Epidemiologic Reviews

Vol. 43, 2021
https://doi.org/10.1093/epirev/mxab008
Advance Access publication:
June 8, 2021

Review

Is the Way Forward to Step Back? Documenting the Frequency With Which Study Goals Are Misaligned With Study Methods and Interpretations in the Epidemiologic Literature

Katrina L. Kezios*

* Correspondence to Dr. Katrina L. Kezios, Room 504, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 W. 168th Street, New York, New York, 10032 (e-mail: klk2131@cumc.columbia.edu)

Accepted for publication September 14, 2021.

In any research study, there is an underlying process that should begin with a clear articulation of the study's goal. The study's goal drives this process; it determines many study features, including the estimand of interest, the analytic approaches that can be used to estimate it, and which coefficients, if any, should be interpreted. Misalignment can occur in this process when analytic approaches and/or interpretations do not match the study's goal; misalignment is potentially more likely to arise when study goals are ambiguously framed. In this study, misalignment in the observational epidemiologic literature was documented and how the framing of study goals contributes to misalignment was explored. The following 2 misalignments were examined: use of an inappropriate variable selection approach for the goal (a "goal–methods" misalignment) and interpretation of coefficients of variables for which causal considerations were not made (e.g., Table 2 Fallacy, a "goal–interpretation" misalignment). A random sample of 100 articles published 2014–2018 in the top 5 general epidemiology journals were reviewed. Most reviewed studies were causal, with either explicitly stated ($n = 13$; 13%) or associational-framed ($n = 71$; 69%) aims. Full alignment of goal–methods–interpretations was infrequent ($n = 9$; 9%), although clearly causal studies ($n = 5$ of 13; 38%) were more often fully aligned than were seemingly causal ones ($n = 3$ of 71; 4%). Goal–methods misalignments were common ($n = 34$ of 103; 33%), but most frequently, methods were insufficiently reported to draw conclusions ($n = 47$; 46%). Goal–interpretations misalignments occurred in 31% ($n = 32$) of the studies and occurred less often when the methods were aligned ($n = 2$; 2%) compared with when the methods were misaligned ($n = 13$; 13%).

associational; causal; directed acyclic graph; generalized linear model; literature review; misalignment; risk factor; study goals

Abbreviations: GLM, generalized linear model; DAG, directed acyclic graph.

INTRODUCTION

Recent discussion in the epidemiologic methods and teaching literature centers around the importance of clearly stating study goals, disentangling the goal of causation from prediction, and clarifying the statistical tools that can address each goal (1–8). This discussion illuminates ways in which mismatches can occur between study goals, methods, and interpretations, which has been synthesized in the present report into the concept of "misalignment." Although misalignments can occur and may cause problems, their pervasiveness has not been documented in the observational epidemiologic literature.

The concept of misalignment

In any research study, there is an underlying research process that should begin with a clear articulation of the study's goal (via statement of the research question or study aims). The study's goal drives this process; it determines many study features, including the estimand of interest, the

Table 1. Summary of the Essential Conceptual Differences in the Analytic Approach for Causation and Prediction in Epidemiologic Studies Published 2014–2018[a]

Essential Difference In	Causation	Prediction[b]
Overarching goal	Identify causes, estimate their effects, and/or understand how they work; counterfactual thinking[c]	To accurately estimate the value of a dependent variable from ≥1 independent variables, also called predictors
Parameter estimate of interest	$\hat{\beta}_X$; the effect estimate for the main exposure of interest[d]	$\hat{Y}$; the predicted outcome (e.g., predicted risk of disease)
Exchangeability required?	Yes	No (not applicable)
Basic variable selection approach	Essential goal: satisfy exchangeability Variables play roles defined in relation to the exposure–disease relationship Variable relationships often are explicated through use of DAGs Variable selection into regression model is based on role in relation to X→Y of interest	Essential goal: create well-fitting models that accurately predict the value of the outcome (e.g., risk of disease) No main exposure Variables have no roles (e.g., they are all predictors of disease) Variables have equal status for selection into prediction model, so selection can be solely data driven

Abbreviation: DAG, directed acyclic graph.

[a] Although developed independently, this summary overlaps substantially with that extensively detailed by Arnold et al. (21) suggesting these are, indeed, key essential differences.

[b] To fully differentiate causation and prediction, here prediction refers only to "passive prediction" ((9), p. 92). Counterfactual or "active prediction"((27), p.44) considers how a change (e.g., manipulation) in 1 variable produces a change in another (prediction of a causal effect, or β) and, therefore, active prediction is considered in this study to be a causal goal.

[c] Counterfactual or active prediction might have a goal of predicting the effects of interventions (prediction of intervention effects).

[d] Generally speaking, for a causal analysis, the estimate of the total effect of an exposure on an outcome is of interest (of course, estimates of other causal estimands may be of interest; e.g., direct effects).

analytic approaches that can be used to estimate it, and how findings should be interpreted (9). Different study goals drive different research processes and, critically, the analytic approaches and interpretations appropriate for one goal may be inappropriate for another (9, 10). Thus, alignment among study goals, methods, and interpretations is important for obtaining valid study inferences.

It follows, then, that misalignments are places in the research process where decisions are made that are inconsistent with the study's goal (Web Figure 1) (available at https://doi.org/10.1093/aje/mxab008). Misalignments are concerning, because when an inappropriate approach is used for a given study goal, it could lead to bias in parameter estimates and contribute to incorrect interpretations of findings (11).

Two primary goals of epidemiologic studies are causation and prediction (6, 10, 12), and these distinct goals lead to fundamentally different decisions in the research process (6, 9, 10, 13–18). In the following sections, essential differences in causal and predictive research processes are briefly laid out (and summarized in Table 1), and 2 ways in which misalignments occur for these goals in the epidemiologic literature are discussed. Of note, where applicable, differences are discussed in the context of generalized linear models (GLMs). GLMs can be used to answer questions of both causation and prediction (6, 10, 17, 19–21), and although this property is not unique to GLMs (21, 22), GLMs are the predominant tool used to analyze epidemiologic data. For these reasons, this discussion focuses on GLMs.

Essential differences in the research process for causation versus prediction

Causation and prediction are distinct study goals that address different questions (Table 1, row 1). Broadly, in observational studies with causal goals, the aims are to identify causes, estimate their effects (6), and/or understand how they work (23, 24). For studies with causal goals, researchers might ask 1) "is X a cause of Y?" or "what would have happened if X had not occurred?" ((25, p.33); or 2) "how, when, in whom, or why did X cause Y?" (24). Causal questions can also be interventionist, asking 3) "what would be the causal effect of manipulating X on Y?" or "what would happen if we set X to x?" ((25, p.32). In contrast, the aim for passive prediction (26, 27), or prediction based simply on correlations (hereafter, "prediction" (see Table 1 footnote)), is to accurately estimate the value of a dependent variable from 1 or more independent variables, also called "predictors." A predictor is any variable that provides information about another (28), regardless of whether the variables have a causal relationship.

For a causal question, a researcher aims to isolate the causal effect of a particular exposure from the effects of variables that create noncausal associations between the exposure and the outcome (Table 1, row 2). In practice, to isolate this causal effect, the assumption of exchangeability (among others, depending on the causal parameter of interest) must be satisfied, and a primary way that researchers think about achieving exchangeability is through adjustment for variables that are common causes of the exposure and the outcome (e.g., confounders), although nonexchangeability can arise from other sources (Table 1, row 3). Tools like directed acyclic graphs (DAGs) are often used to aid in the identification or selection of the minimal set of variables whose control is necessary to satisfy exchangeability (Table 1, row 4). A common approach to adjustment is to condition on these variables in a GLM (29).

In contrast, for questions of prediction, exchangeability is not a requirement (7, 16, 17) and, therefore, adjustment for confounding is irrelevant (Table 1, row 3). Exchangeability is a concept that only applies when considering the difference between an association and a causal effect, and for questions of prediction, the analytic focus is not on the isolation of a particular exposure–disease relationship (Table 1, row 2). Rather, the focus is about identifying or selecting the set of variables that best predict the outcome (30), which can be determined solely by a chosen statistical criterion (e.g., P value, R^2) (Table 1, row 4).

Causal analyses must consider exchangeability (among other assumptions), because the ultimate goal of the analysis is to make causal inferences; causal assumptions must be met to give findings a causal interpretation (6). In practice, variables' associations are often interpreted using language that suggests a causal relationship (specifically, a total effect (31)) between them and the outcome (18) or are discussed in some other manner that implies interest in causality (e.g., discussing policy relevance of finding, etiologic mechanisms (32, 33)). These types of interpretations should only be made for variables for which causal considerations are made in the research process, and these considerations (e.g., exchangeability) are typically only made for the main exposure in a causal model (34). Therefore, for a causal model, it is typically appropriate to interpret just the main exposure coefficient ($\hat{\beta}_x$), which is the best estimate of the causal effect of interest. In contrast, for a prediction model, interpreting predictor coefficients as discussed (i.e., in a manner that implies etiology or causality (17)) should be avoided, because predictors are not selected into the model on the basis of causal considerations (16, 17, 21, 31, 34, 35).

Misalignment in the epidemiologic literature

One way the issue of misalignment has been discussed in the epidemiologic literature, where causal questions predominate (12, 36–45), is the misuse of variable selection approaches appropriate for prediction in analyses meant to answer (explicitly or implicitly) causal questions (6, 9, 20, 34, 41, 46, 47). Because of the different motivations behind variable selection for causal versus predictive analyses (Table 1), using a predictive variable selection approach that is not centered around exchangeability to answer a causal question could result in misestimated exposure effects, for instance, via inclusion of variables whose adjustment could create bias (e.g., mediators, colliders) (Web Figure 1).

Another way that misalignment appears in the epidemiologic literature is when coefficients are interpreted for variables in regression models that are not the target of a causal analysis (i.e., variables for which exchangeability was not considered, such as confounders) (Web Figure 1). This was originally defined in a causal context as the Table 2 Fallacy (31), although this label has since been more broadly applied (11, 48, 49). If the actual, but unacknowledged, goal, as implied or revealed by interpretations (13, 32, 33), is to draw causal inferences for these variables, then causal assumptions must be considered (20); otherwise, the interpretation of these coefficients as effects is likely unwarranted. For example, for a causal study, although DAGs contain all common causes of any 2 variables (6), adjustment is required only for the minimally sufficient set of variables necessary to satisfy exchangeability for the exposure. This makes it likely that the associations between these adjustment variables and the outcome in a regression model are themselves confounded (34). Furthermore, for a confounder (which is included in the model with its mediator, the exposure), for example, the association would represent, at best, an unconfounded direct effect. In this scenario, the interpretation of the confounder coefficient as a total effect could lead to inappropriate study conclusions (11, 48–50).

Could unclear goals contribute to misalignment?

In practice, study goals are often framed using associational language common to both goals (e.g., risk factor) (14, 51) rather than being explicitly declared, particularly if that goal is causal (8, 13, 33). Ambiguously framed research goals could create problems because vague goals may make it hard to ensure an appropriate research process, which depends on the goal, is followed. Compounding this, conventionally, the same statistical tools (e.g., GLMs) are used to examine causal and predictive questions. Because of this overlap, an unclear or ambiguously framed study goal may increase the likelihood that the methods of one approach are inappropriately used to answer questions of the other.

Objectives

Although it is recognized that these misalignments can occur (6, 9, 20, 31, 34, 41, 46, 47), their pervasiveness has not been examined; it is also unclear if and how the statement of study goals contributes to their occurrence. Thus, the primary goals of this study were 2-fold: 1) to document the frequency and framing of different study goals in the observational epidemiologic literature, and 2) to document the frequency with which misalignments arise in these studies, specifically 1) use of an inappropriate variable selection approach for the goal (a goal–methods misalignment) and 2) interpretation of coefficients of variables for which causal considerations were not made (e.g., Table 2 Fallacy, a goal–interpretation misalignment).

METHODS

The article selection and data abstraction processes are described in the following section. Web Appendix 1 and Web Tables 1–4 include details and results from pilot studies performed to evaluate the reliability of both processes.

Table 2. Frequency of Goal–Methods (Mis)alignments in the Total Sample of Studies (n = 103) and Stratified by Each Study Goal, 2014–2018

Study Goal	Total Sample		Goal–Methods (Mis)alignment: Methods Aligned		Methods Misaligned With Goal		Methods Insufficiently Reported to Draw Conclusions		Goal Unclear, Method Alignment Unclear	
	No.	%[a]	No.	%[b]	No.	%[b]	No.	%[b]	No.	%[b]
Clearly descriptive	2	2	0	0	2	100	0	0	0	0
Seemingly causal[c]	71	69	4	6	27	38	40	56	0	0
Clearly causal	13	13	5	38	5	38	3	23	0	0
Clearly predictive	2	2	2	100	0	0	0	0	0	0
Goal unclear	15	15	n/a	n/a	n/a	n/a	4[d]	27	11[e]	73
Total	103[f]	100	11	11	34	33	47	46	11	11

Abbreviation: n/a, not applicable.
[a] Percentages may sum to 99 or 101 due to rounding.
[b] Percentage of the total number of studies including a given goal.
[c] Study goal is framed using associational or implicitly causal language but seeming interest is causation (e.g., a specific exposure–disease (X–Y) relationship is of primary interest, covariate adjustment is performed).
[d] In 4 of the 15 studies with unclear goals, methods were insufficiently reported for conclusions to be drawn even if the goal had been clear.
[e] Among these 11 studies (where methods were sufficiently reported, but because the goal was unclear, methods were unclear): 0 used a clearly causal variable selection approach, 1 used a variable selection method based on associations of the covariates with X and Y, and 10 selected covariates based on their associations with only Y.
[f] A paper could contribute >1 aim if 1) the different aims were distinctly stated, 2) multivariable regression analysis was used to address the additional aim, and 3) the additional aim did not overlap and/or was not an extension of the overarching aim of the study (i.e., was not same goal, same methods used, same implications for interpretations).

Article selection

The goal was to randomly select 100 articles from the epidemiologic literature that were observational epidemiologic studies (52) with a substantive research question examining a health outcome measured at the individual level and that included variable selection in a GLM as a component of the analysis. Importantly, because in this analysis a study's variable selection approach determines misalignment, the study had to be 1) observational (so that the causal studies captured dealt with confounding via analysis vs. study design (e.g., randomization)) and 2) perform variable selection among multiple (>1) independent variables. By randomly selecting among articles meeting these criteria, the search was not limited to studies of a particular goal.

Figure 1 displays the process of creating the article database and screening randomly selected articles to obtain a review set of 100. An initial pool of 5,500 articles was identified in PubMed, restricted to observational studies published in English between January 1, 2014, and December 31, 2018, in the following top 5 general epidemiology journals, according to 2018 impact factors: *International Journal of Epidemiology*, *American Journal of Epidemiology*, *Epidemiology*, *European Journal of Epidemiology*, and *Journal of Epidemiology and Community Health*. Also excluded were editorials (e.g., commentaries, letters), short reports, methodological articles, simulation studies, studies of infectious disease dynamics, genetic association studies, ecological studies, systematic reviews and meta-analyses of aggregate data, and studies in which authors controlled for confounding primarily or in part by design rather than analysis (e.g., randomized controlled trials, quasi-experimental studies). These criteria are consistent with those in prior reviews documenting variable selection approaches in observational epidemiologic studies (47, 53).

The 4,121 remaining articles were downloaded into a database, sorted in random order, and hand-screened beginning with the first entry and working down the file one by one until 4 articles per each 5-year category for each of the 5 journals were randomly selected. Additional details of the article screening and selection strategy are provided in Web Appendix 2. This selection strategy allowed review of only as many articles as needed to reach the target sample size while maintaining the random article order. A total of 251 articles were hand-screened to find 100 that were applicable (a 60% exclusion rate; reasons for exclusion are listed in Web Table 5).

Among the 100 reviewed articles, 103 study goals were identified. An article could contribute more than 1 goal if multiple aims were stated and if 1) the aims were distinct, 2) variable selection was performed in both aims, and 3) the aims did not overlap and/or one was not an extension of the other (i.e., same goal, same methods used, same implications for interpretations). These 103 goals (hereafter, the

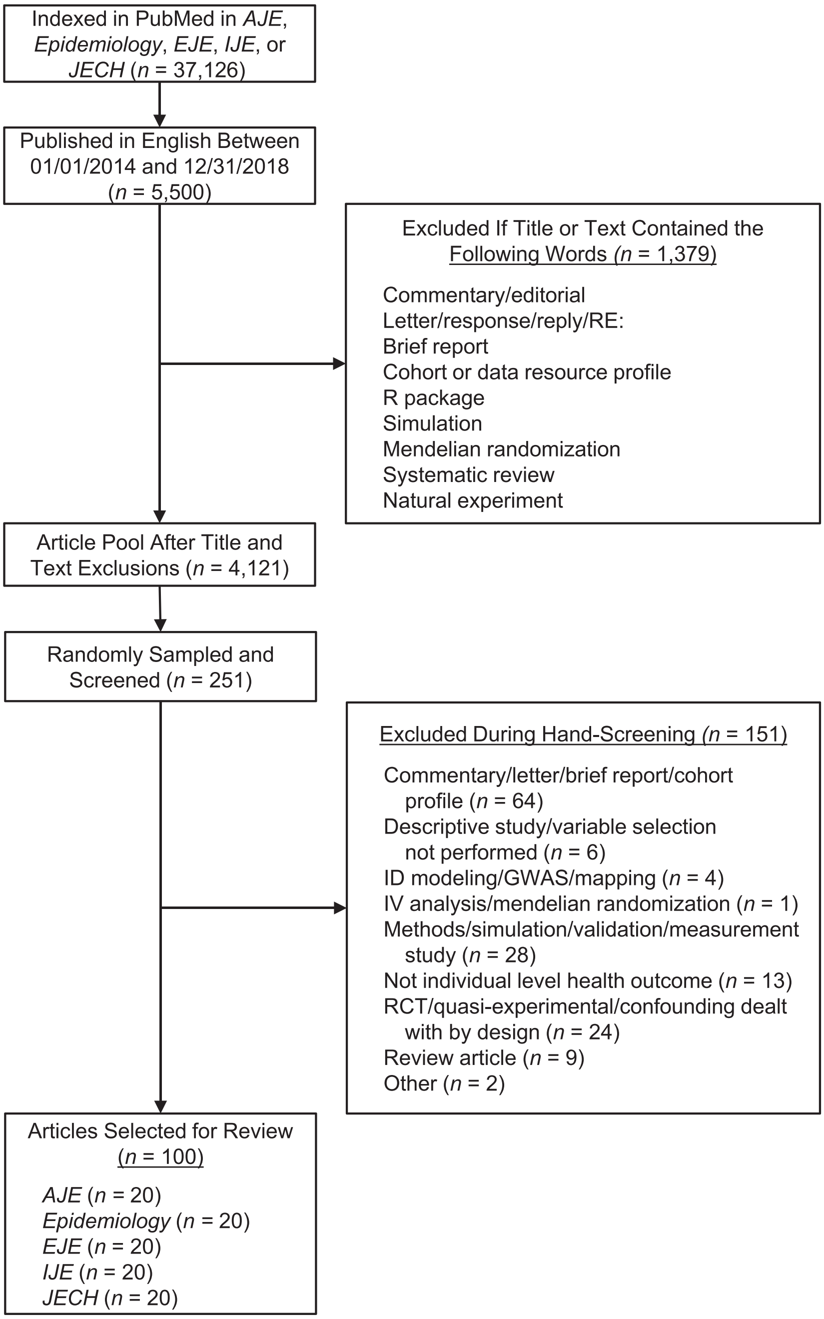

Figure 1. Article exclusion and selection flow chart. The process of creating an article pool from the PubMed search is illustrated. After journal, language, date, and title and text term exclusions, the pool of 4,121 articles was downloaded into a database within which articles were sorted in random order and hand-screened until 4 articles per each 5-year category for each of the 5 journals were selected. To get to the final sample set of 100 articles for review (*n* = 20 articles per journal), 251 articles had to be randomly sampled and screened. This exclusion rate (60%) is nearly identical to that in both pilot studies; for both, 24 articles were screened to obtain a set of 10 articles for review (58% exclusion rate). Web Appendix 1 has more details about the pilot studies; additional details of the article database creation and screening process are provided in Web Appendix 2. AJE, *American Journal of Epidemiology*; EJE, *European Journal of Epidemiology*; IJE, *International Journal of Epidemiology*; JECH, *Journal of Epidemiology and Community Health*; ID, infectious disease; GWAS, genome-wide association study; IV, instrumental variable; RCT, randomized controlled trial.

"total sample" of "studies") constituted the final sample size of the review. In most articles (*n* = 94), authors used GLMs to perform multivariable outcome regression (e.g., linear, logistic); the remaining 9 articles (9%) used a GLM elsewhere (e.g., to estimate propensity scores for a weighted analysis). All articles were deidentified before review.

Measures of interest

Study goals. Because of a priori interest in exploring how the aims of the study drive the research approach, classification of study goals was determined completely by the statement of the study's aim(s). To identify a study's aims, the following definition was used: *A study's aims are usually found in the final paragraph of the Introduction section and are stated in action-oriented language (e.g., "to assess . . . we"; "to examine"; "we estimated"). The aims describe what the authors planned to do in the paper; sometimes hypotheses also are stated.* The following definitions were used to categorize abstracted study aims into different goals.

Clearly causal: A specific exposure–disease (X-Y) relationship is of primary interest and the aim is stated using explicitly causal terms (e.g., *cause*; (causal or treatment) *effect*; *effect*; (hypothetical) *intervention*; *target trial*). Examples include: "to estimate the effect of X on Y"; "does X cause Y?"; and "to identify a potential causal relationship between X and Y."

Clearly predictive: There is no specific exposure of interest and predicting disease or building and/or evaluating a model to predict disease is the explicit goal (e.g., forecasting, prognosis, diagnosis). Examples include: "to develop a [risk] prediction model"; "to build a model to predict Y"; and "to evaluate the performance of a model to predict [risk of] Y."

Association (but seemingly causal): A specific exposure–disease (X-Y) relationship is of primary interest but the aim is not stated using explicitly causal terms. Examples include: "to examine the [independent] association between X and Y"; "is X a risk factor for Y?"; and "are X and Y related/linked?"

Unclear goal: There is no specific exposure of interest and it is unclear if the aim is to predict disease or estimate a causal effect; associational language common to both goals is used to describe aims. Examples include: "to examine determinants of Y"; "early life risk factors for Y"; and "predictors of Y in a population of . . . "

Clearly descriptive: Based on both pilot studies (Web Appendix 1) and the variable selection inclusion criteria (4, 32, 54), descriptive studies were not expected in the article sample. However, in 2 studies, the authors clearly stated that their aim was descriptive; here, there was no main exposure of interest and, distinct from clearly predictive studies, these studies referenced interest in describing the epidemiology of or pattern of certain diseases (with no mention of predictors or building a prediction model).

By these definitions, a seemingly causal goal differs from an unclear goal in that the latter lacks reference to a main exposure–disease relationship of interest. A clearly predictive goal differs from an unclear goal in that the predictive goal expresses explicit interest in predicting outcomes or building a predictive model, whereas an unclear goal is vaguely framed using language shared between goals. Thus, a seemingly causal study goal is possible to define uniquely (associational language, main exposure) whereas a seemingly predictive goal (not uniquely defined) is subsumed under an unclear goal, given overlapping properties (i.e., associational language, no main exposure).

Misalignment 1: classification of goal–methods alignment and misalignment The variable selection approach determined whether methods used in a study were aligned with the study's goal. Two decision points of variable selection were considered: 1) the choice of candidate variables for selection (i.e., the pool of variables; e.g., prior knowledge) and 2) how variables were selected from the pool into regression models (i.e., the selected variables; e.g., 10% change-in-exposure); multiple approaches could be used in a study. The frequency with which no distinction was made between these 2 features was also documented. Information on variable selection procedures was abstracted from articles' Methods section and appendices.

For studies with clearly stated or discernible goals, (mis)alignment categories included: aligned, misaligned, and insufficiently reported. For studies with unclear goals, categories included insufficiently reported and unclear.

Aligned: For a clearly causal or seemingly causal goal, use of a DAG, descriptions consistent with a theory-driven analysis (e.g., common causes, closing backdoor paths), or use of the "traditional approach" were considered aligned variable-selection approaches (6). The traditional approach identifies a pool of variables that are 1) associated with the exposure, 2) associated with the outcome among the unexposed, and 3) not on the causal pathway from exposure to outcome (6). A relaxed version of the traditional approach was used: if either the pool or selected set comprised variables that were associated with X and Y and not on the causal pathway, this counted as aligned. Note that although this approach has its limitations (6) and should not be used as a confounder criterion (it defines, rather, properties of a confounder (12)), because it is commonly taught in epidemiology courses and requires thinking about variable roles, it was considered aligned with a causal goal. For a clearly predictive goal, minimally, an aligned approach focuses on a pool of predictors (no main exposure) from which selection could be based on statistical criteria alone (e.g., P values, Akaike information criterion). Note that this approach would not necessarily lead to a quality prediction model. However, for both goals, only the consistency between the basic variable selection approach and study goal was assessed; the quality of the approach (or variables selected) was not judged.

Misaligned: For a clearly causal or seemingly causal goal, variable selection approaches that lacked consideration for adjustment variables' causal relationship to both X and Y were considered misaligned. There were 2 main instances in which this occurred. In one, either the determination of variable pool or the selection procedure was entirely outcome-focused (e.g., known risk factors of Y; predictors of Y; variables that explain variation in Y). In the other, either the determination of the variable pool or the selection procedure was based on the variables' association with X and Y, but with no mention of excluding variables potentially on the causal pathway. In addition, 6 articles lacked detail about the pool of variables, but in these studies, a data-driven method was used to select adjustment variables; this was also categorized as a misalignment. For a clearly predictive goal, if confounding was discussed or the prediction model included adjustment for confounders, this was considered misaligned. For a clearly descriptive goal, regardless of the variable selection approach, methods were categorized as misaligned because variable selection was performed, which was considered inapplicable for pure description (4, 32, 54).

Insufficiently reported: For all goals, listing the final set(s) of adjustment variables or predictors, without any information about why they were considered or how they were selected, was insufficient to draw conclusions about alignment. In addition, for clearly causal or seemingly causal goals, the following scenarios were categorized as insufficiently reported: 1) reporting use of a DAG-based variable selection approach without either including a figure of the DAG or providing details about decisions underlying its construction, and 2) labelling adjustment variables as confounders or potential confounders, with no additional information about how confounder was defined. In the absence of this information, no assumptions were made about the authors' correct construction and use of DAGs and/or definition of a confounder.

Unclear: When a study had an unclear goal, it was not possible to determine if the methods matched the goal of the study, because the context in which a given variable selection approach is aligned or misaligned depends on the goal. Here, in lieu of determination about alignment, the variable selection procedures used in the study were reported.

Misalignment 2: classification of goal–interpretation alignment and misalignment Goal–interpretation misalignment was determined on the basis of whether the article's authors had any Table 2 Fallacies in the text, defined broadly as presenting in tables and/or figures and/or interpreting or discussing in text (and/or appendices) any $\hat{\beta}$ values for variables other than the main exposure of interest (e.g., confounders in a causal study) (31, 34). Importantly, decisions were only based on whether the correct coefficient was presented or interpreted, not the quality or justifiability of the interpretation (thus, a goal–methods misalignment does not preclude alignment between goals and interpretations).

Also, supplemental analyses were performed to capture information about the implied goal of the study via the language used and discussion of findings (e.g., *effect estimate* reveals interest in causality via effect estimation), even if not explicitly stated so in the study aims. First, use of the terms *effect*, *cause*, and *protect* (including derivatives; e.g., *effect estimate*, *causal association*) in the Methods, Results, and Discussion sections was documented. Any use counted as an implied interest in causality (e.g., "findings suggest the exposure may protect against" counted equally as "a protective effect was observed"). In addition, mention of the following causal concepts in the Discussion section was documented: unmeasured or residual confounding, reverse causation, etiologic explanations for the observed findings, or identification of variables as opportunities for intervention and/or policy recommendations.

Studies with clearly stated or discernible goals were categorized as aligned or misaligned. Studies with unclear goals, were categorized as suggests causal or misaligned.

Aligned: In a clearly causal or seemingly causal study, when only the $\hat{\beta}$ (effect estimate) for the causal estimand of interest was interpreted or discussed (in all cases, minimally, the total effect of the main exposure), the study was categorized as aligned. For a clearly predictive study, reporting and interpreting model performance measures and model estimations of y ($\hat{y}$; e.g., predicted probabilities) constituted aligned interpretations.

Misaligned: For clearly causal or seemingly causal study goals, misalignments occurred when the $\hat{\beta}$ values for adjustment variables (e.g., confounders, even if not labeled as such) were reported in tables and/or interpreted or discussed in text (in addition to or instead of the exposure $\hat{\beta}$). In studies with clearly predictive, clearly descriptive, or unclear goals, misalignments occurred when the $\hat{\beta}$ values for any variable (e.g., predictors, risk factors) were reported in tables and/or interpreted or discussed in text. In contrast, with decisions about goal–methods misalignments, goal–interpretation misalignments could be determined even if goals were unclear, because Table 2 Fallacies are misalignments for all examined study goals.

Suggests causal: When a study had an unclear goal but made no Table 2 Fallacies, the goal implied by the discussion of findings was reported in place of a decision about (mis)alignment. In every case, the study implied an interest in causality via mention of at least 1 of the aforementioned causal concepts.

Statistical analysis

First, the frequency of each study goal, and the overall proportions of goal–methods and goal–interpretation misalignments were documented. (Mis)alignment frequencies were then documented within strata of study goals. In addition, for each goal, the frequency with which causal concepts were discussed and/or causal terms were used was reported. Finally, all possible goal–methods–interpretations patterns were mapped out and their frequencies were reported within goal strata. Supplemental analyses documented misalignment frequencies by journal and year of publication. All analyses are descriptive and were performed in R, version 3.5.3 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Study goals

The primary goal among the 103 reviewed studies was causation (Table 2). As expected, more causal studies had seemingly causal (69%) than clearly causal (13%) goals. Only 2 studies had clearly predictive goals, suggesting this type of study is infrequently published in top epidemiology journals, at least in the period 2014–201 8.

Misalignment 1: goal–methods (mis)alignments:

In the total sample, only 11% of studies' methods were aligned with the goal (Table 2). A third of studies had misaligned methods, and the methods of 46% were insufficiently reported to enable conclusions about alignment; for the remaining 11%, (mis)alignment was unclear because the goal was unclear. Table 3 reports the variable selection

Table 3. Frequency of Different Procedures Used to Determine the Pool of Candidate Variables for Selection and/or Procedures to Select Variables Into the Multivariable Regression Model, in the Total Sample (*n* = 103) and Stratified by Study Goal, 2014–2018

Decisions[a]	Total Sample (*n* = 103)		Stratified by Study Goal									
			Clearly Descriptive (*n* = 2)		Seemingly Causal[b] (*n* = 71)		Clearly Causal (*n* = 13)		Clearly Predictive (*n* = 2)		Goal Unclear (*n* = 15)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Decisions About Candidate Variable Pool[c]												
Prior knowledge[d]	27	26	0	0	20	28	4	31	1	50	2	13
DAG and/or common causes[e]	10	10	0	0	4	6	6	46	0	0	0	0
Traditional approach[f]	1	1	0	0	1	1	0	0	0	0	0	0
Variables associated with X and Y[g]	5	5	0	0	4	6	0	0	0	0	1	7
Variables associated with X or Y[h]	8	7	0	0	7	10	1	8	0	0	0	0
Variables associated with Y[i]	29	28	0	0	15	21	3	23	2	100	9	60
Unspecified[j]	48	47	2	100	38	54	2	15	0	0	6	40
Decisions About Selected Variables[c]												
Unspecified[k]	73	71	2	100	51	72	9	69	0	0	11	73
Change-in-exposure criteria used[l]	11	11	0	0	11	15	0	0	0	0	0	0
Predictor selection method used[m]	12	12	0	0	6	8	2	15	0	0	4	27
Adjust for all[n]	10	10	0	0	5	7	3	23	2	100	0	0

Abbreviation: DAG, directed acyclic graph.

[a] Categories can overlap because a study can report >1 approach was used; thus, column percentages are reported but do not sum to 100% (each column reports the percentage of each approach used out of the total studies for that column).

[b] Study goal is framed using associational or implicitly causal language but seeming interest is causation (e.g., a specific exposure–disease (X-Y) relationship is of primary interest, covariate adjustment is performed).

[c] For 55 studies (53%), no distinction was made between the pool of candidate variables considered for selection and those selected for the regression model. When stratified by study goal, 1 clearly descriptive study (50%), 43 seemingly causal studies (61%), 3 clearly causal studies, (23%) 1 clearly predictive study (50%), and 7 studies with unclear goals (47%) made no distinction between these features.

[d] For example, prior literature, prior knowledge, literature review, a priori, known, suspected, or established (e.g., risk factors, predictors).

[e] For example, a figure of a DAG included and/or pool described causally as common causes, variables needed to close backdoor paths, or variables responsible for treatment assignment.

[f] 1) associated with X and 2) associated with Y, with or without "among the unexposed," and 3) not on the causal pathway.

[g] Mentioned consideration and/or adjustment for variables associated with X and Y but no mention of elimination of variables on the causal pathway.

[h] Mentioned consideration and/or adjustment for variables associated with X or Y and no mention of elimination of variables on the causal pathway.

[i] Mentioned consideration and/or adjustment for variables associated with Y only; no mention of X in discussion of variable adjustment (e.g., predictors of Y, risk factors for Y, variables that explain variance in Y, variables that differed by case–control status); this includes when exposure is the Y in a propensity score model (e.g., predictors of X, n = 1 study of 4 that used propensity scores).

[j] The pool of candidate variables was not specified or no details were provided.

[k] The method by which variables were selected for the model was not specified or no details were provided.

[l] For example, for a 10% change in the exposure estimate, adjustment did not substantially alter estimates, findings, or results.

[m] Outcome-focused (like footnote i), but here the specific selection approach was specified (e.g., forward, backward, or stepwise selection, all covariates with $P < 0.20$ in univariate analysis, significant predictors were retained in the model).

[n] All candidate variables in the pool were adjusted for in the model.

procedures used in these studies (a study could contribute to >1 category). Studies with insufficiently reported methods rarely included information beyond a list of covariates included in the final regression model; when additional details were provided, they were still uninformative (e.g., prior knowledge). The majority of studies (53%) did not

Table 4. Frequency of Goal–Interpretation (Mis)alignments, in the Total Sample of Studies (n = 103) and Stratified by Each Study Goal, 2014–2018

Study Goal	Total Sample		Goal Clear (or Discernible)[a]				Goal Unclear[b]			
			Interpretations Aligned With Goal		Interpretations Misaligned With Goal		Interpretations Suggest Causality of Interest		Interpretations Misaligned Regardless of Goal	
	No.	%[c]	No.	%[d]	No.	%[d]	No.	%[d]	No.	%[d]
Clearly descriptive	2	2	0	0	2	100	n/a	n/a	n/a	n/a
Seemingly causal[e]	71	69	55	77	16	23	n/a	n/a	n/a	n/a
Clearly causal	13	13	9	69	4	31	n/a	n/a	n/a	n/a
Clearly predictive	2	2	1	50	1	50	n/a	n/a	n/a	n/a
Goal unclear	15	15	n/a	n/a	n/a	n/a	6[f]	40	9[f]	60
Total	103	100	65	63	23	22	6	6	9	9

Abbreviation: n/a, not applicable.
[a] N = 88 studies had clear or discernible goals.
[b] N = 15 studies had unclear goals.
[c] Percentages may sum to 99 or 101 due to rounding.
[d] Percentage of the total number of studies citing a given goal.
[e] Study goal is framed using associational or implicitly causal language but seeming interest is causation (e.g., a specific exposure–disease (X-Y) relationship is of primary interest, covariate adjustment is performed).
[f] Of the 15 studies for which goals were unclear, 6 had interpretations that implied a causal goal was of interest (i.e., the interpretations would be aligned with a clearly causal or seemingly causal study goal; e.g., made causal interpretations, references to etiology, and/or intervention or policy recommendations based on findings for XY). For the remaining 9 studies with unclear goals, interpretations were misaligned (i.e., even though the goal was unclear, a misinterpretation was made that would have been misaligned for any goal; i.e., a Table 2 Fallacy).

distinguish between the pool of candidate variables and those selected for the model; when this occurred, best efforts were made to determine if the reported approach, if any, applied more to the determination of the pool of candidate variables or to the selection of variables into models.

(Mis)alignment frequencies were next reported within strata of study goals. However, because most studies had seemingly causal goals, sample size within strata was limited. Among studies with associational-framed but seemingly causal goals, goal–methods alignment occurred in only 6% of articles (Table 2). In contrast, 38% of the 13 studies with clearly causal goals had aligned methods.

Among the 15 studies with unclear goals, 4 had insufficiently reported methods such that an alignment decision would not have been possible regardless (Table 2). The remaining 11 studies had sufficiently reported methods, but because the goal could not be determined, neither could goal–methods (mis)alignment. In most of these 11 articles, authors reported an outcome-focused variable selection approach (Table 3), reflecting misalignment with a causal goal or alignment with prediction.

Misalignment 2: goal–interpretation (mis)alignments

Overall, 63% of studies had goal–interpretation alignment, whereas misalignment occurred in 31% of studies (Table 4). The remaining 6% were studies with unclear goals and, because no Table 2 Fallacy occurred, a (mis)alignment decision was not possible. In the majority of clearly causal (69%) and seemingly causal (77%) studies, only the main exposure was correctly interpreted (Table 4). For the 15 studies with unclear goals, 60% had misaligned interpretations regardless of actual goal (i.e., included a Table 2 Fallacy). In the remaining 40% of studies in which no Table 2 Fallacy occurred, all studies included interpretations that implied that causality was of interest (Tables 4 and 5). Most studies with misaligned interpretations (75%) included both a "Table 2" (i.e., table of all regression model coefficients) and in-text interpretations or discussion of these coefficients (25% simply presented a Table 2 without in-text interpretation).

Goal–methods–interpretations (mis)alignments

Figure 2 aggregates both types of misalignments to show all possible patterns of goal–methods–interpretation (mis)alignments. The proportion of studies from the total sample represented by a given pattern is shown in the penultimate step. For studies with a clearly stated or discernible goal, the breakdown by study goal is reported in the final step. Web Table 6 displays this figure as a table.

The most common pattern was having a clearly stated goal, insufficiently reported methods, and interpretations aligned with goals (34%). Among all studies, 9% had a

Table 5. Frequency With Which Causal Concepts Were Mentioned in the Discussion Section of Articles, Stratified by Study Goal, 2014–2018

Causal Concept Mentioned or Discussed in Discussion Section	Frequency With Which Causal Concept Was Discussed in the Total Sample and by Study Goal[a]											
	Total Sample		Cleary Descriptive		Seemingly Causal		Clearly Causal		Clearly Predictive		Goal Unclear[b]	
	No.	%[c]	No.	%[c]	No.	%[c]	No.	%[c]	No.	%[c]	No.	%[c]
Unmeasured or residual confounding	42	41	0	0	30	42	10	77	0	0	2	13
Reverse causation	19	18	0	0	16	23	1	8	0	0	2	13
Potential etiologic explanations for findings[d]	70	68	2	100	52	73	8	62	0	0	8	53
Identified variables for which policies or interventions should be considered[e]	29	28	1	50	19	27	6	46	0	0	3	20

Click here to enter text.

[a] Sample sizes for each goal were as follows: total sample (n = 103), clearly descriptive (n = 2), seemingly causal (n = 71), clearly causal (n = 13), clearly predictive (n = 2), unclear (n = 15).

[b] In the 6 studies (40%) with unclear goals and in which no interpretational misalignments were made, all discussed ≥1 of the listed causal concepts in their Discussion section.

[c] More than 1 category could be selected per study; percentages are out of the total number of studies that included a given goal.

[d] For example, provided explanations for or pathways or mechanisms through which the findings may have arisen.

[e] Variable simply had to be identified as an opportunity for intervention or policy, but the intervention or policy did not have to be specified.

clearly stated or discernible study goal and no documented misalignments in methods or interpretations. Although numbers are small for comparison, it is worth noting that a higher proportion of clearly causal studies were fully aligned ($n = 5$ of 13; 38%) than seemingly causal studies ($n = 3$ of 71; 4%).

Only 2 studies (2%) with aligned methods had misaligned interpretations. Of the 34 studies with misaligned methods, 20% interpreted findings in line with the study's goal, whereas 13% had goal–interpretation misalignment. These patterns occurred with similar frequency among studies with seemingly causal and clearly causal goals.

Supplemental results

Web Tables 7 and 8 show study goals within strata of journal and year of publication, respectively. Goal–methods (Web Tables 9 and 10) and goal–interpretation (Web Tables 11 and 12) misalignments were also documented by journal and year of publication for all goals combined.

Use of the term *effect* is reported in Web Table 13 (results for *cause* and *protect*, used with much lower frequency, are reported later in this section). Complete avoidance of the term effect was least common in clearly causal studies (8%), more common in seemingly causal studies (27%), and most common (40%) in studies with unclear goals. Authors were also more likely to use the term effect in the Methods section of studies with clearly causal goals (84%) than in studies with seemingly causal (29%) or unclear goals (40%). In contrast, in 44% of seemingly causal studies and 20% of studies with unclear goals, authors avoided the term effect in the Methods section but went on to use it in the Results and/or Discussion section; in only 1 of 13 clearly causal studies (8%) did the authors do the same.

Interestingly, although the word *causal* was rarely used to describe associations (4% of studies; data not shown), the term *protective* was used 4 times as frequently (17% of studies). There appeared to be greater caution against suggesting an association could be causal than implying an exposure could be protective, although protection can be thought of as the causal opposite of causal (55). Of note, in an exploratory analysis, authors described associations as *beneficial* (2%) versus *harmful*, *detrimental*, or *deleterious* (4%) with similar frequency, suggesting the difference in use of causative versus protective is less about the direction of effect and more about specific avoidance of the word *cause* (13).

DISCUSSION

A total of 100 articles were reviewed from the observational epidemiologic literature to document the frequency of different study goals, whether they were clearly stated, and where misalignments arose in these studies. In most of the reviewed articles, authors stated either clearly or seemingly causal goals and, in 38% of the articles ($n = 32$ of 84), authors used a misaligned variable selection approach that, in the vast majority of cases, was entirely outcome focused. This outcome-focused approach risks overadjustment for mediators and/or colliders, potentially biasing estimates of total effects. Among studies with unclear goals and sufficiently reported methods, in 91% (n = 10 of 11) (Table 3), authors used an outcome-focused variable selection approach, even

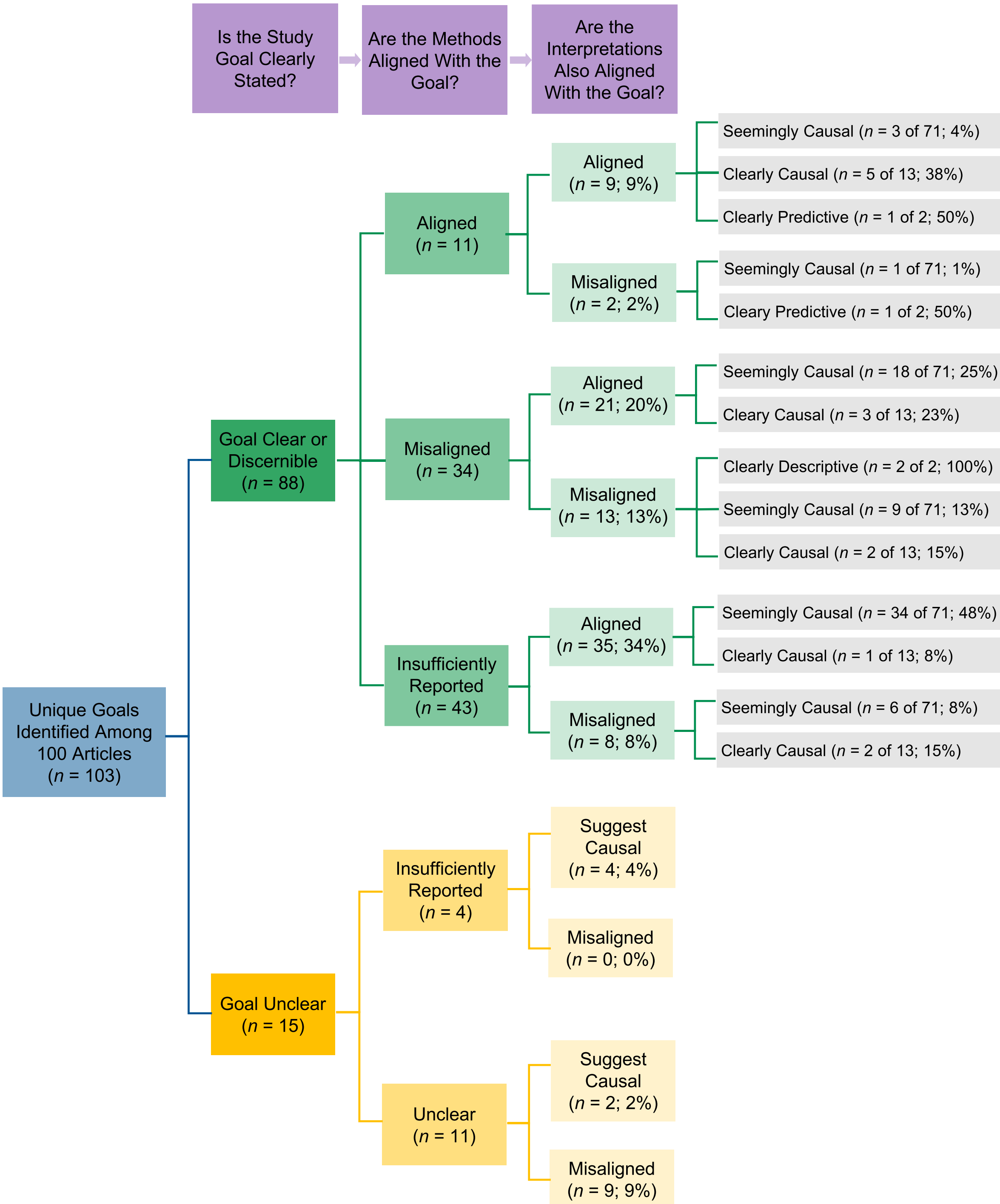

Figure 2. Frequencies of different patterns of goal–methods–interpretations (mis)alignment categories. Both types of (mis)alignments are aggregated to show all possible goal–methods–interpretations patterns. The proportion of studies from the total sample represented by a given pattern is shown in the penultimate step of the diagram. For studies with clearly stated or discernible goals, the breakdown of each pattern by study goal is reported in the final step of the figure.

though all went on to inappropriately interpret or discuss study coefficients (Figure 2).

Fully aligned studies were infrequent ($n = 9$ of 103; 9%), although, for causal goals in particular, full alignment was more frequent among studies with clear aims stated using causal terms like *effect*, *cause*, and/or *intervention* ($n = 5$ of 13; 38%), compared with their associational-framed counterparts ($n = 3$ of 71; 4%). Most frequently, methods were insufficiently reported to draw conclusions (46%). In the majority of studies, authors interpreted the appropriate coefficients for the study goal; specifically, goal–interpretations misalignments occurred in 31% of studies and were less likely to occur when the methods were aligned (2%) than when the methods were misaligned (13%).

Reflections

Although the assessment of (mis)alignment in this study is unique, to the author's knowledge, some findings herein can be compared with those in prior reviews of variable-selection approaches in the epidemiologic literature. For example, in 11% of the articles reviewed here, authors used change-in-estimate procedures and in 12%, authors used a predictor selection approach for variable selection (Table 3). This is consistent with Talbot and Massamba (47), who observed these frequencies as 12% and 14%, respectively. Similarly, Pocock et al. (56) observed that in 15% of their sample of 2001 articles, authors used stepwise methods; Walter and Tiemeier (53) reported a frequency of 20% among epidemiologic articles published in 2009. Although misalignments were not examined specifically in these studies, the similarity in the observed frequencies of these methods (for Talbot and Massamba (47), in predominantly causal studies) suggests that goal–methods misalignment has not improved over time. This similarity with findings of prior reviews also suggests that the article-selection procedures in the present review captured an unbiased sample.

Notably, in a high proportion of studies, authors insufficiently reported their variable selection methods (46%), which is consistent with observations from prior reviews (e.g., Talbot and Massamba (47): 37%, Walter and Tiemeier (53): 35%) (56). Simply put, there was much more *what* than *why* in these articles. Often, selected variables were simply listed, and when more detail was given, it was still uninformative (e.g., *because they are potential confounders*). Here, without knowing the authors' definition of a confounder, a reader is unable to judge the appropriateness of this approach. Rarely was additional detail provided in appendices if it was excluded from the main text.

Given the strong emphasis on causal thinking in modern epidemiologic methods, the lack of detail about the decisions driving variable selection in such a large proportion of mostly causal studies is troubling. Without such details, the reader cannot critically appraise the quality of the findings and justifiability of the interpretations. Furthermore, this article search was conducted among top epidemiology journals, so it is likely a conservative report of the misalignment problem. That is, arguably, finding evidence of misalignments among top epidemiology journals implies that the problem is likely even more ubiquitous among observational studies in the broader health and medical sciences literature; these disciplines lack the same long history that epidemiology has of considering causation in observational studies (6, 12, 25, 57) and potentially place less emphasis on causal thinking in variable selection. However, it is important to caution that the present study is only generalized to the observational epidemiologic literature; alignment could occur more frequently among study designs excluded from this report (e.g., randomized controlled trials, quasi-experimental studies).

The Strengthening the Reporting of Observational Studies in Epidemiology guidelines (58) clearly advise that authors report 1) all "candidate variables" considered for selection, 2) full details on the variable selection procedures used, and 3) details on any procedures taken to trim a list of candidate variables to those included in the final model. Here, authors discussed the pool of candidate variables in 53% of studies and those selected into the model as one and the same (Table 3). When no distinction was made between these features, often the variable selection details provided were more related to the candidate variables (e.g., we adjusted for known risk factors for the outcome) than the selection criteria used (e.g., 10% change-in-estimate, $P < 0.20$). Nonetheless, although advised by Strengthening the Reporting of Observational Studies in Epidemiology, decisions leading to final regression models were rarely reported in these studies.

Finally, the article search unexpectedly captured 2 clearly descriptive studies, although the inclusion criterion *performed variable selection* was assumed to weed these out. This begs the question of what is the purpose of multivariable adjustment in a purely descriptive study? Porta (52) defines a descriptive study as "a study concerned with and designed only to describe the existing distribution of variables without much regard to causal relationships or other hypotheses." If documentation of existing distributions in the data is of interest, adjustment seems unnecessary (54, 59). Adjustment is performed to isolate the independent effect of a particular variable (e.g., exposure) from the effects of variables that create noncausal associations between it and the outcome; it edges studies away from description (reporting observed facts) and toward causation (comparing facts vs. counterfactuals) (32, 59).

Limitations

Although there is an inevitable element of subjectivity in the article review process, much care was taken in the development and piloting of this study (Web Appendix 1), and pilot results were quite consistent with those of the full review (Web Tables 1–4). However, to minimize subjectivity, the quality and/or reasonableness of methods and interpretations were not judged. This review was focused only on whether a study's methods and/or interpretations were appropriate for its goal and did not require judgments about the quality or justifiability of variable selection decisions or interpretations (e.g., *the DAG-driven approach was appropriate for the causal goal*, not *results likely were biased due to exclusion of C from the DAG*). Identifying the inappropriate use of a certain method for a given goal is a decision rooted in objective methodological guidance; although this likely helped minimize subjectivity in (mis)alignment decisions, it is important to acknowledge that alignment does not necessarily indicate quality.

Similarly, goal–methods (mis)alignment could only be judged by the procedures reported in studies. Anything unreported was assumed not to have been done. Many considerations (e.g., exclusion of mediators from the variable pool) may have taken place in the planning and analysis of a study but not ended up in the published article (although best practice states that these decisions be transparently reported (58)). However, even if, for example, every seemingly causal study in which it was reported that adjustment for variables associated with "X and Y" had, in truth, removed mediators from consideration (but not reported this), goal–methods alignment would only increase from 6% to 11%.

In addition, although GLMs are the predominant tool used to analyze epidemiologic data, GLMs are not used for analyses in all epidemiologic studies. Thus, the focus on GLMs in this study may limit the generalizability of the findings; the proportions of misalignments observed herein may not apply to studies in which other analytic methods are used.

Finally, observed results conceivably could have been influenced in part by journal language restrictions and/or word limits (60, 61). The *International Journal of Epidemiology*, *European Journal of Epidemiology*, Epidemiology, and *Journal of Epidemiology and Community Health* have no language restrictions, whereas the *American Journal of Epidemiology* discourages the use of the word *effect* as proxy for *association*, except when describing the estimand. However, the only notable difference in the frequency of study goals across journals was for *Epidemiology*, which had no studies with unclear goals and the highest proportion of studies with clearly causal goals (Web Table 7). Word limits for journals were comparable, and, consistent with this, the proportion of studies with insufficiently reported methods was similar across journals (Web Table 9). Still, a possible solution for the observed high proportion of insufficiently reported variable selection approaches is for journals to offer an additional dedicated manuscript word count to "variable selection decisions" and require that, along with a list of final covariates selected, this decision-making be outlined.

CONCLUSIONS

Recent commentaries reflecting on *ways forward* or the *future of epidemiology* suggest epidemiologists be trained to ask clearer research questions and use the right tools to answer them (1–5). This review takes a step back to systematically assess how study goals are framed and how often misalignments occur in current practice. The findings hopefully will motivate more widespread appreciation of the problem and serve as a jumping-off point for investigating remediation strategies to eliminate misalignments in future practice.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York, United States (Katrina L. Kezios).

This work was supported by grant T32-ES02377 from the National Institute of Environmental Health Sciences.

The data in this review were derived from randomly selected articles in the published literature. All information needed to replicate the analysis are provided in the manuscript: the exact search terms to create the article database from which to randomly sample articles are available in Web Appendix 2; the exact coding schemes required to categorize articles are reported in the Methods section. The data underlying this article will be shared on reasonable request to the corresponding author.

The author thanks Natalie Levy for reviewing articles for the validation study; your time and effort throughout the planning and completion of this project are greatly appreciated, and your insight and advice along the way were invaluable. The author also thanks Dr. Sharon Schwartz, Dr. Lisa Bates, and Dr. Pam Factor-Litvak for their expertise, guidance, critical feedback, and encouragement during the planning and execution of this project.

Conflict of interest: none declared.

REFERENCES

1. Musci RJ, Stuart E. Ensuring causal, not casual inference. *Prev Sci.* 2019;20(3):452–456.
2. Fox MP, Edwards JK, Platt R, et al. The critical importance of asking good questions: the role of epidemiology doctoral training programs. *Am J Epidemiol.* 2020;189(4): 261–264.
3. Harper S. A future for observational epidemiology: clarity, credibility, transparency. *Am J Epidemiol.* 2019;188(5): 840–845.
4. Lesko CR, Keil AP, Edwards JK. The epidemiologic toolbox: identifying, honing, and using the right tools for the job. *Am J Epidemiol.* 2020;189(6):511–517.
5. Lau B, Duggal P, Ehrhardt S, et al. Perspectives on the future of epidemiology: a framework for training. *Am J Epidemiol.* 2020;189(7):634–639.
6. Hernán M, Robins J. *Causal Inference: What If.* Chapman & Hall/CRC: Boca Raton, FL; 2020.
7. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance.* 2019;32(1):42–49.
8. Rohrer JM. Thinking clearly about correlations and causation: graphical causal models for observational data. *Adv Methods Pract Psychol Sci.* 2018;1(1):27–42.
9. Greenland S, Pearce N. Statistical foundations for model-based adjustments. *Annu Rev Public Health.* 2015;36: 89–108.
10. Shmueli G. To explain or to predict? *Stat Sci.* 2010;25(3): 289–310.
11. Ellison G. COVID-19 and the epistemology of epidemiological models at the dawn of AI. *Ann Hum Biol.* 2020;47(6):506–513.
12. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
13. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health.* 2018;108(5):616–619.
14. Huitfeldt A. Is caviar a risk factor for being a millionaire? *BMJ.* 2016;355:i6536.
15. Thornley S. Causation and statistical prediction: perfect strangers or bedfellows. *J Biom Biostat.* 2012;3(7):e115.

16. Waljee AK, Higgins PDR, Singal AG. A primer on predictive models. *Clin Transl Gastroenterol.* 2014;5(1):e44.
17. van Diepen M, Ramspek CL, Jager KJ, et al. Prediction versus aetiology: common pitfalls and how to avoid them. *Nephrol Dial Transplant*. 2017;32(suppl_2):ii1–ii5.
18. Zalpuri S, Middelburg RA, van de Watering L, et al. Association vs causality in transfusion medicine: understanding multivariable analysis in prediction vs etiologic research. *Transfus Med Rev.* 2013;27(2): 74–81.
19. Breiman L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat Sci.* 2001; 16(3):199–231.
20. Prins SJ, Reich A. Can we avoid reductionism in risk reduction? *Theor Criminol.* 2018;22(2):258–278.
21. Arnold KF, Davies V, de Kamps M, et al. Reflections on modern methods: generalized linear models for prognosis and intervention—theory, practice and implications for machine learning. *Int J Epidemiol.* 2021;49(6):2074–2082.
22. Rose S. Intersections of machine learning and epidemiological methods for health services research. *Int J Epidemiol.* 2021;49(6):1763–1770.
23. Pearl J. *Causality*. New York, NY: Cambridge University Press; 2009.
24. Shadish W, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth, Cengage Learning: Belmont, CA; 2002.
25. Pearl J, Mackenzie D. *The Book of Why: The New Science of Cause and Effect*. New York, NY: Penguin Books Limited; 2018.
26. Prince JT. A paradigm for assessing the scope and performance of predictive analytics. *Inf Econ Policy.* 2019; 47:7–13.
27. Greenland S. Causal inference as a prediction problem: assumptions, identification and evidence synthesis. In: Berzuini C, Dawid P, Bernardinelli L, eds. *Causality: Statistical Perspectives and Applications*. Chichester, West Sussex, United Kingdom: John Wiley & Sons; 2012.
28. Salkind N. *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Publications Inc.; 2010.
29. Greenland S, Pearl J. Adjustments and their consequences—collapsibility analysis using graphical models. *Int Stat Rev.* 2011;79(3):401–426.
30. Kleinberg J, Ludwig J, Mullainathan S, et al. Prediction policy problems. *Am Econ Rev.* 2015;105(5):491–495.
31. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol.* 2013;177(4):292–298.
32. Kezios KL, Hayes-Larson E. A clarification on causal questions: we ask them more often than we realize. *Am J Public Health.* 2018;108(8):e4–e4.
33. Joffe M, Gambhir M, Chadeau-Hyam M, et al. Causal diagrams in systems epidemiology. *Emerg Themes Epidemiol.* 2012;9(1):1–1.
34. Lederer DJ, Bell SC, Branson RD, et al. Control of confounding and reporting of results in causal inference studies: guidance for authors from editors of respiratory, sleep, and critical care journals. *Ann Am Thorac Soc.* 2019; 16(1):22–28.
35. Leisman DE, Harhay MO, Lederer DJ, et al. Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med.* 2020;48(5):623–633.
36. Martin W. Making valid causal inferences from observational data. *Prev Vet Med.* 2014;113(3):281–297.
37. Aiello AE, Larson EL. Causal inference: the case of hygiene and health. *Am J Infect Control.* 2002;30(8):503–511.
38. Constantine NA. Regression analysis and causal inference: cause for concern? *Perspect Sex Reprod Health.* 2012;44(2): 134–137.
39. Olsen J, Jensen UJ. Causal criteria: time has come for a revision. *Eur J Epidemiol.* 2019;34(6):537–541.
40. Hernán MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health.* 2004;58(4): 265–271.
41. Vansteelandt S, Bekaert M, Claeskens G. On model selection and model misspecification in causal inference. *Stat Methods Med Res.* 2012;21(1):7–30.
42. Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidimiology.* 2014;25(3):418–426.
43. Pearl J, Glymour M, Jewell NP. *Causal Inference in Statistics: a Primer*. Chichester, West Sussex, United Kingdom: John Wiley & Sons; 2016.
44. Westreich D. *Epidemiology by Design: A Causal Approach to the Health Sciences*. New York, NY: Oxford University Press; 2020.
45. Savitz DA, Wellenius GA. *Interpreting Epidemiologic Evidence: Connecting Research to Applications*. New York, NY: Oxford University Press; 2016.
46. Hernán MA, Hernández-Díaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol.* 2002;155(2):176–184.
47. Talbot D, Massamba VK. A descriptive review of variable selection methods in four epidemiologic journals: there is still room for improvement. *Eur J Epidemiol.* 2019;34(8): 725–730.
48. Westreich D, Edwards J, van Smeden M. Comment on Williamson et al. (OpenSAFELY): the table 2 fallacy in a study of COVID-19 mortality risk factors. *Epidemiology.* 2021;32(1):e1–e2.
49. @EpiEllie. *There's a new paper circulating today about "risk factors" for COVID19 which is getting misinterpreted in a pretty common way: applying conclusions about causation to results obtained via methods designed only for finding correlations. It's time for a #tweetorial!* https://twitter.com/EpiEllie/status/1258607277357006849. Posted: May 7, 2020. Accessed: September 15, 2020.
50. Tennant P, Murray E. The quest for timely insights into COVID-19 should not come at the cost of scientific rigor. *Epidemiology.* 2021;32(1):e2.
51. Schooling CM, Jones HE. Clarifying questions about "risk factors": predictors versus explanation. *Emerg Themes Epidemiol.* 2018;15(1):10.
52. Porta M. *A Dictionary of Epidemiology*. New York, NY: Oxford University Press; 2014.
53. Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. *Eur J Epidemiol.* 2009;24(12): 733–736.
54. Conroy S, Murray EJ. Let the question determine the methods: descriptive epidemiology done right. *Br J Cancer.* 2020;123(9):1351–1352.
55. Weiss NS. Can we protect ourselves from excesses of language? *Am J Epidemiol.* 1998;148(2):131–132.
56. Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ.* 2004;329(7471):883.

57. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health.* 2005;95(S1): S144–S150.
58. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4(10):e297.
59. Kaufman JS. Statistics, adjusted statistics, and maladjusted statistics. *Am J Law Med.* 2017;43(2–3):193–208.
60. JAMA. Network Open. *Instructions for authors.* 2018. https://jamanetwork.com/pages/jama-network-open-instructions-for-authors/?utm_campaign=articlePDF%26utm_medium=articlePDFlink%26utm_source=articlePDF%26utm_content=jamanetworkopen.2018.0268. Accessed March 25, 2019.
61. AMA Style Insider. The use of cause-and-effect language in the JAMA Network Journals. 2017. http://amastyleinsider.com/2017/09/19/use-cause-effect-language-jama-network-journals/. Accessed March 25, 2019.