The Negative Control Approach to Detect and Control for Unobserved Confounding

Eric J Tchetgen Tchetgen
Professor of Biostatistics and Epidemiologic Methods,
Harvard University

National Academies of Sciences, Engineering and Medicine, 3/6/2017

Negative controls in experimental studies

- Non-causal associations in observational studies may be classified into three categories (in addition to chance)1: *mismeasurement* (eg, recall bias), *confounding*, and *biased selection* of individuals into the analysis.

Negative controls in experimental studies

- Non-causal associations in observational studies may be classified into three categories (in addition to chance)1: *mismeasurement* (eg, recall bias), *confounding*, and *biased selection* of individuals into the analysis.
- In experimental biology, the manipulation of experimental conditions prevents many of the noncausal associations that arise in observational studies.

Negative controls in experimental studies

- Non-causal associations in observational studies may be classified into three categories (in addition to chance)1: *mismeasurement* (eg, recall bias), *confounding*, and *biased selection* of individuals into the analysis.
- In experimental biology, the manipulation of experimental conditions prevents many of the noncausal associations that arise in observational studies.
- Nonetheless, experimental biologists routinely question whether they have correctly inferred causal relationships from the results of their experiments.

Negative controls in experimental studies

- Non-causal associations in observational studies may be classified into three categories (in addition to chance)1: *mismeasurement* (eg, recall bias), *confounding*, and *biased selection* of individuals into the analysis.
- In experimental biology, the manipulation of experimental conditions prevents many of the noncausal associations that arise in observational studies.
- Nonetheless, experimental biologists routinely question whether they have correctly inferred causal relationships from the results of their experiments.
- Biologists employ "negative controls" as a means of ruling out possible noncausal interpretations of their results.

Negative controls in experimental studies

- Consider the hypothesis that a particular cytokine enhances the killing of a species of bacteria by neutrophils, a class of white blood cells.

Negative controls in experimental studies

- Consider the hypothesis that a particular cytokine enhances the killing of a species of bacteria by neutrophils, a class of white blood cells.
- An experiment is devised in which neutrophils, bacteria, and growth medium are mixed together.

Negative controls in experimental studies

- Consider the hypothesis that a particular cytokine enhances the killing of a species of bacteria by neutrophils, a class of white blood cells.
- An experiment is devised in which neutrophils, bacteria, and growth medium are mixed together.
- In condition 1, the cytokine is added, and in condition 2, some inert substance such as saline solution is added. After incubation, the bacteria are enumerated and the number of live bacteria compared between conditions 1 and 2.

Negative controls in experimental studies

- Consider the hypothesis that a particular cytokine enhances the killing of a species of bacteria by neutrophils, a class of white blood cells.
- An experiment is devised in which neutrophils, bacteria, and growth medium are mixed together.
- In condition 1, the cytokine is added, and in condition 2, some inert substance such as saline solution is added. After incubation, the bacteria are enumerated and the number of live bacteria compared between conditions 1 and 2.
- fewer live bacteria in condition 1 than in 2 is consistent with the hypothesis that the cytokine enhanced neutrophil-mediated killing.

Negative controls in experimental studies

- Nonetheless, concern remains that something other than cytokine-aided, neutrophil-mediated killing may be responsible.

Negative controls in experimental studies

- Nonetheless, concern remains that something other than cytokine-aided, neutrophil-mediated killing may be responsible.
  - e.g.a contaminant in the cytokine preparation that directly kills bacteria,

Negative controls in experimental studies

- Nonetheless, concern remains that something other than cytokine-aided, neutrophil-mediated killing may be responsible.
  - e.g.a contaminant in the cytokine preparation that directly kills bacteria,
  - or perhaps the cytokine itself kills bacteria,

Negative controls in experimental studies

- Nonetheless, concern remains that something other than cytokine-aided, neutrophil-mediated killing may be responsible.
  - e.g.a contaminant in the cytokine preparation that directly kills bacteria,
  - or perhaps the cytokine itself kills bacteria,
  - or perhaps some other unintended difference between the treated and untreated conditions (e.g., temperature or pH) caused the differential survival of the bacteria.

Negative controls in experimental studies

- Nonetheless, concern remains that something other than cytokine-aided, neutrophil-mediated killing may be responsible.
  - e.g.a contaminant in the cytokine preparation that directly kills bacteria,
  - or perhaps the cytokine itself kills bacteria,
  - or perhaps some other unintended difference between the treated and untreated conditions (e.g., temperature or pH) caused the differential survival of the bacteria.
- Each of these unintended differences is broadly similar to a *confounder* – a characteristic associated with the exposure (presence or absence of the cytokine) and causes the outcome (differences in bacterial counts).

Negative controls in experimental studies

- Experimental biologists address such concerns in two ways.

Negative controls in experimental studies

- Experimental biologists address such concerns in two ways.
  - (i)attempt to eliminate unwanted differences between the compared groups (in the design) and

Negative controls in experimental studies

- Experimental biologists address such concerns in two ways.
  - (i)attempt to eliminate unwanted differences between the compared groups (in the design) and
  - (ii)to measure and account for any unavoidable differences (in the analysis).

Negative controls in experimental studies

- Experimental biologists address such concerns in two ways.
  - (i)attempt to eliminate unwanted differences between the compared groups (in the design) and
  - (ii)to measure and account for any unavoidable differences (in the analysis).
- The second general approach is to perform *negative controls (exposures or outcomes)*: to repeat the experiment under conditions in which it is expected to produce a null result and verify that it does indeed produce a null result.

Negative controls in experimental studies

- Several strategies are employed to design negative controls, such as:

Negative controls in experimental studies

- Several strategies are employed to design negative controls, such as:
  - Leave out an essential ingredient (neutrophils), this negative control should produce fewer bacteria with the cytokine than without.

Negative controls in experimental studies

- Several strategies are employed to design negative controls, such as:
  - Leave out an essential ingredient (neutrophils), this negative control should produce fewer bacteria with the cytokine than without.
  - Inactivate the (hypothesized) active ingredient: add specific antibodies that neutralize the cytokine but have no effect on a contaminant.

Negative controls in experimental studies

- Several strategies are employed to design negative controls, such as:
  - Leave out an essential ingredient (neutrophils), this negative control should produce fewer bacteria with the cytokine than without.
  - Inactivate the (hypothesized) active ingredient: add specific antibodies that neutralize the cytokine but have no effect on a contaminant.
  - Check for an effect that would be impossible by the hypothesized mechanism:species of bacteria completely impervious to the actions of neutrophils.

Negative controls in experimental studies

- Several strategies are employed to design negative controls, such as:
  - Leave out an essential ingredient (neutrophils), this negative control should produce fewer bacteria with the cytokine than without.
  - Inactivate the (hypothesized) active ingredient: add specific antibodies that neutralize the cytokine but have no effect on a contaminant.
  - Check for an effect that would be impossible by the hypothesized mechanism:species of bacteria completely impervious to the actions of neutrophils.
- the list of possible non-causal explanations for an experimental result is almost endless and so is the list of possible negative controls, judgment is required to assess how many such noncausal explanations are plausible and which negative controls are of greatest value in ruling out key threats to valid inference.

Negative controls in observational studies

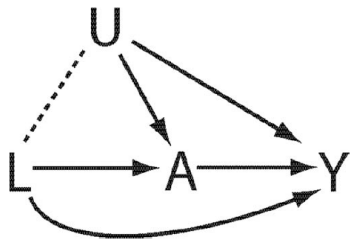


FIGURE 1. Causal diagram for the effect of an exposure of interest (A) on an outcome of interest (Y), with confounders L (assumed measured) and U (assumed uncontrolled) that cause both A and Y. The dashed line between L and U indicates that either may cause the other, and they may share common causes.

- Epidemiologists also sometimes use negative controls to detect confounding and other sources of incorrect causal inference.

Negative controls in observational studies

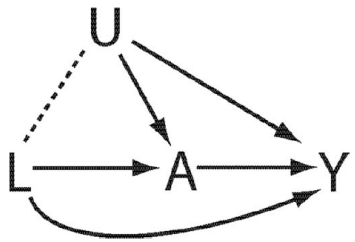


FIGURE 1. Causal diagram for the effect of an exposure of interest (A) on an outcome of interest (Y), with confounders L (assumed measured) and U (assumed uncontrolled) that cause both A and Y. The dashed line between L and U indicates that either may cause the other, and they may share common causes.

- Epidemiologists also sometimes use negative controls to detect confounding and other sources of incorrect causal inference.
- Example 1:Observational analysis of the effect of vaccination in the elderly suggest a remarkably large reduction in one's risk of pneumonia/influenza hospitalization and also in one's risk of all cause mortality in the following season.

Negative controls in observational studies

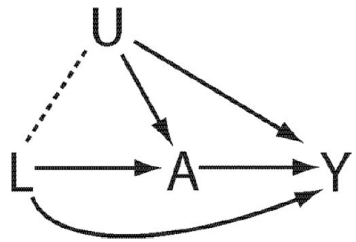


FIGURE 1. Causal diagram for the effect of an exposure of interest (A) on an outcome of interest (Y), with confounders L (assumed measured) and U (assumed uncontrolled) that cause both A and Y. The dashed line between L and U indicates that either may cause the other, and they may share common causes.

- Epidemiologists also sometimes use negative controls to detect confounding and other sources of incorrect causal inference.
- Example 1:Observational analysis of the effect of vaccination in the elderly suggest a remarkably large reduction in one's risk of pneumonia/influenza hospitalization and also in one's risk of all cause mortality in the following season.
- This large observed effect, combined with the lack of measurable vaccine effect in ecological studies, have led to a suspicion that uncontrolled confounding has exaggerated the impact of influenza vaccination on mortality and on pneumonia/influenza hospitalization

Negative controls in observational studies

- Jackson et al (2006) implemented negative control analyses.

Negative controls in observational studies

- Jackson et al (2006) implemented negative control analyses.
- For the first negative control outcome, the authors used the fact that vaccination often begins in autumn, while influenza transmission is often minimal until winter.

Negative controls in observational studies

- Jackson et al (2006) implemented negative control analyses.
- For the first negative control outcome, the authors used the fact that vaccination often begins in autumn, while influenza transmission is often minimal until winter.
- Thus they could assess the risk of pneumonia/influenza hospitalization and all-cause mortality among vaccinated vs. unvaccinated persons both before, during and after influenza season.

Negative controls in observational studies

- Jackson et al (2006) implemented negative control analyses.
- For the first negative control outcome, the authors used the fact that vaccination often begins in autumn, while influenza transmission is often minimal until winter.
- Thus they could assess the risk of pneumonia/influenza hospitalization and all-cause mortality among vaccinated vs. unvaccinated persons both before, during and after influenza season.
- The only biologically plausible mechanism by which influenza vaccine could protect against mortality or pneumonia/influenza hospitalization is by preventing influenza or its consequences; therefore, Jackson and colleagues reasoned that if the effect measured in prior studies were causal, it should be most prominent during influenza season.

Negative controls in observational studies

- If instead it were due to confounding, then the protective effect should be observable immediately after vaccination but before influenza season.

Negative controls in observational studies

- If instead it were due to confounding, then the protective effect should be observable immediately after vaccination but before influenza season.
- In a cohort study analyzed with a Cox proportional hazards model, despite efforts to control for confounding, they observed that the protective effect was actually greatest before, intermediate during, and least after influenza season.

Negative controls in observational studies

- If instead it were due to confounding, then the protective effect should be observable immediately after vaccination but before influenza season.
- In a cohort study analyzed with a Cox proportional hazards model, despite efforts to control for confounding, they observed that the protective effect was actually greatest before, intermediate during, and least after influenza season.
- They concluded that this is evidence that confounding, rather than protection against influenza, accounts for a substantial part of the observed "protection."

Negative controls in observational studies

- If instead it were due to confounding, then the protective effect should be observable immediately after vaccination but before influenza season.
- In a cohort study analyzed with a Cox proportional hazards model, despite efforts to control for confounding, they observed that the protective effect was actually greatest before, intermediate during, and least after influenza season.
- They concluded that this is evidence that confounding, rather than protection against influenza, accounts for a substantial part of the observed "protection."
- The use of this negative control outcome approach is formally similar to the "leave-out-an-essential-ingredient" control described above, as influenza is essential in the proposed causal pathway

Negative controls in observational studies

- Second, Jackson et al. postulated that the protective effects of influenza vaccination, if real, should be limited to outcomes plausibly linked to influenza.

Negative controls in observational studies

- Second, Jackson et al. postulated that the protective effects of influenza vaccination, if real, should be limited to outcomes plausibly linked to influenza.
- They found that influenza vaccination was also "protective" against injury or trauma hospitalization.

Negative controls in observational studies

- Second, Jackson et al. postulated that the protective effects of influenza vaccination, if real, should be limited to outcomes plausibly linked to influenza.
- They found that influenza vaccination was also "protective" against injury or trauma hospitalization.
- This, too, was interpreted as evidence that some of the protection observed for pneumonia/influenza hospitalization or mortality was due to inadequately controlled confounding.

Negative controls in observational studies

- Second, Jackson et al. postulated that the protective effects of influenza vaccination, if real, should be limited to outcomes plausibly linked to influenza.
- They found that influenza vaccination was also "protective" against injury or trauma hospitalization.
- This, too, was interpreted as evidence that some of the protection observed for pneumonia/influenza hospitalization or mortality was due to inadequately controlled confounding.
- This second negative control outcome is formally similar to the "check-for-an-effect-impossible-by-the-hypothesized mechanism" approach described in the experimental setting.

Negative controls in observational studies

- Epidemiologists also sometimes use *negative control exposures* to examine whether observed associations are causal.

Negative controls in observational studies

- Epidemiologists also sometimes use *negative control exposures* to examine whether observed associations are causal.
- Example: One such explicit application of this approach relates to the use of paternal exposures as a negative control exposure for maternal exposures considered to have an intrauterine influence on offspring outcomes (Davey Smith, 2012).

Negative controls in observational studies

- Epidemiologists also sometimes use *negative control exposures* to examine whether observed associations are causal.
- Example: One such explicit application of this approach relates to the use of paternal exposures as a negative control exposure for maternal exposures considered to have an intrauterine influence on offspring outcomes (Davey Smith, 2012).
- Paternal smoking, as a negative control, may show associations with offspring outcomes similar to those of maternal smoking if the associations are generated by shared familial confounding factors or by parental genotypes transmitted to the offspring.

Negative controls in observational studies

- Epidemiologists also sometimes use *negative control exposures* to examine whether observed associations are causal.
- Example: One such explicit application of this approach relates to the use of paternal exposures as a negative control exposure for maternal exposures considered to have an intrauterine influence on offspring outcomes (Davey Smith, 2012).
- Paternal smoking, as a negative control, may show associations with offspring outcomes similar to those of maternal smoking if the associations are generated by shared familial confounding factors or by parental genotypes transmitted to the offspring.
- If, however, there is an intrauterine influence, then only the maternal exposure would be expected to show an independent association with the outcome.

Negative controls in observational studies

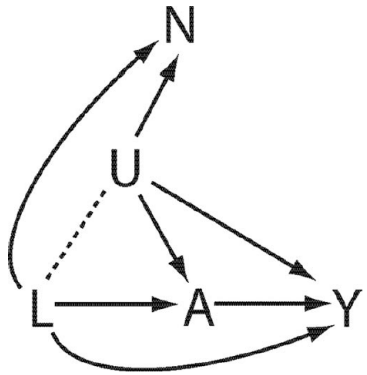


FIGURE 2. Causal diagram showing an ideal negative control outcome N for use in evaluating studies of the causal relationship between exposure A and outcome Y. N should ideally have the same incoming arrows as Y, except that A does not cause N; to the extent this criterion is met, N is called U-comparable to Y.

- Ideally, a negative control outcome (N) should be an outcome such that the set of common causes of exposure A and outcome Y should be as identical as possible to the set of common causes of A and N .

Negative controls in observational studies

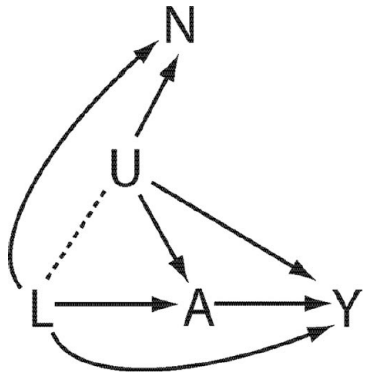


FIGURE 2. Causal diagram showing an ideal negative control outcome N for use in evaluating studies of the causal relationship between exposure A and outcome Y. N should ideally have the same incoming arrows as Y, except that A does not cause N; to the extent this criterion is met, N is called U-comparable to Y.

- Ideally, a negative control outcome (N) should be an outcome such that the set of common causes of exposure A and outcome Y should be as identical as possible to the set of common causes of A and N .
- To the extent that the set of unobserved common causes of A and Y overlaps with the set of unobserved common causes (U) of A and N, we call the negative control outcome N "U-comparable" to Y.

Negative controls in observational studies

- If N and Y are U-comparable outcomes (i.e. with an identical set of common causes that are associated with A), and assuming that N is not caused by A, an association A-N when analyzed according to the same procedure used to analyze A-Y would indicate bias in the association A-Y.

Negative controls in observational studies

- If N and Y are U-comparable outcomes (i.e. with an identical set of common causes that are associated with A), and assuming that N is not caused by A, an association A-N when analyzed according to the same procedure used to analyze A-Y would indicate bias in the association A-Y.
- If N and Y are perfectly U-comparable and N is not caused by A, then a null finding of A-N implies that the A-Y association is not likely biased by the pathways examined through this negative control.

Negative controls in observational studies

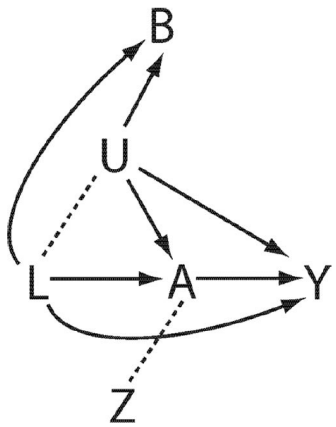


FIGURE 3. Causal diagram showing an ideal negative control exposure B for use in evaluating studies of the causal relationship between exposure A and outcome Y. B should ideally have the same incoming arrows as A; to the extent this criterion is met, B is called U-comparable to A. Z is an instrumental variable of the A-Y relationship and is depicted to illustrate the difference between an instrumental variable and a negative control variable.

- A negative control exposure B should be an exposure such that the common causes of A and Y are as nearly identical as possible to the common causes of B and Y.

Negative controls in observational studies

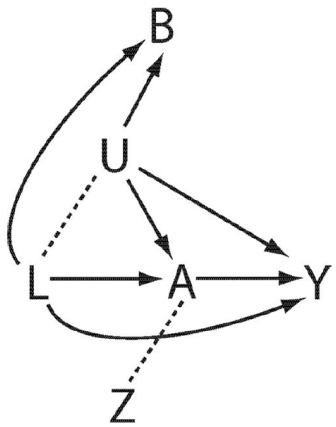


FIGURE 3. Causal diagram showing an ideal negative control exposure B for use in evaluating studies of the causal relationship between exposure A and outcome Y. B should ideally have the same incoming arrows as A; to the extent this criterion is met, B is called U-comparable to A. Z is an instrumental variable of the A-Y relationship and is depicted to illustrate the difference between an instrumental variable and a negative control variable.

- A negative control exposure B should be an exposure such that the common causes of A and Y are as nearly identical as possible to the common causes of B and Y.
- To the extent that the set of unobserved common causes U of A and Y overlaps with the set of unobserved common causes of B and Y, we call the negative control exposure B "U-comparable" to A.

Negative controls in observational studies

- If A and B are perfectly U-comparable and B does not cause Y, then an association B-Y when analyzed according to the same model used to analyze A-Y would indicate bias in the association A-Y.

Negative controls in observational studies

- If A and B are perfectly U-comparable and B does not cause Y, then an association B-Y when analyzed according to the same model used to analyze A-Y would indicate bias in the association A-Y.
- If A and B are perfectly U-comparable and B does not cause Y, then a null finding of A-N means that the A-Y association is unbiased.

Other examples of negative controls

- Exposure: Occupational radon exposure in miners; outcome: lung cancer: Negative control outcome: COPD (Richardson et al, 2014)

Other examples of negative controls

- Exposure: Occupational radon exposure in miners; outcome: lung cancer: Negative control outcome: COPD (Richardson et al, 2014)
- Exposure: flu vaccine, outcome: in‡uenza hospitalization; Negative control exposure:boster tetanus vaccine

Other examples of negative controls

- Exposure: Occupational radon exposure in miners; outcome: lung cancer: Negative control outcome: COPD (Richardson et al, 2014)
- Exposure: flu vaccine, outcome: in‡uenza hospitalization; Negative control exposure:boster tetanus vaccine
- Exposure: Universal Child credit offers benefit to parents with children <6yrs in Canada, Outcome: Family food insecurity, Negative controls: non-eligible families (only with children>6yrs old).

Other examples of negative controls

- Exposure: Occupational radon exposure in miners; outcome: lung cancer: Negative control outcome: COPD (Richardson et al, 2014)
- Exposure: flu vaccine, outcome: in‡uenza hospitalization; Negative control exposure:boster tetanus vaccine
- Exposure: Universal Child credit offers benefit to parents with children <6yrs in Canada, Outcome: Family food insecurity, Negative controls: non-eligible families (only with children>6yrs old).
- Exposure:Air pollution week t, outcome: mortality/hospitalizations week t+1, negative control outcome:mortality/hospitalizations week t-1, negative control exposure: Air pollution week t+1.

Indirect confounding adjustment

Difference-in-differences as a NOC method:

- One common NOC of confounding approach is so-called difference-in-differences" (DID) estimation (Meyer, 1995; Angrist and Krueger, 1999; Blundell and MaCurdy, 2000), which is typically used when

Indirect confounding adjustment

Difference-in-differences as a NOC method:

- One common NOC of confounding approach is so-called difference-in-differences" (DID) estimation (Meyer, 1995; Angrist and Krueger, 1999; Blundell and MaCurdy, 2000), which is typically used when
  - (i) One has observed the outcome pre- and post-exposure for each person (or repeated samples), and

Indirect confounding adjustment

Difference-in-differences as a NOC method:

- One common NOC of confounding approach is so-called difference-in-differences" (DID) estimation (Meyer, 1995; Angrist and Krueger, 1999; Blundell and MaCurdy, 2000), which is typically used when
  - (i) One has observed the outcome pre- and post-exposure for each person (or repeated samples), and
  - (ii) the association of the unobserved confounder with the outcome is assumed equal across exposure groups and constant over time.

Indirect confounding adjustment

Difference-in-differences as a NOC method:

- One common NOC of confounding approach is so-called difference-in-differences" (DID) estimation (Meyer, 1995; Angrist and Krueger, 1999; Blundell and MaCurdy, 2000), which is typically used when
  - (i) One has observed the outcome pre- and post-exposure for each person (or repeated samples), and
  - (ii) the association of the unobserved confounder with the outcome is assumed equal across exposure groups and constant over time.
- Then, the approach entails estimating the effect of exposure by taking a difference between exposure groups of the average change in outcome over time.

Indirect confounding adjustment

Control Outcome Calibration Approach (COCA):

- Suppose that average causal effect (ACE) were known to be additive and constant across population.

Indirect confounding adjustment

Control Outcome Calibration Approach (COCA):

- Suppose that average causal effect (ACE) were known to be additive and constant across population.
- Then for each person, if ACE were known, would could use it to generate the outcome each person would have experienced if possibly contrary to fact, they had not been exposed.

Indirect confounding adjustment

Control Outcome Calibration Approach (COCA):

- Suppose that average causal effect (ACE) were known to be additive and constant across population.
- Then for each person, if ACE were known, would could use it to generate the outcome each person would have experienced if possibly contrary to fact, they had not been exposed.
- This estimated exposure-free outcome is a strong surrogate measure of the unmeasured confounder, and can be used to adjust for confounding between exposure and NCO

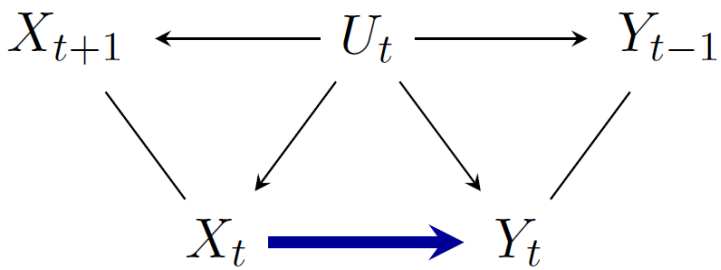
Indirect confounding adjustment

Control Outcome Calibration Approach (COCA):

- Suppose that average causal effect (ACE) were known to be additive and constant across population.
- Then for each person, if ACE were known, would could use it to generate the outcome each person would have experienced if possibly contrary to fact, they had not been exposed.
- This estimated exposure-free outcome is a strong surrogate measure of the unmeasured confounder, and can be used to adjust for confounding between exposure and NCO
- Find all possible values of ACE such that adjusting for the corresponding estimated exposure-free outcome minimizes association between exposure and NCO. See Tchetgen Tchetgen (American Journal of Epidemiology, 2013)

Indirect confounding adjustment

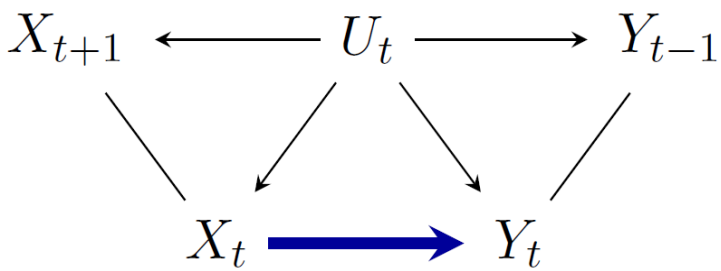
Double Adjustment by Negative Outcome-Negative Exposure (DANONE):



- Ideal for air pollution time series studies: X_t is air pollution measurement at time t and Y_t is health outcome at time t. Then air pollution at time $t+1$ is a negative control exposure and outcome at time $t-1$ is a negative control outcome for unmeasured confounding at time t.

Indirect confounding adjustment

Double Adjustment by Negative Outcome-Negative Exposure (DANONE):



- Ideal for air pollution time series studies: X_t is air pollution measurement at time t and Y_t is health outcome at time t. Then air pollution at time $t+1$ is a negative control exposure and outcome at time $t-1$ is a negative control outcome for unmeasured confounding at time t.
- The causal effect of X_t on Y_t is nonparametrically identified under rank conditions restricting the number of levels the unobserved confounder can take relative to the number of levels of the negative control exposure and outcome. See Miao and Tchetgen Tchetgen