Causal Methods for Observational Data
Amanda Stevenson, University of Texas at Austin Population Research Center, Austin, TX

ABSTRACT
Comparative effectiveness research often uses non-experimental observational data (like hospital discharge records or nationally representative surveys) to draw causal inference about the effectiveness of interventions for health. These ex post inferences require the careful use of specialized statistical methods in order to account for issues like selection bias and unmeasured heterogeneity. This paper briefly discusses the strengths and weakness of some of the most common causal methods for comparative effectiveness evaluation and provides instructions for using SAS® to implement propensity score matching, double difference, instrumental variables methods and regression discontinuity.

INTRODUCTION
Every statistics student knows the phrase, “Correlation is not causation” and most professional analysts and statisticians scrupulously avoid overstating their findings. However, the purpose of much of health research is to understand the effect of an intervention on an outcome and researchers frequently have only observational non-randomized data with which to address their research questions. Therefore, when observational data are used in health outcomes research, explicitly causal models and statistical methods are necessary.

This paper demonstrates how SAS® users can take advantage of the counterfactual model of inference to strengthen the causal implications of their research with observational data. It begins by describing the counterfactual model of inference and providing guidelines for conceptualizing counterfactuals in health outcomes research. It then briefly describes four methods of estimating treatment effects using the counterfactual: propensity score matching, double difference, instrumental variables, and regression discontinuity. Propensity score matching pairs treatment individuals with control individuals who are similar on pretreatment variables, allowing the pairs to act as counterfactuals for each other. Double difference (or difference in differences) requires panel data and uses baseline values to control for time-invariant unobserved heterogeneity. The instrumental variables method relies on exogenous variability in an appropriate third variable to isolate the variation shared by the treatment and the outcome even in the presence of unobserved treatment selection. The regression discontinuity method estimates the treatment effect by comparing treated and untreated individuals in a neighborhood of the eligibility cutoff when eligibility for treatment is assigned relatively strictly based on some known exogenous variable.

THE COUNTERFACTUAL MODEL OF INFERENCE
The counterfactual model asks the question, “If the treatment had been assigned in the opposite direction (i.e. if the treated were the controls and the controls were the treated) how would each individual’s outcome have changed?” This hypothetical alternate outcome is called the counterfactual. The mean difference between the observed outcome and the counterfactual outcome is then the estimated treatment effect. In an experiment, the participants are randomized into control and treatment groups and the groups act as each other’s counterfactual. This is a valid counterfactual model because every individual had the same chance of membership in the treatment group and therefore the difference in group means can only be due to a treatment effect. In contrast, when observational data are used the counterfactual cannot be measured and must always be estimated.

Robert Frost’s poem “The Road Less Travelled” illustrates the fundamentally unmeasurable nature of the counterfactual in observational data1. While the traveler in the poem attributes “all the difference” to the selection of the road less traveled, he also recognizes the impossibility of going back to the point at which the roads diverged and taking the alternate route. In the poem, Frost arrived at the observed outcome and the counterfactual outcome is where he would have ended up had he taken the other road. Estimating the causal treatment effect requires a comparison of the observed and counterfactual outcomes and thus requires an estimate of the counterfactual.

THE COUNTERFACTUAL MODEL AND HEALTH OUTCOMES RESEARCH
Comparative effectiveness research can benefit greatly from causal models because it seeks to compare disease prevention and treatment strategies. These comparisons require the isolation of the causal effect of each strategy or

---

1 This example is given in Angrist 2009.
Applying the counterfactual model to health outcomes research with observational data requires three steps:

1. Develop a substantive understanding of the ideal counterfactual of interest
2. Select a statistical method for estimating the counterfactual of interest
3. Implement the selected statistical method with the available data

In practice, 1 and 2 often occur simultaneously, as our research objectives must conform with the realities of the available data. However, it is still useful to imagine what the ideal counterfactual would be. This exercise can also guide later justifications of causal claims when the results of the study are presented.

Some questions to ask while developing ideal counterfactual in step 1 are:

- Will my results be used to motivate a change in how treatment is assigned? If so, the counterfactual may be treatment for the untreated.
- Will my results be used to evaluate the cost effectiveness of a current treatment assignment scheme? If so, the counterfactual may be control for the treated.
- Am I interested in estimating the causal effect of the treatment on the treated? If so, the counterfactual may be control for the treated.
- Am I interested in estimating the total treatment effect, or the mean effect of treatment on everybody? If so, the counterfactual may be control for the treated and treatment for the controls.

For example, if the aim is to estimate how much people would benefit from lowering an eligibility threshold for a particular treatment, an ideal counterfactual would be the case in which individuals just below the threshold were in the treatment rather than the control group. In this case, regression discontinuity might give a good estimate of the treatment effect in the neighborhood of the current threshold while propensity score matching might estimate only the effect of the treatment on the treated, not what the effect would have been on the untreated had they received treatment.

Understanding how alternate methods work will enable the researcher to select an appropriate method for their ideal counterfactual. The remainder of this paper provides guidelines for completing step 2 by briefly describing four common causal methods from the counterfactual perspective. Also included are notes on how use SAS® to implement step 3, when actual data analysis is conducted.

**PROPENSITY SCORE MATCHING**

Propensity score matching uses observed factors to model the propensity to be in the treatment group and then estimates the treatment effect as the mean difference in differences for pairs of treatment and control individuals with similar propensities. Propensity score matching is a three step process. First propensities are estimated. Second, treated and untreated individuals are matched. Third, the treatment effect is estimated as the mean of the difference in outcomes within the pairs.

**WHEN TO USE PROPENSITY SCORE MATCHING**

From a data perspective, propensity score matching can be used when both baseline characteristics and outcome measures are available for treated and untreated individuals.

Three conditions are necessary for propensity score matching to yield a valid estimate of causal effect (Morgan 2007; Khandker 2010):

1. Unobserved characteristics must not account for treatment receipt.
2. Common Support. The distributions of propensities for treatment in the control and treatment groups must overlap sufficiently to allow the pairing of treatment and control individuals.
3. Conditional Independence Assumption (CIA). Individuals in the treatment group must not benefit from treatment differently than individuals in the control group would have, conditional on propensity to be treated.

**STRENGTHS**

Propensity score matching allows for causal modeling when only cross sectional data are available, since some time-invariant and frequently collected characteristics like education and race might be drivers of propensity for treatment (Khandker 2010).
As a semiparametric method, propensity score matching imposes few assumptions on the functional form of treatment effect (Khandker 2010).

LIMITATIONS

The validity of propensity score matching relies on the untestable assumption that unobserved characteristics do not influence treatment participation. Careful substantive justification of this assumption is warranted and to the degree that this assumption is not met, the estimates from this method can be invalid (Morgan 2007).

Data are required on a substantial set of untreated individuals for common support to be met. Even when common support is relatively good, dropping treatment cases with very high propensities is often necessary (Khandker 2010). This leads to a loss of information about how high propensity individuals react to the treatment and can bias the estimate of treatment effect. For this reason, propensity score matching might not be the best choice if the treatment effect on the treated is the primary aim of the research question.

IMPLEMENTATION IN SAS®

The first step in estimating treatment effect with propensity score matching is the creation of a propensity score for every individual or incident:

```
proc logistic data=observed;
  model tx = <BASELINE VARIABLES>;
  output out = propensity_scores pred = prob_tx;
run;
```

Once the propensity scores are created, the researcher must match the treatment individuals with the control individuals based on the propensity scores. This matching can be conducted using exact matching, nearest neighbor matching, radius matching, weighted distance matching, or kernel matching. A good discussion of the selection of a matching technique is available in Khandker (2010). Researchers should keep the counterfactual model in mind when selecting a matching method, since some methods may yield more or less appropriate matches from a counterfactual perspective. Examining the results of the match by directly viewing the pairs of matched observations and by investigating the dropped treatment observations can illuminate the implications of the assumptions used for matching.

Once the propensity scores are constructed and a matching method is selected, a SAS® macro must be used to match the treatment and control individuals. Several SUG papers with SAS® macros for matching on propensity scores are listed at the end of this paper.

DOUBLE DIFFERENCE IN PANEL DATA

In order to estimate the effect of a treatment, the double difference method compares change over time for members of the treatment group to change over time for members of the control group. Double difference is a popular method among health researchers because it is easily interpretable as “difference in improvement” and because it is procedurally similar to the estimate of treatment effect in a randomized controlled trial. However, the causal interpretation of a double difference estimate of treatment effect must be made with caution when data are observational, since the relationship between selection into treatment and change over time is not accounted for in this model. From a causal perspective, this comparison uses each individual as their own counterfactual, which is fine as long as there is not something about the treatment group that varies with time differently than the control group does. For example, the members of the treatment group should not be getting worse or better faster. This presents a special problem for health researchers using observational data, since doctors may choose to treat individuals with greater chances for improvement (Stukel 2007).

WHEN TO USE DOUBLE DIFFERENCE

From a data perspective, double difference is an option when baseline and subsequent outcome measures are available for the treatment and control groups.

However, two conditions must be met for double difference to yield a valid estimate of causal effect:

1. Treatment must not be correlated with time varying influences on the outcome.
2. Conditional Independence Assumption (CIA). Individuals in the treatment group must not benefit from treatment differently than individuals in the control group would have.
STRENGTHS
Because the double difference method uses each individual as their own control, no time invariant sources of heterogeneity will bias the estimate of treatment effect. For example, if an intervention is supposed to improve blood pressure, all preexisting and time invariant individual-level causes of high blood pressure are controlled for in a double difference model, even if they are correlated with both the treatment and the outcome.

The double difference method allows for an estimate of the treatment effect on the whole population of the treated, since all individuals are included in the analysis.

The double difference method is procedurally similar to methods used in case controlled trials, which makes the results accessible to a wide audience.

LIMITATIONS
The double difference model does not estimate the treatment effect on the untreated directly. If the aim of research is to estimate the global or total treatment effect or to estimate the treatment effect on the untreated, careful justification of the comparability of the treatment and control groups is necessary.

The double difference method does not control for time varying sources of treatment-correlated heterogeneity.

IMPLEMENTATION IN SAS®
The implementation of the double difference method is very simple. A data step and a descriptive procedure are all that is required:

```sas
data outcome;
  set observed;
  individual_diff = pre-post;
run;

proc means data = outcome n mean max min range std;
  var individual_diff;
  title 'Difference In Difference Estimate of Tx Effect';
run;
```

INSTRUMENTAL VARIABLES
An instrumental variable (IV) approach can be used to estimate the causal effect of a treatment when an experiment with randomized controls is unfeasible or impossible and when matching approaches are unsuccessful. The IV approach isolates the covariance of the treatment and outcome variables through the use of an appropriate exogenous variable, an instrument. While IV can be relatively easy to implement and gives unbiased estimates of treatment effect, it can only be used if a good instrument for treatment is present in the data set. A variable is a good instrument if it is correlated with treatment but uncorrelated with any unmeasured predictors of the outcome. For example, income may be predictive of access to some medical treatments but might not be correlated with unmeasured predictors of a given health outcome. The degree to which an instrument is correlated with treatment makes it “strong” or “weak.” Strong instruments can be used to estimate treatment effects directly. Sometimes weak instruments can be used to create upper and lower bounds on treatment effect. While both kinds of instruments are commonly used in econometrics and can be used in comparative health outcomes research, caution must be used with weak instruments, which can add additional bias in many circumstances (Morgan 2007).

The instrumental variable method is a two stage process. First, an appropriate regression method is used to predict treatment using all covariates and the instrument. Second, the predicted value of treatment is used in a second stage regression to predict the outcome.

WHEN TO USE INSTRUMENTAL VARIABLES
From a data perspective it is only possible to use an instrumental variables approach when a good instrument is available. Good instruments must be substantively and empirically demonstrated to be correlated with treatment but not with unmeasured predictors of the outcome.

In addition to a good instrument, the following assumptions must be substantiated for an IV estimate of treatment effect to be a valid causal estimate (Morgan 2007):

1. Exclusion restriction: The instrument must not be correlated with any predictors of the outcome other than treatment assignment.

3. Monotonicity assumption: The sample cannot include both individuals for whom the effect of the instrument on treatment is positive and individuals for whom the effect of the instrument on treatment is negative.

**STRENGTHS**

When the above assumptions are met, the IV method yields local area treatment effect estimates which are easily interpretable as responses to reduced or increased barriers to participation when the instrument is interpretable as a change in the accessibility of treatment (Imbens 1994, Morgan 2007).

Instrumental variables estimates of treatment effect are robust to time varying selection bias. This means that even if individuals in the treatment group differ from individuals in the control group in a way that changes over time, the IV estimate of treatment effect will still be valid.

Using treatment assignment as an instrument can correct for measurement error in observed participation.

**LIMITATIONS**

Good instruments can be hard to find and justifying instruments’ lack of correlation with unmeasured predictors of the outcome can require substantial substantive knowledge.

If the instrument cannot be interpreted as a changing barrier to participation, the causal interpretation of the IV estimate of treatment effect is ambiguous (Morgan 2007).

Different instruments will yield different estimates of treatment effect, which can be unsettling to some researchers and readers.

**IMPLEMENTATION IN SAS®**

Instrumental variables methods can be implemented several ways in SAS® and may be used as part of more complex models, including time series methods and difference in difference methods. For example, the SYSLIN procedure with the 2SLS option can estimate simultaneous systems of linear equations in two stages as IV requires. For accessibility and illustration, here is a simple two-stage IV estimation process using the LOGISTIC procedure to model probability of treatment based on the instrument and covariates and the REG procedure to model the outcome based on the prediction from the first stage and covariates:

```sas
proc logistic data = observed;
model tx = IV <COVARIATES>;
output out = IVpreds
predicted = IV_sub;
run;

proc reg data=IVpreds;
model outcome=IV_sub <COVARIATES>;
run;
```

**REGRESSION DISCONTINUITY**

When programmatic rules are applied to determine eligibility for a given treatment, those rules can be used to compare treated and untreated individuals within the neighborhood of the eligibility threshold. Regression discontinuity methods conduct this comparison by fitting separate regressions with equal slopes but unequal intercepts for the control and treatment groups. The difference in the intercepts is then interpreted as the treatment effect in the neighborhood of the eligibility threshold (Khandker 2010). In regression discontinuity, the arbitrariness of the eligibility threshold justifies using treatment and control group members in the neighborhood of the threshold as counterfactuals for each other.

**WHEN TO USE REGRESSION DISCONTINUITY**

From a data perspective, regression discontinuity can be used when eligibility criteria are used for treatment assignment.

In order for regression discontinuity to yield a valid causal estimate of treatment effect the following four conditions must be met:
1. The discontinuity at the eligibility threshold must only be in the outcome measure, not also in some other categorical or continuous covariate. This is usually verified through visual inspections of plots of the data (Khandker 2010).

2. Eligibility rules must have been strictly applied, since otherwise unmeasured heterogeneity may account for treatment receipt and outcome (Khandker 2010).

3. Conditional Independence Assumption (CIA). Individuals in the treatment group must not benefit from treatment differently than individuals in the control group would have.

4. The treatment eligibility threshold must not have substantive meaning. In other words, those who are eligible for the treatment must not differ substantially and meaningfully from those who are ineligible for the treatment. For example, if the cutoff has great clinical or biomedical meaning the regression discontinuity estimate of treatment effect is not valid.

**STRENGTHS**

Regression discontinuity can be conducted when the data do not support other causal methods of treatment effect estimation because it only requires eligibility measures and outcome measures for the control and treatment groups.

Because it only estimates the treatment effect at the actual threshold for treatment eligibility, regression discontinuity imposes no assumptions on the distribution of the treatment effect.

**LIMITATIONS**

If eligibility rules were inconsistently applied, treatment is likely to be correlated with unmeasured sources of heterogeneity and the validity of the treatment effect estimate is undermined.

The regression discontinuity estimate only describes the treatment effect in the neighborhood of the cutoff. Therefore, if the researcher is seeking an estimate of the treatment effect for a large group, this method is insufficient.

Extensive robustness checks are necessary to validate the model.

**IMPLEMENTATION IN SAS®**

Implementing regression discontinuity in Base SAS® requires a macro, which is beyond the scope of this paper. However, before writing a macro to implement the RD method, the researcher should conduct some exploratory analyses of the data. Begin with a plot of the data using the SGPLOT procedure with the reg statement.

```sas
proc sgplot data=observed;
  reg x=criterion y=outcome / group=tx;
run;
```

A logical next step is to fit two separate regressions using the REG procedure, one for the treated group and the other for the control group.

```sas
proc reg data=observed (where=(tx=1));
  reg x=criterion y=outcome;
run;
proc reg data=observed (where=(tx=0));
  reg x=criterion y=outcome;
run;
```

The estimated regression equations should then be compared in order to determine if regression discontinuity makes sense. If the coefficients on the treatment selection criterion are not similar between the two regression equations, regression discontinuity may still be performed, but implementation and interpretation may be more complex (Shahidur 2010).

**CONCLUSION**

The counterfactual model of causal inference can guide health outcomes researchers in two important ways:

1. Selection of an appropriate statistical method. Using the counterfactual model of inference forces the researcher to formulate an ideal counterfactual. This exercise leads to precise knowledge about what needs to be estimated statistically in order to estimate a causal effect of treatment. The goal is always to estimate “what might have been” and to compare it to what was observed.
2. Justification of causal claims. The counterfactual model can guide researchers’ efforts to justify and support causal claims in the presentation of research.

REFERENCES


ACKNOWLEDGMENTS

The author is supported by a pre-doctoral training fellowship from the National Institute of Child Health and Human Development and by NICHD grant 5R24HD042849-10.

RECOMMENDED READING

- For optimized matching macros for use with propensity score matching methods, please see Marcelo Coca-Perraillon’s 2007 SAS Global Forum paper “Local and Global Optimal Propensity Score Matching”: [http://www.biostat.umn.edu/~will/6470stuff/Class25-10/SAS%20matching%202007.pdf](http://www.biostat.umn.edu/~will/6470stuff/Class25-10/SAS%20matching%202007.pdf)

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Amanda J. Stevenson  
Population Research Center  
The University of Texas at Austin  
1 University Station, G1800  
Austin, TX 78712  
E-mail: stevenam@prc.utexas.edu  
Web: [http://www.utexas.edu/cola/centers/prc/](http://www.utexas.edu/cola/centers/prc/)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.
Other brand and product names are trademarks of their respective companies.