Methods for the Selection of Covariates in Nutritional Epidemiology Studies: A Meta-Epidemiological Review

Dena Zeraatkar,[1] Kevin Cheung,[2] Kirolos Milio,[2] Max Zworth,[2] Arnav Gupta,[3] Arrti Bhasin,[1] Jessica J Bartoszko,[1] Michel Kiflen,[1,4] Rita E Morassut,[5] Salmi T Noor,[1] Daeria O Lawson,[1] Bradley C Johnston,[6] Shrikant I Bangdiwala,[1,4] and Russell J de Souza [iD][1,4]

[1]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada; [2]Department of Medicine, McMaster University, Hamilton, Ontario, Canada; [3]Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada; [4]Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada; [5]Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; and [6]Department of Community Health and Epidemiology, Faculty of Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

ABSTRACT

Background: Observational studies provide important information about the effects of exposures that cannot be easily studied in clinical trials, such as nutritional exposures, but are subject to confounding. Investigators adjust for confounders by entering them as covariates in analytic models.

Objective: The aim of this study was to evaluate the reporting and credibility of methods for selection of covariates in nutritional epidemiology studies.

Methods: We sampled 150 nutritional epidemiology studies published in 2007/2008 and 2017/2018 from the top 5 high-impact nutrition and medical journals and extracted information on methods for selection of covariates.

Results: Most studies did not report selecting covariates a priori (94.0%) or criteria for selection of covariates (63.3%). There was general inconsistency in choice of covariates, even among studies investigating similar questions. One-third of studies did not acknowledge potential for residual confounding in their discussion.

Conclusion: Studies often do not report methods for selection of covariates, follow available guidance for selection of covariates, nor discuss potential for residual confounding. *Curr Dev Nutr* 2019;3:nzz104.

Introduction

The majority of studies published in nutrition are observational in design (1). Observational studies can provide important information about the comparative effectiveness and safety of interventions or exposures that cannot be easily studied in clinical trials owing to ethical or feasibility reasons. Typically, nutritional interventions and exposures are difficult to study in randomized trials. Usually, participants are unable to adhere to dietary interventions long enough to reasonably expect to observe an effect on health outcomes (2–5). The rigorous randomized trials that maintain adherence through housing of participants in controlled environments and the provision of meals are costly and cannot recruit large enough numbers of participants or study clinically important outcomes that require long durations of follow-up (6). Hence, observational studies are likely to remain an important source of evidence on the relation between nutrition and health.

Despite their high prevalence in nutrition, as well as other fields, observational studies are subject to potential confounding bias, whereby there is a distortion of the observed effect of the exposure of interest due to its association with other factors that affect the outcome (7). The risk of

bias from known confounders can be mitigated through design considerations, such as restricting the study sample to ≥1 levels of confounding factors or matching, as well as through statistical methods, such as stratified analyses or adjustment through generalized linear models. Generalized linear models are currently the most common method by which data from observational studies are analyzed, owing to their flexibility (8).

Unlike clinical trials that commonly operate under strict standards at every step of data analysis, including the preparation of detailed data analysis plans before the review of data by investigators, when researchers conduct statistical analyses of observational studies, they have a great deal of discretion over data analysis methods (9). A major, but often overlooked, source of discretion is the choice of covariates to include in analytic models (10). Covariates are variables that are included in analytic models to adjust for confounding and to produce more precise effect estimates (11). By including covariates that inflate the effect of the exposure of interest and excluding covariates that deflate the effect estimate, an analyst's model-building procedure may be heavily influenced by the possibility of obtaining statistically significant or interesting results.

Selecting covariates a priori reduces opportunity for *P* value shopping, whereby the choice of covariates is influenced by the possibility of obtaining statistically significant results (12–17). Although authors should strive to be comprehensive in the covariates included in analytic models, sometimes, any additional covariate not part of the correct model specification can increase bias (18, 19) and decrease precision (12, 14, 19). Current guidance suggests that the choice of covariates should be primarily guided by empirical evidence or theoretical knowledge of suspected or established confounding factors and adequately justified in study reports (12–16, 20). Use of variable selection algorithms that do not incorporate subject matter knowledge may lead to inclusion of variables that increase bias (12, 14, 21–23). Given the potential for residual confounding in observational studies, authors should also be cautious in their interpretation of results and discuss the possibility of residual confounding (20, 24). Inappropriately exaggerating the certainty of findings from observational studies that are at risk of residual confounding may be misleading and perceived as sensationalism, overall diminishing trust in research.

To our knowledge, no work has been done to evaluate the reporting and credibility of methods for selection of covariates, to assess consistency in choice of covariates among studies reporting on the same outcome and similar exposures, or to describe how authors discuss potential for residual confounding in nutritional epidemiology studies. Furthermore, it is unclear whether the reporting of methods used to select covariates has improved over time. The selection of confounders in nutritional epidemiology can be more challenging than in other areas owing to the very high correlation observed between nutritional exposures, which makes teasing apart the effects of individual exposures difficult (25).

Objectives

The objectives of this investigation are as follows:

1) describe the reporting of methods for selection of covariates in a sample of nutritional epidemiology studies:

   a) estimate the proportion of studies that describe how covariates are selected for analysis;
   b) describe methods by which authors select covariates for analysis;
   c) assess consistency in choice of covariates among studies reporting on similar exposures and outcomes;
   d) describe how authors discuss potential for residual confounding; and
2) evaluate whether methods for selection of covariates and reporting of methods have evolved from 2007/2008 to 2017/2018.

Methods

The protocol for this study was registered on the Open Science Framework website (https://osf.io/4p7jx).

Search strategy

Our search strategy was developed with the help of an experienced research librarian and is presented in **Supplemental Appendix 1**. We randomly sampled 150 nutritional studies from the top 5 general medicine and nutrition journals, based on the h5-index (Google Scholar, 2019), with start of publication preceding 2007: *New England Journal of Medicine*, *Lancet*, *British Medical Journal* (*BMJ*), *Journal of the American Medical Association* (JAMA), and *Annals of Internal Medicine*; and *The American Journal of Clinical Nutrition*, *British Journal of Nutrition*, *Clinical Nutrition*, *Nutrition Journal*, and *The Journal of Nutrition*. A sample size of 150 studies was selected to allow estimation of the prevalence of even uncommon methods by which authors of nutritional studies select covariates (i.e., methods used in ∼5% of studies) with acceptable precision (i.e., ±3.5%) (26).

Our aim was to sample 50 studies from general medicine journals and 100 studies from nutrition journals, with an equal number of studies being sampled from 2007/2008 and 2017/2018 and from each general medicine journal and nutrition journal. Because we did not identify a sufficient number of eligible studies from general medicine journals, we also searched their associated subjournals (*BMJ Open*, *Lancet Public Health*, etc.). To reach our target sample size, we oversampled studies from 2017/2018 and from select journals that published nutritional epidemiology more frequently (i.e., *BMJ*, *The American Journal of Clinical Nutrition*, *Clinical Nutrition*, and *The Journal of Nutrition*).

Study selection

Teams of 2 reviewers independently screened studies for eligibility. Reviewers resolved discrepancies by discussion or consultation with an expert nutrition researcher and methodologist. We included observational studies that reported on the association between ≥1 nutritional exposures and patient-important health outcomes using generalized linear models. We defined nutritional exposures as foods or food chemicals that are typically consumed through the diet (excluding fortification and supplementation) or measures of dietary patterns and patient-important health outcomes as direct measures of mortality, morbidity, and quality of life. We excluded systematic reviews, commentaries, and diagnostic and prognostic studies in which the main aim was prediction modelling (e.g., What is the risk of cancer in North American women >40 y of age who regularly consume coffee?) and not

causal inference (e.g., Does drinking coffee increase risk of cancer?). Studies with <100 participants were excluded because nonrandomized studies with few participants are typically experimental or mechanistic studies, rather than epidemiological studies.

Data extraction

Teams of 2 reviewers, working independently, extracted the following information from each study, focusing on the primary analytic model: research question, study design, methods used to control for effects of covariates, methods used to select covariates for analysis, list of covariates addressed through the study design or analysis, presentation of results, and authors' discussion of potential for confounding bias. We defined the primary model as the model on the association between the primary exposure and outcome of interest that authors most frequently discussed in the study abstract and the results and discussion sections of the article and that was primarily used to guide authors' conclusions. This was usually (but not always) the most adjusted model. If a primary exposure and outcome of interest were not specified, we assumed that the primary exposure and outcome were the exposure and outcome for which results were first presented in the study abstract. For studies that used matching or propensity score methods, we treated matching variables and variables used to create propensity scores the same as covariates included in analytic models. Reviewers resolved discrepancies through discussion or by consultation with an expert nutrition researcher and methodologist. If a study cited a protocol, reviewers retrieved the protocol and reviewed for additional relevant information.

Data synthesis and analysis

We present descriptive statistics (frequencies and percentages) to describe reporting of methods for selection of covariates and authors' discussion of potential for residual confounding. We stratified results by year of publication and type of journal. For outcomes for which we included >10 studies (all-cause mortality and type 2 diabetes), we assessed consistency in choice of covariates among studies. We grouped studies by the types of exposures (i.e., micronutrients, macronutrients, foods, food groups, and dietary patterns) and present the number and proportion of studies that included each covariate in their primary analytic model. We do not present covariates related to study design (e.g., study center) because inconsistency in adjustment for these covariates would be expected given design variations across studies. We also indicate studies for which adjustment for certain covariates is not applicable (e.g., baseline diabetes status if participants with diabetes were not eligible for inclusion in the study) and studies that considered additional covariates in secondary models.

Results

Study characteristics

Supplemental Appendix 2 presents details of study selection and **Table 1** presents study characteristics. We included 36 studies from general medicine journals and 114 studies from nutrition journals. Fifty-three studies were published in 2007/2008 and 97 in 2017/2018. **Supplemental Appendix 3** lists the included studies.

The majority of studies described cohort designs. The median number of participants included in each study was slightly >8000. The majority of studies reported on dietary patterns. The most commonly reported outcome was all-cause mortality, followed by digestive cancers and type 2 diabetes. Other outcomes investigated included measures of weight, BMI, waist circumference, psychological profiles, and cognitive measures.

Most studies used regression methods to control for the effects of covariates and one-third used a combination of regression methods and stratification, most commonly by sex. Other methods included a combination of regression methods and individual or frequency matching. One study did not control for the effects of any covariates in its analysis (27). In this study, participants' change in weight was evaluated before and after Ramadan, with participants acting as their own controls. The most commonly used analytic model was the Cox proportional hazards model, followed by logistic regression and multivariable linear regression. Over two-thirds of studies reported statistically significant associations between the primary exposure and outcome of interest. Characteristics of studies were similar between studies published in 2007/2008 and in 2017/2018, except a larger proportion of studies published in 2017/2018 reported statistically significant effects.

Reporting of methods for selection of covariates

Table 2 presents details on reporting of methods for selection of covariates. Examples of these methods are presented in **Supplemental Appendix 4**. A small minority of studies reported selecting all covariates a priori (i.e., before the review of data by investigators) and a very small minority reported selecting some (but not all) covariates a priori. Among studies that reported that all covariates were selected a priori, all were published in 2017/2018. None of the included studies reported that covariates were not selected a priori.

Two-thirds of studies did not report the methods or criteria used to select covariates for analysis. The proportion of studies was similar between studies published in 2007/2008 and in 2017/2018. Among studies that did report methods for selection of covariates, most studies reported selecting established or suspected risk factors for the outcome of interest. Other criteria used included known or suspected factors associated with the exposure, factors associated with both the exposure and the outcome, or factors associated with either the exposure or the outcome. A very small minority of studies reported using Directed Acyclic Graphs to select potential confounders (factors associated with the exposure that also act on the outcome) for adjustment (28).

Over three-quarters of studies did not cite any sources to justify their choice of covariates. This was similar between studies published in 2007/2008 and in 2017/2018. Among studies that cited sources, most cited other epidemiological studies, systematic reviews, or authoritative documents (e.g., World Cancer Research Fund report on risk factors for cancer). A small minority of studies reported conducting de novo literature searches to identify relevant confounding factors. Among studies that cited additional literature to justify their choice of covariates, the additional literature most often addressed risk factors for the outcome of interest.

After selecting covariates for analysis, approximately one-third of studies reported using data-driven methods to narrow down their pool of covariates for inclusion in the final analytic model. The proportion

TABLE 1 Study characteristics[1]

	2007/2008 (n = 53)	2017/2018 (n = 97)	All articles (n = 150)
Journal			
Annals of Internal Medicine	2 (3.8)	3 (3.1)	5 (3.3)
British Medical Journal	4 (7.5)	11 (11.3)	15 (10.0)
Journal of the American Medical Association	3 (5.7)	5 (5.2)	8 (5.3)
Lancet	1 (1.9)	6 (6.2)	7 (4.7)
New England Journal of Medicine	0 (0)	1 (1.0)	1 (0.7)
The American Journal of Clinical Nutrition	14 (26.4)	10 (10.3)	24 (16.0)
British Journal of Nutrition	14 (26.4)	10 (10.3)	24 (16.0)
Clinical Nutrition	0 (0)	24 (24.7)	24 (16.0)
The Journal of Nutrition	14 (26.4)	10 (10.3)	24 (16.0)
Nutrition Journal	1 (1.9)	17 (17.5)	18 (12.0)
Study design			
Cohort	30 (56.6)	67 (69.1)	97 (64.7)
Case-control	8 (15.1)	5 (5.2)	13 (8.7)
Nested case-control	0 (0)	2 (2.1)	2 (1.3)
Case-cohort	1 (1.9)	0 (0)	1 (0.7)
Cross-sectional	14 (26.4)	23 (23.7)	37 (24.7)
Participants, n	5823 [1864–26,238]	11,879 [2121–88,184]	8072 [2035–62,461]
Primary exposures investigated			
Micronutrient	10 (18.9)	10 (10.3)	20 (13.3)
Macronutrient	6 (11.3)	13 (13.4)	19 (12.7)
Food	9 (17.0)	23 (23.7)	32 (21.3)
Food group	10 (18.9)	13 (13.4)	23 (15.3)
Dietary pattern	18 (34.0)	38 (39.2)	56 (37.3)
Primary outcomes investigated			
All-cause mortality	4 (7.5)	16 (16.5)	20 (13.3)
Cardiovascular mortality	1 (1.9)	2 (2.1)	3 (2.0)
Cardiovascular disease	1 (1.9)	3 (3.1)	4 (2.7)
Stroke	0 (0)	3 (3.1)	3 (2.0)
Myocardial infarction	3 (5.7)	3 (3.1)	6 (4.0)
Brain cancer and tumors of the spinal cord	1 (1.9)	1 (1.0)	2 (1.3)
Digestive cancers	7 (13.2)	11 (11.3)	18 (12.0)
Endocrine-related cancers	0 (0)	1 (1.0)	1 (0.7)
Female cancers	3 (5.7)	1 (1.0)	4 (2.7)
Prostate cancer	4 (7.5)	1 (1.0)	5 (3.3)
Overall cancer mortality	1 (1.9)	0 (0)	1 (0.7)
Type 2 diabetes	4 (7.5)	9 (9.3)	13 (8.7)
Other	24 (45.3)	46 (47.4)	70 (46.7)
Methods for controlling for effects of covariates			
Regression methods	31 (58.5)	52 (53.6)	83 (55.3)
Combination of regression methods and stratification	14 (26.4)	39 (40.2)	53 (35.3)
Combination of regression methods and individual matching	4 (7.5)	3 (3.1)	7 (4.7)
Combination of regression methods and frequency matching	4 (7.5)	2 (2.1)	6 (4.0)
None	0 (0)	1 (1.0)	1 (0.7)
Analytic model			
Multivariable linear regression	5 (9.4)	9 (9.3)	14 (9.3)
Logistic regression	21 (39.6)	30 (30.9)	51 (34.0)
Cox proportional hazards model	22 (41.5)	51 (52.6)	73 (48.7)
ANOVA methods	2 (3.8)	4 (4.1)	6 (4.0)
Poisson regression	2 (3.8)	0 (0)	2 (1.3)
Other	1 (1.9)	3 (3.1)	4 (2.7)
Reported statistically significant association between the primary exposure and outcome of interest?			
Yes	35 (66.0)	82 (84.5)	117 (78.0)
No	18 (34.0)	15 (15.5)	33 (22.0)

[1]Values are n (%) or median [IQR].

TABLE 2 Reporting of methods for selection of covariates[1]

	2007/2008 (n = 53)	2017/2018 (n = 97)	All articles (n = 150)
Reported whether covariates were selected a priori?			
Some (but not all) covariates were selected a priori	1 (1.9)	1 (1.0)	2 (1.3)
All covariates were selected a priori	0 (0)	7 (7.2)	7 (4.7)
Not reported	52 (98.1)	89 (91.8)	141 (94.0)
Reported methods for selection of covariates for analysis?			
Reported criteria for selection of all covariates	9 (17.0)	21 (21.7)	30 (20.0)
Reported criteria for selection of some covariates	10 (18.9)	15 (15.5)	25 (16.7)
Not reported	34 (64.2)	61 (62.9)	95 (63.3)
Among studies that reported methods for selection of covariates, covariates were selected from:[2]			
Factors known or suspected to be associated with the exposure	2 (3.8)	3 (3.0)	5 (3.3)
Known or established risk factors for the outcome	13 (24.5)	26 (26.8)	39 (26.0)
Factors known or suspected to be associated with both the exposure and outcome	1 (1.9)	4 (4.1)	5 (3.3)
Factors known or suspected to be associated with either the exposure or the outcome	2 (3.8)	0 (0)	2 (1.3)
Confounders (factors associated with the exposure that also act on the outcome) as identified by Directed Acyclic Graphs	0 (0)	4 (4.1)	4 (2.7)
Other	4 (7.5)	2 (2.1)	6 (4.0)
Not reported	34 (64.2)	61 (62.9)	95 (63.3)
Sources cited to support choice of covariates?[2]			
Systematic review	1 (1.9)	5 (5.2)	6 (4.0)
Authoritative document (e.g., World Cancer Research Fund report)	0 (0)	4 (4.1)	4 (2.7)
Narrative review	0 (0)	1 (1.0)	1 (0.7)
Epidemiological study	9 (17.0)	11 (11.3)	20 (13.3)
De novo literature search conducted by authors	1 (1.9)	9 (9.3)	10 (6.7)
Methodology article	0 (0)	1 (1.0)	1 (0.7)
No source cited	44 (83.0)	76 (78.4)	120 (80.0)
Reported use of data-driven methods for selection of covariates for inclusion in final analytic model?			
Reported use of data-driven methods for selection of all covariates for inclusion in final analytic model	6 (11.3)	8 (8.3)	14 (9.3)
Reported use of a combination of data-driven and hypothesis-driven methods to select covariates for inclusion in final analytic model	11 (20.8)	15 (15.4)	26 (17.3)
Did not report using any data-driven methods to select covariates	36 (67.9)	74 (76.3)	110 (73.3)
Among studies that reported use of data-driven methods for selection of covariates, covariates were selected based on:[2]			
If their inclusion appreciably changed the effect estimate of the primary exposure (change-in-estimate criterion)	11 (20.8)	11 (11.3)	22 (14.7)
P value in the final analytic model	2 (3.8)	1 (1.0)	3 (2.0)
P value in univariate model with the exposure as the dependent variable	1 (1.9)	3 (3.1)	4 (2.7)
P value in univariate model with the outcome as the dependent variable	3 (5.7)	6 (6.1)	9 (6.0)
Backward elimination	1 (1.9)	0 (0)	1 (0.7)
Stepwise selection	0 (0)	2 (2.1)	2 (1.3)
Magnitude of correlation with exposure	0 (0)	1 (1.0)	1 (0.7)
Whether inclusion reduced the SE of the effect estimate of the primary exposure	1 (1.9)	1 (1.0)	2 (1.3)
Model fit[3]	0 (0)	1 (1.0)	1 (0.7)
Some description provided but unclear which specific method was used	4 (7.5)	3 (3.1)	7 (4.7)
Did not report using any data-driven methods to select covariates	36 (67.9)	74 (76.3)	110 (73.3)
Conducted quantitative bias analysis to evaluate impact of potential unadjusted/unmeasured confounders on results?			
Yes, according to methods described by Lin et al. (29)	0 (0)	1 (1.0)	1 (0.7)
Yes, according to methods described by Ding and VanderWeele (30)	0 (0)	1 (1.0)	1 (0.7)
Yes, by constructing a hypothetical confounder	1 (1.9)	0 (0)	1 (0.7)
No	52 (98.1)	95 (97.9)	147 (98.0)

[1]Values are n (%).
[2]Categories are not mutually exclusive.
[3]Specific measure of model fit used not reported.

of studies was similar between studies published in 2007/2008 and in 2017/2018. Among studies that reported using data-driven methods, the majority used a combination of hypothesis- and data-driven methods, whereby some covariates were forced to be included in the model based on prior knowledge. Common data-driven methods for selection of covariates included the change-in-estimate criterion, P values in either univariate or multivariate models, and stepwise procedures. Studies that used the change-in-estimate criterion included

TABLE 3 Interpretation of results in nutritional epidemiology studies[1]

	2007/2008 (n = 53)	2017/2018 (n = 97)	All articles (n = 150)
Acknowledge potential for residual confounding?			
Yes	31 (58.5)	67 (69.1)	98 (65.3)
No	22 (41.5)	30 (30.9)	52 (34.7)
Acknowledge existence of unadjusted confounders?[2]			
Yes, because the confounder was not measured	4 (7.5)	11 (11.3)	15 (10.0)
Yes, for other reasons	0 (0)	4 (4.1)	4 (2.7)
Yes, no reason is provided	6 (11.3)	9 (9.3)	15 (10.0)
No, the authors do not acknowledge the existence of confounders not included in the analysis	43 (81.1)	74 (76.3)	117 (78.0)
Acknowledge measurement error as a potential source of confounding?			
Yes	4 (7.5)	11 (11.3)	15 (10.0)
No	49 (92.5)	86 (88.7)	135 (90.0)
Described likelihood of residual confounding affecting results?			
Likely	1 (1.9)	2 (2.1)	3 (2.0)
Possible	23 (43.4)	62 (63.9)	85 (56.7)
Unlikely	6 (11.3)	1 (1.0)	7 (4.7)
Not possible	0 (0)	0 (0)	0 (0)
Not discussed	23 (43.4)	32 (33.0)	55 (36.7)
Report a causal link between exposure and outcome?			
Yes	6 (11.3)	6 (6.2)	12 (8.0)
No	47 (88.7)	91 (93.8)	138 (92.0)

[1]Values are *n* (%).
[2]Categories are not mutually exclusive.

covariates in the final analytic model if their inclusion changed the effect estimate of the exposure of interest by an appreciable amount (e.g., 5%, 10%). Studies that used *P* values to screen covariates either excluded covariates from the final model if their *P* values were not sufficiently low (e.g., $P < 0.10$) or included covariates in the final model if their *P* values in univariate models with either the outcome or exposure of interest as the dependent variable were sufficiently low. Stepwise procedures included backward elimination or stepwise regression. A small minority of studies reported using some data-driven methods but did not provide an adequate description to classify the methods used.

Three studies, 1 published in 2007/2008 and 2 published in 2017/2018, conducted quantitative bias analysis to evaluate the impact of potential unadjusted confounders. Two of the studies used methods by Lin et al. (29) and Ding and VanderWeele (30) and the third explored the sensitivity of results to a hypothetical confounder.

Consistency in selection of covariates for similar outcomes and exposures
Supplemental Appendix 5 presents matrixes of covariates included in the final analytic models of 20 and 13 studies reporting on all-cause mortality and diabetes, respectively. Studies adjusted for a total of 72 and 62 unique covariates for all-cause mortality and diabetes, respectively. We did not find any studies that adjusted for exactly the same set of covariates and we found substantial inconsistency, even among studies investigating similar types of exposures. We categorized covariates as participant characteristics, measures of socioeconomic status, health-related covariates, medications, family history of diseases, dietary characteristics, and dietary patterns. The

median number of studies that adjusted for each covariate was 1 out of 20 studies for all-cause mortality and 1 out of 13 studies for diabetes.

Interpretation of results in light of potential for confounding
Table 3 presents information on authors' interpretation of their results. Examples are presented in Supplemental Appendix 4. One-third of studies did not acknowledge the potential for residual confounding in their discussion. Among studies that discussed the likelihood of residual confounding, most studies described the likelihood as possible and a very small minority of studies described the likelihood as unlikely. A quarter of studies identified unadjusted confounders that may have influenced results. This was most often because an important confounder was not measured in the study. The vast majority of studies did not discuss the possibility of measurement error in covariates leading to residual confounding. There was a slight improvement in the interpretation of results from 2007/2008 to 2017/2018, with slightly more references from the latter years acknowledging the potential for residual confounding, the existence of unadjusted confounders, and the potential for measurement error being a source of residual confounding.

Type of journal
Results stratified by type of journal (general medicine compared with nutrition) are presented in **Supplemental Appendix 6**. Reporting and methods used to select covariates and discussion of residual confounding were similar between studies published in general medicine and nutrition journals.

Discussion

Main findings

Our investigation of a sample of 150 nutritional epidemiology studies suggests that the reporting of methods for selection of covariates does not adhere to available guidance. In addition, methods used to select covariates are often suboptimal. Selecting covariates a priori reduces opportunity for the choice of covariates to be influenced by the possibility of obtaining statistically significant or interesting results (9, 12–17). Our investigation shows that authors of nutritional epidemiology studies very rarely report selecting covariates a priori. The STROBE Statement recommends authors to make clear "which confounders were adjusted for and why they were included" (20). However, only one-third of studies reported criteria for selection of covariates for analysis.

Among studies that described criteria for selection of covariates for analysis, one-quarter selected known or suspected risk factors for the outcome as covariates for analysis. A few also considered factors suspected to be associated with the exposure. Including all factors suspected to be associated with the exposure or the outcome as covariates may be occasionally problematic because it may lead to more covariates than needed to adequately adjust for confounding (12). Within even the largest studies, including too many covariates may lead to the breakdown of conventional fitting methods, such as maximum likelihood, and lead to data sparsity in which there are too few subjects at crucial combinations of covariates, with consequent inflation of effect estimates (15, 19, 31). In addition, selecting factors highly predictive of the exposure can produce multicollinearity, and hence unnecessarily wide CIs and potentially inflated effect estimates (12, 14, 15). For example, adjustment for instrumental variables, variables that are predictive of exposure but have no causal association with the outcome, has been shown to decrease precision and increase bias in the presence of residual confounding (32–35).

Despite widespread endorsement in the literature, very few studies reported using causal diagrams to select covariates (12, 14, 28, 36–38). Causal diagrams are constructed to display the analysts' best understanding of factors associated with the exposure and outcome of interest. They are used to identify all potential confounders for inclusion in the model, while also ensuring that unnecessary variables or variables whose inclusion in the model may increase bias are not included. For example, causal diagrams can be used to identify colliders, variables with ≥ 2 antecedent causes that lie within the pathway between the exposure and outcome of interest, that are often confused with confounders, but adjustment for which biases results (18). Causal diagrams can also reduce the potential for inclusion of intermediary variables in analytic models.

More than one-quarter of studies reported using data-driven methods, alone or in combination with hypothesis-driven methods, to narrow the pool of covariates for inclusion in the final analytic model. Among studies that reported using data-driven methods, the 3 most common methods were stepwise procedures, screening covariates based on P values in univariate or multivariate models, and the change-in-estimate criterion.

Stepwise procedures and screening covariates based on P values can achieve a parsimonious model but may select weak confounders over strong confounders if weaker confounders are more strongly correlated with the exposure or the outcome (12, 15, 21, 39). For example, a variable that is highly predictive of the outcome but unrelated to the exposure may be selected over a confounder that is only moderately correlated with both the exposure and outcome. Even when a given variable is significantly associated with the exposure or outcome, statistical significance is not informative about the magnitude of correlation and hence potential for confounding (21, 39). In addition, these methods ignore problems with preliminary testing and produce P values for the exposure effect that are too small and CIs that are too narrow (12, 15, 22, 40, 41). They also ignore the causal structure of the data by treating confounders and colliders equally (12, 14, 22).

Selecting covariates based on the change-in-estimate criterion is generally preferred over other methods, but as with other data-driven methods, this method ignores theoretical and empirical understanding of important confounders and relies heavily on the available data, in which causal relations may or may not be evident (12, 42). In addition, as with other data-driven methods, this method may produce P values that are too small because it ignores preliminary testing (12, 15, 22, 40, 41).

Very few studies reported conducting quantitative bias analysis to evaluate the robustness of results to potential unadjusted confounders. Authors may not be motivated to present quantitative bias analyses that are unfavorable because they may reduce the perceived validity of findings. Authors may also lack knowledge of these methods or the statistical expertise for their implementation or they may not find these analyses informative. Because the potential for residual confounding is a direct function of the effect size and corresponding CIs, it has also been argued that quantitative bias analysis contributes little information beyond what is already typically reported in reports of epidemiology studies (43).

We found a lack of consistency in choice of covariates, even among studies investigating the same outcome and similar types of exposures. This may be an artifact of inconsistency of methods used to select covariates, a general lack of knowledge by authors of the causal structure of the research questions being investigated, or lack of availability of important variables in the data. Many epidemiological studies are initiated decades before the analysis of data, at which point understanding of the causal structure of the problem being investigated may have been poor and so measurement of variables considered important now may have been omitted at the time of study inception. In addition, many studies are used to investigate secondary research questions while the variables measured may have only been tailored for the primary research question.

We also found more than one-quarter of studies to not acknowledge the potential for residual confounding in their discussion of study results. Some studies even made causal inferences, suggesting authors have a low level of appreciation of the potential for confounding bias in observational studies. It is possible that authors of nutritional epidemiology studies might not acknowledge potential for residual confounding bias because they may see it as an inevitable limitation of observational studies that is too well-known to merit discussion. However, observational studies are too often misconstrued by readers and sensationalized by the media (44–46). More conspicuous consideration of residual confounding bias in reports of nutritional epidemiology studies may lead to more cautious interpretation of findings by readers.

Relation to previous work

Previous studies have evaluated the quality of reporting of methods for handling confounders, as well as authors' interpretation of results in light of potential for residual confounding, in the general medicine and epidemiological literature (24, 47–52). Similarly to our findings, previous studies have found that choice of covariates is most often not justified and that many studies lack a satisfactory discussion of limitations related to confounding bias (24, 47–52). This suggests that these issues are not unique to nutrition but are also prevalent in observational studies from other fields.

Strengths and limitations

The strengths of our study include our duplicate screening and extraction of data for accuracy, as well as inclusion of a representative sample of high-impact nutritional epidemiology studies from both general medicine and specialized nutrition journals. We provide a comprehensive picture of the reporting of methods used to control for the effects of confounders, the consistency in choice of covariates among studies reporting on the same outcome and similar exposures, as well as how authors interpret their findings in light of the potential for residual confounding.

The most major limitation of this investigation is our inability to ascertain the methods used to select covariates when they were not explicitly described in the article or protocol of a study. Given word limitations of journals, authors may not be able to report on all aspects of their methods. However, it could be argued that studies that undertake a systematic effort to identify important covariates will likely report on such efforts and although methods used may be valid despite poor reporting, lack of reporting leaves readers unable to gauge validity. In addition, authors are able to provide detailed descriptions of their model-building procedures in supplementary materials, the content of which is not typically limited by journals. None of the studies included in our sample provided additional details on model-building procedures in supplementary materials.

We did not evaluate the validity of the methods used for the measurement (e.g., validity and reliability of the instruments used to measure covariates) and operationalization (e.g., functional form, whether and how continuous variables were categorized or dichotomized, whether time-dependency of covariates was appropriately considered) of covariates in studies because this was considered outside the scope of this investigation. However, it should be acknowledged that inclusion of important covariates is not sufficient for valid analysis. Important covariates should also be measured via valid tools and appropriately operationalized in analytic models.

Our investigation of the consistency in choice of covariates was focused on covariates included in the primary analytic model of each study. Although a minority of studies also considered additional covariates in secondary models, results from these models were not used to guide authors' conclusions and so authors likely did not deem these additional covariates as being important in estimating the association between the exposure and outcome of interest. Furthermore, our investigation likely underestimates inconsistency in choice of covariates across studies because we combined similar types of covariates (e.g., different measures of physical activity were grouped together, including nonleisure, aerobic, nonaerobic, and moderate-to-vigorous intensity). Finally, we could not fulfill our originally intended sampling strategy because we were unable to identify a sufficient number of eligible studies from some journals. Despite this, our sample is still representative of high-impact nutritional epidemiology studies to which other researchers and the public are regularly exposed.

Implications

Given that observational studies are likely to continue to predominate in nutrition, appropriately dealing with confounding bias is essential to being able to draw valid inferences from these studies. In addition, to avoid being misleading, reports of nutritional epidemiology studies should only provide a cautious interpretation of their results, given the magnitude of associated uncertainty.

Our study shows that there is a need to encourage authors of nutritional epidemiology studies to select covariates a priori using the best available theoretical or empirical evidence on important confounders of the relation being investigated, to transparently describe the criteria used to select covariates for analysis, and to discourage use of suboptimal methods for covariate selection. We acknowledge that recommended guidance for covariate selection may not be applicable in scenarios in which there may be little known about the causal structure of the outcome of interest and factors related to the exposure (12–16, 20). In these situations, authors may have to rely on data-driven model-building procedures. However, in these situations, we encourage use of more sophisticated techniques, like shrinkage and penalized regression, that have been shown to outperform traditional data-driven methods (53–55). In addition, analytic criteria that will be used to select covariates should be transparent and prespecified. One way in which this can be done is through protocols or statistical analysis plans. We caution that no methodology is perfect in all scenarios and that modeling methods should be documented in enough detail that readers can interpret results in light of the strengths and weaknesses of the methods used. Further, authors should interpret their results cautiously and provide a discussion of potential for residual confounding to avoid being misleading.

Conclusions

Improper omission and indiscriminate inclusion of covariates in statistical models can lead to compromised inferences (12–15, 19, 21). Selecting important confounders as covariates, minimizing their impact through appropriate design and statistical considerations, and acknowledging remaining uncertainty due to potential residual confounding is an integral component of inference-making in epidemiology. Our review shows that nutritional epidemiology studies do not adhere to available guidance for reporting of methods for selection of covariates and that methods used to select covariates are often suboptimal. Perhaps as a result, there is inconsistency in choice of covariates even among studies evaluating the same outcome and similar types of exposures. We also found reports of nutritional epidemiology studies to lack adequate discussion of potential for residual confounding. We encourage authors, peer reviewers, journal editors, and research funders to be more mindful of these issues.

and DOL: screened the studies and extracted the data; DZ: analyzed the data; DZ and RJdS: wrote the manuscript; DZ, BCJ, RJdS, and SIB: provided feedback on the manuscript; and all authors: read and approved the final manuscript.

References

1. Ortiz-Moncada R, González-Zapata L, Ruiz-Cantero MT, Clemente-Gómez V. Priority issues, study designs and geographical distribution in nutrition journals. Nutr Hosp 2011;26(4):784–91.

2. Chlebowski RT, Blackburn GL, Buzzard IM, Rose DP, Martino S, Khandekar J, York RM, Jeffery RW, Elashoff RM, Wynder EL. Adherence to a dietary fat intake reduction program in postmenopausal women receiving therapy for early breast cancer. The Women's Intervention Nutrition Study. J Clin Oncol 1993;11(11):2072–80.

3. Zazpe I, Sanchez-Tainta A, Estruch R, Lamuela-Raventos RM, Schröder H, Salas-Salvado J, Corella D, Fiol M, Gomez-Gracia E, Aros F, et al. A large randomized individual and group intervention conducted by registered dietitians increased adherence to Mediterranean-type diets: the PREDIMED study. J Am Diet Assoc 2008;108(7):1134–44.

4. Inelmen EM, Toffanello ED, Enzi G, Gasparini G, Miotto F, Sergi G, Busetto L. Predictors of drop-out in overweight and obese outpatients. Int J Obes 2005;29(1):122–8.

5. Douketis J, Macie C, Thabane L, Williamson D. Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. Int J Obes 2005;29(10):1153–67.

6. Hébert JR, Frongillo EA, Adams SA, Turner-McGrievy GM, Hurley TG, Miller DR, Ockene IS. Perspective: randomized controlled trials are not a panacea for diet-related research. Adv Nutr 2016;7(3):423–32.

7. Porta M. A dictionary of epidemiology. 6th ed. New York: Oxford University Press; 2014.

8. Real J, Forné C, Roso-Llorach A, Martínez-Sánchez JM. Quality reporting of multivariable regression models in observational studies: review of a representative sample of articles published in biomedical journals. Medicine (Baltimore) 2016;95(20):e3653.

9. Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, Bahník Š, Bai F, Bannard C, Bonnier E, et al. Many analysts, one data set: making transparent how variations in analytic choices affect results. Adv Methods Pract Psych Sci 2018;1(3):337–56.

10. Patel CJ, Burford B, Ioannidis JP. Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. J Clin Epidemiol 2015;68(9):1046–58.

11. Everitt BS, Skondal A. The Cambridge dictionary of statistics. 4th ed. New York: Cambridge University Press; 2014.

12. Greenland S, Pearce N. Statistical foundations for model-based adjustments. Annu Rev Public Health 2015;36:89–108.

13. Lederer DJ, Bell SC, Branson RD, Chalmers JD, Marshall R, Maslove DM, Ost DE, Punjabi NM, Schatz M, Smyth AR, 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):22–8.

14. Sauer BC, Brookhart MA, Roy J, VanderWeele T. A review of covariate selection for non-experimental comparative effectiveness research. Pharmacoepidemiol Drug Saf 2013;22(11):1139–45.

15. Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. Int J Epidemiol 2016;45(2):565–75.

16. Thomas L, Peterson ED. The value of statistical analysis plans in observational research: defining high-quality research from the start. JAMA 2012;308(8):773–4.

17. Braga LHP, Farrokhyar F, Bhandari M. Practical tips for surgical research: confounding: what is it and how do we deal with it? Can J Surg 2012;55(2):132–8.

18. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003;14(3):300–6.

19. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ 2016;352:i1981.

20. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4(10):e296.

21. Vansteelandt S, Bekaert M, Claeskens G. On model selection and model misspecification in causal inference. Stat Methods Med Res 2012;21(1):7–30.

22. Hurvich CM, Tsai C-L. The impact of model selection on inference in linear regression. Amer Statist 1990;44(3):214–17.

23. Hernan MA. The C-word: scientific euphemisms do not improve causal inference from observational data. Am J Public Health 2018;108(5):616–19.

24. Hemkens LG, Ewald H, Naudet F, Ladanie A, Shaw JG, Sajeev G, Ioannidis JP. Interpretation of epidemiologic studies very often lacked adequate consideration of confounding. J Clin Epidemiol 2018;93:94–102.

25. Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. J Epidemiol Community Health 2014;68(11):1096–100.

26. Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. Arch Orofac Sci 2006;1:9–14.

27. Ali Z, Abizari AR. Ramadan fasting alters food patterns, dietary diversity and body weight among Ghanaian adolescents. Nutr J 2018;17(1):75.

28. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10(1):37–48.

29. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998;54(3):948–63.

30. Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. Epidemiology 2016;27(3):368–77.

31. Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. Am J Epidemiol 2000;151(5):531–9.

32. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. Developing a protocol for observational comparative effectiveness research: a user's guide. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

33. Bhattacharya J, Vogt WB. Do instrumental variables belong in propensity scores? Cambridge, MA: National Bureau of Economic Research; 2007.

34. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol 2006; 163(12):1149–56.

35. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Joffe MM, Glynn RJ. Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am J Epidemiol 2011;174(11):1213–22.

36. Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. Respirology 2014;19(3):303–11.

37. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. Biometrics 2011;67(4):1406–13.

38. Staplin N, Herrington WG, Judge PK, Reith CA, Haynes R, Landray MJ, Baigent C, Emberson J. Use of causal diagrams to inform the design and interpretation of observational studies: an example from the Study of Heart and Renal Protection (SHARP). Clin J Am Soc Nephrol 2017;12(3): 546–52.

39. Dales LG, Ury HK. An improper use of statistical significance testing in studying covariables. Int J Epidemiol 1978;7(4):373–6.

40. Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol 1989;129(1):125–37.

41. Draper NB, Guttman I, Lapczak L. Actual rejection levels in a certain stepwise test. Commun Stat Theory 1979;8(2):99–105.

42. Maldonado G, Greenland S. Simulation study of confounder-selection strategies. Am J Epidemiol 1993;138(11):923–36.

43. Ioannidis JPA, Tan YJ, Blum MR. Limitations and misinterpretations of E-values for sensitivity analyses of observational studies. Ann Intern Med 2019;170(2):108–11.

44. Selvaraj S, Borkar DS, Prasad V. Media coverage of medical journals: do the best articles make the news? PLoS One 2014;9(1):e85355.

45. Zeraatkar D, Obeda M, Ginsberg JS, Hirsh J. The development and validation of an instrument to measure the quality of health research reports in the lay media. BMC Public Health 2017;17(1):343.

46. Kininmonth AR, Jamil N, Almatrouk N, Evans CE. Quality assessment of nutrition coverage in the media: a 6-week survey of five popular UK newspapers. BMJ Open 2017;7(12):e014633.

47. Groenwold RH, Hoes AW, Hak E. Confounding in publications of observational intervention studies. Eur J Epidemiol 2007;22(7):413–15.

48. Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. Ann Epidemiol 2008;18(10):746–51.

49. Müllner M, Matthews H, Altman DG. Reporting on statistical methods to adjust for confounding: a cross-sectional survey. Ann Intern Med 2002;136(2):122–6.

50. Pocock SJ, Collier TJ, Dandreo KJ, de Stavola BL, Goldman MB, Kalish LA, Kasten LE, McCormack VA. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ 2004;329(7471):883.

51. Pouwels KB, Widyakusuma NN, Groenwold RH, Hak E. Quality of reporting of confounding remained suboptimal after the STROBE guideline. J Clin Epidemiol 2016;69:217–24.

52. Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. Eur J Epidemiol 2009;24(12):733.

53. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. Am J Epidemiol 2008;167(5):523–9; discussion 530–1.

54. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Stat Med 2015;34(23):3133–43.

55. Van Houwelingen J. Shrinkage and penalized likelihood as methods to improve predictive accuracy. Stat Neerl 2001;55(1):17–34.