Control Outcomes and Exposures for Improving Internal Validity of Nonrandomized Studies

*Stacie B. Dusetzina, M. Alan Brookhart, and
Matthew L. Maciejewski*

Objective. Control outcomes and exposures can improve internal validity of nonrandomized studies by assessing residual bias in effect estimates. Control outcomes are those expected to have no treatment effect or the opposite effect of the primary outcome. Control exposures are treatments expected to have no effect on the primary outcome. We review examples of control outcomes and exposures from prior studies and provide recommendations for conducting and reporting these analyses.

Data Sources and Study Design. Review in Google Scholar and Medline of research studies employing control outcomes or exposures. We abstracted publication year, control outcome, control exposure, primary outcome, primary exposure, control outcome/exposure effect, proposed source of bias, and causal criteria.

Principal Findings. There is inconsistent terminology for these concepts, making study identification challenging. Six of 11 studies found null associations between treatments and negative control outcomes/exposures, providing greater confidence that the primary study findings were not biased. Five studies found unexpected associations, suggesting bias in the primary association.

Conclusions. The rigor of nonrandomized studies can be improved with inclusion of control outcomes and exposures for bias detection. Given ongoing concern about clinical and policy inferences from nonrandomized studies, we recommend adoption of these measurement tools.

Key Words. Control outcomes, control exposures, falsification tests, nonequivalent outcomes, nonequivalent exposures

There is an ongoing debate about the evidentiary value of nonrandomized study results. Much of this has stemmed from the controversy surrounding the Women's Health Initiative trial of hormone replacement therapy and risk for coronary heart disease (Manson et al. 2003). Although other efforts have demonstrated that well-conducted nonrandomized studies can generate

results equivalent with randomized trials (Benson and Hartz 2000; Concato, Shah, and Horwitz 2000; Furlan et al. 2008; Hernan et al. 2008), this perspective is far from universally shared because several meta-analyses have reported substantial discordance between randomized and nonrandomized studies (Ioannidis et al. 2001; Ioannidis 2005).

Efforts have been made in recent years to address concerns regarding the internal validity of nonrandomized studies, including the widespread adoption of new-user designs (Ray 2003), use of sample restriction (Schnee-weiss et al. 2007), and use of statistical methods such as propensity scores, instrumental variables, and marginal structural models. These statistical tools are often a last resort, because internal validity threats of selection bias (also known as unobserved confounding) were not addressed via study design or measurement strategies. Further, any assessment of causality in a nonrandomized study relies on assumptions about statistical models and their specification that must be guided by subject-matter knowledge (Robins 2001; Hernan et al. 2002). Researchers using secondary data for comparative effectiveness research have to address limitations regarding a lack of information on known confounders (since known confounders may be unmeasured) and potential gaps in subject-matter knowledge that reduce the likelihood of estimating causal effects without bias (Brookhart et al. 2010b).

Nevertheless, researchers may have adequate subject-matter knowledge to be able to identify outcomes that are *not* expected to change in response to the intervention of interest, which have been referred to as control outcomes or nonequivalent outcomes. Inclusion of control outcomes in nonrandomized studies can be a potentially useful strategy for detecting selection bias by expanding the measurement set beyond outcomes expected to change in response to the exposure or treatment of interest. Researchers may also be able to identify treatments that are known to be unrelated to the primary outcome (analogous to placebos). These control exposures are additional tools that researchers could use to assess selection bias, as is routinely done in the

Address correspondence to Matthew L. Maciejewski, Ph.D., Center for Health Services Research in Primary Care, Department of Veterans Affairs; Division of General Internal Medicine, Department of Medicine, Duke University Medical Center, Durham, NC 27705; e-mail: mlm34@duke.edu. Stacie B. Dusetzina, Ph.D., is with the Division of Pharmaceutical Outcomes and Policy, UNC Eshelman School of Pharmacy; Department of Health Policy and Management, Gillings School of Global Public Health; Lineberger Comprehensive Cancer Center; Cecil G. Sheps Center for Health Services Research, all at University of North Carolina at Chapel Hill, Chapel Hill, NC. M. Alan Brookhart, Ph.D., is with the Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill; Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, NC.

economics literature (Basker 2005; Holmlund, McNally, and Viarengo 2010; Rothstein 2010). Unfortunately, control outcomes and control exposures are under-utilized in nonrandomized comparative effectiveness research, although their use and advantages in this area have been described previously (Brookhart et al. 2010a; Lipsitch, Tchetgen Tchetgen, and Cohen 2010; Prasad and Jena 2013).

The purpose of this paper is to introduce control outcomes and exposures to a wider audience by summarizing illustrative examples from prior studies, which may provide readers with insights for identification of control outcomes and exposures in their own work. We conclude with recommendations for the conduct and reporting of studies that employ control outcomes or exposures, and present a framework for identifying them using Sir Austin Bradford Hill's factors for assessing causation (Hill 1965). We expect this overview to be of interest to researchers, manuscript reviewers, and grant reviewers seeking to improve the rigor and internal validity of comparative effectiveness research studies using non-randomized study designs.

METHODS

To identify published studies that employed control outcomes or exposures, we searched MEDLINE (via PubMed) and Google Scholar for manuscripts that included terms related to control outcomes, control exposures, falsification endpoints, falsification tests, and nonequivalent outcomes or exposures. We also included articles previously known or produced by the study team that were not identified in our prior search since there are no Medical Subject Heading (MeSH) terms for control outcomes or control exposures. We then manually searched bibliographies and works citing selected articles and consulted with colleagues to guide further study selection.

Identified studies were retained if they clearly identified the control outcome or exposure on the basis of its stated purpose for inclusion in the analysis. From our search we identified 11 studies that utilized control outcomes, control exposures, or both (Table 1). For each of these studies, we abstracted the following information: year of publication, control outcome, control exposure, primary study endpoint, primary study exposure/treatment, primary study effect measure, control outcome effect, control exposure effect, the proposed source of bias (if named), and the causal criteria (if known). We used this

Table 1: Summary of Published Studies Using Nonequivalent Outcomes

Citation	Nonequivalent Outcome	Nonequivalent Exposure	Primary Study Endpoint	Exposure or Treatment	Primary Study Effect Measure	Nonequivalent Outcome Effect	Nonequivalent Exposure Effect	Proposed Source of Bias	Causal Criteria
Redelmeier, BMJ 2005	Posturgical noncardiac complications (wound infection, ileus, pneumonia, aspiration, respiratory failure, renal failure, delirium)	N/A	Death or AMI	Atenolol versus metoprolol	Relative risk from logistic regression	No differences in posturgical noncardiac complications by beta-blocker (atenolol vs. metoprolol)	N/A	Unobserved confounding	Unstated
Mauri, Circulation 2008	Mortality within 2 days of stenting	N/A	Mortality within 2 years of stenting	Drug-eluting or bare metal stents	Percentage difference in mortality rates	Small absolute difference in 2-day mortality (0.45% vs. 0.68%, $p = .10$)	N/A	Unobserved confounding	Unstated
Rasmussen, JAMA 2007	Cancer-related hospital admissions	Calcium channel blockers	Long-term mortality post-MI	Adherence to statins and beta-blockers	Hazard ratio from Cox model	No increase in cancer-related admissions by statin or beta-blocker adherence level	Adherence to calcium channel blockers was not associated w/mortality endpoints, as expected	Healthy user bias	No biological plausibility

Continued

Table 1. *Continued*

<i>Citation</i>	<i>Nonequivalent Outcome</i>	<i>Nonequivalent Exposure</i>	<i>Primary Study Endpoint</i>	<i>Exposure or Treatment</i>	<i>Primary Study Effect Measure</i>	<i>Nonequivalent Outcome Effect</i>	<i>Nonequivalent Exposure Effect</i>	<i>Proposed Source of Bias</i>	<i>Causal Criteria</i>
Maciejewski, Health Affairs. 2010	ARBs, cholesterol absorption inhibitors	N/A	Medication adherence to ACEIs, BBs, CCBs, statins, diuretics	VBID (lower copays)	OR and predicted adherence showed that VBID avoided a 2–3% drop in adherence at 1 year	Null effect as expected	N/A	Unobserved confounding	Unstated
McClellan, JAMA. 1994	Mortality within 1 day of acute myocardial infarction (AMI)	N/A	Mortality within 4 years of AMI	Cardiac catheterization	Percentage difference in mortality rates	Significant difference at 1 day, which was unexpected	N/A	Unobserved confounding due to differential access to cath-ing hospitals	Access to cath-ing hospitals

Continued

Table 1. Continued

Citation	Nonequivalent Outcome	Nonequivalent Exposure	Primary Study Endpoint	Exposure or Treatment	Primary Study Effect Measure	Nonequivalent Outcome Effect	Nonequivalent Exposure Effect	Proposed Source of Bias	Causal Criteria
Jackson, Int J Epidemiol. 2006	Injury or trauma hospitalization, also possibly admits for IHD, CHF and CVD (but not clearly stated as such)	Flu shot in pre-flu season	Mortality and flu/pneumonia hospitalization in flu season	Flu shot	Relative risk from Cox model showed that flu shot was protective versus all-cause mortality and pneumo/flu admits in flu season	Flu shot significantly protective versus injury/trauma admit in pre-flu season and in flu season	Flu shot significantly protective against all-cause mortality and flu/pneumonia hospitalization in pre-flu season	Confounding by health status	Unstated
Brookhart, Am J Epidemiol. 2007	Bone mineral density test, screening mammography, PSA, FOBT, flu shot, pneumo shot	N/A	N/A		Hazard ratio from Cox model		N/A	Healthy user bias	Unstated

Continued

Table 1. *Continued*

<i>Citation</i>	<i>Nonequivalent Outcome</i>	<i>Nonequivalent Exposure</i>	<i>Primary Study Endpoint</i>	<i>Exposure or Treatment</i>	<i>Primary Study Effect Measure</i>	<i>Nonequivalent Outcome Effect</i>	<i>Nonequivalent Exposure Effect</i>	<i>Proposed Source of Bias</i>	<i>Causal Criteria</i>
Patrick, Value Health. 2011	Preventive services (bone min density, PSA, mammogram, pneumonia shot), clinical outcomes (asthma, burns, GI bleeds, skin infect)	N/A	MI, death, nursing home admit	Statin initiation	Rate ratio from Cox model for preventive, various methods for clinical outcomes	Significant difference in preventive (mammogram, flu shot) and clinical outcomes (asthma, burns, falls, fractures, motor vehicle accident)	N/A	Healthy user bias	Unstated
Jena, J Gen Intern Med. 2013	Dx of osteoarthritis, chest pain, UTI, DVT, skin infection, and RA	N/A	Community-acquired pneumonia	PPI fill	Proportion in quarter with Dx of CAP showed that PPI users had higher incidence rates than nonusers	PPI associated with higher rates of all unexpected Dx	N/A	Confounding by indication, disease severity	No biological plausibility

information to summarize results from eight studies that used control outcomes, two studies that used control exposures, and one study that used both.

RESULTS

Studies Using Control Outcomes

Control outcomes can be either negative (i.e., outcomes known to be unaffected by the treatment under study) or positive (i.e., outcomes known to be affected by treatment). The nine studies that used negative control outcomes chose outcomes that were not anticipated to be related to exposure (McClellan, McNeil, and Newhouse 1994; Redelmeier, Scales, and Kopp 2005; Jackson et al. 2006; Brookhart et al. 2007; Rasmussen, Chong, and Alter 2007; Mauri et al. 2008; Maciejewski et al. 2010; Patrick et al. 2011; Jena, Sun, and Goldman 2013) and were sorted into two groups. Four studies found that the association of the treatment with the negative control outcome was null (as expected) and provided greater confidence that the treatment effect/association with the primary outcome was not biased by unobserved confounding (Redelmeier, Scales, and Kopp 2005; Rasmussen, Chong, and Alter 2007; Mauri et al. 2008; Maciejewski et al. 2010). Five studies found that the treatment was *unexpectedly* associated with the negative control outcomes, which raised concerns about residual bias in the treatment effect/association on the primary outcome (McClellan, McNeil, and Newhouse 1994; Jackson et al. 2006; Brookhart et al. 2007; Patrick et al. 2011; Jena, Sun, and Goldman 2013).

Studies by Redelmeier, Mauri, Rasmussen, and Maciejewski were represented by this first group. Redelmeier, Scales, and Kopp (2005) compared the risk of acute myocardial infarction (AMI) or death among patients receiving either atenolol or metoprolol following elective surgery. They included several postsurgical noncardiac complications (wound infection, ileus, pneumonia, aspiration, respiratory failure, renal failure, delirium) as negative control outcomes and found no differences in these negative controls by beta-blocker received. Mauri et al. (2008) compared mortality, myocardial infarction (MI), and target-vessel revascularization within 2 years among patients receiving a drug-eluting stent or a bare metal stent. Mortality during the first 2 days following stent placement was the negative control outcome, since benefits of using one therapy over the other would not be expected in the immediate postsurgery period. They found no differences in 2-day mortality as expected, so concluded that the association between treatment and primary outcome was unconfounded.

Rasmussen, Chong, and Alter (2007) compared long-term mortality post-AMI among patients with high, intermediate, and low levels of adherence to statins or beta-blockers. They included cancer-related hospital admissions as a control outcome as adherence to statins or beta-blockers was not hypothesized to increase this risk. As expected, they found no association between treatment adherence and cancer hospitalizations, suggesting that healthy adherer bias was not likely to be influencing their findings. Finally, Maciejewski et al. (2010) evaluated the impact of a value-based insurance design scheme (lower prescription drug copayments) on refill adherence to angiotensin-converting-enzyme (ACE) inhibitors, beta-blockers, calcium channel blockers, statins, and diuretics, comparing patients in plans that did and did not implement copayment changes. Angiotensin receptor blockers and cholesterol absorption inhibitors were the negative control outcomes because copayment changes for these drugs were comparable between patients in the two arms. As expected, they found no difference in adherence to these drugs, so concluded that the association between copayment reduction and increased refill adherence to the other drugs was unconfounded.

The studies by McClellan, Jackson, Brookhart, Patrick, and Jena were represented by the second group in which the treatment was *unexpectedly* associated with the negative control outcomes. The study by McClellan, McNeil, and Newhouse (1994) evaluated the impact of cardiac catheterization on mortality within 4 years following an AMI. They utilized mortality at 1 day post-AMI as a negative control outcome, hypothesizing that effects appearing on the first day post-AMI were unlikely to be related to catheterization and revascularization but to other aspects of treatment that correlated with the procedures. Unlike the null finding of a similar “early effect of treatment” from the Mauri study, they found a significant difference in 1-day mortality, indicating residual confounding.

Next, Jackson et al. (2006) evaluated the impact of influenza vaccination on mortality and influenza/pneumonia-related hospitalizations during influenza season. They used hospitalizations for injury or trauma during influenza season as their primary negative control outcomes. In a novel use of timing as a negative control outcome, they also used mortality and influenza/pneumonia-related hospitalizations in the pre-influenza season as a negative control outcome because influenza vaccination was not expected to affect these outcomes before influenza season started. They found that influenza vaccination was negatively associated with (i.e., protective against) hospitalizations for injuries and trauma during influenza season and with mortality and influenza/pneumonia-related hospitalizations in the pre-influenza season. These unex-

pected findings suggested that the association between influenza vaccination and mortality and influenza/pneumonia-related hospitalizations during influenza season was likely confounded.

The two papers by Brookhart et al. (2007) and Patrick et al. (2011) used the occurrence of burns, asthma, and gastrointestinal bleeding as negative control outcomes when evaluating the effect of statins and statin adherence. They argued that there is no biologically plausible rationale or causal pathway through which statins would impact the likelihood of these negative control outcomes (Patrick et al. 2011). Thus, significant associations between statins and these negative control outcomes could suggest residual confounding in the relationship between statins and other outcomes of interest (e.g., mortality) in prior studies (Aronow et al. 2001; Stenestrand and Wallentin 2001). They found significant differences in multiple control outcomes, including increased preventative services use (bone mineral density testing, fecal-occult blood tests, mammography, and influenza and pneumonia vaccinations) and clinical outcomes (asthma, burns, falls, fractures, motor vehicle accidents, wounds, gastrointestinal bleeding, skin infections) for patients who were adherent to statins, suggesting that other outcomes were likely confounded via healthy adherer/healthy user bias.

Finally, Jena, Sun, and Goldman (2013) evaluated the incidence of community acquired pneumonia among patients who did and did not use proton pump inhibitors (PPIs). They selected several negative control outcomes that had no biologically plausible relationship with PPIs, including osteoarthritis, chest pain, urinary tract infections, deep vein thrombosis, skin infections, and rheumatoid arthritis. They found associations between each of the selected outcomes and PPIs, suggesting that there was possible confounding by indication or disease severity that was unaccounted for in their study.

It should be noted that these five papers, in which the treatment was *unexpectedly* associated with the negative control outcomes did not then attempt to statistically adjust for this evidence of unobserved confounding. Instead, each of these papers noted that the treatment effect on the primary outcome was likely biased, and that future work was needed to improve upon these estimates given significant treatment effects/associations in the primary outcome.

Studies Using Control Exposures

An additional approach to improving the internal validity of nonrandomized studies is to use control exposures, which are treatments that are expected to have no effect on the outcome of interest (analogous to a placebo). We

identified three studies in our review that used negative control exposures (Table 2). Zaadstra et al. (2008) evaluated childhood infections that could be possible causes of multiple sclerosis. To address the possibility of recall bias in their patient survey, they used several negative control exposures, including broken arms, concussions, and tonsillectomy. Two of the negative control exposures (concussions and tonsillectomy) were associated with the later development of multiple sclerosis, which they interpreted as evidence of recall bias since neither childhood event was plausibly related to multiple sclerosis.

Dusetzina et al. (2013) used aromatase inhibitor initiators as the negative control exposure when evaluating the effect of an FDA label change targeting drug interaction risks between strong CYP2D6 inhibitor antidepressants and tamoxifen. In this example, the drug interaction risk exists only for tamoxifen-treated patients as aromatase inhibitors are metabolized outside of the CYP2D6 pathway. Changes in the use of strong inhibitor antidepressants (the primary outcome) among tamoxifen users could then be estimated while controlling for broader changes in antidepressant use over time among women using endocrine therapy. The authors observed greater decreases in strong CYP2D6 inhibitor antidepressant use among individuals prescribed tamoxifen, suggesting that the label change related to CYP2D6 risk resulted in selective prescribing of therapies and not general reductions in strong inhibitor antidepressants.

Finally, Rasmussen, Chong, and Alter (2007) evaluated the impact of varying levels of adherence to statins and beta-blockers on long-term mortality following AMI. The authors used adherence to calcium channel blockers as a control exposure as these treatments have no biologically plausible effect on post-AMI mortality. They compared three levels of adherence to each therapy and found patients with intermediate and lower adherence to statins were at increased risk of mortality as compared with patients with high adherence. Similar trends were observed for patients taking beta-blockers. They found no association between adherence to calcium channel blockers and mortality, suggesting that healthy user bias was minimized in their study and the protective effect of high statin adherence against mortality was unlikely to be confounded.

DISCUSSION

Control outcomes and exposures can improve the internal validity of nonexperimental studies because in certain situations they can detect confounding

Table 2: Summary of Published Studies Using Nonequivalent Exposures

Citation	Nonequivalent Outcome	Nonequivalent Exposure	Primary Study Endpoint	Exposure or Treatment	Primary Study Effect Measure	Nonequivalent Outcome Effect	Nonequivalent Exposure Effect	Proposed Source of Bias	Causal Criteria
Zaadsma, Mult Scler. 2008	N/A	Broken arm, concussion, tonsillectomy	Multiple sclerosis	Childhood infections (rubella, chicken pox, mono, measles, mumps)	Odds ratios from logistic regression	N/A	Patients with MS had higher rates of concussion and tonsillectomy	Recall bias	Unstated
Dusetzina, Breast Cancer Res Treat. 2013	N/A	Aromatase inhibitors	Co-prescribing of antidepressants and endocrine therapy	Tamoxifen	Risk ratios from binomial regression using a difference-in differences approach	N/A	Greater decline in strong inhibitor antidepressant use among tamoxifen users than aromatase inhibitor users	Unobserved confounding, temporal changes in antidepressant use	No biological plausibility
Rasmussen, JAMA. 2007	Cancer-related hospital admissions	Calcium channel blockers	Long-term mortality post-MI	Adherence to statins and beta-blockers	Hazard ratio from Cox model	No increase in cancer-related admissions by statin or beta-blocker adherence level	Adherence to calcium channel blockers was not associated with mortality endpoints, as expected	Healthy user bias	No biological plausibility

and selection bias via measurement of whether the treatment effect in the outcome of interest is confounded by unexpected factors (Brookhart et al. 2010a; Lipsitch, Tchetgen Tchetgen, and Cohen 2010). This direct assessment is an appealing feature of using control outcomes as measurement tools because they may be easier to implement in practice than statistical tools like instrumental variables. Of the 11 peer-reviewed studies we summarize that incorporated control outcomes or exposures to evaluate bias, six of them found a null association between treatment and the selected negative control outcomes or exposure. The null association where expected provided greater confidence in the validity of significant associations of interest (also where expected). However, five studies found unexpected associations for the selected control outcomes or exposures, which suggest that the primary association of interest may be biased.

Inclusion of control outcomes in nonrandomized studies is also appealing when there is interest in making causal claims because control outcomes impose a higher threshold for rejecting the null hypothesis than exists for nonrandomized studies without them. In nonrandomized studies that include control outcomes or exposures, the null hypothesis of no association between treatment and primary outcome of interest can only be rejected if treatment is significant in the outcome equation of interest and one fails to reject the null in a control outcome or exposure. Control outcomes can also be used to assess the potential for recall bias in surveys, although we found few examples of this application. Given their utility, control outcomes have been recommended in guidance on the design of prospective nonrandomized studies (Berger et al. 2012, 2014).

Proposed Framework for Identifying Control Exposures or Outcomes

Given the theoretical utility of these measurement tools, one must identify a potentially valid control outcome or control exposure a priori. This requires subject-matter knowledge and consideration of the causal criteria in the specific analysis under consideration. Application of formal criteria to this task can aid in identification of valid control outcomes or control exposures, particularly when there is an interest in estimating the causal association between an exposure and primary outcome.

Some of the best known epidemiologic criteria for evaluating the cause and effect relationship were summarized by Hill (1965). The criteria outlined for identifying a causal relationship between an exposure and outcome can be readily adapted and used as a framework for identifying control exposures or

outcomes. In particular, Hill's criteria for (biological) plausibility, temporality, specificity, consistency, and analogy seem most appropriate for this application (Table 3). For medical product evaluations, the criteria of temporality and biological plausibility may be the most familiar to researchers.

We found that these two criteria were most commonly employed by researchers for justification for the selection of controls, with the Jackson study providing a particularly thoughtful application of the temporality and biological plausibility criteria. If vaccination is expected to reduce influenza/pneumonia-related hospitalizations only during influenza season, then one might reasonably assume that vaccination should have no impact on these hospitalizations in the pre-influenza season due to biological implausibility and timing criteria. As a further robustness check for unobserved confounding, Jackson and colleagues also assessed a second set of outcomes that did not have biological plausibility for association with vaccination: hospitalizations for injury or trauma during influenza season. The only plausible way in which influenza/pneumonia vaccination could affect hospitalizations for injury or trauma would be through unobserved confounding. However, Jackson and col-

Table 3: Bradford-Hill Criteria to Consider When Identifying Control Outcomes or Exposures

<i>Criteria</i>	<i>Description</i>
Strength	The larger the association, the more likely that the association is causal. However, a small association does not mean that there is not a causal effect as expected treatment effects in medicine are often small.
Consistency	Findings have been replicated by other researchers and/or in different samples.
Specificity	The more specific an association between a factor and an effect is, the greater the probability of a causal relationship. Causation is likely if the association is identified under specific circumstances and that there is no other likely explanation.
Temporality	Cause precedes effect; if there is an expected delay between the cause and expected effect, then the effect must occur after that delay.
Biological Gradient	For exposures that follow a dose-response curve, greater exposure should generally lead to greater incidence of the effect. In some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.
Plausibility	A plausible biological mechanism between cause and effect is helpful.
Coherence	Coherence between epidemiological and laboratory findings increases the likelihood of an effect. Results need to be interpreted in light of existing data and known facts of the natural history and disease biology.
Experiment	Reducing exposure to the risk factor reduces the likelihood of the outcome.
Analogy	Exposures with similar mechanisms of action may result in similar outcomes.

leagues did not then re-analyze their primary outcomes and these control outcomes after accounting for unobserved confounding via covariate adjustment or statistical methods. Only through this additional step would it be possible to know whether the association between influenza/pneumonia vaccination and the control outcomes converged to the null after initial detection of unobserved confounding. Nonetheless, the analysis by Jackson was a creative application of the timing and biological plausibility criteria for identification of control outcomes.

Although not all criteria are appropriate for control identification in every setting, this framework may be a useful starting place for identifying controls. Importantly, proper control selection requires that the researcher understand the mechanism of potential confounding and that he or she selects a control that is subject to the same confounding mechanism but is not impacted by the treatment of interest. Suppose a researcher is studying the effect of statins on mortality following MI and is concerned that this treatment effect may be prone to healthy user bias. To test this possibility, the researcher should select a control outcome that has been associated with patient health behaviors but that is not influenced by statin use. Selecting a control outcome or exposure that is unrelated to patient health behaviors (for example, kidney stones or diverticulitis) (Dormuth et al. 2009) does not provide a robust test because these control outcomes are not likely impacted by the confounding mechanism (healthy user bias) that is potentially biasing the effect of statins on post-MI mortality. Thus, a null finding of statins on these control outcomes would provide the researcher with a false sense of confidence that the effect of statins on post-MI mortality was not subject to healthy user bias. Clinical judgment—particularly knowledge of the disease process—will often be critical for selection of control outcomes or control exposures that serve their stated purpose.

Recommendations for Reporting

We suggest that researchers using control outcomes and exposures should explicitly identify these measures in the methods sections of their manuscripts and include a rationale for their inclusion. Further, researchers should report results of all *a priori* selected controls, regardless of their consistency or lack thereof with the investigators' hypothesis.

Next, we recommend that terminology be standardized to improve the recognition of these measures. Terms that have been previously employed to describe these measures include "falsification endpoints," "non-equivalent

controls/exposures,” and “control outcomes/exposures.” We recommend the use of “control outcomes” or “control exposures” with specification of the proposed direction of the effect (e.g., negative or positive). Further, we recommend the creation of a MeSH term for improving the identification of the use of these measurement tools in nonrandomized studies. This would benefit researchers since these measurement tools are likely to be used increasingly over time to improve the rigor and internal validity of comparative effectiveness research. Easier identification of studies using control outcomes and exposures will allow for further evaluation of the adoption of these tools and assessments of the quality of reporting of their use.

We identified a limited number of papers for inclusion in this review since these measures tend to be reported within the manuscript text and not in fields used for keyword identification within PubMed. As a result, we may have excluded other papers that faithfully applied control outcomes or control exposures to great effect. As these measurement tools become more widely used and easier to systematically identify, it will be useful to examine ways in which residual confounding has been addressed upon its identification in studies that find evidence of confounding via these tools.

CONCLUSION

Control outcomes and exposures are important tools for evaluating the internal validity of nonrandomized study findings. Routine use of controls will increase the rigor of studies by helping to identify studies where residual confounding is a concern (when control findings are not consistent with the researcher’s hypothesis) or they may act as confirmation of study findings (when control findings are consistent with the researcher’s hypothesis). Their use will create a higher threshold for rejecting the null hypothesis in an association of interest by requiring rejection of the null in the outcome equation of interest and failure to reject the null in a control outcome or exposure. Given the ongoing concern about clinical and policy inferences from nonrandomized studies, it seems reasonable to more widely employ these measurement tools.

This paper was developed for researchers conducting nonrandomized comparative effectiveness research who are unfamiliar with these nonequivalent outcomes. As these studies undergo continued scrutiny and investigators need to increase the validity of their nonrandomized studies, the use of tools to increase validity will be important.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This work was supported by the Office of Research and Development, Health Services Research and Development Service, Department of Veterans Affairs. Dr. Dusetzina is supported by the NIH Building Interdisciplinary Research Careers in Women's Health (BIRCWH) K12 Program and the North Carolina Translational and Clinical Sciences Institute (UL1TR001111). Dr. Brookhart (MAB) receives investigator-initiated research funding from the National Institutes of Health (R01 AG042845, R21 HD080214, R01 AG023178) and through contracts with the Agency for Healthcare Research and Quality's DEcIDE program and the Patient Centered Outcomes Research Institute. Dr. Maciejewski was supported by a Research Career Scientist award from the Department of Veterans Affairs (RCS 10-391) and received investigator-initiated research funding from the Department of Veterans Affairs, the National Institutes of Health (R01 DK097165), the Agency for Healthcare Research and Quality (R01 HS023085, R01 HS023099), and the Robert Wood Johnson Health Care Financing and Organization Initiative (70922). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the University of North Carolina at Chapel Hill, the Department of Veteran Affairs, or Duke University. MAB has received research support from Amgen for unrelated projects and has served as an unpaid scientific advisor for Amgen, Merck, Pfizer, and GSK. He receives consulting fees from World Health Information Science Consultants and RxAnte, Inc.

Disclosure: None.

Disclaimer: None.

REFERENCES

- Aronow, H. D., E. J. Topol, M. T. Roe, P. L. Houghtaling, K. E. Wolski, A. M. Lincoff, R. A. Harrington, R. M. Califf, E. M. Ohman, N. S. Kleiman, M. Keltai, R. G. Wilcox, A. Vahanian, P. W. Armstrong, and M. S. Lauer. 2001. "Effect of Lipid-Lowering Therapy on Early Mortality after Acute Coronary Syndromes: An Observational Study." *Lancet* 357 (9262): 1063–8.
- Basker, E. 2005. "Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion." *Review of Economics and Statistics* 87 (1): 174–83.
- Benson, K., and A. J. Hartz. 2000. "A Comparison of Observational Studies and Randomized, Controlled Trials." *New England Journal of Medicine* 342 (25): 1878–86.

- Berger, M. L., N. Dreyer, F. Anderson, A. Towse, A. Sedrakyan, and S. L. Normand. 2012. "Prospective Observational Studies to Assess Comparative Effectiveness: The ISPOR Good Research Practices Task Force Report." *Value Health* 15 (2): 217–30.
- Berger, M. L., B. C. Martin, D. Husereau, K. Worley, J. D. Allen, W. Yang, N. C. Quon, C. D. Mullins, K. H. Kahler, and W. Crown. 2014. "A Questionnaire to Assess the Relevance and Credibility of Observational Studies to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report." *Value Health* 17 (2): 143–56.
- Brookhart, M. A., A. R. Patrick, C. Dormuth, J. Avorn, W. Shrank, S. M. Cadarette, and D. H. Solomon. 2007. "Adherence to Lipid-Lowering Therapy and the Use of Preventive Health Services: An Investigation of the Healthy User Effect." *American Journal of Epidemiology* 166 (3): 348–54.
- Brookhart, M. A., A. R. Patrick, W. H. Shrank, and C. R. Dormuth. 2010a. "Validating Studies of Adherence through the Use of Control Outcomes and Exposures." *American Journal of Hypertension* 23 (2): 110.
- Brookhart, M. A., T. Sturmer, R. J. Glynn, J. Rassen, and S. Schneeweiss. 2010b. "Confounding Control in Healthcare Database Research: Challenges and Potential Approaches." *Medical Care* 48 (6 Suppl): S114–20.
- Concato, J., N. Shah, and R. I. Horwitz. 2000. "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs." *New England Journal of Medicine* 342 (25): 1887–92.
- Dormuth, C. R., A. R. Patrick, W. H. Shrank, J. M. Wright, R. J. Glynn, J. Sutherland, and M. A. Brookhart. 2009. "Statin Adherence and Risk of Accidents: A Cautionary Tale." *Circulation* 119 (15): 2051–7.
- Dusetzina, S. B., G. C. Alexander, R. A. Freedman, H. A. Huskamp, and N. L. Keating. 2013. "Trends in Co-Prescribing of Antidepressants and Tamoxifen among Women with Breast Cancer, 2004-2010." *Breast Cancer Research and Treatment* 137 (1): 285–96.
- Furlan, A. D., G. Tomlinson, A. A. Jadad, and C. Bombardier. 2008. "Methodological Quality and Homogeneity Influenced Agreement between Randomized Trials and Nonrandomized Studies of the Same Intervention for Back Pain." *Journal of Clinical Epidemiology* 61 (3): 209–31.
- Hernan, M. A., S. Hernandez-Diaz, M. M. Werler, and A. A. Mitchell. 2002. "Causal Knowledge as a Prerequisite for Confounding Evaluation: An Application to Birth Defects Epidemiology." *American Journal of Epidemiology* 155 (2): 176–84.
- Hernan, M. A., A. Alonso, R. Logan, F. Grodstein, K. B. Michels, W. C. Willett, J. E. Manson, and J. M. Robins. 2008. "Observational Studies Analyzed Like Randomized Experiments: An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease." *Epidemiology* 19 (6): 766–79.
- Hill, A. B. 1965. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58: 295–300.
- Holmlund, H., S. McNally, and M. Viarengo. 2010. "Does Money Matter for Schools?" *Economics of Education Review* 29 (6): 1154–64.

- Ioannidis, J. P. 2005. "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research." *Journal of the American Medical Association* 294 (2): 218–28.
- Ioannidis, J. P., A. B. Haidich, M. Pappa, N. Pantazis, S. I. Kokori, M. G. Tektonidou, D. G. Contopoulos-Ioannidis, and J. Lau. 2001. "Comparison of Evidence of Treatment Effects in Randomized and Nonrandomized Studies." *Journal of the American Medical Association* 286 (7): 821–30.
- Jackson, L. A., M. L. Jackson, J. C. Nelson, K. M. Neuzil, and N. S. Weiss. 2006. "Evidence of Bias in Estimates of Influenza Vaccine Effectiveness in Seniors." *International Journal of Epidemiology* 35 (2): 337–44.
- Jena, A. B., E. Sun, and D. P. Goldman. 2013. "Confounding in the Association of Proton Pump Inhibitor Use with Risk of Community-Acquired Pneumonia." *Journal of General Internal Medicine* 28 (2): 223–30.
- Lipsitch, M., E. Tchetgen Tchetgen, and T. Cohen. 2010. "Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies." *Epidemiology* 21 (3): 383–8.
- Maciejewski, M. L., J. F. Farley, J. Parker, and D. Wansink. 2010. "Copayment Reductions Generate Greater Medication Adherence in Targeted Patients." *Health Affairs (Millwood)* 29 (11): 2002–8.
- Manson, J. E., J. Hsia, K. C. Johnson, J. E. Rossouw, A. R. Assaf, N. L. Lasser, M. Trevisan, H. R. Black, S. R. Heckbert, R. Detrano, O. L. Strickland, N. D. Wong, J. R. Crouse, E. Stein, and M. Cushman. 2003. "Estrogen Plus Progestin and the Risk of Coronary Heart Disease." *New England Journal of Medicine* 349 (6): 523–34.
- Mauri, L., T. S. Silbaugh, R. E. Wolf, K. Zelevinsky, A. Lovett, Z. Zhou, F. S. Resnic, and S. L. Normand. 2008. "Long-Term Clinical Outcomes after Drug-Eluting and Bare-Metal Stenting in Massachusetts." *Circulation* 118 (18): 1817–27.
- McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272 (11): 859–66.
- Patrick, A. R., W. H. Shrank, R. J. Glynn, D. H. Solomon, C. R. Dormuth, J. Avorn, S. M. Cadarette, H. Mogun, and M. A. Brookhart. 2011. "The Association between Statin Use and Outcomes Potentially Attributable to an Unhealthy Lifestyle in Older Adults." *Value Health* 14 (4): 513–20.
- Prasad, V., and A. B. Jena. 2013. "Prespecified Falsification end Points: Can They Validate True Observational Associations?" *Journal of the American Medical Association* 309 (3): 241–2.
- Rasmussen, J. N., A. Chong, and D. A. Alter. 2007. "Relationship between Adherence to Evidence-Based Pharmacotherapy and Long-Term Mortality after Acute Myocardial Infarction." *Journal of the American Medical Association* 297 (2): 177–86.
- Ray, W. A. 2003. "Evaluating Medication Effects Outside of Clinical Trials: New-User Designs." *American Journal of Epidemiology* 158 (9): 915–20.
- Redelmeier, D., D. Scales, and A. Kopp. 2005. "Beta Blockers for Elective Surgery in Elderly Patients: Population Based, Retrospective Cohort Study." *British Medical Journal* 331 (7522): 932.

- Robins, J. M. 2001. "Data, Design, and Background Knowledge in Etiologic Inference." *Epidemiology* 12 (3): 313–20.
- Rothstein, J. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125 (1): 175–214.
- Schneeweiss, S., A. R. Patrick, T. Sturmer, M. A. Brookhart, J. Avorn, M. Maclure, K. J. Rothman, and R. J. Glynn. 2007. "Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison with Randomized Trial Results." *Medical Care* 45 (10 Supl 2): 131–42.
- Stenestrand, U., and L. Wallentin. 2001. "Early Statin Treatment Following Acute Myocardial Infarction and 1-Year Survival." *Journal of the American Medical Association* 285 (4): 430–6.
- Zaadstra, B. M., A. M. Chorus, S. van Buuren, H. Kalsbeek, and J. M. van Noort. 2008. "Selective Association of Multiple Sclerosis with Infectious Mononucleosis." *Multiple Sclerosis* 14 (3): 307–13.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.