relied upon to extract estimates that can support causal interpretations,[1] especially given the growing availability of high quality observational data such as electronic health records or patient registry databases. However, there is a lack of conceptual framework and practical tools for stakeholders (i.e. physicians and

researcher) for conducting PCOR/CER using CI method. In this article, we describe an effort to develop a web-based tool to implement the process of conducting CER based on a framework of emulating an hypothetical randomized study.

2. The Conceptual Framework

Core patient-centered outcomes research (PCOR) questions are causal: should I do thing A or thing B to improve my health? To conduct causal inference (CI) studies for PCOR questions, two *key* steps are required:

1. Formulating a well-defined *causal* question that is relevant to patients and useful for decision making.
2. Providing a valid answer using the best available data and CI analysis methods.

Getting both steps right is crucial: a poorly formulated causal question may not lead to an actionable answer even when using high-quality data and cutting edge methodology; an inappropriate analytical approach or incomplete data may lead to a biased answer even for a well defined causal question.
Step 1 (formulating the causal question) critically requires input from PCOR *stakeholders*, including patients, clinicians, patient advocates, and policymakers. Stakeholders will need to request guidance from statisticians as they move from Step 1 to Step 2. Step 2 (providing a valid answer) critically requires input from *researchers* with expertise in comparative effectiveness research (CER). In practice, there will typically be an **iterative** process to complete these steps, e.g., the causal question may need to be modified to match the available data, and thus sound PCOR necessitates the close collaboration of stakeholders and researchers throughout the entire research process. Stakeholders and researchers need to engage in both steps of CI research. (**Figure 1**). Specifically,
Step 1: Formulating a <u>well-defined causal question</u>. In principle, a well-defined causal question to compare actionable interventions could be answered by a randomized experiment in which all participants adhere perfectly to the strategies specified in the study protocol. (In practice, such randomized experiment may be costly, infeasible, or unethical, which is why we resort to the analysis of existing data.) Using observational data to formulate a causal research question is complex. Therefore, a method to ensure a well-defined causal question is to design the ideal, hypothetical randomized trial that stakeholders would like to conduct if practical and ethical constraints did not exist. We refer to this trial as the 'hypothetical' *target trial*. Formulating a well-defined causal question or scientific hypothesis requires a detailed and clear specification of the protocol of the target trial. It is important to meet the requirements of internal validity, feasibility, timeliness, and relevance.

Step 2: Providing a valid answer. The choice of data analysis technique(s) follows from the research question defined in Step 1, understanding of data, confounding and possible biases. The tool will guide researchers in the choice and application of appropriate CER techniques, especially so-called 'causal modeling' (CI) techniques often perceived as unduly complex or challenging.

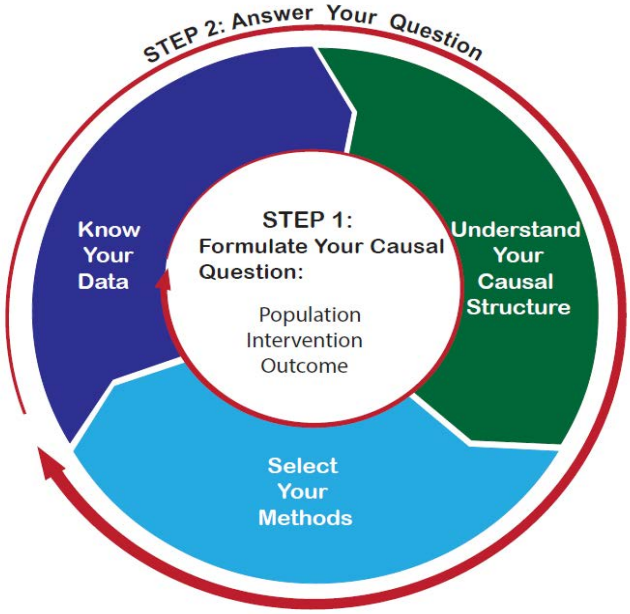

Figure 1: **Conceptual Framework to Conduct Causal Inference Based on Oberservational Data**

3. Major Functionality of CERBOT

3.1. Flowchart

To conduct CI research to compare dynamic interventions, the first key step should be formulating a well-defined causal question that is relevant to patients and useful for decision making. A method to ensure a well-defined causal question is to design the ideal, hypothetical randomized clinical trial (RCT) that stakeholders would like to conduct if practical and ethical constraints did not exist. It becomes problematic if you cannot imagine such an RCT to address your research question. A detailed and clear unambiguous specification of the protocol of this target trial (TT) precludes ill-defined questions and spurious conclusions. The choice of analytical methods is dictated by the specific research question. CERBOT will guide users to 1) construct a well-defined research question by specifying the TT; and 2) identify and use appropriate analytical methods. **Figure 2** shows the CERBOT flowchart illustrating the process of specifying and emulating the TT. To design a TT to answer your CER question, six main components (modules) are recommended to construct: 1). eligibility criteria of patient population, 2) treatment strategies, 3) assignment procedures, 4) follow-up, 5) study outcomes, and 6) causal effects of interests. Each module represents a unique component of the TT. First, you specify (delineate) each component. Second, you review your data to assess the feasibility of your specifications (also called emulate). Otherwise you must re-specify until emulation is possible. This iterative process will result in a summary TT that can implemented with available data. The final output also including selection of analytic methods based on specified components.

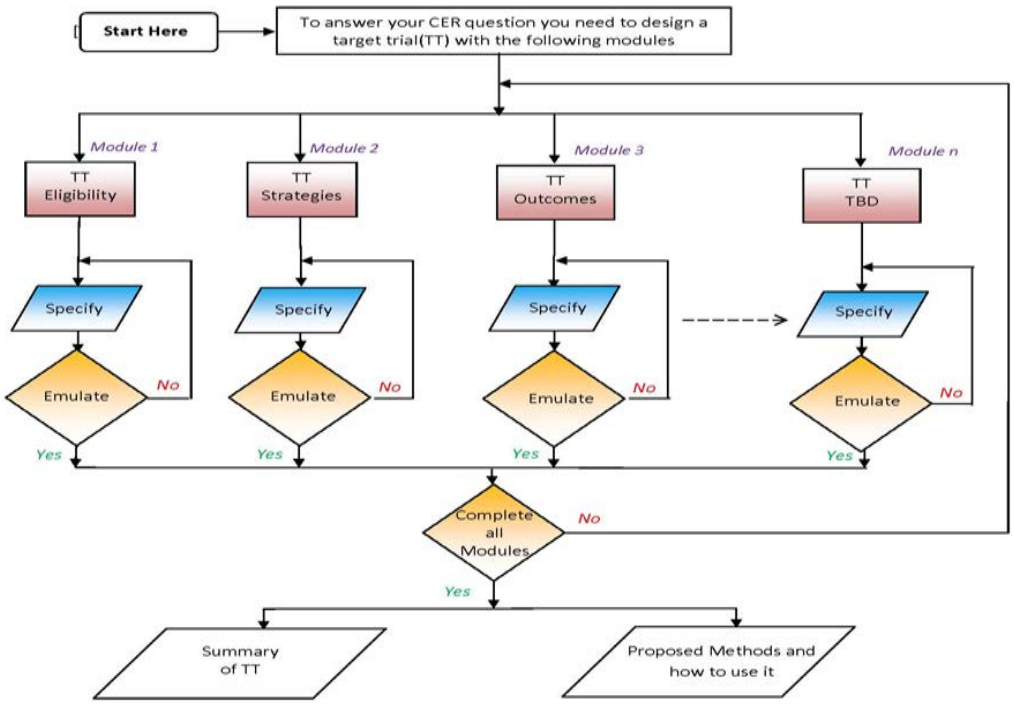

Figure 2: CERBOT flowchart

3.2. Modules

Table 1a-1g below list detailed descriptions of each module corresponding to protocol component of the TT.

Table 1a. Module 1. *Population Eligibility*
Define exactly which patients should be included in your TT

Eligibility Criteria	***Description***
Inclusion criteria	Are they broad enough to represent population of interest?
Personal	
Age	
Sex	
Race	
Ethnicity	
Specific disease	
Disease of a particular severity or duration	
Medical condition	
Lab results	
Diagnostic methods	
Particular features of clinical history	
Other	

Enrollment period	
Setting	
Geographic location	
Other	Any other inclusion criteia you can think of?
Exclusion criteria	Are they restrictive enough to narrow confounding? impact on the generalizability of the study?
Health status or any clinical conditions	
Clinical/laboratory indicators	
Drugs	
Devices	
Surgery	
Cancer or other chronic debilitating disease	
Alcohol and/or drug dependence	
Setting	
Geographic location	
Other	Any other exclusion criteria you can think of?
Study Baseline	When all eligibility criteria are met
Patient subgroups	Defined by characteristics such as age, sex, race, disease stage

* There will be 1-5 examples focusing on different types of interventions. These examples serve 2 purposes: 1) to illustrate the process of using the modules; and 2) to show what the final output looks like by going through these modules. Users can choose to which example to see at any time (otherwise they are not shown on the screen).

Table 1b. Module2: Treatment Strategies

Define and characterize the set of interventions to be compared	
should be directly linked to the study question being addressed. Treatment arm, control arm (current best care, standard/usual care, minimal intervention).	
Number of treatment strategies to be compared	
Define 1st treatment strategy	
Strategy label	
Number of intervention variables	must be actionable and implementable. Cannot be a biomarker
Variable name	Specify
One time	
Sustained (longitudinal)	
Fixed time for treatment change	e.g. no change after baseline, change at 6 month
Variable change based on evolving variables	
Clinically indicated	unethical not to change
Not clinically indicated	ethical to/not to change
Intervention period	Schedules of interventions
Start	

End	
route and mode of administration	
Any other specifics about when or how to intervene	
Setting	
Define 2nd treatment strategy	Do the same thing as Strategy 1

Table 1c. Module 3: Assignment procedures

Participants will be randomly assigned to either strategy at baseline, and will be aware of the strategy they have been assigned to. To emulate the random assignment of the strategies at baseline, we need to adjust for all measured confounding factors.

Baseline confounders	*Adjustment for baseline confounders*
Patient demographics and sociaeconomic status; confounding factors related to patient outcomes and/or related to both treatment strategies and patient outcomes.	Data source, definition, measurement, missingness, and validity of these confounding factors in your existing data
Comparability(exchangeability) achieved through randomization	Regression, matching, stratification, propensity score, inverse-probability-weighting, g-estimation, doubly-robust methods, machine learning tools

Table 1d. Module 5: Follow-up

Specification	*Emulation*
Start (time zero of follow-up, or baseline): when an eligible individual initiates a treatment strategy. Eligibility criteria need to be met at or before baseline.	If eligibility criteria are met at multiple times, there're two unbiased choices of time zero: 1) us the first eligible time or a random eligible time; b) use every elibible time by emulating multiple nested trials, each of them with a different start of follow-up
Grace period: time between baseline and initiation of treatment strategies	Must allow for an analogous grace period measured from baseline. An individual's observational data often is consistent with more than one treatment strategy during grace period. There're two unbiased choices: 1) randomly assign individule to one of the two or more potential strategies; 2) create multiple copies of this individual (clones) assigned with different strategies
Loss-to-follow up (drop outs)	Define, how to measure

Table 1e. Module 5: Outcomes

Succinctly and precisely define the most relevant clinical outcomes	
Number of study outcomes	
Outcome 1	Most clinically relevant ,Direct measure of clinical benefit, patient-censored?
Primary	

Secondary	Should be limited
Composite or not	Specify what it's comprised of
Surrgate outcome?	aka intermediate endpoint. Try to avoid because it may not be a true predictor of clinical benefit
Justification	why the main disease outcome of interest is not being used
Timing of assessment	
Start of follow up	
Length of follow up	Do you have enough follow up in your data?
Censoring	dropout, loss to follow up
Measurement	
Dichotomous, categorical, continuous	
time to first occurrence	
rate of occurrence	
slope of rate over time	
proportions	
positive or negative changes	

Table 1f. Causal contrasts of interest

Specification	Emulation
Intention-to-treat: The effect of being assigned to treatment, regardless of treatment received	Requires adjustment for baseline confounders and time-varying selection bias due to loss to follow-up
Per-protocol: The effect of receiving the treatment regimes specified in the study protocol	Artificially censor patients at deviation from originally assigned treatment strategy. Requires adjustment for baseline confounders, time-varying confounding, and selection bias due to artificial censoring
As-treated: The effect of receiving treatment regimes other than the ones specified in the study protocol	Requires adjustment for baseline confounders and time-varying confounding

To accomplish our mission, each module will be accompanied with a detailed manual, case examples, FAQs, check lists, videos, and links to relevant literature and references as appropriate and necessary. A forum section (or commenting/blog) will also be included in the website for both stakeholders and researchers to ask questions, discuss issues and become part of the larger PCOR community conducting CER studies. Finally, CERBOT will allow researchers to create a group or research circle to share contents generated and saved by CERBOT. Ultimately, the CI website will take complex and challenging Toolkit content and present it in a way that is user friendly and valuable to both stakeholders and researchers.

4. CERBOT Output

After submitting information of each model component, CERBOT user will be provided with a report summarizing the specified research question as well as recommended methods including standard statistical methods and new methods developed for CI using observational data. Over the past two decades, many exciting new developments in CI methods have evolved, such as marginal structural models (MSM),**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,**Error! Bookmark not defined.**,[19,20,24] structural equation models (SEM), [2,3,4] G-estimation of structural nested models, the parametric g-formula, [5,6] instrumental variables (IVs), [7,8,9] principle stratification, [10,11,12] propensity scoring (and matching),[13,14,15]and the so-called doubly-robust estimation (DBR).[16,17] These statistical tools have become more appealing to researchers who conduct PCOR analyses, [18,19,20,21,22,23,24,25] including ourselves, [26,27,28,29,30] because these methods are superior to conventional methods in addressing complex confounding issues commonly found in observational data.

There is no 'best' method. The choice of a statistical method is always dependent upon research question. design, causal quantity you are interested in, quality of data, and assumptions made. There are always alternative methods to address the same questions. Sensitivity analyses are always required due to the trade-offs for each method regarding bias and variance, more or less necessary assumptions. CERBOT allows users view the components associated with each method included in CERBOT (regardless the method is the selected or recommended method or not). These pages include: a) basic description of the method; b) steps to implement it; c) tools and resources to help you implementing it; d) common errors and fixes in applications; d) examples; e) sensitivity analysis; and f) limitations.

Acknowledgements

Research reported in this article was funded through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1303-6031). We thank our advisory committee member for their valuable contribution to the development of this tool.

Disclaimer

The statements presented in this work are solely the responsibility of the author(s) and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors or Methodology Committee.

References

[1] PCORI (Patient-Centered Outcomes Research Institute) Methodology Committee. 2013. "The PCORI Methodology Report." pcori.org/research-we-support/research-methodology-standards.

[2] Simpson VL, Hyner GC, Anderson JG. Lifestyle behavior change and repeat health risk appraisal participation: a structural equation modeling approach. Am J Health Promot. 2013 Nov-Dec;28(2):128-35. doi: 10.4278/ajhp.120306-QUAN-128. Epub 2013 Apr 26

[3] Trzeciakowski JP, Gardiner L, Parrish AR. Effects of environmental levels of cadmium, lead and mercury on human renal function evaluated by structural equation modeling. Toxicol Lett. 2014 Apr 21. pii: S0378-4274(14)00158-1. doi: 10.1016/j.toxlet.2014.04.006.

[4] Hoyle RH, Gottfredson NC. Sample Size Considerations in Prevention Research Applications of Multilevel Modeling and Structural Equation Modeling. Prev Sci. 2014 Apr 23.

[5] Joffe MM, Yang WP, Feldman H. G-estimation and artificial censoring: problems, challenges, and applications. Biometrics. 2012 Mar;68(1):275-86. doi: 10.1111/j.1541-0420.2011.01656.x. Epub 2011 Sep 23

[6] Joffe MM Structural nested models, g-estimation, and the healthy worker effect: the promise (mostly unrealized) and the pitfalls. Epidemiology. 2012 Mar;23(2):220-2. doi: .1097/EDE.0b013e318245f798

[7] Hogan JW, Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. Stat Methods Med Res. 2004 Feb;13(1):17-48.

[8] Lu M. The productivity of mental health care: an instrumental variable approach. J Ment Health Policy Econ. 1999 Jun 1;2(2):59-71.

[9] O'Malley AJ, Elwert F, Rosenquist JN, Zaslavsky AM, Christakis NA. Estimating peer effects in longitudinal dyadic data using instrumental variables. Biometrics. 2014 Apr 29. doi: 10.1111/biom.12172

[10] Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002 Mar;58(1):21-9.

[11] Frangakis CE, Rubin DB, An MW, MacKenzie E. Principal stratification designs to estimate input data missing due to death. Biometrics. 2007 Sep;63(3):641-9; discussion 650-62.

[12] Shrier I, Kaufman JS, Platt RW, Steele RJ. Principal stratification: a broader vision. Int J Biostat. 2013 Oct 11;9(2):307-13. doi: 10.1515/ijb-2013-0045.

[13] Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med. 2014 Mar 15;33(6):1057-69. doi: 10.1002/sim.6004.

[14] Austin PC, Schuster T. The performance of different propensity score methods for estimating absolute effects of treatments on survival outcomes: A simulation study. Stat Methods Med Res. 2014 Feb 3

[15] Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat Med. 2014 Mar 30;33(7):1242-58. doi: 10.1002/sim.5984.

[16] Rotnitzky A, Lei Q, Sued M, Robins JM. Improved double-robust estimation in missing data and causal inference models. Biometrika. 2012 Jun;99(2):439-456. Epub 2012 Apr 29

[17] Tchetgen Tchetgen EJ, Robins JM, Rotnitzky A. On doubly robust estimation in a semiparametric odds ratio model. Biometrika. 2010 Mar;97(1):171-180. Epub 2009 Dec 8.

[18] Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men. Epidemiology 2000; 11(5):561-570.

[19] Hernan MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. Stat Med 2002; 21: 1689-1709

[20] Cook NR, Cole SR, Hennekens CH. Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the physicians' health study. Am J Epid 2002 155(11):1045-1053.

[21] Choi HK, Hernán MA, Seeger JD, Robins JM, Wolfe F, Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study, Lancet 2002, 359:1173-1177.

[22] Cole SR, Hernan MA, Robins JM, Anastos K, Chmiel J, Detels R, Ervin C, Feldman J, Greenblatt R, Kingsley L, Lai S, Young M,Cohen M, Muñoz A. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. Am J Epidemiol. 2003;158(7):687-94.

23 Sterne JA; Hernán MA; Ledergerber B; Tilling K, Weber R, Sendi P, Rickenbach M, Robins JM, Egger M. Swiss HIV Cohort Study. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. Lancet 2005;366(9483):378-84

24 Cole SR, Hernán MA, Margolick JB, Cohen MH, Robins JM. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. Am J Epidemiol. 2005;162(5):471-8.

25 Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic & Clinical Pharmacology & Toxicology 2006;98 (3):237–242.

[26] Zhang Y, Thamer M, Cotter D, Kaufman J, Hernán MA The estimated effect of epoetin dose on survival among elderly hemodialysis patients in the United States. Clin J Am Soc Nephrol. 2009;4(3):638-44. PMCID:PMC2653651

[27] Thamer M, Hernán MA, Zhang Y, Cotter D, Michelle P. Relationship between prednisone, lupus activity and permanent organ damage. J Rheumatol. 2009;36(3):560-4

[28] Cotter D, Thamer M, Zhang Y, Kaufman J, Hernan MA. The effect of epoetin dose on hematocrit. Kidney Int. 2008;73(3):347-53

29 Zhang Y, Thamer M, Kaufman J, Cotter DJ, Hernán MA. High doses of epoetin do not lower mortality and cardiovascular risk among elderly hemodialysis patients with diabetes. Kidney Int. 2011; 80(6): 663-669

[30] Zhang Y, Thamer M, Kaufman J, Cotter D, Hernan M. Comparative Effectiveness of Two Anemia Management Strategies for Complex Elderly Dialysis Patients Medical Care; 2014 Mar;52 Suppl 3:S132-9. doi: 10.1097