

Practice of Epidemiology

Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Miguel A. Hernán* and James M. Robins

* Correspondence to Dr. Miguel A. Hernán, Department of Epidemiology, 677 Huntington Avenue, Boston, MA 02115 (e-mail: miguel_hernan@post.harvard.edu).

Initially submitted December 9, 2014; accepted for publication September 8, 2015.

Ideally, questions about comparative effectiveness or safety would be answered using an appropriately designed and conducted randomized experiment. When we cannot conduct a randomized experiment, we analyze observational data. Causal inference from large observational databases (big data) can be viewed as an attempt to emulate a randomized experiment—the target experiment or target trial—that would answer the question of interest. When the goal is to guide decisions among several strategies, causal analyses of observational data need to be evaluated with respect to how well they emulate a particular target trial. We outline a framework for comparative effectiveness research using big data that makes the target trial explicit. This framework channels counterfactual theory for comparing the effects of sustained treatment strategies, organizes analytic approaches, provides a structured process for the criticism of observational studies, and helps avoid common methodologic pitfalls.

big data; causal inference; comparative effectiveness research; target trial

Large observational databases are often used to answer questions about comparative effectiveness or safety. These databases, which are frequently labeled as “big data” (1), typically include many variables measured in many people.

Observational analyses of big databases, however, are not the preferred choice for comparative effectiveness research. We resort to observational analyses of existing data because the randomized trial that would answer our causal question—the target trial—is not feasible, ethical, and timely. One can often fruitfully regard causal inference from big data as an attempt to emulate a target trial (2). If the emulation is successful, the analysis of the observational data yields the same effect estimates (except for random variability) as the target trial would have yielded had the latter been conducted.

Though the concept of a target trial is implicit in many big data analyses, the target trial itself is rarely characterized. In this paper, we outline a framework for comparative effectiveness research using big data that revolves around the explicit description and emulation of the target trial. This framework channels the existing counterfactual theory for comparing the effects of point treatments (3) and sustained treatment strategies (4–6), organizes analytic approaches dispersed throughout the literature, provides a structured process for the criticism of observational studies, and helps avoid common methodologic pitfalls.

Suppose we want to estimate the effect of estrogen plus progestin hormone therapy on the 5-year risk of breast cancer among postmenopausal women. Table 1 lists 7 key components of the target trial protocol: the eligibility criteria, treatment strategies being compared (including their start and end times), assignment procedures, follow-up period, outcome of interest, causal contrast(s) of interest, and analysis plan. Several of these components are part of the widespread population, intervention, control, and outcome approach to the formulation of clinical questions (7). In the next sections, we outline some of the main obstacles for the emulation of the trial protocol using observational data and some strategies to partly overcome those obstacles.

EMULATING THE TARGET TRIAL USING OBSERVATIONAL DATA

Suppose we use a large database of health care claims to emulate the target trial of hormone therapy and breast cancer in this population, as described in Table 1. Because the data were not collected for research purposes, data codes may be inconsistent or ambiguous. For example, the code “breast cancer” might have been recorded in the database when a woman was diagnosed with breast cancer or simply when her physician suspected breast cancer and ordered a diagnostic

Table 1. A Summary of the Protocol of a Target Trial to Estimate the Effect of Postmenopausal Hormone Therapy on the 5-Year Risk of Breast Cancer

| Protocol Component | Description |
|------------------------------|--|
| Eligibility criteria | Postmenopausal women within 5 years of menopause between the years 2005 and 2010 and with no history of cancer and no use of hormone therapy in the past 2 years. |
| Treatment strategies | Refrain from taking hormone therapy during the follow-up. Initiate estrogen plus progestin hormone therapy at baseline and remain on it during the follow-up unless you are diagnosed with deep vein thrombosis, pulmonary embolism, myocardial infarction, or cancer. |
| Assignment procedures | Participants will be randomly assigned to either strategy at baseline and will be aware of the strategy to which they have been assigned. |
| Follow-up period | Starts at randomization and ends at diagnosis of breast cancer, death, loss to follow-up, or 5 years after baseline, whichever occurs first. |
| Outcome | Breast cancer diagnosed by an oncologist within 5 years of baseline. |
| Causal contrasts of interest | Intention-to-treat effect, per-protocol effect |
| Analysis plan | Intention-to-treat effect estimated via comparison of 5-year cancer risks among individuals assigned to each treatment strategy. Per-protocol effect estimation requires adjustments for pre- and postbaseline prognostic factors associated with adherence to the strategies of interest. All analyses will be adjusted for pre- and postbaseline prognostic factors associated with loss to follow-up (57). This analysis plan implies that the investigators prespecify and collect data on the adjustment factors. |

test. Hence, researchers unfamiliar with the data must both consult with knowledgeable data users and conduct data validation studies.

Let us assume that our particular database has passed many high-quality validation studies (e.g., confirmation that 95% of “breast cancer” codes correspond to true diagnoses in medical records) so that the emulation of the target trial can proceed. In this section, we outline the main components of the emulation except for one: the choice of the time zero of follow-up (or baseline), which is deferred to the next section.

Eligibility criteria

Our observational analysis should apply the same eligibility criteria used in the target trial (see Table 1). Therefore, we impose the restriction that, at baseline, women must have been included in the database for long enough to apply the exclusion criteria (at least 2 years) and early enough to contribute 5 or more years of active follow-up. Note that the eligibility criteria cannot include restrictions based on post-baseline events (e.g., “include only individuals who ever used therapy during the follow-up”), which may introduce bias in the analysis of both randomized trials and observational data (8) and which cannot be applied at the time of randomization in a true randomized trial.

Emulating the desired eligibility criteria can still be problematic, as the following examples illustrate. First, suppose that we are interested in a target trial in which all potential participants are required to have a baseline mammography and then those with breast calcifications are excluded. Emulating this target trial may be impossible if our observational database only includes information on whether a mammography was performed (for billing purposes) but not the findings. We then must decide whether emulating a different target trial that includes women with calcifications also addresses a question of interest.

Second, suppose we want to emulate a randomized trial in which individuals will be followed via their contacts with the health care system. This target trial would only include individuals who can be expected to remain actively engaged with

their health care providers during the follow-up period. Although expected engagement can be explicitly defined and assessed during the recruitment process of a true trial (e.g., by asking the question, “Do you plan to move or change jobs within the next 2 years?”), this eligibility criterion is often elusive in observational analyses. A common strategy to emulate this criterion is to restrict the analysis to individuals who have been in regular contact with the health care system before baseline (e.g., those who attended regular check-ups or filled any prescriptions within the 2 previous years) in the hope that they will remain in contact thereafter. Note that we cannot simply exclude individuals whose claims are no longer found in the database some time after baseline (perhaps because they changed insurers). Rather, we must regard such individuals as lost to follow up (i.e., censored).

Treatment strategies

The target trial emulated using observational data will typically be a pragmatic trial, that is, one in which treatment strategies are compared under the usual conditions in which they will be applied (9, 10). For instance, we cannot emulate a placebo-controlled trial with tight monitoring and enforcement of adherence to the study protocol.

To emulate our target trial, we identify individuals in the database who meet all of the eligibility criteria. We then assign them to the trial strategy or strategies that are consistent with their baseline data. In our example, eligible women who did not start hormone therapy will be coded as having initiated the first strategy, and eligible women who did start estrogen plus progestin therapy will be coded as having initiated the second strategy. Note that otherwise eligible individuals who did not start any of the strategies of interest are considered ineligible for the target trial emulation and excluded from the observational analysis (in the presence of effect modification, this exclusion means we are choosing not to estimate the effect in the entire population of eligible women; a somewhat analogous situation arises in truly randomized trials restricted to women who wish to participate). In our example,

women who started estrogen only therapy will not participate in the emulation even if they meet all of the eligibility criteria.

Comparisons of initiators of the various treatment strategies under investigation (sometimes referred to as new-user designs) are 1 simple way to avoid biases due to the selection of individuals who meet eligibility criteria that are defined after the initiation of a treatment strategy and therefore are possibly affected by the strategy itself (11). For example, the comparison of current (prevalent) users, who had initiated therapy months or years before baseline, with never users may have contributed to the failure to identify the early effect of estrogen plus progestin therapy on the risk of coronary heart disease in observational studies (12, 13). Because the therapy caused a short-term increase in risk, the group of prevalent users might have been relatively depleted of susceptible women. The ultimate problem is that a comparison of current users with never users does not correspond to any contrast between treatment strategies that could, even in principle, be compared in a randomized trial.

Assignment procedures

We can only emulate target trials without blind assignment, which is the standard design of pragmatic trials, because individuals in the data set and their health care workers are usually aware of the treatments that participants receive. This is not necessarily a limitation if the goal is comparing the effects of realistic treatment strategies in individuals who are aware of their care.

To emulate the random assignment of strategies at baseline, we need to adjust for all confounding factors required to ensure comparability (exchangeability) of the groups defined by initiation of the treatment strategies (14). The adjustment for baseline confounders may be performed via matching (perhaps on the propensity score), stratification or regression, standardization or inverse probability weighting, g-estimation, or doubly robust methods. For a basic description of these methods, see Hernán and Robins (14).

If the observational database does not contain sufficient information on baseline confounders or if we fail to identify them, successful emulation of the target trial's random assignment is not possible. Confounding bias may be especially serious when emulating target trials that, like ours, compare an active treatment with no treatment (or usual care) rather than with another active treatment (15, 16).

Although it is generally impossible to determine whether the emulation failed because of uncontrolled confounding, indirect approaches may alert about possible unmeasured confounding. One such approach is emulating a target trial with "reversed" strategies, for example, a trial in which hormone therapy users are assigned to the strategies of "continue using therapy" or "stop using therapy" (13). Incompatible or surprising effect estimates (e.g., a decreased risk both when initiating therapy in our original target trial and when discontinuing therapy in the reversed target trial) suggest that at least 1 of the 2 emulations failed to ensure a fair comparison. As an aside, in both target trials individuals with a common treatment history at baseline (nonusers in the original trial and users in the reversed trial) are compared, which is the basic idea of g-estimation (17).

A second approach is to consider outcome controls (18) for which no causal effect is expected, for example, brain cancer

or pneumonia in our hormone therapy example. If the confounders for the study and control outcomes are sufficiently similar, then the use of outcome controls can help detect confounding. One can also consider control outcomes for which the magnitude of the effect is nonzero but approximately known. Analogously, one could use treatment controls by considering treatment strategies with indications similar to the ones under study but for which no effect is expected.

Other approaches to ameliorating unmeasured confounding rely on extracting information from sources previously considered impractical for large-scale research. For example, novel technologies for natural language processing and advanced image processing might eliminate the need for manual, labor-intensive review of medical records. Machine-learning tools and other computer science techniques might also help investigators search for combinations of variables that improve confounding adjustment compared with traditional methods (19–21).

Outcome

We would use the database to identify women with a diagnosis of breast cancer during the follow-up. Independent outcome validation is often warranted, because several studies have shown that lack of outcome validation may result in misleading effect estimates (22–25). We often would prefer to emulate a target trial with systematic and blind ascertainment of the outcome to ensure that knowledge of treatment status does not influence a doctor's decision to look for the outcome. In our example, such differential ascertainment may result in an increased incidence of breast cancer diagnosis among hormone users even in the absence of a biological effect.

Nonetheless, because doctors will generally be aware of the treatment received by the individual, we cannot use observational data to emulate a target trial with systematic and blind outcome ascertainment except when outcome ascertainment cannot be affected by treatment history (e.g., if the outcome is death and is independently ascertained from a death registry). Note, however, that if we were interested in comparing the effects of different hormone treatment strategies on the rate of breast surgery and thus on the need for breast surgeons, no difficulty would arise because the target trial to be emulated would have unblinded ascertainment.

Causal contrast(s) of interest

Several causal effects can be of interest in true randomized trials (26). Two common ones are the intention-to-treat effect (i.e., the comparative effect of being assigned to the treatment strategies at baseline, regardless of whether the individuals continue following the strategies after baseline) and the per-protocol effect (i.e., the comparative effect of following the treatment strategies specified in the study protocol). Often, both effects are of interest (27). If the intention-to-treat and per-protocol effects are of interest in the target trial, we would try to estimate analogs of both effects from our observational data.

Analysis plan

To estimate the intention-to-treat effect in an actual randomized trial, we would conduct an intention-to-treat analysis

to compare the outcomes of the groups assigned to each treatment strategy. An intention-to-treat analysis, however, is rarely possible in observational analyses of existing data. In our example, the closest observational analog of the intention-to-treat analysis is a comparison of initiators of the different treatment strategies, assuming adequate adjustment for baseline confounders. A comparison of initiators parallels the intention-to-treat analysis in target trials in which assignment and initiation of the treatment strategies always occur together at baseline, regardless of whether individuals continue on the strategies after baseline. In our example, if we had data on prescription (rather than dispensing) of therapy, a comparison of groups according to whether they did or did not receive a prescription of therapy at baseline would be somewhat more analogous to the intention-to-treat analysis in the target trial.

To estimate the per-protocol effect in both true randomized trials and emulated trials like ours, adjustment for baseline and postbaseline confounding is necessary when the treatment strategies under study are sustained over time. Because postbaseline prognostic factors associated with subsequent adherence to the strategies may be affected by prior adherence, Robins's g-methods are generally required, even in the absence of unmeasured confounding and model misspecification. See Robins and Hernán (6) for a review of these methods.

Furthermore, in the presence of selection bias due to loss to follow-up, adjustment for postbaseline factors might also be needed to validly estimate both intention-to-treat effects and per-protocol effects in both actual trials and observational analyses that emulate a target trial. When the postbaseline adjustment factors are affected by the treatment strategies themselves, g-methods are generally needed.

DEFINING TIME ZERO

Successful emulation of a target trial requires a proper definition of time zero of follow-up in the observational data, also referred to as baseline. Eligibility criteria need to be met at that point but not later; study outcomes begin to be counted after that point but not earlier.

In our target trial, the natural start of follow-up is the time when the treatment strategy is assigned, which often occurs either shortly before or at the same time as initiation of the treatment strategy. Starting after randomization could result in selection bias because all outcome cases between randomization and time zero would be excluded from the analysis.

With observational data, the best way to emulate time zero of the target trial is to define time zero to be the time when an eligible individual initiates a treatment strategy. However, implementation of this criterion is not straightforward because the eligibility criteria can be met at many different times for the same individual. Consider the following scenarios.

Eligibility criteria are met at a single time

In the first scenario, follow-up starts at the time the eligibility criteria are met, which may vary across individuals. Some examples include:

- a study to compare high versus low epoetin doses among patients who have undergone 3 months of hemodialysis (follow-up starts after 3 months of hemodialysis) (28);

- a study to compare immediate initiation of antiretroviral therapy when the CD4 cell count first drops below 500 cells/ μ L versus delayed initiation in patients infected with human immunodeficiency virus (follow-up starts the first time an individual's CD4 cell count drops below 500 cells/ μ L) (29); and
- a study to compare various lifestyle interventions to reduce the 20-year risk of diabetes between June 1982 and June 2002 (follow-up starts in June 1982) (30).

Eligibility criteria are met at multiple times

In a second scenario, eligibility criteria are met at multiple times. An example is below:

- Consider a study to compare initiation versus no initiation of estrogen plus progestin therapy among postmenopausal women with no history of chronic disease and no use of hormone therapy during the previous 2 years. A woman who meets these eligibility criteria continuously between the ages of 51 and 65 years would be eligible for the target trial at multiple times during her lifetime—that is, she has multiple times that can qualify as time zero. When should her follow-up start in the observational study (at age 51, 52, 53 . . .)?

Two unbiased choices of time zero with multiple eligible times are: 1) a single eligible time (e.g., the first eligible time or a random eligible time) and 2) all eligible times or a large subset thereof. The second strategy requires emulating multiple nested trials, each of them with a different start of follow-up (31–33). The number of nested trials depends on the frequency with which data on treatment and covariates are collected.

- If there is a fixed schedule for data collection at prespecified times (e.g., every 2 years, like in the Nurses' Health Study), then we can emulate a new trial starting at each prespecified time (13).
- If there is subject-specific, possibly unscheduled, data collection (e.g., electronic medical records), then we can choose a short fixed time unit (e.g., a month) and emulate a new trial starting at each time unit (34).

From a statistical standpoint, the second strategy can be more efficient than the first. However, because individuals may be included in multiple emulated trials, appropriate adjustment of the usual variance estimator is required.

SPECIFYING A GRACE PERIOD

A pragmatic trial often is designed to allow for the constraints faced by decision makers in practice. For example, once a patient and her clinician decide that the patient should initiate hormone therapy, it may take several weeks to complete the clinical tests (e.g., a bone density scan and a lipid panel) and administrative procedures required before treatment initiation. Therefore, the trial protocol might specify that a women assigned to the strategy "initiate hormone therapy" is allowed a 1-month grace period so that she is considered compliant with the protocol if she initiates therapy within a month

from randomization. If we designed a trial with a strategy requiring instant initiation of hormone therapy at randomization without a grace period, the trial would then include strategies that could not be successfully implemented in clinical practice.

In emulating a target trial that includes grace periods using observational data, we must allow for an analogous grace period measured from time zero. The use of a target trial with a grace period not only ensures that the strategies remain realistic but also increases the number of people in the observational database whose data can be used to emulate the target trial.

A consequence of having a grace period is that, for the duration of the grace period, an individual's observational data is consistent with more than 1 strategy. In our example, the introduction of a 3-month grace period implies that the strategies are redefined as "initiate therapy within 3 months of eligibility" versus "never initiate therapy." Therefore, an individual who starts therapy in month 3 after baseline has data consistent with both strategies during months 1 and 2. Had she died during those 2 months, to which strategy of the target trial would we have assigned her? Whenever an individual's data at baseline are consistent with initiation of 2 or more treatment strategies, 1 possibility is to randomly assign her to 1 of them.

Another possibility is to create 2 exact copies of this individual—clones—in the data and assign each of the 2 clones to a different strategy (35–38). Clones are then censored at the time when their data stop being consistent with the strategy to which they were assigned. For example, if the individual starts therapy in month 3, the clone assigned to "never initiate therapy" would be censored at that time. The potential bias introduced by this likely informative censoring needs to be corrected by appropriate adjustment for time-varying factors (e.g., via inverse probability weighting (39)). Importantly, if the individual had died in month 2, then both clones would have died and therefore the death would have been assigned to both strategies. This double allocation of events prevents the bias that could arise if events that occurred during the grace period were systematically assigned to 1 of the 2 strategies only.

A consequence of cloning and censoring is that an intention-to-treat effect cannot be emulated because each individual may have been assigned to many or even all strategies at baseline. Therefore, a contrast based on baseline assignment (i.e., an intention-to-treat analysis) will compare groups with essentially identical outcomes. Analyses with a grace period at baseline are geared towards estimating a per-protocol effect of a target trial.

DISCUSSION

The target trial approach is consistent with a formal counterfactual theory of causality (4) yet avoids the theory's often unfamiliar mathematical notation and concepts. Further, the approach provides an organizing principle for causal inference methods that implicitly rely on counterfactual reasoning (e.g., new-users design, negative outcome controls), establishes a link between methods for the analysis and the reporting of observational studies and randomized trials (2, 40), naturally leads to analytic approaches that prevent apparent paradoxes and common biases (41), facilitates a systematic

methodologic evaluation of observational studies (42, 43) and the transportability of their estimates (44–46), and may help explain between-studies differences (47) as different observational analyses may be emulating different target trials.

As an illustration of the target trial approach, we chose a relatively simple emulation of static treatment strategies related to hormone therapy. The advantages of the approach, however, become clearer when emulating target trials that compare dynamic strategies that are sustained over time and that adjust treatment to the evolving characteristics of patients (28, 29, 48). Also, for brevity, we focused on follow-up studies, but the target trial approach can be extended to case-control sampling designs (49–51). Investigators will first need to define the follow-up study from which cases and controls were sampled and then the target trial that the follow-up study emulates. Finally, we did not consider target trials with interference (52–54) or crossover, and we ignored effects that may result from the scaling up of the treatment strategies outside of the studied populations.

Effect estimates from observational data are well defined when one is able to map the observational analysis into a particular target trial. However, we will rarely be able to emulate the ideal trial in which we are most interested. As discussed above, a number of compromises will have to be made regarding eligibility criteria, strategies to be compared, etc. The specification of the protocol of the target trial will typically be an iterative process during which we will learn which particular target trials may be reasonably supported by the observational data (55). Of all those possible target trials, we will choose the one that is closest to the ideal trial that we would have liked to conduct to answer our question. We will then be able to outline a protocol, present a flow chart, summarize how the observational data set is used to emulate the target trial, and explain how the target trial differs from the ideal trial (28, 48).

An explicit target trial approach is also advantageous to improve the quality of big data. When investigators can influence how data are being actually recorded, a target trial approach helps them identify critical data items for comparative effectiveness research and articulate a compelling rationale to modify data structuring or recording practices. When investigators from different institutions use a Common Data Model (56), an explicit target trial approach may assist them in the development and evolution of the structure and contents of their data model.

The term "big data" has been a branding success compared with the previously used term "large observational databases." Other things being equal, big data is better than small data. Indeed the sheer size and increasing availability of big data facilitates the emulation of target trials. Yet, it is also important to understand the limitations of large observational databases. Keenly aware of these limitations, epidemiologists may become reasonably concerned when big data is discussed in the lay press as an alternative to randomized trials (1). The certification of big data as research grade will generally require harmonization and standardization procedures that accommodate time-varying clinical workflows, idiosyncratic coding practices, and changes of software versions. These procedures require intimate knowledge of the data set and may need to be followed by costly validation studies and indirect validation

approaches, such as comprehensive internal consistency checks and comparisons across data sets.

Because many decisions need to be made in the absence of randomized trials, it is important to adopt a sound approach to the design and analysis of observational studies. Making the target trial explicit is one step in that direction. Big data—and the increasingly sophisticated tools used for analysis—may not always suffice to appropriately emulate our ideal trial. Even so, the target trial approach allows us to systematically articulate the tradeoffs that we are willing to accept. This explicit approach, in combination with subject-matter expertise, epidemiologic and methodologic proficiency, and innovative computer science tools, seems our best bet to maximize the societal benefits of big data for causal inference.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Miguel A. Hernán, James M. Robins); Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Miguel A. Hernán, James M. Robins); and Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts (Miguel A. Hernán).

This research was partly funded by National Institutes of Health grants R01 AI102634 and P01 CA134294.

Conflict of interest: none declared.

REFERENCES

- Greenwood V. Can big data tell us what clinical trials don't? *The New York Times*. October 3, 2014. <http://www.nytimes.com/2014/10/05/magazine/can-big-data-tell-us-what-clinical-trials-dont.html>. Accessed September 4, 2015.
- Hernán MA. With great data comes great responsibility: publishing comparative effectiveness research in epidemiology. *Epidemiology*. 2011;22(3):290–291.
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688–701.
- Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect [published erratum appears in *Math Model*. 1987;14:917–921]. *Math Model*. 1986; 7(9-12):1393–1512.
- Robins JM. Addendum to “a new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect” [published erratum appears in *Comput Math Appl*. 1989;18:477]. *Comput Math Appl*. 1987;14(9-12):923–945.
- Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, et al, eds. *Longitudinal Data Analysis*. New York, NY: Chapman and Hall/CRC Press; 2008:553–599.
- Richardson WS, Wilson MC, Nishikawa J, et al. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12–A13.
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5): 615–625.
- Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2009;62(5):464–475.
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis*. 1967;20(8):637–648.
- Ray WA. Evaluating medication effects outside of clinical trials: New-user designs. *Am J Epidemiol*. 2003;158(9): 915–920.
- Hernán MA, Robins JM. Observational studies analyzed like randomized experiments: best of both worlds. *Epidemiology*. 2008;19(6):789–792.
- Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–779.
- Hernán MA, Robins JM. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC; 2016.
- Glynn RJ, Knight EL, Levin R, et al. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology*. 2001;12(6):682–689.
- Schneeweiss S, Patrick AR, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care*. 2007;45(10 suppl 2):S131–S142.
- Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. *Health Services Research Methodology: A Focus on AIDS*. Washington, DC: US Public Health Service National Center for Health Services Research; 1989:113–159.
- Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383–388.
- Gruber S, Logan RW, Jarrín I, et al. Ensemble learning of inverse probability weights for marginal structural modeling in large observational datasets. *Stat Med*. 2015;34(1):106–117.
- Schneeweiss S, Rassen JA, Glynn RJ, et al. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4): 512–522.
- Polley E, Rose S, Van der Laan M. Super learning. In: Van der Laan MJ, Rose S, eds. *Targeted Learning*. New York, NY: Springer; 2012:43–66.
- Saunders KW, Dunn KM, Merrill JO, et al. Relationship of opioid use and dosage levels to fractures in older chronic pain patients. *J Gen Intern Med*. 2010;25(4):310–315.
- García Rodríguez LA, Ruigómez A. Case validation in research using large databases. *Br J Gen Pract*. 2010;60(572): 160–161.
- Ives DG, Fitzpatrick AL, Bild DE, et al. Surveillance and ascertainment of cardiovascular events. The Cardiovascular Health Study. *Ann Epidemiol*. 1995;5(4):278–285.
- Hernán MA, Jick SS, Olek MJ, et al. Recombinant hepatitis B vaccine and the risk of multiple sclerosis: a prospective study. *Neurology*. 2004;63(5):838–842.
- Hernán MA, Robins JM. Observational studies analyzed like randomized trials, and vice versa. In: Gatsonis C, Morton S, eds. *Methods in Comparative Effectiveness Research*. Boca Raton, FL: Chapman & Hall/CRC. In press.
- Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1): 48–55.
- Zhang Y, Thamer M, Kaufman J, et al. Comparative effectiveness of two anemia management strategies for complex elderly dialysis patients. *Med Care*. 2014;52(zuppl 3): S132–S139.

29. HIV-CAUSAL Collaboration, Cain LE, Logan R, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Ann Intern Med.* 2011;154(8):509–515.
30. Danaei G, Pan A, Hu FB, et al. Hypothetical midlife interventions in women and risk of type 2 diabetes. *Epidemiology.* 2013;24(1):122–128.
31. van der Laan MJ, Petersen ML, Joffe MM. History-adjusted marginal structural models and statically-optimal dynamic treatment regimens. *Int J Biostat.* 2005;1(1):Article 4.
32. Petersen ML, Deeks SG, Martin JN, et al. History-adjusted marginal structural models for estimating time-varying effect modification. *Am J Epidemiol.* 2007;166(9):985–993.
33. Robins JM, Hernán MA, Rotnitzky A. Effect modification by time-varying covariates. *Am J Epidemiol.* 2007;166(9):994–1002; discussion 1003–1004.
34. Danaei G, Rodríguez LA, Cantero OF, et al. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res.* 2013;22(1):70–96.
35. Cain LE, Robins JM, Lanoy E, et al. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int J Biostat.* 2010;6(2):Article 18.
36. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part II: proofs of results. *Int J Biostat.* 2010;6(2):Article 9.
37. Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content. *Int J Biostat.* 2010;6(2):Article 8.
38. van der Laan MJ, Petersen ML. Causal effect models for realistic individualized treatment and intention to treat rules. *Int J Biostat.* 2007;3(1):Article 3.
39. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In: Jewell N, Dietz K, Farewell V, eds. *AIDS Epidemiology—Methodological Issues.* Boston, MA: Birkhäuser; 1992:297–331.
40. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med.* 2010;7(3):e1000251.
41. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol.* 2008;167(4):492–499.
42. Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs* 2012. Washington, DC: The National Academies Press; 2012.
43. Sterne JAC, Higgins JPT, Reeves BC. The ROBINS-I tool (Risk Of Bias In Non-randomized Studies - of Interventions). Version 7. <https://www.riskofbias.info>. Accessed March 10, 2016.
44. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology.* 2011;22(3):368–377.
45. Bareinboim E, Pearl J. External validity: from do-calculus to transportability across populations. *Stat Sci.* 2014;29(4):579–595.
46. Robins JM, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. *Stat Med.* 2008;27(23):4678–4721.
47. Madigan D, Stang PE, Berlin JA, et al. A systematic statistical approach to evaluating evidence from observational studies. *Annu Rev Stat Appl.* 2014;1:11–39.
48. Garcia-Albeniz X, Chan JM, Paciorek A, et al. Immediate versus deferred initiation of androgen deprivation therapy in prostate cancer patients with PSA-only relapse. An observational follow-up study. *Eur J Cancer.* 2015;51(7):817–824.
49. Rose S, van der Laan MJ. A targeted maximum likelihood estimator for two-stage designs. *Int J Biostatistics.* 2011;7(1):Article 17.
50. King G, Zeng L, Chow S-C. *Inference in Case Control Studies, in Encyclopedia of Biopharmaceutical Statistics.* 3rd ed. New York, NY: Marcel Dekker; 2010.
51. Robins JM. Comment on “choice as an alternative to control in observational studies” by Paul Rosenbaum. *Stat Sci.* 1998;14(3):281–293.
52. Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res.* 2012;21(1):55–75.
53. Hudgens MG, Halloran ME. Toward causal inference with interference. *J Am Stat Assoc.* 2008;103(482):832–842.
54. Sobel ME. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J Am Stat Assoc.* 2006;101(476):1398–1407.
55. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res.* 2012;21(1):31–54.
56. Curtis LH, Weiner MG, Boudreau DM, et al. Design considerations, architecture, and use of the Mini-Sentinel distributed data system. *Pharmacoepidemiol Drug Saf.* 2012;21(suppl 1):23–31.
57. 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.