Title
Negative controls to detect selection bias and measurement bias in epidemiologic studies

Journal
Epidemiology, 27(5)

Author
Arnold, Benjamin F

Publication Date
2016

Peer reviewed

OPEN

Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies

Benjamin F. Arnold, Ayse Ercumen, Jade Benjamin-Chung, and John M. Colford, Jr

Abstract: Biomedical laboratory experiments routinely use negative controls to identify possible sources of bias, but epidemiologic studies have infrequently used this type of control in their design or measurement approach. Recently, epidemiologists proposed the routine use of negative controls in observational studies and defined the structure of negative controls to detect bias due to unmeasured confounding. We extend this previous study and define the structure of negative controls to detect selection bias and measurement bias in both observational studies and randomized trials. We illustrate the strengths and limitations of negative controls in this context using examples from the epidemiologic literature. Given their demonstrated utility and broad generalizability, the routine use of prespecified negative controls will strengthen the evidence from epidemiologic studies.

(*Epidemiology* 2016;27:637–641)

Negative controls are used in laboratory science to help detect problems with the experimental method. In epidemiologic studies, a negative control outcome acts as a surrogate for the actual outcome—the negative control should be subject to the same potential sources of bias as the outcome but is not caused by the exposure of interest. Negative control exposures are conceptually the same, but defined relative to the actual exposure. Lipsitch et al.[1] defined the structure of negative controls to detect unmeasured confounding and described by way of example how negative controls could be used to detect selection bias and measurement (information) bias. Here, we define the causal structure of negative controls with respect to selection bias[2] and measurement bias,[3] and illustrate their use with published examples.

NEGATIVE CONTROLS TO DETECT SELECTION BIAS

For clarity, we have focused on structures of selection bias under the null (no effect of exposure) and have focused on four structures we would expect to be most relevant to epidemiologic research (Fig. 1).[2] Selection bias occurs when the analysis conditions on a third variable C that is a common descendant of exposure A and outcome Y or a common descendant of unmeasured causes of either A or Y or both, denoted U_A or U_Y.[2] Defining C as the combination of censoring mechanisms during enrollment, follow-up, and analysis, standard epidemiologic measures are limited to the stratum of $C = 0$ (uncensored, available data). We denote negative control exposures as N_A and negative control outcomes as N_Y—they could be dichotomous, categorical, or continuous.

A common form of selection bias can result from conditioning on a common descendant of the exposure and outcome (Fig. 1A). For example, in case–control designs where selection into the study (C) conditions on the outcome ($Y \rightarrow C$), selection bias results if the exposure affects participant selection ($A \rightarrow C$) differentially by case/control status. This bias structure could also occur in the re-analysis of a case–control study for a secondary outcome Z, which is intermediate between the exposure and outcome: $A \rightarrow Z \rightarrow Y \rightarrow C$. This design is used in genetic epidemiology studies that repurpose costly genomic measures A and look at their association with additional outcomes Z.[4,5] Negative control outcomes or exposures to detect this type of bias would need to similarly affect participant selection ($N_A \rightarrow C$, Table, example 1 or $N_Y \rightarrow C$, Table, example 2).

A second form of selection bias can occur in cross-sectional or retrospective studies when the outcome Y and an unmeasured cause of the exposure U_A affect study enrollment C (Fig. 1B). This bias can be detected by using a negative control exposure that shares the same unmeasured parent of the exposure ($U_A \rightarrow N_A$, Table, example 3) or a negative control outcome that similarly affects enrollment ($N_Y \rightarrow C$).

A third form of selection bias can occur when a study conditions on a common descendant of the exposure and an

Submitted 22 January 2015; accepted 9 May 2016.
From the Division of Epidemiology, University of California, Berkeley, CA.
Benjamin F. Arnold and Ayse Ercumen contributed equally to this study.
This study was funded in part by National Institutes of Health Grants 1R01HD078912, 1R21HD076216, 1R03HD076066, and 1K01AI119180.
The authors report no conflicts of interest.
SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).
Correspondence: Benjamin F. Arnold, Division of Epidemiology, University of California, Berkeley, 101 Haviland Hall, MC7358, Berkeley, CA 94720. Email: benarnold@berkeley.edu.

ISSN: 1044-3983/16/2705-0637
DOI: 10.1097/EDE.0000000000000504

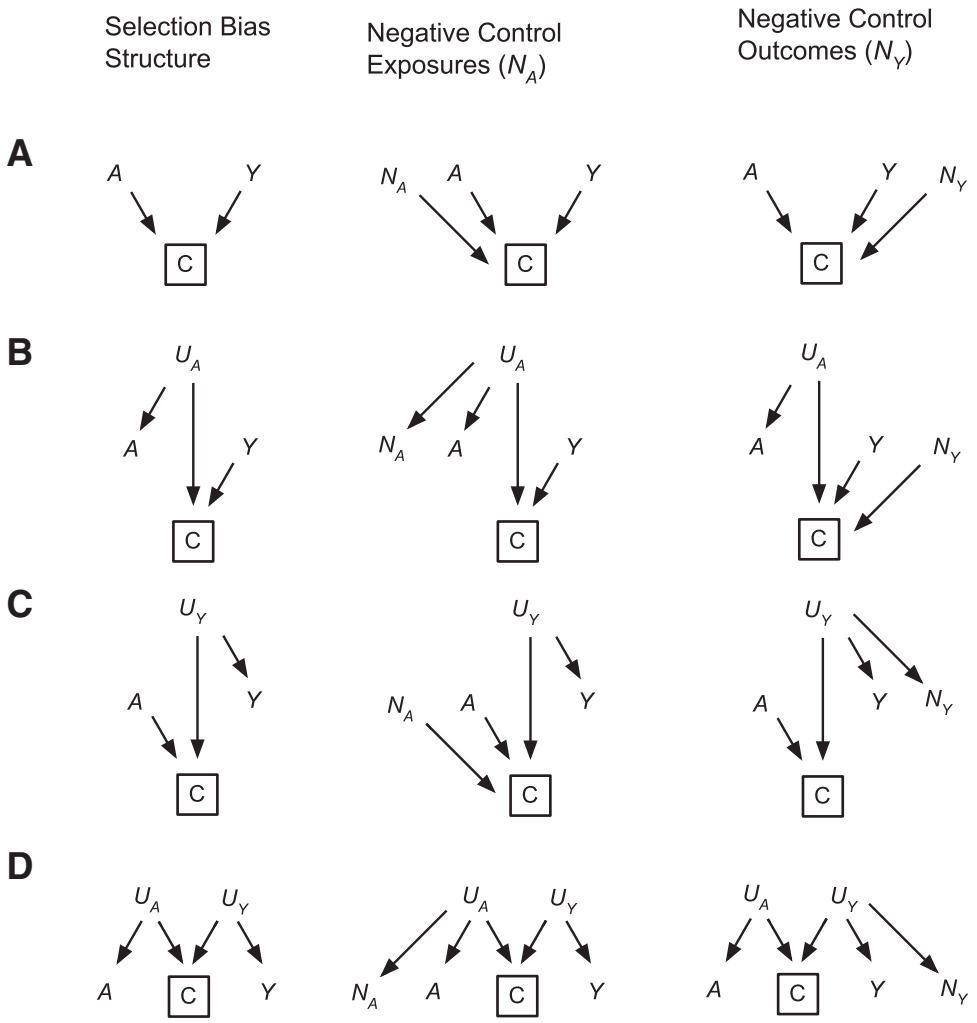


FIGURE 1. Simplified causal diagrams of selection bias for exposure *A* and outcome *Y* along with negative control exposures (N_A) and outcomes (N_Y). In all four structures, selection bias results from conditioning on *C*, a common descendant of (A) exposure *A* and outcome *Y*, (B) cause of exposure U_A and outcome *Y*, (C) exposure *A* and cause of outcome U_Y, or (D) cause of exposure U_A and cause of outcome U_Y.

unmeasured cause of the outcome (Fig. 1C). In per-protocol analyses of randomized trials, investigators limit the analysis to individuals that complied with their respective group assignments. Bias results if compliance (C) is determined by treatment assignment (A) and by unmeasured characteristics (U_Y) that affect both individuals' willingness to comply with their assigned treatment and their outcome.[6] For example, if individuals assigned to treatment who comply with their regimen are more health conscious than noncompliers, a naive per-protocol analysis could overestimate the benefits of the treatment. Figure 1C also applies to selection bias in prospective studies if exposure A and an unmeasured cause of the outcome U_Y affect loss to follow-up C. A negative control outcome that shares the same unmeasured parent as the outcome ($U_Y \rightarrow N_Y$, Table, example 4) or a negative control exposure that similarly affects enrollment, loss to follow-up or compliance ($N_A \rightarrow C$) can be used to detect this type of selection bias.

Finally, selection bias can occur if a cause of the exposure U_A and a cause of the outcome U_Y both affect enrollment C (Fig. 1D). One example is volunteer bias in cohort studies,[2] where individuals' underlying characteristics might affect their exposures and health outcomes as well as their decision to enroll in the study. This bias could be detected by using a negative control outcome that shares the same parent as the actual outcome ($U_Y \rightarrow N_Y$) or a negative control exposure that shares the same parent as the actual exposure ($U_A \rightarrow N_A$); however, we are unaware of a study that has used negative controls for this bias structure.

NEGATIVE CONTROLS TO DETECT MEASUREMENT BIAS

Many studies have measurement error so that investigators observe Y^*, which is an error-prone version of the outcome Y.[3] For example, if Y is an enteric infection that

TABLE. Examples of Studies that Have Used Negative Controls to Detect Selection or Measurement Bias Following Bias Structures in Figures 1 and 2

Example	Bias Structure	Design	Exposure (A)	Outcome (Y)	Potential Source of Bias	Negative Control*
Selection bias						
1. Ruckart et al.[8]	Figure 1A	Retrospective case–control	Chemical drinking water contaminant exposure during trimesters 1–2	Hematopoietic cancers, neural tube defects and oral clefts	Selective study enrollment among cases versus controls	Exposure (N_A) Chemical drinking water contaminant exposure during trimester 3
2. Ivers et al.[9]	Figure 1A	Retrospective case–control	Oral cholera vaccine	Diarrhea stool sample positive for cholera	Selective study enrollment among cases versus controls	Outcome (N_Y) Noncholera diarrhea
3. De Groot et al.[10]	Figure 1B	Retrospective case–control	Use of ACE inhibitors, statins, and proton pump inhibitors	Community-acquired pneumonia	Selective study enrollment among hospitalized patients	Exposure (N_A) Selective serotonin reuptake inhibitors
4. Danaei et al.[11]	Figure 1C	Prospective cohort	Statin use	Diabetes	Loss to follow-up that is affected by exposure	Outcome (N_Y) Peptic ulcers
Measurement bias						
5. Ercumen et al.[12]	Figure 2A	Randomized controlled trial	Safe storage and chlorination of drinking water	Reported diarrhea	Differential reporting error due to courtesy bias	Outcome (N_Y) Skin rash, toothache
6. Colford et al.[13]	Figure 2A	Prospective cohort	Swimmer exposure to *Enterococcus* levels in water	Reported diarrhea	Differential reporting error, unmeasured confounding	Exposure (N_A) *Enterococcus* levels assigned to nonswimmers
7. Zaadstra et al.[14]	Figure 2B	Retrospective case–control	Viral infections in early childhood	Multiple sclerosis	More accurate exposure recall among cases	Exposure (N_A) Broken arm, concussion, and tonsillectomy
8. Khush et al.[15]	Figure 2B	Prospective cohort	Fecal indicator bacteria in drinking water	Reported diarrhea	Exposure measured concurrently with outcome could be influenced by the outcome itself	Outcome (N_Y) Cough, congestion/coryza

The eAppendix (http://links.lww.com/EDE/B56) includes a more detailed discussion of each example.
*The rationale for each negative control listed in the table was an effect that would be impossible by the hypothesized mechanism, with one exception: for Colford et al.[13] (example 6), the rationale was to leave out an essential ingredient (water exposure).

causes diarrhea, Y^* could be caregiver-reported diarrhea symptoms. In the diagrams, we assume U_Y accounts for all other unmeasured causes of Y^* beyond Y. Similarly, A^* can be subject to unmeasured sources of error U_A. We focus our definitions on differential measurement errors (Fig. 2) because they are most likely to cause bias and the consequent bias is often the least predictable.[3] For parsimony, we have not provided formal definitions of negative controls under more complex (or simple nondifferential) measurement error scenarios, but in principle Figure 2 could be extended to accommodate them—for example, removing edges $Y \rightarrow U_A$ or $A \rightarrow U_Y$ defines negative controls for independent, nondifferential errors.

Differential outcome measurement error occurs when A influences the measured outcome Y^* through U_Y (Fig. 2A). In an unblinded study, physician follow-up (U_Y) may be increased in treated patients compared with the untreated ($A \rightarrow U_Y$), and selective follow-up causes differential measurement error of Y. Differential outcome reporting can also bias observational studies and unblinded trials with subjectively reported outcomes,[7] where participant knowledge of their exposure or treatment assignment could influence reporting. An ideal negative control outcome for this scenario shares a common source of correlated measurement error (U_Y) with the true outcome (Table, example 5). Negative control exposures for differential outcome measurement error also exist—placebo drugs in clinical trials are a classic example. Negative control exposures that act like a placebo can be devised for observational studies (Table, example 6).

Differential exposure measurement error is possible when the exposure A is measured concurrently with or after the occurrence of the outcome Y (Fig. 2B) and is of greatest concern in retrospective or cross-sectional studies. For example, retrospective case–control studies can be biased if they

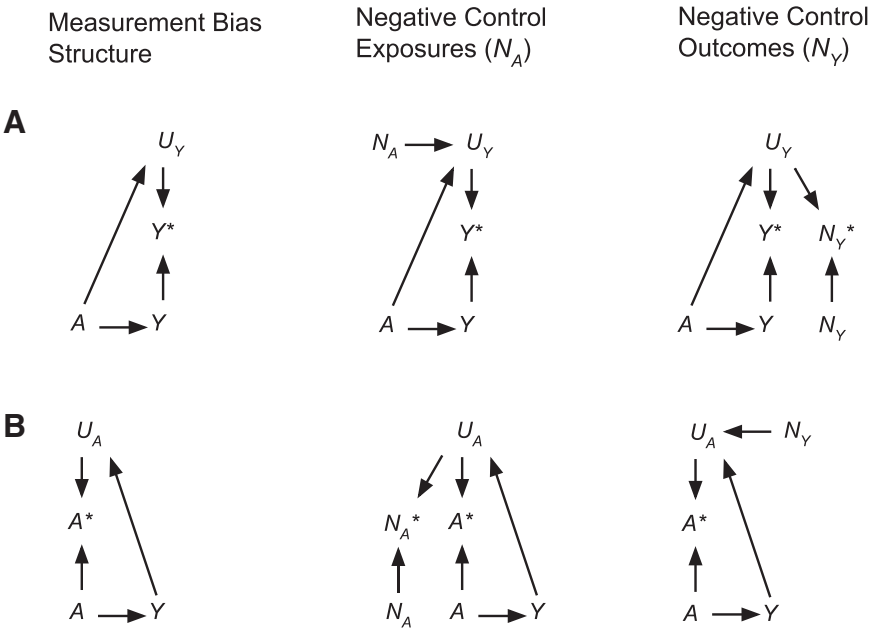


FIGURE 2. Simplified causal diagrams of differential measurement error for an exposure *A* that causes outcome *Y*. The basic structures for outcome measurement error (A) and exposure measurement error (B) are summarized along with negative control exposures (N_A) and outcomes (N_Y). U_Y represents other causes of the measured value of *Y** and U_A represents other causes of the measured value of *A**.

rely on self-reported exposures *A** as a proxy for true exposures *A* and cases remember exposures more accurately than controls. Negative controls for exposure measurement error need to share correlated errors (U_A) with the exposure (Table, example 7). The bias described in Figure 2B can also occur if an outcome is measured concurrently with an exposure, where the measured exposure (*A**) is used as a proxy for the same measure at a time in the past that is relevant for causing disease (*A*) (Table, example 8).[8–15]

DISCUSSION

We defined the structure of negative controls to detect common forms of selection and measurement bias in observational studies and randomized trials. The examples in the Table illustrate many recent applications, and the structural definitions in Figures 1 and 2 generalize to further applications we have not discussed—for example, Figure 1C describes the structure for healthy worker bias[2] and healthy user/adherer bias.[16] For extensions beyond the detection of bias, recent efforts have used negative controls in sensitivity analyses to quantify the magnitude of bias from unobserved confounding,[17] as a tool to remove bias in standardized mortality ratios,[18] and as a basis for large-scale empirical calibration of *P* values in drug safety studies.[19] We envision similar extensions for the types of negative controls defined here.

Negative controls have some limitations that arise in practice. Lipsitch et al.[1] characterized negative controls as a "blunt tool" to detect bias in the context of confounding, and that characterization is equally apt in the context of selection and measurement bias. Negative controls often lack specificity in the type of bias that they detect—many examples in the Table illustrate this limitation (Discussion in the eAppendix, http://links.lww.com/EDE/B56). Moreover, negative controls may identify the presence of bias but cannot in general determine its direction or magnitude without additional assumptions.[1] Another limitation that many negative controls share is that they often fail to provide a definitive test of the absence of bias.[1,20] All of these limitations coalesce into a common challenge for selecting negative controls: a control must meet its assumed structural definition, otherwise it can be an insensitive or inappropriate diagnostic for bias. Thus, the ability of a negative control to adequately detect bias ultimately relies on the plausibility of (often untestable) assumptions encoded in its causal diagram. Finally, prespecification of primary outcome and exposure definitions helps prevent the selective presentation of favorable results, and prespecification and complete reporting of negative controls would prevent similar problems.[20]

Selection bias or measurement bias threaten nearly every epidemiologic study design. Given their demonstrated utility and broad generalizability, the routine use of negative controls will help detect selection bias and measurement bias in epidemiologic studies.

REFERENCES

1. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21:383–388.
2. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
3. Hernán MA, Cole SR. Invited commentary: causal diagrams and measurement bias. *Am J Epidemiol*. 2009;170:959–962; discussion 963.
4. Tchetgen Tchetgen EJ. A general regression framework for a secondary outcome in case-control studies. *Biostatistics*. 2014;15:117–128.
5. Jiang Y, Scott AJ, Wild CJ. Secondary analysis of case-control data. *Stat Med*. 2006;25:1323–1339.
6. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9:48–55.

7. Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336:601–605.
8. Ruckart PZ, Bove FJ, Maslia M. Evaluation of exposure to contaminated drinking water and specific birth defects and childhood cancers at Marine Corps Base Camp Lejeune, North Carolina: a case-control study. *Environ Health*. 2013;12:104.
9. Ivers LC, Hilaire IJ, Teng JE, et al. Effectiveness of reactive oral cholera vaccination in rural Haiti: a case-control study and bias-indicator analysis. *Lancet Glob Health*. 2015;3:e162–e168.
10. de Groot MC, Klungel OH, Leufkens HG, van Dijk L, Grobbee DE, van de Garde EM. Sources of heterogeneity in case-control studies on associations between statins, ACE-inhibitors, and proton pump inhibitors and risk of pneumonia. *Eur J Epidemiol*. 2014;29:767–775.
11. Danaei G, García Rodríguez LA, Fernandez Cantero O, Hernán MA. Statins and risk of diabetes: an analysis of electronic medical records to evaluate possible bias due to differential survival. *Diabetes Care*. 2013;36:1236–1240.
12. Ercumen A, Naser AM, Unicomb L, Arnold BF, Colford JM, Jr, Luby SP. Effects of source- versus household contamination of tubewell water on child diarrhea in rural Bangladesh: a randomized controlled trial. *PLoS One*. 2015;10:e0121907.
13. Colford JM Jr, Schiff KC, Griffith JF, et al. Using rapid indicators for Enterococcus to assess the risk of illness after exposure to urban runoff contaminated marine water. *Water Res*. 2012;46:2176–2186.
14. Zaadstra BM, Chorus AM, van Buuren S, Kalsbeek H, van Noort JM. Selective association of multiple sclerosis with infectious mononucleosis. *Mult Scler*. 2008;14:307–313.
15. Khush RS, Arnold BF, Srikanth P, et al. H2S as an indicator of water supply vulnerability and health risk in low-resource settings: a prospective cohort study. *Am J Trop Med Hyg*. 2013;89:251–259.
16. Brookhart MA, Patrick AR, Dormuth C, et al. Adherence to lipid-lowering therapy and the use of preventive health services: an investigation of the healthy user effect. *Am J Epidemiol*. 2007;166:348–354.
17. Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. *Am J Epidemiol*. 2014;179:633–640.
18. Richardson DB, Keil AP, Tchetgen Tchetgen E, Cooper G. Negative control outcomes and the analysis of standardized mortality ratios. *Epidemiology*. 2015;26:727–732.
19. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med*. 2014;33:209–218.
20. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA*. 2013;309:241–242.

Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies

Benjamin F. Arnold, Ayse Ercumen, Jade Benjamin-Chung, John M. Colford, Jr

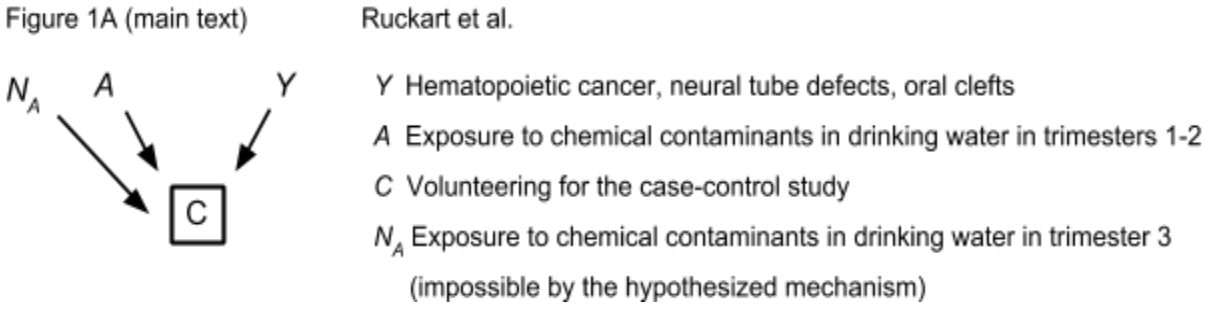
eAPPENDIX

This supplement includes more detailed discussion of each example listed in the main text Table.

Example 1: Ruckart et al. Negative Control Exposure for Selection Bias

A case-control study that examined the effect of in-utero exposure to chemical contaminants in drinking water on the risk of childhood hematopoietic cancers, neural tube defects and oral clefts[1] provides an example of negative control exposures to detect the type of selection bias depicted in Figure 1A. The study launched a media campaign that encouraged families to contact investigators if they conceived a child while living in the study area during the drinking water contamination period. The investigators enrolled cases from respondents that reported one of three outcomes of interest, and randomly selected controls from the remaining respondents. Monthly average levels of water contaminants in the study area were determined using groundwater modeling, and linked to participants through their residential address. Exposure categories were defined based on average concentrations during the exposure window of interest, with the "unexposed" group defined as individuals with no residential exposure to the chemical of interest during that window. Selection bias could arise in this study if exposed cases were more likely to respond to the media campaign than exposed controls because of heightened

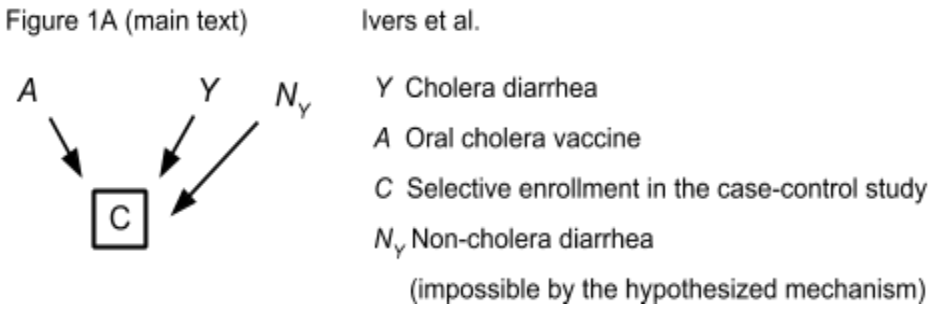
awareness about the health sequelae of the exposure (*A* affects *C* through increased awareness, differentially by case status *Y*). To detect this potential bias, investigators included non-relevant exposure periods as negative control exposures N_A (Figure 1A); they hypothesized that being exposed to drinking water chemicals during the third trimester of pregnancy could not plausibly cause neural tube defects and oral clefts given the timing of the formation of these organ systems. The investigators found no association between the negative control exposures and these two outcomes of interest, which helped rule out selection bias.



Example 2: Ivers et al. Negative Control Outcome for Selection Bias

Ivers et al.[2] conducted a treatment-facility based case-control study to measure the effectiveness of oral cholera vaccine against cholera in rural Haiti. Investigators were concerned that selective presentation at treatment facilities could bias the measure of effectiveness through selection bias following the structure of Figure 1A. To test whether this bias was present, investigators conducted two parallel case-control studies. The primary study had a case definition of a diarrheal stool that tested positive for cholera (*Y*); the second study used a negative control outcome case definition of a diarrheal stool that tested negative for cholera (N_Y). Both studies used an identical control sampling strategy: they enrolled four geographically matched controls for each case, under the assumption that exposure to cholera vaccine was

geographically clustered. Since the oral cholera vaccine has no protective efficacy against non-cholera diarrhea (an effect that would be impossible by the hypothesized mechanism), any association between cholera vaccine and non-cholera diarrhea would be due to selection bias or other unmeasured confounding. The case-control study estimated the protective efficacy of the vaccine of 58% (13%-80%) against cholera-positive diarrhea and there was no association with non-cholera diarrhea. This lent additional credibility to the study's findings. This general strategy of conducting parallel case-control studies with the auxiliary case-control sample based on a negative control outcome has been used in other vaccine effectiveness studies, including haemophilus influenzae type b (Hib) vaccine.[3]



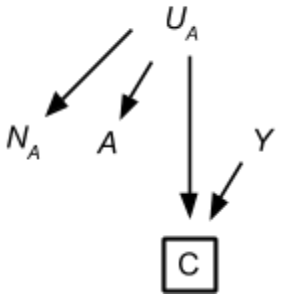
Example 3: de Groot et al. Negative Control Exposure for Selection Bias

A set of case-control studies investigated the association between community-acquired pneumonia and the use of angiotensin-converting-enzyme-inhibitors, statins, and proton pump inhibitors.[4] The investigators hypothesized that selection bias would result if they selected cases from hospitalized pneumonia patients rather than non-hospitalized pneumonia patients. Referral to the hospital is influenced by characteristics like age and comorbidities so patients that are referred and enrolled in the study are likely to be older and have comorbid symptoms; however, these symptoms U_A are also associated with the use of the drugs under study, giving rise to

selection bias ($Y \rightarrow C$ and $U_A \rightarrow C$ through hospital referrals). The investigators included the use of selective serotonin reuptake inhibitors (SSRI) as a negative control exposure N_A that likely shared the same parent factors as the exposures of interest but could not biologically cause pneumonia (Figure 1B). The investigators found significant associations between SSRI use and community-acquired pneumonia, suggesting selection bias could have influenced their other observed associations.

This example illustrates a limitation of negative controls: the assumptions about unobserved variables U encoded in the DAGs are often untestable. In Figure 1B if a directed edge also exists between $U_A \rightarrow Y$, then U_A would also meet the definition of an unmeasured confounder. Bias from unmeasured confounding in addition to selection bias is plausible in the pneumonia example, and illustrates how in practice it may be difficult to distinguish whether bias results from a single structure or from some mixture of them – this limitation surfaces repeatedly in the examples we describe.

Figure 1B (main text) de Groot et al.



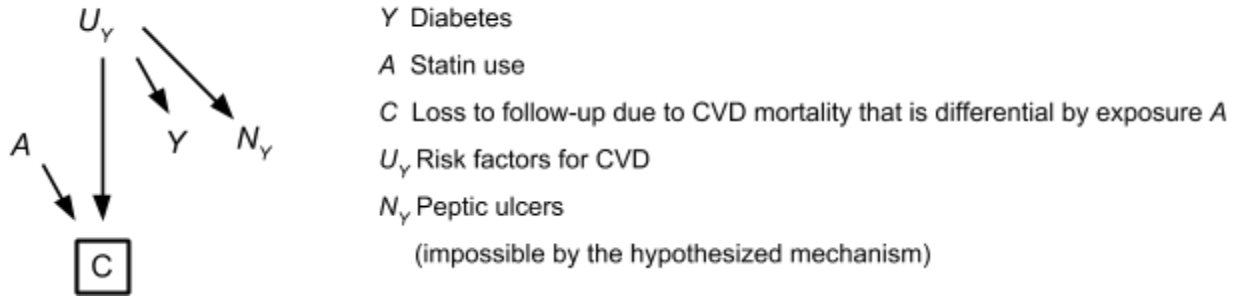
Y Community acquired pneumonia
A Use of ACE inhibitors, statins, & proton pump inhibitors
C Selective enrollment among hospitalized patients
U_A Co-morbidities that cause hospitalization *C* and drug exposures *A*
N_A Selective serotonin reuptake inhibitors
(impossible by the hypothesized mechanism)

Example 4: Danaei et al. Negative Control Outcome for Selection Bias

Danaei et al.[5] used electronic health records to emulate a hypothetical trial on statin use and diabetes, and used a negative control outcome N_Y to detect selection bias from selective loss

to follow-up that could be affected by exposure A (Figure 1C). The investigators hypothesized that because statin use reduces the risk of death from cardiovascular disease, untreated individuals would be more likely to be lost to follow-up due to death from cardiovascular disease (CVD). Furthermore, loss to follow-up would also be affected by an individual's cardiovascular disease risk factors, which are also shared risk factors for diabetes U_Y, thus artificially inflating the risk of diabetes among survivors treated with statins. To detect this bias, investigators included peptic ulcers as a negative control outcome N_Y that is not associated with statins -- an effect that would be impossible by the hypothesized mechanism -- but plausibly shares some of the same risk factors U_Y. The analysis showed no association between statin use and ulcers, lending credibility to the observed association between statins and diabetes.

Figure 1C (main text) Danaei et al.

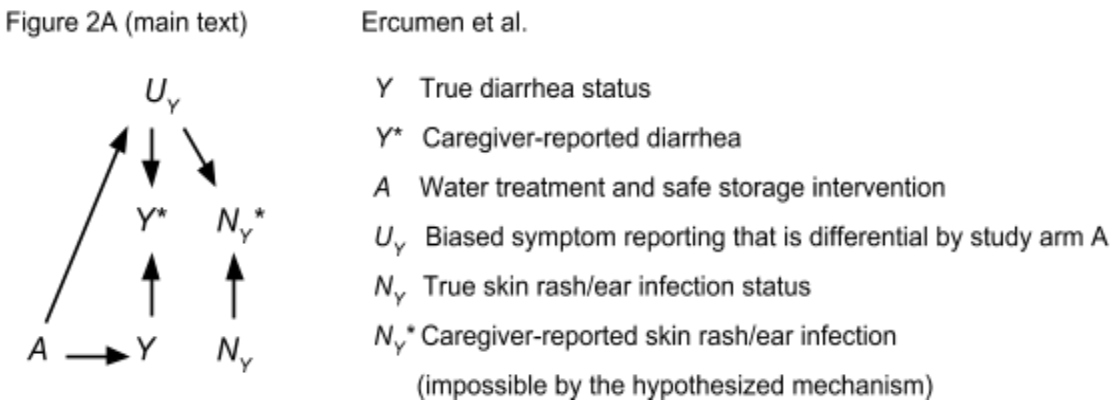


Example 5: Ercumen et al. Negative Control Outcome for Outcome Measurement Error

Ercumen et al.[6] conducted a randomized trial to measure the effect of safe drinking water storage and chlorination on caregiver-reported diarrhea among children used this type of negative control. The intervention reduced diarrhea by 36%, but there was concern that caregiver-reported diarrhea could be subject to courtesy bias if caregivers in the intervention arm differentially under-reported diarrhea to please investigators because the behavior change

component of the intervention emphasized improved child health.[7] If present, the differential measurement error would bias the effect of $A \rightarrow Y^*$ away from the null. To help rule out this possibility, the trial measured caregiver-reported skin rash and ear infections alongside diarrhea symptoms (N_{Y*}), which were thought to be similarly susceptible to differential reporting bias, but could not plausibly be reduced by improved drinking water quality (no direct path from $A \rightarrow N_{Y*}$) -- a check for an effect that would be impossible by the hypothesized mechanism. The trial found no evidence for reductions in either negative control outcome, lending additional credibility to the diarrhea results.

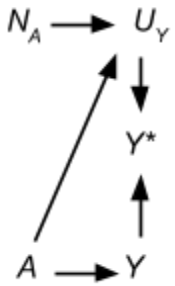
The use of skin rash and ear infections as negative controls in this example assumes that courtesy bias affects negative control reporting N_Y^* in a similar way as it affects diarrhea reporting Y^*; that is, a common source (or correlated sources) of measurement error exists. If instead errors were independent, $U_Y \rightarrow Y^*$ and $\rightarrow N_Y^*$, then it could be possible to see a null result for negative control outcomes and still have bias due to differential reporting of diarrhea. This could occur if participants understood that the water quality intervention was expected to reduce diarrhea, but not ear infections. The example underscores the importance of the assumptions encoded in the DAG -- which are usually unmeasured -- and highlights how negative controls do not necessarily prove the absence of bias when studies "pass" a test (though failing a test is always cause for concern).[8,9]



Example 6: Colford et al. Negative Control Exposure for Outcome Measurement Error

Colford et al.[10] conducted a prospective cohort study to measure the association between *Enterococcus* fecal indicator bacteria in ocean water and incident diarrhea among swimmers at a California beach. The study measured diarrhea using participant-reported symptoms that were potentially subject to reporting errors. Among swimmers who swallowed water, the study estimated that a $\log_{10}$ increase in *Enterococcus* concentration increased the odds of diarrhea by 1.74 (95% CI: 1.25, 2.43). The study also measured incident diarrhea among individuals who were present at the beach, but who did not enter the ocean. By leaving out the essential ingredient -- ingestion of enteric pathogens in ocean water -- there should have been no relationship between *Enterococcus* concentration matched to non-swimmers and incident diarrhea, and indeed that was the case (OR=1.00, 95% CI= 0.76, 1.32). The negative control exposure analysis helped rule out potential sources of unmeasured confounding as well as potential differential reporting errors.

Figure 2A (main text)



Colford et al.

Y True diarrhea status

Y^* Self-reported diarrhea

A Enterococcus concentration in ocean water for swimmers

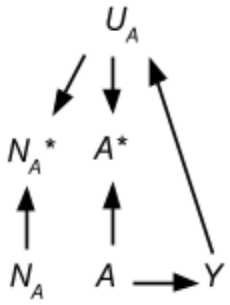
U_Y Biased symptom reporting that is differential by exposure A

N_A Enterococcus concentration in ocean water matched to non-swimmers (leaving out essential ingredient: ingestion of pathogens in ocean water)

Example 7: Zaadstra et al. Negative Control Exposure for Exposure Measurement Error

Zaadstra et al.[11] conducted a retrospective case-control study that measured the association between early childhood illness and multiple sclerosis using this type of negative control exposure. The investigators were concerned that multiple sclerosis cases might remember childhood medical events more accurately than community-based controls, and so in addition to measuring a history of viral infections of interest investigators included in their questionnaire negative control exposures -- broken arm, concussion, and tonsillectomy -- that could not plausibly cause multiple sclerosis. When investigators found significantly elevated odds of multiple sclerosis associated with the negative control exposures, it led to concern about differential recall bias.

Figure 2B (main text)



Zaadstra et al.

Y Multiple sclerosis

A True childhood viral infection status

A^* Reported childhood viral infections

U_A Biased symptom recall that is differential by case status Y

N_A True broken arm, concussion or tonsillectomy in childhood

N_A^* Reported broken arm, concussion or tonsillectomy in childhood (impossible by the hypothesized mechanism)

Example 8. Khush et al. Negative Control Outcome For Exposure Measurement Error

Khush et al.[12] conducted a prospective cohort study in India to measure the association between drinking water quality and child diarrhea. They used a negative control outcome to detect bias from differential exposure measurement error in their analysis (Figure 2B). The analysis matched monthly water samples to a child's diarrhea status at the time of sample collection, with the implicit assumption that water quality measured at the time of diarrhea assessment reflected water quality during the relevant exposure period, 3-10 days earlier. If caregivers boiled water in response to the onset of diarrhea ($Y \rightarrow U_A$), then measured water quality (A^*) would be better than the water quality during the relevant exposure period 3-10 days before diarrhea measurement (A), and this in turn could bias the estimated association toward the null. To help test for this potential bias, the investigators repeated the analysis using cough and congestion/coryza as negative control outcomes, which were unlikely to be caused by fecal water contamination. In this context, caregivers were more likely to boil water if a child had any illness, including cough and congestion/coryza ($N_Y \rightarrow U_A \rightarrow A^*$). The study found an association between water quality and the negative control outcomes. This finding called into question the associations estimated between fecal indicator bacteria concentrations and diarrhea in the study.

Figure 2B (main text)

Khush et al.

- Y Child diarrhea
- A Drinking water quality during relevant period, 3-10 days before diarrhea assessment
- A^* Drinking water quality measured at time of diarrhea assessment
- U_A Water treatment (e.g., boiling) that is differential by outcome Y
- N_Y Cough, congestion/coryza (impossible by the hypothesized mechanism)

ONLINE APPENDIX REFERENCES

1. Ruckart PZ, Bove FJ, Maslia M. Evaluation of exposure to contaminated drinking water and specific birth defects and childhood cancers at Marine Corps Base Camp Lejeune, North Carolina: a case-control study. *Environ Health*. 2013;12:104.

2. Ivers LC, Hilaire IJ, Teng JE, et al. Effectiveness of reactive oral cholera vaccination in rural Haiti: a case-control study and bias-indicator analysis. *Lancet Glob Health*. 2015;3(3):e162-e168.

3. Shapiro ED. Case-control studies of the effectiveness of vaccines: validity and assessment of potential bias. *Pediatr Infect Dis J*. 2004;23(2):127-131.

4. de Groot MCH, Klungel OH, Leufkens HGM, van Dijk L, Grobbee DE, van de Garde EMW. Sources of heterogeneity in case-control studies on associations between statins, ACE-inhibitors, and proton pump inhibitors and risk of pneumonia. *Eur J Epidemiol*. 2014;29(10):767-775.

5. Danaei G, García Rodríguez LA, Fernandez Cantero O, Hernán MA. Statins and risk of diabetes: an analysis of electronic medical records to evaluate possible bias due to differential survival. *Diabetes Care*. 2013;36(5):1236-1240.

6. Ercumen A, Naser AM, Unicomb L, Arnold BF, Colford JM Jr, Luby SP. Effects of source- versus household contamination of tubewell water on child diarrhea in rural Bangladesh: a randomized controlled trial. *PLoS One*. 2015;10(3):e0121907.

7. Schmidt W-P, Cairncross S. Household Water Treatment in Poor Populations: Is There Enough Evidence for Scaling up Now? *Environ Sci Technol*. 2009;43(4):986-992.

8. Lipsitch M, Tchetgen ET, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388.

9. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA*. 2013;309(3):241-242.

10. Colford JM, Schiff KC, Griffith JF, et al. Using rapid indicators for Enterococcus to assess the risk of illness after exposure to urban runoff contaminated marine water. *Water Res*. 2012;46(7):2176-2186.

11. Zaadstra BM, Chorus AMJ, van Buuren S, Kalsbeek H, van Noort JM. Selective association of multiple sclerosis with infectious mononucleosis. *Mult Scler*. 2008;14(3):307-313.

12. Khush RS, Arnold BF, Srikanth P, et al. H2S as an Indicator of Water Supply Vulnerability and Health Risk in Low-Resource Settings: A Prospective Cohort Study. *Am J Trop Med Hyg*. 2013;89(2):251-259.