Supplementary Information

Title: The reporting of study limitations and hedging in articles with and without causal diagrams

Taísa Rodrigues Cortes*[1], Ismael Henrique Silveira[2], Michael Maia Schlussel[3] , Guilherme Loureiro Werneck[1,4] and Claudio José Struchiner[1,5.]

1. Institute of Social Medicine, State University of Rio de Janeiro, Rio de Janeiro, Brazil.
2. Institute of Collective Health, Federal University of Bahia, Salvador, Brazil.
3. UK EQUATOR Centre, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, Oxfordshire, UK.
4. Institute for Public Health Studies, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.
5. School of Applied Mathematics, Getúlio Vargas Foundation, Rio de Janeiro, Brazil.

File S2. Causal models

For the development of our causal DAGs, we adopted the procedures proposed by Ferguson et al., (2020), Petersen and van der Laan, (2014) and Tennant et al., (2021).

a) **Mapping:** constructing a fully connected (saturated) model based on a review of the literature:

First, we conducted a literature review to identify factors potentially related to the communication of causal assumptions and study limitations. As a general hypothesis, we assumed that attitudes to the reporting of causal assumptions and uncertainty in scientific articles are influenced by *perception*-related factors and communication-related *incentives* and *constraints*. Previous research indicates that several individual factors, such as personality traits, domain-specific knowledge, and beliefs about knowledge, may influence our perception of uncertainty (Sorrentino and Roney, 1999; Shuper et al., 2004; Bendixen and Rule, 2004). Other individual characteristics, such as the number of authors and the nationality and expertise of the author team, may also influence attitudes and expectations about how research should be conducted and communicated, and these have been associated with the quality of reports in the biomedical literature (Jin et al., 2018).

Previous research also suggests that the expression of uncertainty in academic articles may vary according to author language (see Kourilova-Urbanczik, 2012). Authors whose native language is not English may transfer the regular discursive patterns of their native language and culture to their English-language articles, which can lead to differences and failures in the application of rhetorical resources, such as hedging (Varttala, 2001).

Besides time other constraints (such as journal submission forms and space limits), the communication of causal assumptions and uncertainty in research articles can also be affected by the (real or perceived) risks associated with reporting those. For instance, Puhan et al. (2012) argued that, in the current academic culture, many authors might not wish to discuss study limitations because they perceive a transparency threshold beyond which manuscript acceptance declines. Researcher expectations regarding how bias and uncertainty will be perceived and used by the media, politicians, and other scientists can also influence the manner in which they communicate their findings (Gustafson & Rice, 2020; Wiedemann et al., 2021). This concern appears to be reinforced by instances in which doubt has served the interests of sponsors, such as tobacco companies, food producers, and fossil fuel companies (Oreskes and Conway, 2010; Elliot and Steel, 2017). Moreover, given the empirical evidence indicating that, in some societies, confidence in science is currently declining (Price and Peterson, 2016), the perception of risks associated with reporting research bias and uncertainty may be exacerbated. On the other hand, recent findings suggest that researchers' self-criticism may have positive effects on the public's trust in scientists (Altenmüllern et al., 2021).

After identifying the main potential causes of exposure and outcome reported in the literature, we constructed a (highly connected) DAG (Figure S2-a), in which all variables are connected by arrows, with the exception of those connections considered highly implausible (Ferguson et al., 2020).

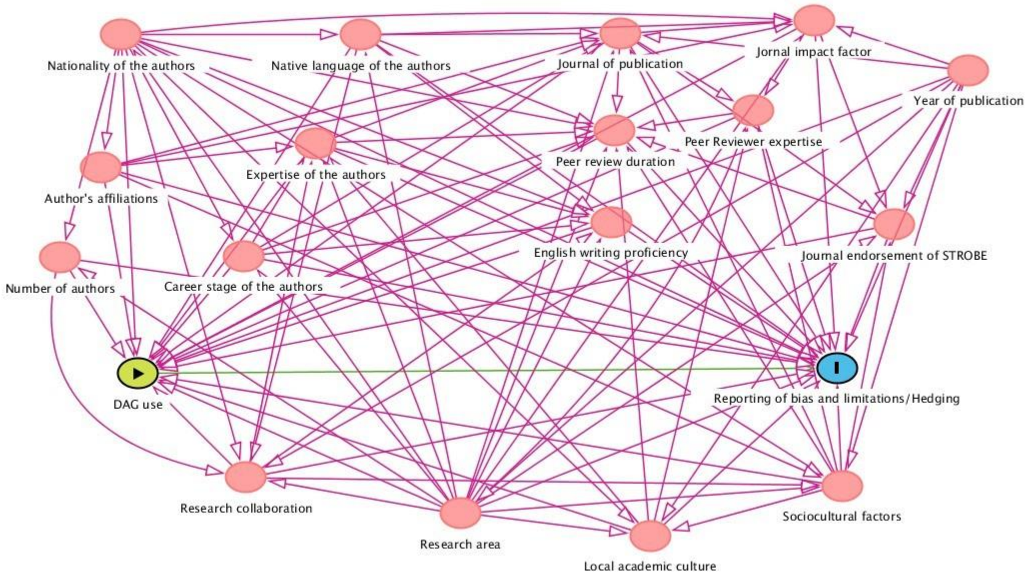

Figure S2-a. Directed acyclic graph representing the outcome variable as blue node, the exposure as yellow node and confounders as red nodes.

b) Translation and Integration: revision of each directed relationship and recombination of node/variables

After mapping the main potential causes of the exposure and outcome, we then examined each directed relationship displayed in the previous DAG. Guidelines for general evaluation included consideration of directionality (whether the direction of the arrow is plausible or should be reversed), whether there is evidence (of null effect) for deleting an arrow, and the possibility of node/variable recombination to reduce model complexity (Ferguson et al., 2020). At this stage, all variables were regarded as measured for the entire population (no sample selection, missing data, etc).

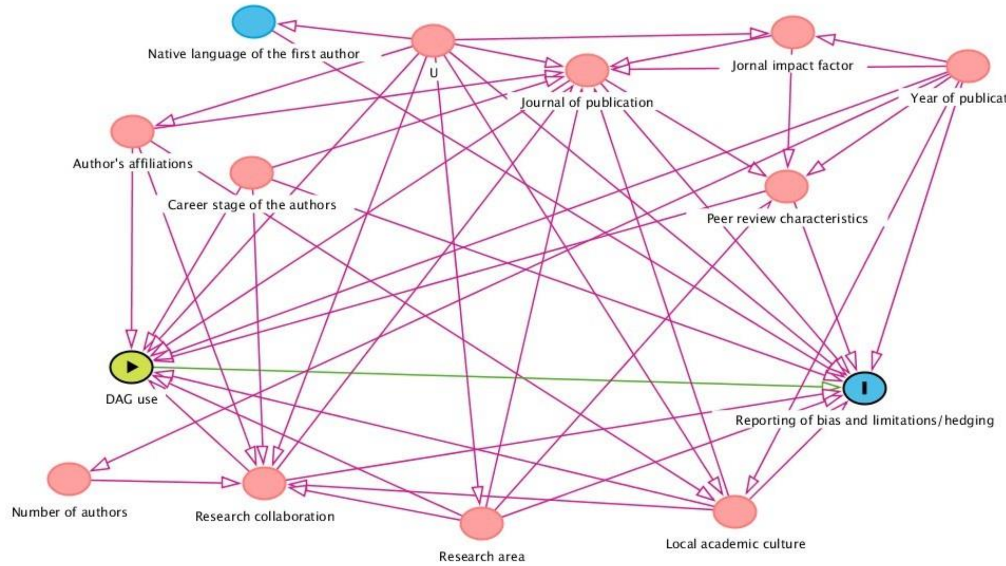

Figure S2-b. Directed acyclic graph representing the outcome variable as blue node, the exposure as yellow node and confounders as red nodes.

This "ideal" DAG (Figure S2-b), in which all variables are measured, entails one minimally sufficient adjustment set for estimating the total effect of DAG use on the Reporting of study bias and limitations, and hedging:

- *Career stage of the authors, Journal of publication, Local academic culture, Peer review characteristics, Research area, Research collaboration, U, Year of publication.*

c) Final model: specify the observed data and their connection to the causal DAG

After identifying the background knowledge and developing an "ideal" DAG, we defined the connection between the observed data and the model structure (Petersen and van der Laan, 2014). At this stage, we considered not only the aspects related to confounding and measurement error but also the potential determinants of study participation or data selection (for example, which variables are related to study selection and exclusion criteria, missingness, etc.).

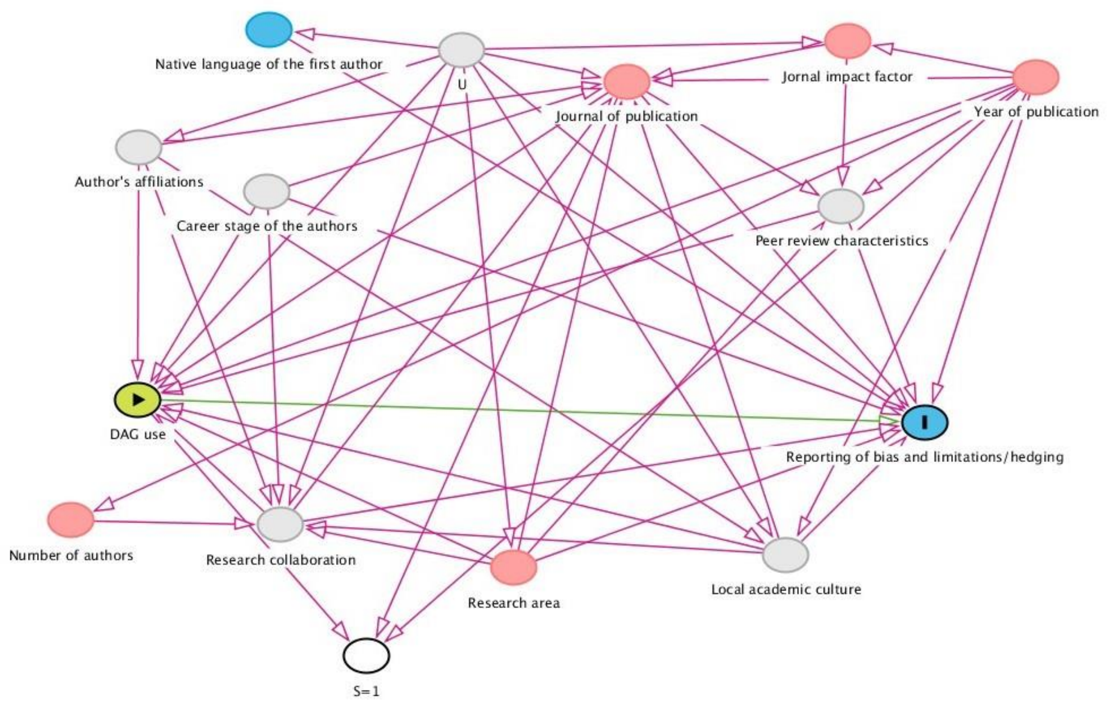

Figure 1 Directed acyclic graph for the relationship between DAG use and reporting of study limitations in scientific articles. The yellow node represents the exposure of interest, while the blue nodes indicates the outcome (or the ancestor of the outcome). Gray nodes represent unmeasured factors. The selection white node S=1 indicates that the analysis is conditional on the inclusion in the matched study.

Our final DAG implies that the causal effect of DAG use on the outcomes is non-identifiable. In other words, the total effect cannot be (distinctly) estimated from the causal model (the causal structure and available data), due to backdoor paths between exposure and outcome that include only unobserved variables, such as the paths *Dag use ← Career stage of the authors → Reporting of bias and limitations*, and *Dag use ← Local academic culture → Reporting of bias and limitations*.

We therefore considered a subset of variables for confounding adjustment, which included some of the matching variables and other observed confounding factors or their proxies: *journal impact factor, research area, publication year, and native laguage of the first author*.

While balance matching makes the exposure independent of the matching variables (journal, year of publication and issue), this independence can be compromised by ignoring these variables when making other adjustments (Sjolander & Greenland, 2013). However, adjusting by publication journal would require multiple dummy variables which, given our sample size, may lead to sparse data bias. Therefore, we decided to adjust by journal impact factor as opposed to journal publication (and issue). A matched analysis, such as conditional logistic regression, would be an option for managing sparse data and was explored as sensitivity analysis.

Several other simplifying assumptions were made in our final analysis. For instance, in our adjusted models, the native language of the first author (predicted by their name) was considered a proxy, not only for English writing proficiency, but also for local academic culture and other related sociocultural factors (teaching model, academic incentives, etc.) which could influence both exposure and outcomes. However, a more appropriate metric would be to determine the affiliation and nationality of all the co-authors based on affiliation data acquired over their careers (Fanelli et al., 2015).

Our main analysis also assumes that the influences of the peer review process and editorial policy, such as review length, and reviewer and editor characteristics, can be measured by journal impact factor, research area, and year of publication. Moreover, the number of co-authors was viewed as a proxy for research collaboration.

At this stage, the consistency between the DAG and the study data was evaluated using conditional independence tests (Textor et al., 2017). However, only a portion of the model is testable, due to its high connectivity and the presence of unobserved variables. In addition, multiple causal structures are compatible with the same set of conditional independences. Due to the matching design, cancellation (unfailfulness) of certain paths is also expected. The model in figure 1 implies the following conditional independences:

Testable implications	**p-value**
Jornal impact factor ⊥ Number of authors | Year of publication	0.063
Native language of the first author ⊥ Number of authors	0.112
Native language of the first author ⊥ Year of publication	0.233
Number of authors ⊥ Research area	0.883
Research area ⊥ Year of publication	0.174

References

Altenmüller MS, Nuding S, Gollwitzer M. No harm in being self-corrective: Self-criticism and reform intentions increase researchers' epistemic trustworthiness and credibility in the eyes of the public. Public Underst Sci. 2021 Nov;30(8):962–76.

Bendixen, L. D., & Rule, D. C. An integrative approach to personal epistemology: A guiding model. Educational Psychologist, 2004. 39(1), 69-80.

Elliott KC, & Steel D (Eds.). Current controversies in values and science. Routledge, Taylor & Francis Group, 2017.

Fanelli D, Costas R, Larivière V. Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. Wray KB, editor. PLoS ONE. 2015 Jun 17;10(6):e0127556

Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, et al. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. International journal of epidemiology. 2020;49(1):322–9

Gustafson A, Rice RE. A review of the effects of uncertainty in public science communication. Public Underst Sci. 2020 Aug;29(6):614–33.

Jin Y, Sanger N, Shams I, Luo C, Shahid H, Li G, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years?A systematic review of reviews: an update. JMDH. 2018 Sep; Volume 11:495–510.

Kourilova-Urbanczik M. Some linguistic and pragmatic considerations affecting science reporting in English by non-native speakers of the language. Interdiscip Toxicol. 2012; 5(2):105-115. doi:10.2478/v10102-012-0018-1

Oreskes, N., & Conway, E. M.. Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming. 2011. Bloomsbury Publishing USA.

Petersen ML, & van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. Epidemiology. 2014. *25*(3), 418.

Price, A. M., & Peterson, L. P. Scientific progress, risk, and development: Explaining attitudes toward science cross-nationally. International Sociology, 2016. 31(1), 57-80.

Puhan MA, Akl EA, Bryant D, Xie F, Apolone G, Riet G. Discussing study limitations in reports of biomedical studies- the need for more transparency. Health Qual Life Outcomes. 2012;10(1):23.

Shuper PA, Sorrentino RM, Otsubo Y, Hodson G, Walker AM. A Theory of Uncertainty Orientation: Implications for the Study of Individual Differences Within and Across Cultures. Journal of Cross-Cultural Psychology. 2004 Jul;35(4):460–80.

Sjölander A, & Greenland S. Ignoring the matching variables in cohort studies–when is it valid and why?. Statistics in medicine, 2013. 32(27), 4696-4708.

Sorrentino RM, Roney CJR. The uncertain mind: Individual differences in facing the unknown. Psychology Press, 1999

Tennant PW, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. International journal of epidemiology. 2021;50(2):620–32.

Textor J, Hardt J, shKnüppel S. DAGitty: A Graphical Tool for Analyzing Causal Diagrams. Epidemiology. 2011 Sep;22(5):745.

Varttala, T. Hedging in scientifically oriented discourse: Exploring variation according to discipline and intended audience. 2001. Doctoral Dissertation, University of Tampere. https://urn.fi/urn:isbn:951-44-5195-3

Wiedemann P, Boerner FU, Freudenstein F. Effects of communicating uncertainty descriptions in hazard identification, risk characterization, and risk protection. Murakami M, editor. PLoS ONE. 2021 Jul 13;16(7):e0253762.