BJU Int 2023; **132:** 4–8 doi:10.1111/bju.16028

BJUI
BJU International

Review

Guidelines for reporting observational research in urology: the importance of clear reference to causality

Andrew J. Vickers[1], Melissa Assel[1], Rodney L. Dunn[2], Emily C. Zabor[3], Michael W. Kattan[3], Maarten van Smeden[4] and Darren Dahly[5]

[1]*Memorial Sloan Kettering Cancer Center, New York, NY,* [2]*University of Michigan, Ann Arbor, MI,* [3]*Cleveland Clinic, Cleveland, OH, USA,* [4]*Utrecht University, Utrecht, The Netherlands, and* [5]*University College Cork, Cork, Ireland*

This article is co-published in the journals: European Urology (10.1016/j.eururo.2023.04.027); Urology (10.1016/j.urology.2023.04.007); Journal of Urology (10.1097/JU.0000000000003479) and BJU International (10.1111/bju.16028)

Keywords

Causality, Urology, observational research, guidelines, reporting

As statistical editors and reviewers, we have noticed that observational research papers submitted for publication often share a common set of problems with respect to the issue of causality. Causality can be conceptualised in counterfactual terms: would the outcome be different if we change what we do? For instance, we say that smoking causes cancer because if we reduce smoking, we reduce cancer; conversely, even though ice cream sales closely track the rate of shark bites, we avoid a causal conclusion because banning ice cream would not make ocean swimming any safer.

Here we suggest guidelines for reporting of observational studies where investigators calculate statistical associations between an exposure, such as a drug, a change in surgical technique or a lifestyle factor like smoking or diet, and an outcome, such as cancer diagnosis or cancer progression.

The core issue is that authors all too often avoid any explicit reference to causal mechanisms and causal inference but then still try to draw causal conclusions. We repeatedly see investigators declare somberly in the methods section that a study aims merely to 'derive statistical associations' but then report results or make recommendations that clearly imply causality: a drug is said to 'reduce risk', a surgical technique to 'improve outcome' or a recommendation made that patients be counselled to change lifestyle so as to 'prevent cancer'. Then in response to criticism, authors sometimes deny in their rebuttal letters (and elsewhere) that their language implies a causal conclusion. Assessing statistical association is easy and determining causation is hard; yet, while associations are sometimes of research interest, understanding causality is necessary to provide solutions that can improve patient outcomes.

Our anecdotal experience has recently been corroborated by systematic reviews of the language used by authors of observational studies: 'one section might be carefully phrased in terms of association while the other presented causal language'; 'some authors (inappropriately) jumped to recommending acting on the findings' despite reporting only statistical associations [1,2]. Of particular note, remarkably few papers explicitly used the word 'cause'. One well-known commentary argues that avoidance of causal language has simply become a reflexive response to the difficulties of causal inference, though, naturally, avoiding a subject does not aid scientific progress [3].

We will use a hypothetical example to avoid singling out individual authors for what is such a common problem, although the example is closely modelled on a published paper from outside the field of urology. Imagine a study on statins and rates of overdiagnosis following PSA screening. A typical report would introduce the hypothesis in terms of whether 'the use of statins is *associated* with rates of prostate cancer overdiagnosis'. The authors might report lower rates of overdiagnosis in statin users and then rapidly switch to causal language: 'statin use shows *protective effects* on overdiagnosis. Low-grade cancers in particular *were reduced* with the use of statins'. A caveat is typically then buried deep in the discussion section ('we cannot fully establish a causal relationship') and the conclusion is generally vague (a 'decrease in overdiagnosis with the use of statins'), something which protects the authors from criticism while allowing them to imply that their paper has causal implications. There is no attempt to identify causal pathways, let alone conduct analyses to explore those pathways.

Causal pathways are critical because they determine the implications of the findings for clinical practice. If statins reduce overdiagnosis directly, for instance, through an antineoplastic mechanism, we might more strongly recommend statins to men who are undergoing PSA

screening. The clinical consequences would be very different if the effects of statins are indirect. For instance, if statins reduce overdiagnosis by lowering PSA levels below biopsy thresholds, then the issue for clinical practice is not statin use but the appropriate level of PSA that leads to biopsy. Similarly, we might consider the association between statins and obesity and between obesity and overdiagnosis, suggesting confounding as explanation for the study finding. If so, we would ignore statin use altogether and make screening decisions depending on obesity.

Once we have thought through causal pathways, we can design appropriate statistical analyses. In our statin example, it is plausible that the apparent effects of statins might in fact be due to PSA haemodilution related to obesity. In brief, men on statins are more likely to be obese, obese men have greater blood volume, and so a given amount of PSA released from the prostate in nanogrammes will be divided by a greater number of millilitres; this leads to a lower PSA level (measured in ng/mL) in obese men and thus reduces the chance of biopsy for elevated PSA. To address this in a causal inference framework, we could include obesity as a covariate in a multivariable model testing the association between statin use and overdiagnosis, with an interaction term between statin use and PSA levels. Inclusion of obesity as a covariate addresses whether obesity is confounding the association between statin use and overdiagnosis, whereas inclusion of the interaction term would address whether statin use is only associated with a reduction in overdiagnosis among those with lower PSA levels. Alternatively, it could be that statins could reduce overdiagnosis by having a direct effect on PSA levels. To test this hypothesis, we could compare PSA levels based on statin use and then model the effect of statin use on biopsy rates after adjustment for PSA, perhaps with an interaction term to see if the effect of statin use differs based on PSA level. The exact details and methods of these analysis are not critical here: the key point is to illustrate how causal mechanisms are identified and then analyses developed accordingly.

Methods for conducting analyses investigating causality have, unsurprisingly, received detailed attention in the methodological literature [4,5]. Indeed, epidemiologists have spent decades thinking about causality and developing methodologies to establish causal inference, from the well-known Bradford Hill criteria [6] to the Rubin potential outcomes framework [7] to directed acyclic graphs [8], a method of visualising causal pathways so as to help determine the appropriate statistical analysis. The epidemiological literature also includes critiques of methods for causal inference, e.g., urging more cautious application of propensity scores [9] or advising against the use of E-values, a statistic purported to give an estimate of confounding [10].

In light of these considerations, we are introducing a set of guidelines for papers published in our urology journals. These guidelines are for observational studies where there is any reasonable question about causality. Table 1 gives specific examples of language to use and avoid.

1. Authors should be judicious and explicit in their use of causal language. This refers not only to the use of words ‘cause’ or ‘causal’, but to words that imply causality—such as ‘effect’, ‘reduce’, ‘increase’ or ‘impact’—or recommendations that depend on a causal claim, such as that ‘patients should avoid’, ‘doctors should use’ or ‘efforts should be made to increase’. This is absolutely **not** a recommendation to avoid causal language, but to be explicit about causality [3]. That said, one term that should probably be avoided is ‘risk factor’, as this has an uncertain meaning [11].
2. Causality should be discussed in the context of practical action. We think about causes of health states because we want to intervene to improve health. For instance, we know that smoking causes lung cancer (among other diseases) and so we advise patients to quit smoking. Coffee is associated with lung cancer, but this association is mostly likely fully explained by confounding by smoking [12] rather than a direct causal effect, and hence we do not make practical attempts to limit coffee consumption to prevent cancer. Thinking in terms of practical action can sharpen causal thinking. For instance, if the causal mechanism for the association between obesity and aggressive prostate cancer is haemodilution, that would lead to very different practical recommendations than if the causal mechanism is alterations in testosterone metabolism [13].
3. There should be explicit reference to causality in the ‘Introduction’ section. This should be related directly to the study question. It is, of course, appropriate in some cases not to have a causal hypothesis, but this should be made explicit, and a rationale given. For instance, in a study of smoking and surgery outcomes, the authors might state: ‘we will not investigate the causal pathways between smoking and postoperative recovery as our purpose is to allow clinicians to counsel patients about their expected postoperative course’. Characterising a study as descriptive or predictive should have a good rationale and should not be used just as a short-cut to avoid grappling with causal inference.
4. Where the purpose of the study involves exploration of causal mechanisms:
    a. Describe possible causal pathways in the ‘Methods’ section. Although this can be done formally, e.g., using directed acyclic graphs, it is also reasonable to describe causal pathways using ordinary language in the main text. In many cases this section can be relatively brief: take, for instance, an observational study comparing two different treatment modalities. In other cases, authors will need to describe pathways in more detail,

Table 1 Examples of language to use and avoid.

Scenario	Language to use	Language to avoid
Causality not investigated or not demonstrated	• 'Associated with …' • 'Causal mechanisms will require further evaluation'	• 'Affects'/'Impacts'/'Increases'/'Reduces' • 'Surgeons should avoid …'
Introduction section of an observational study	• 'This is a preliminary study to assess the association between statins and risk so as to motivate subsequent research on causal mechanisms, were such an association to be found' • 'We are interested in the variants as prognostic markers, and will not be investigating causal pathways here' • 'We hypothesised that any association between smoking and prostate cancer death would be causally related either to lower uptake of PSA screening in smokers or to overall poorer access to care in this group' • 'The purpose of the study is to determine which surgical approach results in a lower complication rate' (Note: this implies a causal relationship) • 'We aimed to assess the effects of neoadjuvant therapy' (Note: this implies a causal relationship)	• 'We aimed to assess the relationship between…' without discussing whether causality or causal mechanisms will be addressed • 'Our aim was to evaluate the association between neoadjuvant therapy and oncological outcome' (Note: this avoids causality when only the causal question is interesting here)
Covariates not well balanced between groups	• 'As there was a difference in comorbidity between the two surgical approaches, we evaluated the association between comorbidity and surgical outcome' • 'Although statistically significant, the number of high-grade tumours was only slightly lower in the high-dose group and is unlikely to explain the very large difference in oncological outcome' • 'The baseline risk of patients receiving blood transfusions was so much higher than those who did not receive transfusion that any causal inference is unsound. We recommend that questions of transfusion approaches in this population can only be addressed by randomised trials'	• 'Case mix differences do not explain the superior outcomes in the surgery arm because stage, grade and PSA were included in a multivariable model' • 'Our propensity score approach simulates a randomised trial meaning that the estimate for the difference between groups is not subject to confounding' • 'Our E-value is high and therefore unmeasured confounding is unlikely'
Discussion section	• 'There are two possible causal mechanisms for the reported association other than an effect of the drug. The first is … which we evaluated by … and found … The second is … which we evaluated by … reporting … Hence, we conclude that a causal effect of the drug is the best explanation of our findings' • 'The difference between groups is larger than could be reasonably explained by the alternative surgical approaches, suggesting considerable unmeasured or residual confounding' • 'We found an association between X and Y. The causal pathways between X and Y require further elucidation'	• 'We conclude that statins are associated with lower cancer risk' (Note: conclusions should include recommendations for research or clinical practice, such as that the causal mechanisms should be explored in subsequent studies)

carefully describing mediators, confounders, and colliders.

b. Describe statistical methods to address causality. This will often focus on confounding. Authors should go beyond a brief reference to 'adjusting' in a multivariable model and should describe the rationale for their choice of specific covariates in the context of causal pathways. These need not be lengthy, take, for instance, 'In our study comparing different treatments for early-stage prostate cancer, it is plausible that disease aggressiveness might affect choice of treatment. Accordingly, we included stage, grade, and PSA as covariates in the multivariable model'. However, in some cases, discussion of analyses to control for confounding might take several paragraphs, as authors would need to describe relationships between measured covariates and pathways of confounding. For instance, diet is a possible confounder in an epidemiological study of exercise and urological cancer. Brief reference to diet being one of a list of covariates would be insufficient: authors would need to assess how well their measure reflects confounding by diet, particularly with respect to when diet was measured relative to the likely time course of carcinogenesis. Where the causal question is one of mediation, a mediation analysis should be considered [14]. Authors should also note that confounding is not the only threat to causal inference, and other causal criteria should be explored. One criterion is that exposed and unexposed participants should be represented across the distribution of confounders (the 'positivity' assumption). For instance, it would be hard to draw conclusions about the effects of chemotherapy if all (or almost all) chemotherapy patients were Stage 3 or 4 whereas the patients not receiving chemotherapy were all Stage 1 or 2. Another criterion is that exposed

participants have similar levels (or dose) of exposure—this is why studies on smoking look at pack-years rather than smoking status—and also that the exposure of one participant does not affect the outcome of an unexposed participant (the reason why studies of vaccine effectiveness are often done at the community level). This criterion is known as the 'Stable Unit Treatment Value Assumption' or 'SUTVA'.

c. Assess control of confounding in the 'Results' section. As discussed in the guidelines for reporting statistics in urology [15], authors should avoid assuming that once a multivariable or a propensity approach has been implemented, confounding is no longer a concern. One good 'rule-of-thumb' is that 'differences in measured confounders imply differences in unmeasured confounders', where the latter includes both measurement error in covariates (sometimes called 'residual' confounding) and variables not included in the analysis. For instance, in a study with a survival endpoint, researchers will not record all comorbidities (such as a rare genetic disorder); moreover, a measured comorbidity (say, diabetes) will often be recorded as present or absent, even though the severity of diabetes can vary between patients. Accordingly, a key analytical step is to compare confounders between exposure groups, e.g., in a table 1. If, for instance, the prevalence of diabetes in higher in one group than another, it is also likely that the severity of diabetes will be higher in that group. As a result, including diabetes in a multivariable model will not completely control for differences between groups in the case where diabetes severity is associated with survival. Authors might also evaluate how well the measured covariates predict outcome. For instance, the higher the discrimination, the lower the likelihood of unmeasured confounding. That said, we are suspicious of analyses that aim to assess confounding numerically, such as the E-value [10], or evaluation of a binary, unmeasured covariate [16]: the influence of confounding on judgements of causality cannot in our view be reduced to one or two numbers.

d. Carefully assess causal inference and causal estimates in the 'Discussion' section. The discussion section should evaluate causal inference in light of the causal pathways described in the 'Introduction' and 'Methods' sections, the findings reported in the 'Results' section and the relevant scientific literature, whether comparable observational studies or basic science research. Control of confounding and other sorts of bias should be a major consideration. Authors should also reflect on the size of the causal estimate. For instance, in one study, high coffee consumption was associated with a >50% reduction in the risk of lethal prostate cancer [17]. This is far higher than the chemopreventive effects of pharmaceutical agents targeting specific carcinogenic pathways, e.g., tamoxifen and raloxifene reduce the risk of breast cancer by about 40%, and is thus implausible.

e. Draw implications for research and/or clinical practice in light of causal findings. Authors should avoid general references to 'association' in their conclusion section and instead make recommendations for research or clinical practice. Such recommendations need to be explicit, such as mechanistic investigation of causality as a research implication or use of a treatment as a clinical practice implication. Implications also need to be specific. Authors should avoid vague calls for further research and instead give details of how such research should be conducted given the particular findings of the current study.

In conclusion, the current literature too often dances around causal issues, avoiding direct reference and tiptoeing back and forth between cautious language about association ('correlation does not imply causation') and incautious language that strongly implies a causal relationship ('patients should avoid X'). We propose these guidelines to ensure that observational research papers clearly reference causality when describing study aims, and, when the aim is causal, to keep causality first and foremost in design, analysis, interpretation, and conclusions. Following these guidelines will help ensure causal conclusions are applied judiciously and communicated clearly so as to help clinicians and patients make better decisions about health.

Funding

This work was supported in part by the National Institutes of Health/National Cancer Institute (NIH/NCI) with a Cancer Center Support Grant to Memorial Sloan Kettering Cancer Center (P30 CA008748), and a Specialized Programs of Research Excellence (SPORE) grant in Prostate Cancer to Dr H. Scher [P50-CA92629], the Sidney Kimmel Center for Prostate and Urologic Cancers.

Disclosure of Interests

None declared.

References

1 **Parra CO**, **Bertizzolo L**, **Schroter S**, **Dechartres A**, **Goetghebeur E**. Consistency of causal claims in observational studies: a review of papers published in a general medical journal. *BMJ Open* 2021; 11: e043339

2 **Haber NA**, **Wieten SE**, **Rohrer JM et al.** Causal and associational linking language from observational research and health evaluation literature in practice: a systematic language evaluation. *Am J Epidemiol* 2022; 191: 2084–97.

3 **Hernan MA**. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health* 2018; 108: 616–9

4 **Rohrer JM**. Thinking clearly about correlations and causation: graphical causal models for observational data. *Adv Methods Pract Psychol Sci* 2018; 1: 27–42

5 **Goetghebeur E**, **le Cessie S**, **De Stavola B**, **Moodie EE**, **Waernbaum I**. Formulating causal questions and principled statistical answers. *Stat Med* 2020; 39: 4922–48

6 **Hill AB**. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58: 295–300

7 **Little RJ**, **Rubin DB**. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health* 2000; 21: 121–45

8 **Tennant PWG**, **Murray EJ**, **Arnold KF et al.** Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol* 2021; 50: 620–32

9 **Harhay MO**, **Donaldson GC**. Guidance on statistical reporting to help improve your chances of a favorable statistical review. *Am J Respir Crit Care Med* 2020; 201: 1035–8

10 **MacLehose RF**, **Ahern TP**, **Lash TL**, **Poole C**, **Greenland S**. The importance of making assumptions in bias analysis. *Epidemiology* 2021; 32: 617–24

11 **Huitfeldt A**. Is caviar a risk factor for being a millionaire? *BMJ* 2016; 355: i6536

12 **Galarraga V**, **Boffetta P**. Coffee drinking and risk of lung cancer-a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2016; 25: 951–7

13 **Klaassen Z**, **Howard LE**, **Moreira DM**, **Andriole GL Jr**, **Terris MK**, **Freedland SJ**. Association of obesity-related hemodilution of prostate-specific antigen, dihydrotestosterone, and testosterone. *Prostate* 2017; 77: 466–70

14 **Vickers AJ**, **Steineck G**. Prognosis, effect modification, and mediation. *Eur Urol* 2018; 74: 243–5

15 **Assel M**, **Sjoberg D**, **Elders A et al.** Guidelines for reporting of statistics for clinical research in urology. *Eur Urol* 2019; 75: 358–67

16 **Rosenbaum PR**, **Rubin DB**. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J R Stat Soc B Methodol* 1983; 45: 212–8

17 **Wilson KM**, **Kasperzyk JL**, **Rider JR et al.** Coffee consumption and prostate cancer risk and progression in the health professionals follow-up study. *J Natl Cancer Inst* 2011; 103: 876–84

Correspondence: Andrew Vickers, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave, New York, NY 10017, USA.

e-mail: vickersa@mskcc.org