Received: 14 October 2022 Revised: 2 March 2023 Accepted: 21 March 2023
DOI: 10.1002/pds.5623

Pragmatic considerations for negative control outcome studies to guide non-randomized comparative analyses: A narrative review

Sara N. Levintow[1,2] | Carrie M. Nielson[3] | Rohini K. Hernandez[3] | Alexander Breskin[1,2] | David Pritchard[2] | Timothy L. Lash[4] | Kenneth J. Rothman[5] | David Gilbertson[6] | Paul Muntner[7] | Cathy Critchlow[3] | M. Alan Brookhart[2,8] | Brian D. Bradbury[3]

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

[2]NoviSci, a Target RWE Company, Durham, North Carolina, USA

[3]Center for Observational Research, Amgen, Thousand Oaks, California, USA

[4]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

[5]RTI Health Solutions, Research Triangle Institute, Research Triangle Park, North Carolina, USA

[6]Chronic Disease Research Group, Hennepin Healthcare Research Institute, Minneapolis, Minnesota, USA

[7]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, Alabama, USA

[8]Department of Population Health Sciences, Duke University, Durham, North Carolina, USA

Correspondence
Sara N. Levintow, Department of Epidemiology, 135 Dauer Drive, 2101 McGavran-Greenberg Hall, CB #7435, Chapel Hill, NC 27599-7435, USA.
Email: levintow@email.unc.edu

Funding information
Amgen, Inc.

Abstract

Purpose: This narrative review describes the application of negative control outcome (NCO) methods to assess potential bias due to unmeasured or mismeasured confounders in non-randomized comparisons of drug effectiveness and safety. An NCO is assumed to have no causal relationship with a treatment under study while subject to the same confounding structure as the treatment and outcome of interest; an association between treatment and NCO then reflects the potential for uncontrolled confounding between treatment and outcome.

Methods: We focus on two recently completed NCO studies that assessed the comparability of outcome risk for patients initiating different osteoporosis medications and lipid-lowering therapies, illustrating several ways in which confounding may result. In these studies, NCO methods were implemented in claims-based data sources, with the results used to guide the decision to proceed with comparative effectiveness or safety analyses.

Results: Based on this research, we provide recommendations for future NCO studies, including considerations for the identification of confounding mechanisms in the target patient population, the selection of NCOs expected to satisfy required assumptions, the interpretation of NCO effect estimates, and the mitigation of uncontrolled confounding detected in NCO analyses. We propose the use of NCO studies prior to initiating comparative effectiveness or safety research, providing information on the potential presence of uncontrolled confounding in those comparative analyses.

Conclusions: Given the increasing use of non-randomized designs for regulatory decision-making, the application of NCO methods will strengthen study design, analysis, and interpretation of real-world data and the credibility of the resulting real-world evidence.

KEYWORDS
comparative effectiveness, comparative safety, healthcare databases, negative control, pharmacoepidemiology, residual confounding

Key Points

- Negative control outcome (NCO) studies can serve as a diagnostic tool for assessing uncontrolled confounding in real-world data before conducting comparative effectiveness or safety analyses.
- Key steps in NCO studies are to define confounding mechanisms in the patient population, identify NCOs affected by the same confounding structure as the treatment under study and outcome of interest, and estimate the effects of the treatment on the selected NCOs.
- The extent to which uncontrolled confounding is detected in the NCO analysis guides decisions to employ study design or analytic approaches to reduce bias and to proceed with comparative analyses.

1 | INTRODUCTION

The use of large healthcare databases to conduct post-marketing drug effectiveness and safety studies has led to advancements in data standards and methods. Widespread adoption of robust designs and analytic approaches, such as active comparator new user designs, propensity score methods, and marginal structural models, have become well-established in pharmacoepidemiology to address uncontrolled confounding arising in non-randomized studies.[1–5] The growing recognition of the potential for healthcare databases to supplement clinical trials in the evaluation of drug effectiveness and safety, coupled with advances in robust design and analysis, has led regulatory agencies to develop guidance on the use of real-world evidence in decision-making.[6–9] Despite these advances, a common concern among pharmacoepidemiologists and regulators is bias resulting from unmeasured or mismeasured confounders. Validation studies, data linkage, and quantitative bias analysis are essential tools to mitigate these concerns by probing and addressing the assumptions needed to ensure study validity. Additional confidence in results can be obtained by assessing the potential direction and magnitude of uncontrolled confounding using negative control outcome (NCO) methods that rely on an alternate set of assumptions.

An NCO is a variable that (a) has no plausible mechanism by which it can be caused by the treatment under study and (b) is expected to be affected by the same confounding structure as the treatment contrast for the outcome of interest.[10,11] If condition (a) is met, then an observed non-null association beyond what might be expected by chance between the treatment and NCO, after controlling for measured confounders, should reflect uncontrolled confounding rather than a causal relationship. If conditions (a) and (b) are met, then an association between treatment and an NCO reflects uncontrolled confounding between the treatment and outcome. Notably, neither condition (a) nor (b) can be evaluated using data; they must be assumed based on existing knowledge of causal mechanisms.[12] NCOs were used informally in epidemiology before a formal theory was advanced. For example, when postmenopausal estrogen use was found to be associated with lower all-cause mortality, an argument for confounding by lifestyle factors was made by demonstrating the association between estrogen use and lower mortality by accidents, suicide, and homicide—outcomes with no causal relationship to hormone replacement therapy.[13]

Studies of NCOs are increasingly used to detect and, in some cases, reduce or remove bias due to uncontrolled confounding, arising as a function of unmeasured or mismeasured confounders.[14–17] A benefit of randomization is its ability to balance all risk factors between treatment groups (in expectation, as imbalances may still occur by chance). However, in medical practice, patients are prescribed medicines based on clinical indications, and comparative analyses can only expect to achieve balance on measured and controlled confounders. When comparing interventions in non-randomized cohorts built from healthcare databases, the standard practice is to describe patient characteristics in the full cohort, calculate propensity scores from variables available in the database, and assess imbalances after propensity score matching or weighting. The threat of uncontrolled confounding is a particular concern in cohorts built on databases that lack information on key confounding variables, such as lifestyle and health behaviors, clinical biomarkers, and imaging results. Nonetheless, if these unmeasured confounders are correlated with those that are measured, standard analytic methods may result in estimates with negligible bias.[18] However, for new therapies that serve as second line or later treatment choices, channeling bias may occur, in which drugs with similar indications are selectively prescribed to groups of patients with varying disease prognoses; this bias can be common, with confounders difficult to ascertain or quantify in any design.[19]

The objective of this narrative review is to propose considerations for the pragmatic use of NCOs in guiding study design decisions and highlight two recent studies that inform these recommendations. Comprehensive reviews of prior work involving NCOs have already been conducted[14,15]; instead, we focus on practical steps to be taken in future research and highlight as examples two recently conducted studies.[20,21] These studies were selected to illustrate use of a variety of NCOs corresponding to several confounding mechanisms in different therapeutic areas. The first assessed the comparability of disease risk for patients initiating osteoporosis medications[20] and the second for those initiating lipid-lowering therapies[21] (Table 1). We propose that an NCO study can serve as a diagnostic

TABLE 1 Summary of two recent studies illustrating applications of NCOs.

Treatments under study	Outcomes of interest	Confounding mechanisms	NCOs selected	Unexpected associations	Potential biases
Osteoporosis treatments (denosumab, intravenous ZA, oral BPs)	Risk of osteoporotic fractures	Confounding by indication, frailty, or health-seeking behaviors	Decubitus ulcer, dementia, transfusion, accident, wellness visit, influenza vaccine, herpes zoster vaccine, pelvic screening, colon cancer screening, Mohs surgery, visual test	*Wellness visit*: Higher risks for denosumab vs. oral BPs, ZA vs. oral BPs. *Influenza vaccine*: Higher risk for ZA vs. oral BPs. *Herpes zoster vaccine*: Lower risk for denosumab vs. oral BPs.	Comparisons of denosumab and ZA to oral BPs may be confounded by health-seeking behaviors.
Lipid-lowering treatments (statins, ezetimibe, PCSK9i)	Risk of atherosclerotic cardiovascular disease events	Confounding by indication, frailty, or health-seeking behaviors	Decubitus ulcer, accident, fracture, cancer, wellness visit, visual test, influenza vaccine, herpes zoster or pneumococcal vaccine, colon cancer screening, non-melanoma skin cancer or Mohs surgery	*Decubitus ulcer*: Lower risks for PCSK9i vs. ezetimibe, PCSK9i vs. high-intensity statin. *Accident*: Lower risks for PCSK9i vs. ezetimibe, PCSK9i vs. high-intensity statin. *Fracture*: Lower risks for PCSK9i vs. ezetimibe, PCSK9i vs. high-intensity statin. *Influenza vaccine*: Higher risks for PCSK9i vs. high-intensity statin.	Comparisons of PCSK9i to statins and ezetimibe may be confounded by frailty and health-seeking behaviors.

Abbreviations: BPs, bisphosphonates; NCOs, negative control outcomes; PCSK9i, proprotein convertase subtilisin/kexin type 9 inhibitors; ZA, zoledronic acid.

for assessing uncontrolled confounding that informs the decision to proceed with comparative effectiveness or safety analyses.

2 | PRACTICAL CONSIDERATIONS FOR NCO STUDIES

To arrive at a set of results from NCO studies that can inform the decision to proceed to comparative analyses, one should consider three steps (Figure 1).

2.1 | Step 1: Describe expected confounding mechanisms in the target patient population

The relevant mechanisms are identified using knowledge of prescriber, patient, and disease characteristics that drive treatment choice and influence the outcome of interest.[12,22] Describing potential confounding mechanisms enables NCOs to be selected, as outlined in Step 2. Confounders relevant to healthcare database research can be organized into the following domains[23]:

2.1.1 | Confounding by indication or contraindication, often synonymous with channeling bias

Treatment selection is affected by factors associated with disease severity or the outcome of interest. A patient's response to prior therapies, the presence of or risk for a contraindicated condition, or their likelihood of benefit may guide the treatment choice, and these factors could independently affect the outcome. For example, newer injected therapies for osteoporosis (denosumab or intravenous zoledronic acid [ZA]) may be preferentially administered to patients with more severe disease and then appear to be associated with worse outcomes, compared with longer-established medications such as oral bisphosphonates (BPs).[20]

2.1.2 | Confounding by frailty, related to confounding by cognitive or functional impairment

Patients who have difficulty performing activities of daily living due to physical frailty or impairments in cognitive functioning may be less likely to visit a health care provider or pharmacy, and prescribers may

also be less likely to prescribe intensive therapy to these patients. For example, newer lipid-lowering therapies including proprotein convertase subtilisin/kexin type 9 inhibitor (PCSK9i) antibodies require injection while statins are taken orally; if patients prescribed PCSK9i antibodies are less likely to have physical or cognitive impairment, and these frailties are associated with poor outcomes, then prescription of PCSK9i antibodies compared with statins would appear to be associated with better outcomes.[21]

2.1.3 | Confounding by health-seeking behavior or access to care

Treatment selection is associated with greater access to and higher use of preventive health services. Patients on newer medications (denosumab or ZA for osteoporosis; PCSK9i for cardiovascular disease) may have higher engagement with the healthcare system due to more comprehensive insurance coverage or other factors. Their greater use of preventive services may lead to a lower risk of adverse outcomes, regardless of any effect of treatment.[20,21]

As illustrated in Figure 2, directed acyclic graphs (DAGs) are useful aids for showing the assumed confounding mechanisms to be addressed in NCO studies.[24] DAGs show assumed relationships among all known measured or unmeasured confounding variables and (a) treatment choice, (b) outcome of interest, and (c) NCO. The DAG should indicate which confounders are included in the propensity score model for the estimand(s) of interest.

2.2 | Step 2: Select NCOs

An NCO must (a) have no plausible causal relationship to the exposure, (b) share the same confounding structure as the outcome of interest, and (c) be measurable and sufficiently common in the target population to allow for effect estimation.[10,11] Although NCOs are often events unrelated to the outcome of interest, the first criterion allows for an NCO to be the same event type if it occurs at a time that rules out a causal relationship. For example, in a study on the effect of drug-eluting stents on 2-year mortality, an NCO was chosen to be mortality within 2 days of stenting, which is too soon to observe the biologic effect of the procedure.[25] If there is an unmeasured confounder that would affect the treatment and outcome of interest, then it likely also affects the NCO.

NCOs should link directly to the hypothesized confounding mechanisms. For example, diagnoses of decubitus ulcer—also called pressure ulcers or bedsores—occur primarily in people with conditions that limit their mobility.[26] In older populations, they are a marker of frailty and were selected as an NCO in the two aforementioned studies[20,21] to detect uncontrolled confounding by this mechanism. Receipt of influenza vaccination can be considered a proxy for higher healthcare engagement and use of preventive services; it was chosen as an NCO by both studies to capture the potential for confounding by health-seeking behaviors.[27] Figure 2 illustrates the potential for bias due to an unmeasured confounder (functional impairment), even after controlling for a measured confounder (age). Despite no causal mechanism for the treatment choice (PCSK9i vs. statin) to affect the NCO (accident risk), a spurious association results from the open

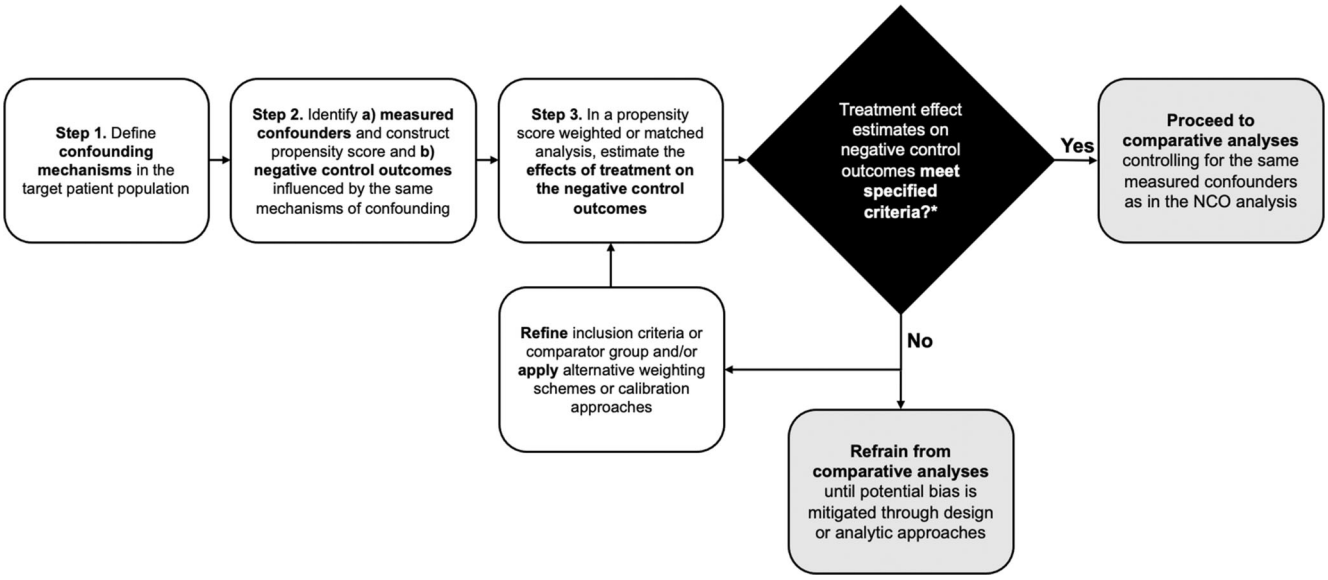

FIGURE 1 Incorporation of negative control outcomes (NCO) into a study design and analysis before proceeding with comparative effectiveness or safety research. *As further described in the text, criteria refer to the conditions under which NCO results suggest negligible residual confounding. They may concern the magnitude and precision of point estimates or apply a Bayesian framework that characterizes the potential bias as a distribution (evaluating whether potential bias remains within specified bounds). If multiple NCOs are tested, decision rules should be formulated to address the possibility that some but not all effect estimates meet specified criteria.

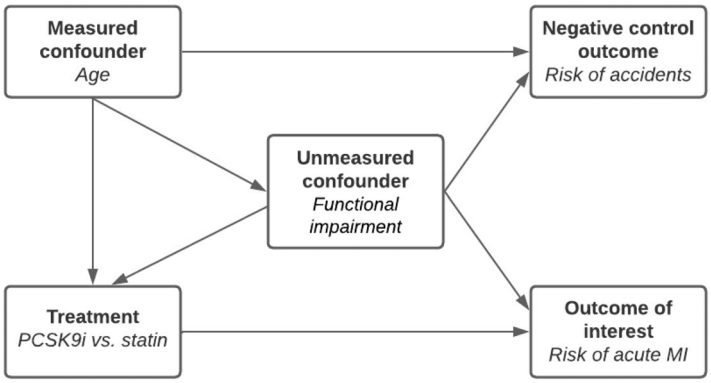

FIGURE 2 Directed acyclic graph showing the hypothesized relationships of measured and unmeasured confounders to the treatment, negative control outcome, and outcome of interest. Despite controlling for the measured confounder (age), the unmeasured confounder (functional impairment) induces an association between the treatment (PCSK9i vs. statin) and the negative control outcome (risk of accidents), indicating the potential for bias when comparing the outcome of interest (risk of acute MI) among treatment groups. For simplicity, the arrow is omitted from age directly to the MI outcome. MI, myocardial infarction; PCSK9i, proprotein convertase subtilisin/kexin type 9 inhibitors.

pathway from treatment to NCO through the unmeasured confounder. This non-causal association indicates the potential for bias when estimating the effect of treatment on the outcome of interest (risk of acute MI).

2.3 | Step 3: Determine whether treatment effects on the NCO indicate uncontrolled confounding

The effect of treatment on the NCO can be assessed using any valid estimator after accounting for measured confounders. Regardless of the analytic approach, the design and analysis of the NCO study should follow as closely as possible that of the study for the outcome of interest to ensure that detected or undetected confounding in the NCO study will also be present in a comparative analysis.

2.3.1 | Interpreting NCO effect estimates

Determining the extent to which uncontrolled confounding is a concern, based on NCO results, can inform whether to proceed with a comparative analysis. If NCO results are to serve as a diagnostic, criteria should be pre-specified to identify the conditions under which the results of the NCO study suggest negligible risk of substantial uncontrolled confounding. Suggested criteria may include the presence of all or a high proportion of results in which the point estimate associating treatment with the NCO lies within a window that indicates low magnitude of association (e.g., risk ratio 0.8–1.2) with a relatively narrow 95% confidence interval. A Bayesian framework may also be used to characterize the magnitude of bias (i.e., uncontrolled confounding) as a posterior distribution.[28,29] The criteria for proceeding with a comparative analysis could then be defined by bounds on the degree of bias to be tolerated (e.g., probability <5% that the magnitude of bias exceeds 1 NCO event out of 100). We recommend that criteria are specified based on subject matter expertise and expected estimates of treatment effects on the outcome of interest in the specific population.

Multiple NCOs are often tested to thoroughly evaluate various sources of uncontrolled confounding. When resulting effect measures range in size and precision, interpreting the set of results is not straightforward. One approach is to apply the pre-specified criteria to sets of NCOs linked to each confounding domain (e.g., markers of frailty or health-seeking behaviors). Rather than aggregating results from all NCOs, which likely represent a variety of confounding mechanisms, conclusions should be drawn separately for each domain. We note that employing a high number of NCOs may increase likelihood of a spurious finding, and investigators should formulate decision rules that allow for disparate results across NCOs in the same domain.

2.3.2 | Mitigating uncontrolled confounding uncovered by NCO analyses

If an NCO analysis suggests the possibility of substantial uncontrolled confounding, several design and analytic approaches can be employed (Table 2). Iterating through mitigation efforts and NCO analyses may resolve concerns of uncontrolled confounding.

Design-driven responses include refining inclusion criteria or the comparator group definition to improve balance of confounders. For example, in the NCO study of patients initiating lipid-lowering

TABLE 2 Approaches to mitigate uncontrolled confounding detected by NCO study.

Design	Analysis
Revise cohort inclusion criteria *Example*: Restrict to patient subset with more balanced covariates and reduced risk of residual confounding.	*Change the estimand* *Example*: Use an alternative weighting scheme (such as SMR weights instead of IPT weights) to improve covariate balance.
Refine comparator definition *Example*: Proceed with comparative analyses only for treatment contrasts that meet criteria for negligible confounding.	*Calibrate results* *Example*: Directly adjust effect estimates and confidence intervals based on control exposures and/or outcomes.

Abbreviations: IPT, inverse probability of treatment; NCOs, negative control outcomes; SMR, standardized mortality ratio.

therapies, results suggested patients with less frailty and higher healthcare utilization were more likely to be prescribed a PCSK9i compared with patients initiating statins or ezetimibe. The risk of decubitus ulcer among PCSK9i initiators was consistently lower (risk ratios 0.4–0.7), while risks of vaccinations tended to be higher (risk ratios 1.1–1.2 for some contrasts).[21] A subsequent NCO study can determine whether bias may be sufficiently reduced through more stringent inclusion criteria to balance patients' health status and healthcare access. Another design-driven approach would be to refine the comparator definition. For example, in the NCO study of osteoporosis treatments, one comparator pair—denosumab and ZA—was considered sufficiently comparable. However, another pair—new users of denosumab and oral BPs—showed evidence of confounding by health-seeking behaviors.[20] Follow-up NCO studies can be used to determine whether confounding persists when the comparator is restricted to patients switching oral BPs rather than initiating any oral medication for the first time, as those switching are hypothesized to be more similar to patients on injected treatments. It is important to ensure that all NCO analyses and design refinements are made before any treatment effect on the primary outcome is estimated.

If a review of the study design does not suggest possibilities for confounding mitigation, analytic approaches can be tested. Alternative weighting schemes can be used to estimate a treatment effect less subject to confounding. For example, NCO results may indicate confounding of the average treatment effect after applying inverse probability of treatment weights, which attempt to balance covariates across treatment groups using their distribution in the overall population. Instead, standardized mortality ratio weights could be used to estimate the average treatment effect in the treated; that is, the effect of treatment if all patients had the distribution of covariates in the treated group. This may achieve balance on the same confounders or require a smaller set, as only the untreated patients are weighted to resemble the treated (and not vice versa).[30] If changing the estimand is not feasible or still indicates bias, several methods have been proposed to reduce or remove bias by calibrating the effect estimate.[31–33] In contrast to the approach presented in this article, these methods are used after conducting a comparative analysis; they directly adjust effect estimates and confidence intervals by leveraging an alternative set of causal assumptions and incorporating control exposures and/or outcomes.[34–37] Studies of this sort have used semi-automated selection of control variables and adjustment of treatment effect estimates based on associations with control variables.[38,39]

3 | DISCUSSION

As non-randomized research in large healthcare databases is increasingly used to supplement conventional randomized trials in the evaluation of medical interventions, assessing the potential for uncontrolled confounding will remain important. We propose use of NCO methods to guide the decision to conduct comparative effectiveness or safety research, with the goal of strengthening study design, analysis, and interpretation of real-world data. Two recently completed studies reviewed in this article demonstrate implementation of NCOs in claims-based data sources and use of results to inform decisions to proceed with comparative studies.

3.1 | Considerations for identifying confounders and selecting NCOs

The studies of osteoporosis and lipid-lowering treatments demonstrate use of multiple NCOs for each domain of confounding, attempting to capture different mechanisms and the various ways in which they manifest in the study population. However, a limitation of this approach is that interpretation can be challenging if the results are conflicting. In the study of osteoporosis medications, NCO effect estimates corresponding to confounding by health-seeking behaviors were in different directions for new users of denosumab compared to oral BPs. Relatedly, confounding structures can differ across patient populations and between strata within a population. This heterogeneity raises the possibility that a null NCO association with treatment observed in one cohort may not apply to another population. Anticipating this possibility, in the NCO study of lipid-lowering therapies, analyses were initially stratified by age group and calendar year to evaluate possible effect modification by age or time. It is advisable to repeat NCO analyses in each cohort and, if effect measure modification is hypothesized, in each stratum of the modifying variable. We also note that our recommendations for selecting and interpreting NCOs draw from specific examples of osteoporosis and lipid-lowering treatments potentially affecting acute health events in older adults and may not generalize to all confounding scenarios.

3.2 | Considerations for interpreting NCOs and the potential for bias

Although NCO analyses can provide evidence of the presence or absence of uncontrolled confounding, like most studies, including conventional randomized trials, they rely on assumptions that may be untestable.[31] A key assumption is that the confounding paths between the treatment and outcome of interest are the same as those between treatment and NCO. If this assumption is violated, a true NCO may not be sensitive to uncontrolled confounding. Although this assumption cannot be tested in observational data, it reinforces the importance of specifying anticipated confounders as completely as possible in the propensity score model. Further, because NCO studies are often ancillary to a primary exposure-outcome analysis, precision may be insufficient to arrive at reliable conclusions about NCO associations. When the NCO incidence is low or study size is small, NCO studies may not yield effect estimates of sufficient precision to adequately assess the presence of uncontrolled confounding. For example, although decubitus ulcer is strongly associated with frailty, its prevalence is low in ambulatory patient populations. Large sample sizes are needed to ensure a sufficient number of outcomes such that an association between treatment and ulcer outcome can be reliably estimated and uncontrolled confounding by frailty, if present, can be detected. Conversely, it is possible to find no association between the

treatment and NCO, even in large studies, if there are multiple sources of opposing bias. Finally, like all studies relying on healthcare databases to build cohorts, ascertainment of key variables may be subject to measurement error and further bias effect estimates.

3.3 | Future directions for NCOs to inform comparative studies

Recommendations for future NCO studies complement guidance regarding adequate measurement of confounding variables in real-world data. Confounding variables should be measured to the extent possible (e.g., by linking claims and electronic health record data), and assessments of validity for potentially mismeasured confounders remain important.[6] After accounting for measured confounders, NCO studies can uncover evidence of uncontrolled confounding and inform the decision to proceed with a comparative analysis. Future research could employ a temporal approach, particularly for new drugs, in which an NCO study is refreshed regularly with new data to monitor the potential for confounding over time. If results indicate uncontrolled confounding, design and analytic approaches can be evaluated before the plan for a comparative analysis is abandoned. Calibration of effect estimates using NCO results, in conjunction with control exposures and/or outcomes, has the potential to reduce bias.[31–37] As the design and analysis of NCO studies advance in support of reliable real-world evidence, study protocols should pre-specify criteria to identify the conditions under which NCO results suggest negligible risk of uncontrolled confounding. Given the increased potential for non-randomized research to be used for regulatory decision-making, greater transparency of protocols governing decisions to move forward with comparative analyses is warranted.[40] In addition, recommendations for reporting results of NCO studies should be built into guidance for pharmacoepidemiology publications, as this would help reviewers to assess the likelihood of uncontrolled confounding as an explanation for reported associations.[41,42]

In conclusion, NCO studies are a useful tool to assess the threat of confounding by unmeasured and mismeasured variables, which remains a chief concern in healthcare database research. Proactively performing NCO analyses as a diagnostic in advance of comparative analyses may strengthen study design, analysis, and interpretation of real-world data, thereby strengthening the rigor and reproducibility of the resulting real-world evidence in informing regulatory decisions.

ACKNOWLEDGMENTS

This study was funded by Amgen, Inc.

CONFLICT OF INTEREST STATEMENT

Sara N. Levintow receives consulting fees from Target RWE, where she was previously an employee. Carrie M. Nielson was an employee of Amgen, Inc. and held stock in Amgen, Inc. while she worked on this study. Rohini K. Hernandez is an employee of and owns stock in Amgen, Inc. Alexander Breskin was an employee of and held stock in Target RWE while he worked on this study. David Pritchard is an employee of and owns stock in Target RWE. Timothy L. Lash is a member of the Amgen Methods Council, for which he receives travel support and consulting fees. He is supported in part by the National Library of Medicine (R01LM013049, PI Timothy L. Lash). A PhD student at Emory University is supported by a training agreement with Amgen, Inc. Paul Muntner has received research grants and consulting fees from Amgen, Inc. Cathy Critchlow is an employee of and owns stock in Amgen, Inc. M. Alan Brookhart serves on scientific advisory committees for Amgen, Astellas/Seagen, Atara Biotherapeutics, Brigham and Women's Hospital, Kite, Gilead, NIDDK, and Vertex; he receives consulting fees and owns equity in Target RWE. Brian D. Bradbury is an employee of and owns stock in Amgen, Inc.

ETHICS STATEMENT

The studies reviewed in this article were approved by the Chesapeake Institutional Review Board.

ORCID

Sara N. Levintow https://orcid.org/0000-0001-5558-526X

REFERENCES

1. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158(9):915-920.
2. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol*. 2006 Jun;163(12):1149-1156.
3. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep*. 2015 Dec;2(4):221-228.
4. Stürmer T, Wang T, Golightly YM, Keil A, Lund JL, Funk MJ. Methodological considerations when analysing and interpreting real-world data. *Rheumatology*. 2020;59(1):14-25.
5. Shiba K, Kawahara T. Using propensity scores for causal inference: pitfalls and tips. *J Epidemiol*. 2021;31(8):457-463.
6. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Oncology Center for Excellence. Real-World Data: assessing electronic health records and medical claims data to support regulatory decision-making for drug and biological products. *Draft Guidance for Industry*; 2021.
7. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Data standards for drug and biological product submissions containing real-world data. *Draft Guidance for Industry*. 2021.
8. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Oncology Center for Excellence. Real-world data: assessing registries to support regulatory decision-making for drug and biological products. *Draft Guidance for Industry*. 2021.
9. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. *Draft Guidance for Industry*. 2021.
10. Arnold BF, Ercumen A. Negative control outcomes: a tool to detect bias in randomized trials. *JAMA*. 2016;316:2597-2598.

11. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010 May;21(3):383-388.
12. Robins JM, Wasserman L. On the impossibility of inferring causation from association without background knowledge. In: Glymour P, Cooper G, eds. *Computation, Causation, and Discovery*. AAAI Press; 1999:305-321.
13. Petitti D, Perlman J, Sidney S. Noncontraceptive estrogens and mortality: long-term follow-up of women in the Walnut Creek Study. *Obstet Gynecol*. 1987;70:289-293.
14. Dusetzina SB, Brookhart MA, Maclejewski ML. Control outcomes and exposures for improving internal validity of nonrandomized studies. *Health Serv Res*. 2015;50(5):1432-1451.
15. Shi X, Miao W, Tchetgen ET. A selective review of negative control methods in epidemiology. *Curr Epidemiol Rep*. 2020 Dec;7(4):190-202.
16. Htoo PT, Measer G, Orr R, et al. Evaluating confounding control in estimations of influenza antiviral effectiveness in electronic health plan data. *Am J Epidemiol*. 2022;191(5):908-920.
17. Etievant L, Sampson JN, Gail MH. Increasing efficiency and reducing bias when assessing HPV vaccination efficacy by using nontargeted HPV strains. *Biometrics*. 2022:1-12. doi:10.1111/biom.13663
18. Barberio J, Ahern TP, Maclehose RF, et al. Assessing techniques for quantifying the impact of bias due to an unmeasured confounder: an applied example. *Clin Epidemiol*. 2021;13:627-635.
19. Lobo FS, Wagner S, Gross CR, Schommer JC. Addressing the issue of channeling bias in observational studies with propensity scores analysis. *Res Social Adm Pharm*. 2006 Mar;2(1):143-151.
20. McGrath LJ, Spangler L, Curtis JR, et al. Using negative control outcomes to assess the comparability of treatment groups among women with osteoporosis in the United States. *Pharmacoepidemiol Drug Saf*. 2020;29(8):854-863.
21. Levintow SN, Orroth KK, Breskin A, et al. Use of negative control outcomes to assess the comparability of patients initiating lipid-lowering therapies. *Pharmacoepidemiol Drug Saf*. 2022;31(4):383-392.
22. Hernán MA, Hernández-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155(2):176-184.
23. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010 Jun;48(6 Suppl):S114-S120.
24. Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021 May 17;50(2):620-632.
25. Mauri L, Silbaugh TS, Wolf RE, et al. Long-term clinical outcomes after drug-eluting and bare-metal stenting in Massachusetts. *Circulation*. 2008;118(18):1817-1827.
26. Jaul E, Barron J, Rosenzweig JP, Menczel J. An overview of co-morbidities and the development of pressure ulcers among older adults. *BMC Geriatr*. 2018;18(1):305.
27. Eurich DT, Achtymichuk KA, Johnson JA, Minhas-Sandhu JK, Lin M. Development and validation of an index score to adjust for healthy user bias in observational studies. *J Popul Ther Clin Pharmacol*. 2017;24(3):e79-e89.
28. McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Stat Med*. 2007;26(11):2331-2347.
29. McCandless LC, Gustafson P, Levy AR, Richardson S. Hierarchical priors for bias parameters in Bayesian sensitivity analysis for unmeasured confounding. *Stat Med*. 2012;31(4):383-396.
30. Sarvet AL, Wanis KN, Stensrud MJ, Hernán MA. A graphical description of partial exchangeability. *Epidemiology*. 2020 May 1;31(3):365-368.
31. Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. *Am J Epidemiol*. 2014;179(5):633-640.
32. Sofer T, Richardson DB, Colicino E, Schwartz J, Tchetgen Tchetgen EJ. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Stat Sci*. 2016;31(3):348-361.
33. Miao W, Geng Z, Tchetgen Tchetgen EJ. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*. 2018;105(4):987-993.
34. Flanders WD, Strickland MJ, Klein M. A new method for partial correction of residual confounding in time-series and other observational studies. *Am J Epidemiol*. 2017 May 15;185(10):941-949.
35. Miao W, Tchetgen Tchetgen E. Invited commentary: bias attenuation and identification of causal effects with multiple negative controls. *Am J Epidemiol*. 2017;185(10):950-953.
36. Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci U S A*. 2018;115(11):2571-2577.
37. Schuemie MJ, Ryan PB, Pratt N, et al. Large-scale evidence generation and evaluation across a network of databases (LEGEND): assessing validity using hypertension as a case study. *J Am Med Inform Assoc*. 2020;27(8):1268-1277.
38. Suchard MA, Schuemie MJ, Krumholz HM, et al. Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *Lancet*. 2019;394(10211):1816-1826.
39. Lane JCE, Weaver J, Kostka K, et al. Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. *Lancet Rheumatol*. 2020;2(11):e698-e711.
40. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1033-1039.
41. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:k3532.
42. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*. 2021;372:m4856.

How to cite this article: Levintow SN, Nielson CM, Hernandez RK, et al. Pragmatic considerations for negative control outcome studies to guide non-randomized comparative analyses: A narrative review. *Pharmacoepidemiol Drug Saf*. 2023;32(6):599-606. doi:10.1002/pds.5623