IEA International Epidemiological Association

Cancer

Vibration of effects in epidemiologic studies of alcohol consumption and breast cancer risk

Lingzhi Chu,[1] John PA Ioannidis,[2,3,4,5] Alex C Egilman,[6] Vasilis Vasiliou,[1] Joseph S Ross[7,8,9,10] and Joshua D Wallach [1,6]*

[1]Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA, [2]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA, [3]Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA, [4]Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA, [5]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA, USA, [6]Collaboration for Research Integrity and Transparency (CRIT), Yale Law School, New Haven, CT, USA, [7]Section of General Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA, [8]National Clinician Scholars Program, Yale School of Medicine, New Haven, CT, USA, [9]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA and [10]Center for Outcomes Research and Evaluation, Yale-New Haven Health System, New Haven, CT, USA

*Corresponding author. Department of Environmental Health Sciences, Yale School of Public Health, 60 College Street, 4th Floor, Room 411, New Haven, CT 06510, USA. E-mail: joshua.wallach@yale.edu

Abstract

Background: Different analytical approaches can influence the associations estimated in observational studies. We assessed the variability of effect estimates reported within and across observational studies evaluating the impact of alcohol on breast cancer.

Methods: We abstracted largest harmful, largest protective and smallest (closest to the null value of 1.0) relative risk estimates in studies included in a recent alcohol–breast cancer meta-analysis, and recorded how they differed based on five model specification characteristics, including exposure definition, exposure contrast levels, study populations, adjustment covariates and/or model approaches. For each study, we approximated vibration of effects by dividing the largest by the smallest effect estimate [i.e. ratio of odds ratio (ROR)].

Results: Among 97 eligible studies, 85 (87.6%) reported both harmful and protective relative effect estimates for an alcohol–breast cancer relationship, which ranged from 1.1 to 17.9 and 0.0 to 1.0, respectively. The RORs comparing the largest and smallest estimates in value ranged from 1.0 to 106.2, with a median of 3.0 [interquartile range (IQR) 2.0–5.2]. One-third (35, 36.1%) of the RORs were based on extreme effect estimates with at least three different model specification characteristics; the vast majority (87, 89.7%) had different exposure definitions or contrast levels. Similar vibrations of effect were observed

when only extreme estimates with differences based on study populations and/or adjustment covariates were compared.

Conclusions: Most observational studies evaluating the impact of alcohol on breast cancer report relative effect estimates for the same associations that diverge by >2-fold. Therefore, observational studies should estimate the vibration of effects to provide insight regarding the stability of findings.

Key Messages

- Different analytical approaches can influence the associations observed in observational studies.
- Three-quarters of the observational studies evaluating the impact of alcohol on breast cancer risk reported extreme relative effect estimates for the same associations that diverged by over 2-fold.
- Approximately one-third of the extreme effect estimates reported in each study had at least three different main model specification characteristics, including exposure definitions, exposure contrast levels, study populations, adjustment covariates and/or model approaches.
- To provide insight regarding stability and generalizability of findings, observational studies should approximate the vibration of effects across a range of different analytical approaches.

Introduction

Alcohol consumption, which has been associated with dozens of acute and chronic diseases, is considered a leading risk factor for global disease burden.[1] Whereas the adverse effects of heavy drinking are well documented,[2,3] the impact of low to moderate consumption is complicated, and studies have suggested both protective and harmful impacts on health, depending on the volume and pattern of consumption.[1] Due to the controversial relationship between alcohol and health (especially when alcohol is consumed at low levels), and the wide prevalence of alcohol consumption,[4] research in this area is of great interest for public health.[1] However, research evidence often comes from observational studies, which do not ensure causality and have several innate study design limitations.

Observational studies are susceptible to confounding, which can distort the relationship between an exposure and outcome.[5–7] Not all authors studying the same exposure–outcome relationships will measure and/or consider the same potential adjusting variables. Furthermore, different model specification characteristics, including the selection of exposure contrasts and reference groups, use of variable transformations and/or the handling of outliers, as well as population and outcome definitions, can have an impact on the results observed in a study.[8,9] For instance, focusing on certain subgroups, like the relationships between alcohol consumption and risk of death in different age groups, can lead to different conclusions.[10] The presence of financial biases in some studies (e.g. those funded by industry)[11] or allegiance bias for specific theories (including white hat bias[12]) may fuel the choice of analyses and results that fit to some specific agenda.

The variability of effect estimates due to these alternative analytical approaches has been referred to as the 'vibration of effects' (VoE),[9] and previous studies have presented approaches to quantify vibration of effects, including taking the ratio of the largest vs the smallest effect on the same association with alternative analytical selections (i.e. vibration ratio).[9,13] Ideally, raw data should be used to generate the full distribution of effect estimates that can be obtained within a study,[13] but raw data are still rarely available for observational studies.[14,15] However, publications reporting different effects on the same association within the same paper can also be used to examine the difference between the extremes of reported effect estimates that have been obtained in the same study using different analytical approaches. It is possible that many different analytical options are pursued and compared after the data have been explored, but only a few are eventually isolated and reported when a paper is published. Reported findings may or may not suffer from selective reporting bias. Meta-analyses traditionally try to identify, summarize and compare effect estimates from studies with relatively similar exposures, populations and adjustment variables,

even if, in the absence of individual-level data, these analyses may still differ across included studies. Vibration of effects analyses based on reported extreme effect estimates offer a partial view of the full vibration of effects that would consider all possible analyses that could be done with various datasets on the same questions.

Currently, little is known about how different model specifications can influence the observed associations between alcohol and health outcomes. One area where the association has been particularly unclear is the relationship between alcohol consumption and breast cancer risk. Some studies have suggested a J-shaped association,[16,17] indicating that at low alcohol consumption levels, there is no association or protective effects, but at higher levels, the risk increases as the amount of alcohol consumption increases. However, other studies have indicated null,[18,19] weak[20] or monotonically increasing[1] relationships. To evaluate the extent that effect estimates for the relationship between alcohol and breast cancer can vary, we approximated the vibration of effects for reported results by comparing the largest and the smallest relative effect estimates reported in each study across a large sample of published observational studies.

Methods

Data identification and eligibility

We identified all 102 observational studies on alcohol use and breast cancer included in the 2016 Global Burden of Disease, Injuries, and Risk Factors Study (GBD), which included meta-analyses of alcohol use and 23 different health outcomes.[1] Considering that the GBD authors performed a recent and comprehensive search of PubMed, the Global Health Data Exchange and references of published meta-analyses to identify cohort or case-control studies reporting three common relative measures of risk (i.e. odds ratio, hazard ratio and relative risk) on breast cancer and dose–response amounts on alcohol consumption,[1] we did not perform a new, separate meta-analysis. Duplicates and studies without relative effect estimates on broadly-defined alcohol consumption and breast cancer were excluded.

Data extraction

One researcher (L.C.) screened the full text of all articles and recorded the title, year of publication, journal name, study design (i.e. case-control vs cohort study), cohort name, study country, population subgroup, study period, age range, sample size, number of breast cancer cases and source of funding (i.e. governmental and/or other non-profit organizations only, including industry or none reported). We also determined the 2017 impact factor of each publication's journal in Journal Citation Reports. No

information was recorded for journals without a 2017 impact factor. All uncertainties were discussed with the senior author (J.D.W.).

For each study, we identified and extracted the largest harmful, largest protective and smallest (closest to the null value of 1.0) relative risk estimates corresponding to alcohol exposure and breast cancer related outcomes. For example, if a study reported relative effect estimates of 1.1, 2.3, 0.99, 0.7, the largest harmful, largest protective and smallest values would be 2.3, 0.7 and 0.99, respectively. Reported relative risk estimates of broadly-defined alcohol consumption and breast cancer were all considered eligible, regardless of whether measures of precision, such as confidence intervals or *P*-values, were provided. All relative risk estimates were standardized to reflect a comparison of higher alcohol exposure vs lower (or no) alcohol exposure. For each estimate, we recorded the type of estimate (e.g. odds ratio, relative risk or hazard ratio) and recorded: (i) how alcohol exposure was defined (exposure definition), (ii) how alcohol exposure was measured (exposure contrast levels), (iii) which covariates were included in the multivariate model (adjustment covariates), (iv) the study population considered (subgroups), and (v) which model approach was used. When reported, we also abstracted the corresponding *P*-values and 95% confidence intervals. When extreme estimates with the exact same magnitude were identified in a study, we randomly selected one estimate.

Data analysis

Descriptive statistics were conducted to summarize the characteristics of eligible studies, including study design, study area and number of breast cancer cases.

Calculation of vibration ratio

All identified relative risks were assumed to be interchangeable, and thus metrics that were not odds ratios were considered to be good approximations to the odds ratio, which is a reasonable assumption because breast cancer incidence is relatively uncommon.[21] To illustrate the range of effect estimates reported in individual studies on the association of alcohol and breast cancer, all extreme effect estimates were presented in a forest plot using the 'ggplot2' package in R (version 3.5.2; The R Project for Statistical Computing), with the largest harmful, largest protective and smallest (closest to the null value of 1.0) relative risk estimates from the same eligible studies in the same row for direct comparison. Then, we estimated the vibration ratio by the relative odds ratio [ratio of odds ratios (RORs)], which is obtained by dividing the largest estimate in value by the smallest estimate in value in each study. When studies only reported harmful estimates, the

largest harmful estimate was divided by the smallest harmful estimate closest to the null. For studies with only protective estimates, we divided the smallest protective estimate closest to the null by the largest protective estimate in magnitude. The ROR represents how much reported relative risk estimates for the same exposure–outcome relationship within a study change based on different model specifications. The forest plots for the RORs were created using the 'forestplot' package in R. For each ROR, we then recorded which of the five main model specification differences there were between the largest and the smallest estimates and recorded the number of RORs >1.2, 1.5, 2.4 and 10.0.

We visualized potential relationships between RORs and certain study features using linear regression, and calculated Pearson correlation coefficients for RORs with continuous study characteristics (journal impact factor and log-transformed number of breast cancer cases). Kruskal-Wallis tests with a confidence level of 0.005[22] were used to assess the relationship between the RORs and study type, metric type, study area and funding source type.

Sensitivity analysis

To evaluate the consistency of the observed vibration of effects, analyses were repeated (i) excluding studies where the extreme estimates differed based on exposure definition or contrast levels (pre-specified) and (ii) including only studies where the extreme estimates only differed in exposure definition or contrast levels (*post hoc*). As suggested during peer review, we also calculated the 95% confidence intervals for the RORs. Considering that the extreme effect estimates were from the same study, we calculated the variance of each ROR assuming correlations of 0.5, 0.75 and 1.0. RORs were combined using the DerSimonian and Laird procedure for random effects.

Results

Study characteristics

The GBD alcohol and breast cancer meta-analysis referenced 102 studies, of which 1 was a duplicate and 4 did not report any relative effect estimates for broadly-defined alcohol consumption and breast cancer associations. The 97 eligible studies were published between 1984 and 2012, in journals with a median impact factor of 4.55 (IQR 3.61–7.36) (Table 1, Supplementary Table S1, available as Supplementary data at *IJE* online). There were 80 studies (82.5%) funded by governmental and/or other nonprofit organizations. Nearly two-thirds (59, 60.8%) were

Table 1. Summary of study characteristics

Characteristic	Case-control *n* (%)	Cohort *n* (%)	Total *n* (%)
Total studies	59 (60.8)	38 (39.2)	97 (100.0)
Publication year			
<1990	11 (18.6)	2 (5.3)	13 (13.4)
1990–1999	21 (35.6)	9 (23.7)	30 (30.9)
2000–2009	19 (32.3)	20 (52.6)	39 (40.2)
≥2010	8 (13.6)	7 (18.4)	15 (15.5)
Impact factor			
≤3	14 (23.7)	6 (15.8)	20 (20.6)
>3–5	19 (32.3)	15 (39.5)	34 (35.1)
>5–10	19 (32.3)	8 (21.1)	27 (27.8)
>10	7 (11.9)	9 (23.7)	16 (16.5)
Funding source			
Governmental and/or other nonprofit organizations only	49 (83.1)	31 (81.6)	80 (82.5)
Including industry	2 (3.4)	2 (5.3)	4 (4.1)
None reported	8 (13.6)	5 (13.2)	13 (13.4)
Study area			
North America	27 (45.8)	18 (47.4)	45 (46.4)
Europe	24 (40.7)	11 (28.9)	35 (36.1)
Asia	3 (5.1)	6 (15.8)	9 (9.3)
Other	5 (8.5)	3 (7.9)	8 (8.2)
Sample size			
Median (IQR)	1943 (1107-3188)	55 387 (14 167–103 631)	4622 (1618–22 200)
Number of cases			
Median (IQR)	859 (444-1571)	483 (248–1274)	740 (349–1508)

case-control studies, with a median sample size of 1943 (IQR 1107–3188); 38 (39.2%) were cohort studies, with a median sample size of 55 387 (IQR 14 167–103 631). The vast majority of the studies were conducted in North America (45, 46.4%) and Europe (35, 36.1%).

Distribution of extreme effect estimates

Of the 97 eligible studies, 85 (87.6%) reported both harmful and protective relative risk estimates, 11 (11.3%) reported only harmful estimates, and 1 (1.0%) reported only protective estimates (Fig. 1; the figure with 95% confidence intervals is provided in Supplementary Figure S1, available as Supplementary data at IJE online). Among the 11 studies reporting only harmful estimates, there was one study with only one eligible effect estimate. Across all 97 studies, the largest harmful and protective estimates ranged from 1.1 to 17.9 [median 2 (IQR 1.5–3.2)] and 0.0 to 1.0 [median 0.7 (IQR 0.5–0.8)], respectively. There were six (6 of 85, 7.1%) studies where the largest harmful and protective estimates were reported without a confidence interval or P-value. The smallest (closest to the null value of 1.0) reported estimates ranged from 0.8 to 1.4 [median 1.0 (IQR 1.0–1.0)].

Vibration of effects

The RORs, which were obtained by dividing the largest and smallest effect estimates in value in each of the 97 eligible studies, ranged from 1.0 to 106.2 (Fig. 2), with a median of 3.0 (IQR 2.0–5.2; see summary ROR and 95% confidence interval in Supplementary Figure S2a–c, available as Supplementary data at IJE online). There were 94 (96.9%) RORs that were >1.2, 87 (89.7%) that were >1.5, 65 (67.0%) that were >2.4, and 9 (9.3%) that were >10.0. Among the 97 RORs, 35 (36.1%) were based on extreme effect estimates with at least three different main model specification characteristics (examples in Table 2). Although the vast majority (87, 89.7%) of the extreme effect estimates reported in each study had different contrast levels (e.g. <3, 3–7, 7+ drinks of liquor per week and ex-drinkers vs <3, 3–7, 7+ drinks of beer per week and ex-drinkers), there were 43 (44.3%) with different exposure definitions, 46 (47.4%) with different population subgroups, 35 (36.1%) with different covariates and 1 (1.0%) with a different model approach. Approximately one-third (34, 35.1%) of RORs were based on extreme effect estimates where the only differences were the exposure definition and/or contrast levels. One (1.0%) ROR was equal to 1.0 and had all the same model specifications because there was only one eligible effect estimate reported in that study.

RORs were not associated with journal impact factor, number of cases, study type, study area, measure type and funding source (Fig. 3). However, among case-control studies, the RORs decreased as impact factor and number of breast cancer cases increased, and the corresponding Pearson correlation coefficients were −0.199 (95% confidence interval −0.439 to 0.067) and −0.089 (95% confidence interval −0.338 to 0.171), respectively. For cohort studies, there was no clear pattern of changes with impact factor (−0.039, 95% confidence interval −0.359 to 0.288) and with number of cases (0.151, 95% confidence interval −0.178 to 0.449).

There were 34 studies (34 of 97, 35.1%) where the RORs were based on extreme effect estimates that differed only on exposure definition and/or exposure contrast, and the median ROR was 2.00 (IQR 1.65–2.89, Supplementary Fig. S3, available as Supplementary data at IJE online). Among the 9 (9 of 97, 9.3%) RORs based on extreme effect estimates with only different study populations and/or adjustment covariates, the median ROR was 3.69 (IQR 3.17–9.00) (Fig. 4).

Discussion

Our analysis found a wide vibration of effects within and across 97 individual observational studies evaluating the impact of alcohol consumption on breast cancer risk. Nearly all studies reported both harmful and protective relative risk estimates for broadly-defined alcohol consumption and breast cancer associations, and nearly three-quarters had extreme estimates that diverged more than 2-fold. Approximately one-third of the extreme effect estimates reported in each study had at least three different main model specification characteristics. Although the vast majority of extreme effect estimates had different exposure definitions or contrast levels, similar vibration of effects were observed when only extreme effect estimates with differences based on study populations and/or adjustment covariates were compared. Vibration of effects were smaller when only extreme effect estimates with differences based on exposure definitions and/or exposure contrasts were compared. These findings suggest that whereas certain analytical and modelling choices, reflecting different types of alcohol and/or doses, can result in genuine differences, it is possible that many different analytical options, with different results, are pursued and selectively reported. Therefore, individual reported relative risk estimates from observational studies should be interpreted with caution.

Within and across studies evaluating the impact of alcohol on breast cancer, there are multiple factors that can contribute to vibration of effects. In our evaluation, we found that many extreme effect estimates had different exposure definitions and contrast levels. Previous studies have outlined difficulties in measuring alcohol
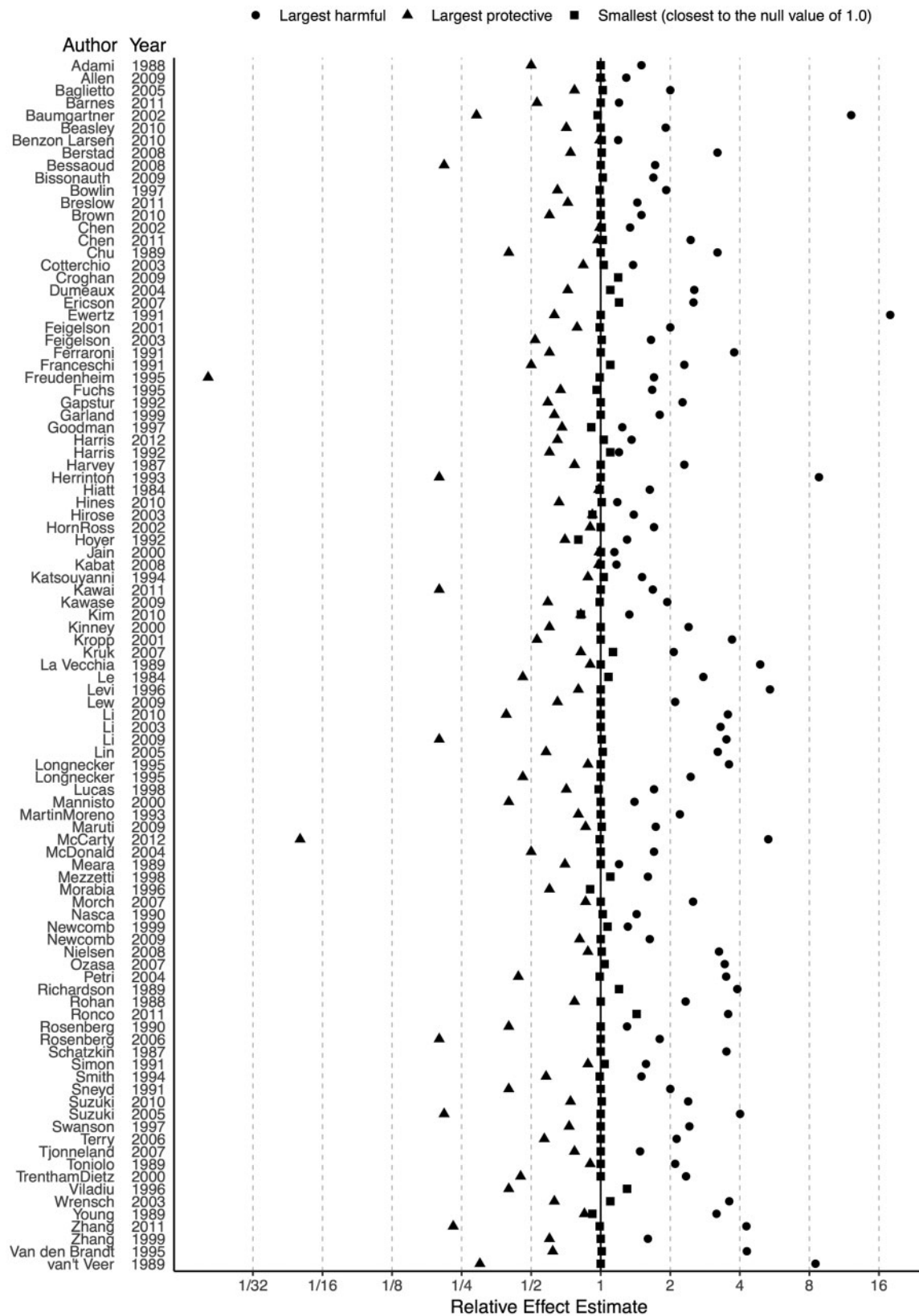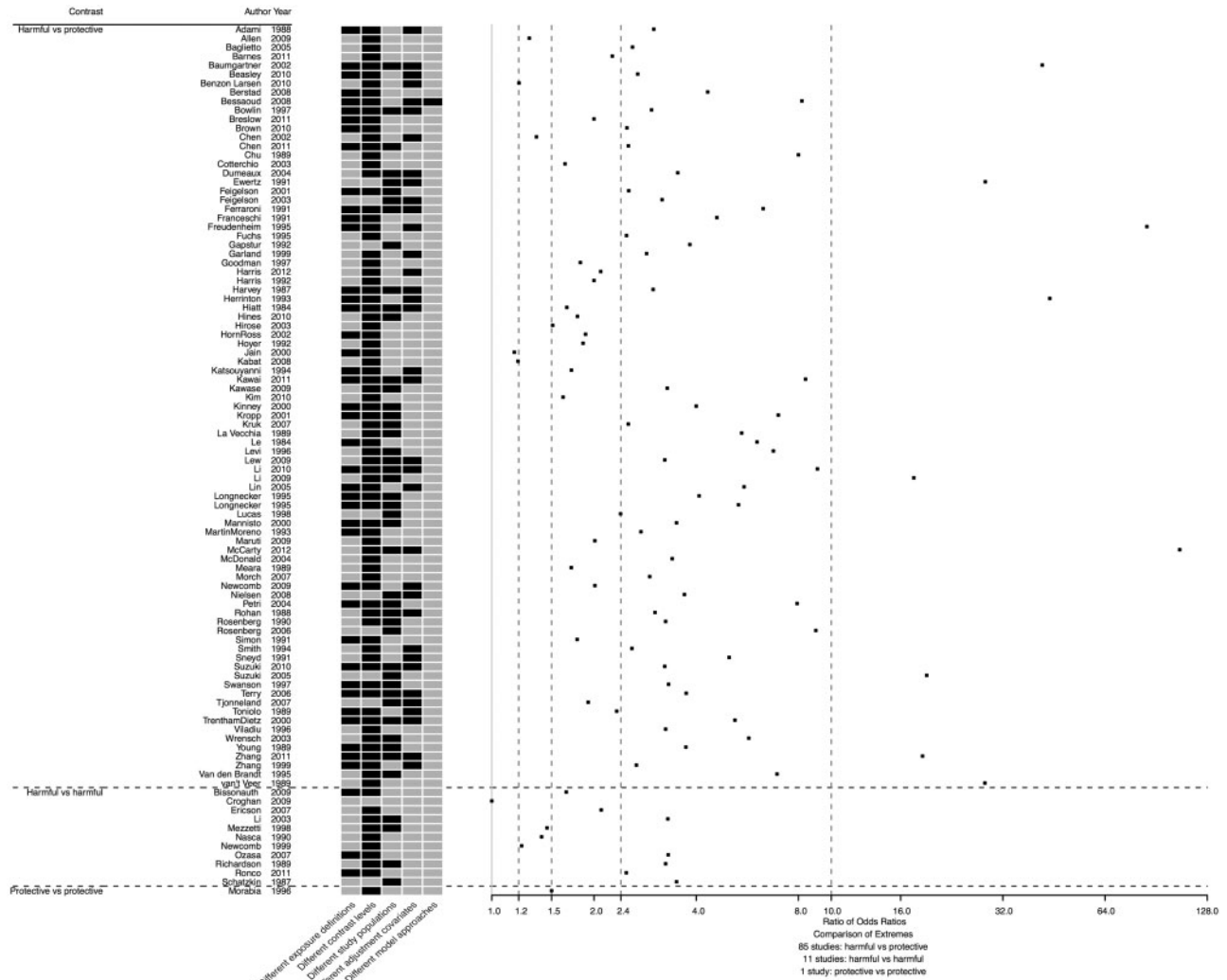
Figure 1. Largest harmful, largest protective and smallest (closest to the null value of 1.0) relative risk estimates in the 97 observational studies evaluating the effect of alcohol consumption on breast cancer risk.

Figure 2. Scatter plot of ratio of odds ratios for the 97 observational studies evaluating the effect of alcohol consumption on breast cancer risk, with a map of model specification differences. Black cells indicate observed differences.

consumption (e.g. using different beverage types, grams per day or units per week),[25] selecting categories of consumption and establishing reference groups.[26] For instance, some analyses may use a combined reference group of 'never drinker' and 'former drinker'. However, when the 'former drinker' group includes 'sick quitters', who may have stopped drinking due to poor health outcomes, harmful outcomes can be observed in the reference group.[27] Other evaluations have suggested that the J-shaped curve for alcohol consumption and health-related outcomes disappears after accounting for 'sick quitters' bias.[28,29] Although it can be expected that the highest levels of alcohol consumption will lead to more harmful health outcomes (e.g. larger effect estimates),[1] other model characteristics, such as choice of adjusting variables and subgroups can alter observed estimates substantially, and can contribute to both harmful or protective associations within the same analyses.[13,30,31]

There are some potential limitations in our study. First, with 97 eligible studies, our results may not be generalizable to all observational studies evaluating the impact of alcohol (let alone other postulated risk factors) on health outcomes. The potential impact of vibration of effects needs to be carefully considered and dissected in diverse fields of observational epidemiology.[32–34] Second, considering that our evaluation was based on studies identified by a recent meta-analysis, we may have missed some eligible studies. However, the GBD researchers carefully searched two main databases as well as the references of previously published meta-analysis, and it is unlikely that additional articles would influence our overall findings. Third, we focused on the extreme reported estimates, and ideally raw data should be used to generate the full distribution of effect estimates that can be obtained within a study.[13] If anything, our estimates are underestimates of the potential vibration of effects that can be achieved, but

Table 2. Two illustrative examples of extreme estimates and relative odds ratio calculations; BMI, body mass index
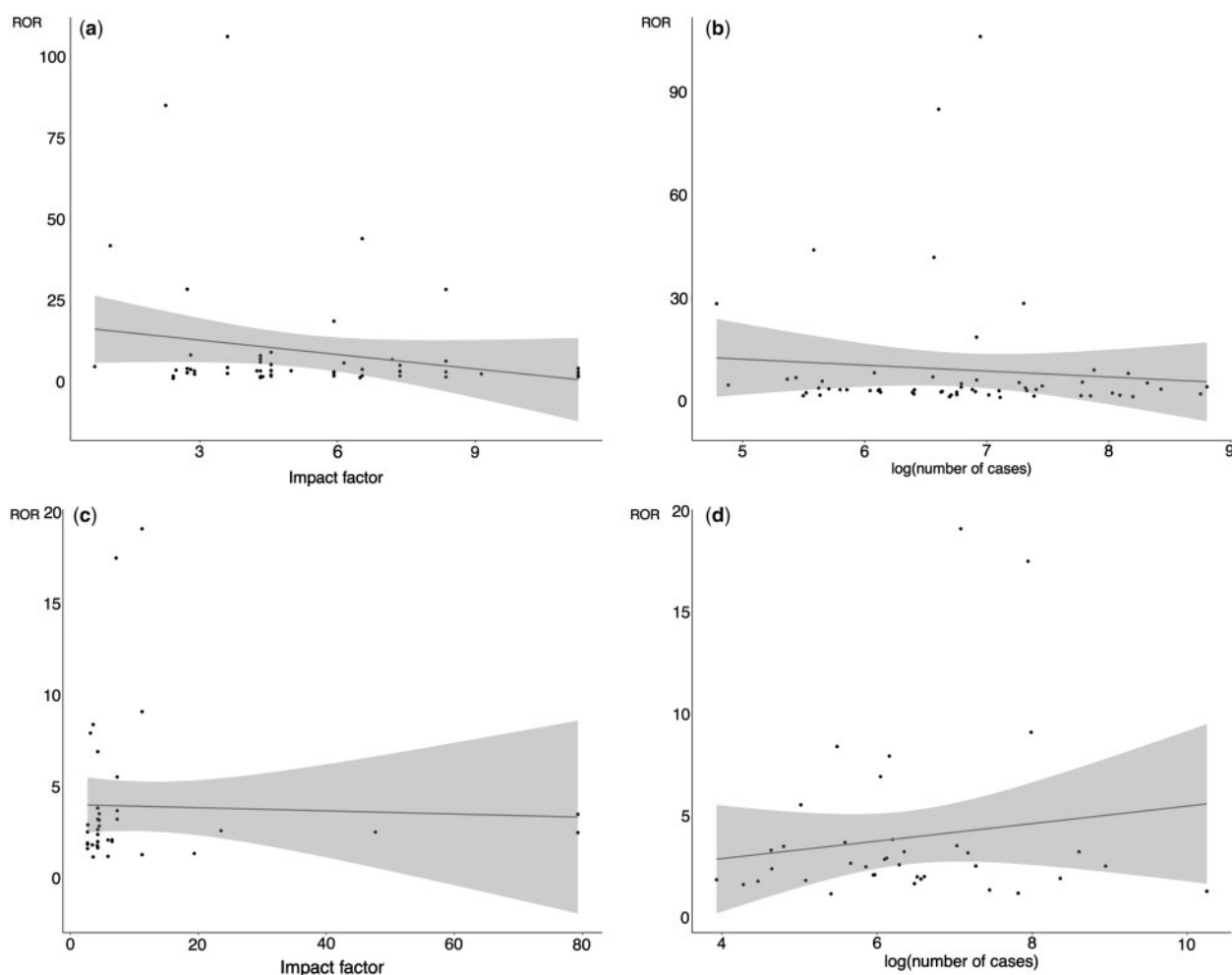
Author, year	Extreme effect estimate	Exposure definition	Exposure categories	Contrast levels	Subgroups	Adjustment covariates	Effect estimate	Measure type	Ratio of odds ratio
23	Largest harmful	Beer, bottles/day	0, 1, 2+	2+ vs 0	None	None	1.5	Odds ratio	3.00
	Largest protective	Total alcohol, g/day	0, 0.1–1.2, 1.3–4.9, 5.0–14.9, 15+	15+ vs 0	None	Education, age at menarche, age at first full-term pregnancy, parity, menopause, history of operation for benign breast disease, family history of breast cancer, total duration of oral contraceptive use, smoking, and the consumption of alcoholic beverages other than those analysed	0.5	Odds ratio	
24	Largest harmful	Average daily alcohol intake	Abstain, <4g/day, ≥4 g/day	≥4 g/day vs abstain	BMI ≤22.89 kg/m²	Age	2.26	Relative risk	3.83
	Largest protective	Average daily alcohol intake	Abstain, <4g/day, ≥4 g/day	≥4 g/day vs abstain	Age at first livebirth (years) ≤19	Age	0.59	Relative risk	

they have the advantage that they represent real, published analyses rather than possible, but unpublished analyses. Fourth, considering our focus on extreme effect estimates, we did not expect many of the same adjustment variables to be considered across different analyses. Therefore, we did not attempt to identify adjustment variables or analyse patterns according to the variables examined. Lastly, some vibration may reflect legitimate changes based on exposure definitions and exposure contrast, but it is difficult to disentangle genuine differences from study-level biases. For instance, our findings were consistent after excluding RORs that were more likely to represent genuine differences that could result from exposure definitions and exposure contrasts, but RORs were smaller when only extreme effect estimates with differences based on exposure definitions and/or exposure contrasts were compared. Further studies, using raw data, should evaluate the attributed vibration of each component in the analyses.

To provide insight regarding the stability of claimed association and minimize selective reporting, authors should clarify their selected model specifications, including exposure definition, contrast levels, adjustment covariates and population subgroups. In particular, directed acyclic graphs (DAGs) could be used to discuss measured and unmeasured confounders prior to conducting analyses.[35] Observational studies should report the median and range of relative risk estimates and P-values across a large number of sensitivity analyses. Presenting the pattern of the vibration of effects when different assumptions are made can offer a broad view of the impact of sensitivity analyses. Furthermore, results from vibration of effects analyses can be used to inform causal inference, such as disentangling the relationships between various exposures definitions, covariates and outcomes, and guiding Mendelian randomization and natural experiment studies.[36]

The large vibration of effects means that very different results can be obtained based on what analytical and modelling choices are made. The vibration is typically much larger than the usual effect sizes of relative risks reported in alcohol studies of breast cancer. This suggests that the potential analytical noise is much greater than the potential signals; it is often debated as to whether they are null, protective or consistently harmful.[1,37] Ideally, to contain these inadvertent degrees of freedom in the analyses, pre-registration with fully detailed specification of the analyses should be considered.[38,39] However, for most epidemiological studies to-date, pre-registration is either not done or done in spurious ways, e.g. studies are registered after they are completed, which obviously offers no guarantee of any protection from these biases.[40] There is also debate about whether pre-registration for many observational studies is even feasible.[41,42] Furthermore, when

Figure 3. Scatter plot of ratio of odds ratios with (a) impact factor in case-control studies, (b) log(number of cases) in case-control studies, (c) impact factor in cohort studies, and (d) log(number of cases) in cohort studies.

datasets are already collected and analyses can be done at any time, it is difficult to ensure that the analyses have not already been explored when a study is seemingly pre-registered. Nevertheless, pre-registration of methods should still be a legitimate option if done before data collection and/or before any access to the collected data is granted for analysis. The sharing of raw data used for analyses will also increase transparency, thereby allowing investigators to better understand the impact of using different exposure definitions.

Given that these practices are rarely if ever adopted to-date in the field of alcohol and cancer risk assessment, one has to be careful about making strong statements about the validity of the published estimates of risk. In the presence of very strong opinions and beliefs in the field of alcohol exposure and cancer research, there is a risk that the literature may be shaped by the opinions of researchers, reviewers and editors, picking the results of analyses that fit best their preconceived theories. In that case, the published, seemingly objective quantitative data may still

reflect mostly subjective expert opinions, and the synthesis of data may really represent a form of expert vote-counting instead of rigorous quantitative synthesis. Therefore, observational studies should estimate the vibration of effects to provide insight regarding the stability of findings.
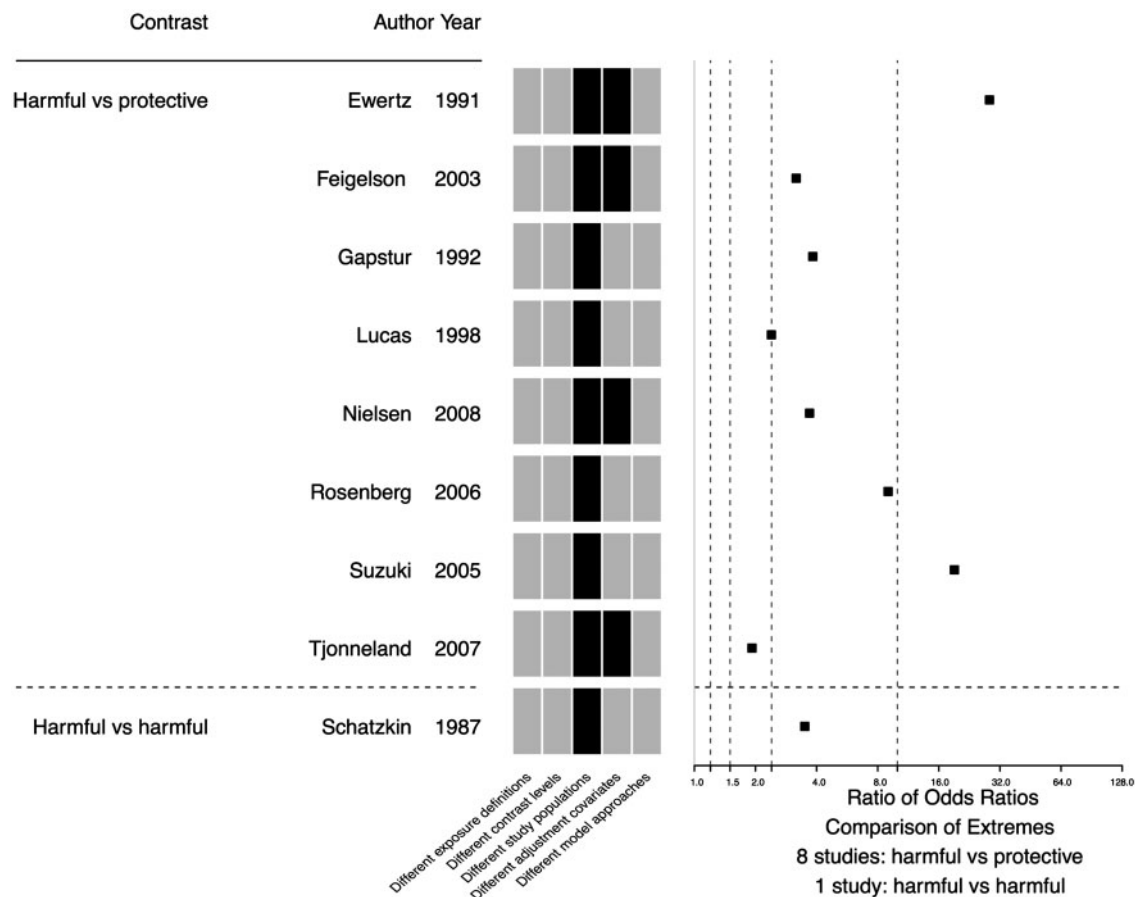
Supplementary data

Supplementary data are available at *IJE* online.

Funding

Author contributions

J.D.W. initiated the study. L.C. analysed the data. All authors interpreted the results. L.C. extracted the data. L.C. and J.D.W. wrote the first draft and all authors made revisions on the article. All

Figure 4 Scatter plot of ratio of odds ratios from 9 studies where the extreme effect estimates differed only on study populations, adjustment covariates and/or model approaches, with a map of model specification differences. Black cells indicate observed differences.

authors read and approved the final version of the article. J.D.W. supervised the research. All authors had full access to all the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. J.D.W. is the guarantor. L.C. and J.D.W. have checked the references for accuracy and completeness.

References

1. Griswold MG, Fullman N, Hawley C *et al*. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2018;**392**:1015–35.
2. O'Keefe JH, Bybee KA, Lavie CJ. Alcohol and cardiovascular health: the razor-sharp double-edged sword. *J Am Coll Cardiol* 2007;**50**:1009–14.
3. Ezzati M, Vander Hoorn S, Lopez AD *et al*. Comparative quantification of mortality and burden of disease attributable to selected risk factors. In: Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL (eds). *Global Burden of Disease and Risk Factors, Vol*. **2**. Oxford: Oxford University Press, 2006, pp. 241–396.
4. Ahrnsbrak R, Bose J, Hedden S, Lipari R, Park-Lee E. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2016 National Survey on Drug Use and Health*. Rockville, MD: Center for Behavioral Health Statistics

and Quality, Substance Abuse and Mental Health Services Administration, 2017.

5. Psaty BM, Koepsell TD, Lin D *et al*. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999;**47**:749–54.

6. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;**359**:248–52.

7. Hemkens LG, Ewald H, Naudet F *et al*. Interpretation of epidemiologic studies very often lacked adequate consideration of confounding. *J Clin Epidemiol* 2018;**93**:94–102.

8. Young SS, Karr A. Deming, data and observational studies. *Significance* 2011;**8**:116–20.

9. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;**19**:640–8.

10. White IR, Altmann DR, Nanchahal K. Alcohol consumption and mortality: modelling risks for men and women at different ages. *BMJ* 2002;**325**:191.

11. Friedman LS, Richter ED. Relationship between conflicts of interest and research results. *J Gen Intern Med* 2004;**19**:51–6.

12. Cope MB, Allison DB. White hat bias: examples of its presence in obesity research and a call for renewed commitment to faithfulness in research reporting. *Int J Obes* 2010;**34**:84.

13. 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**:1046–58.

14. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible research practices and transparency across the biomedical literature. *PLoS Biol* 2016;**14**:e1002333.

15. Wallach JD, Boyack KW, Ioannidis JP. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLoS Biol* 2018;**16**:e2006930.

16. Williams LA, Olshan AF, Hong C-C *et al*. Alcohol intake and breast cancer risk in African American Women from the AMBER Consortium. *Cancer Epidemiol Biomarkers Prev* 2017;**26**:787–94.

17. Kropp S, Becher H, Nieters A, Chang-Claude J. Low-to-moderate alcohol consumption and breast cancer risk by age 50 years among women in Germany. *Am J Epidemiol* 2001;**154**:624–34.

18. Kinney AY, Millikan RC, Lin YH, Moorman PG, Newman B. Alcohol consumption and breast cancer among black and white women in North Carolina (United States). *Cancer Causes Control* 2000;**11**:345–57.

19. Kawai M, Minami Y, Kakizaki M *et al*. Alcohol consumption and breast cancer risk in Japanese women: the Miyagi Cohort study. *Breast Cancer Res Treat* 2011;**128**:817–25.

20. Freudenheim JL, Marshall JR, Graham S *et al*. Lifetime alcohol consumption and risk of breast cancer. *Nutrition and Cancer* 1995;**23**:1.

21. Golubnitschaja O, Debald M, Yeghiazaryan K *et al*. Breast cancer epidemic in the early twenty-first century: evaluation of risk factors, cumulative questionnaires and recommendations for preventive measures. *Tumor Biol* 2016;**37**:12941–57.

22. Ioannidis JP. The proposal to lower P value thresholds to. 005. *JAMA* 2018;**319**:1429–30.

23. Adami HO, Lund E, Bergström R, Meirik O. Cigarette smoking, alcohol consumption and risk of breast cancer in young women. *Br J Cancer* 1988;**58**:832–7.

24. Gapstur SM, Potter JD, Sellers TA, Folsom AR. Increased risk of breast cancer with alcohol consumption in postmenopausal women. *Am J Epidemiol.* 1992;**136**:1221–31.

25. Mukamal JK. Overview of the risks and benefits of alcohol consumption. In: Elmore JG, Kunins L (ed). *UpToDate*. Waltham, MA: UpToDate Inc. https://www.uptodate.com/contents/overview-of-the-risks-and-benefits-of-alcohol-consumption (June 2018, date last accessed).

26. Naimi TS, Stockwell T, Zhao J *et al*. Selection biases in observational studies affect associations between 'moderate'alcohol consumption and mortality. *Addiction* 2017;**112**:207–14.

27. Hensing G, Holmgren K, Mårdby A-C. Harmful alcohol habits were no more common in a sample of newly sick-listed Swedish women and men compared with a random population sample. *Alcohol Alcohol* 2011;**46**:471–7.

28. Stockwell T, Zhao J, Panwar S, Roemer A, Naimi T, Chikritzhs T. Do "moderate" drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. *J Stud Alcohol Drugs* 2016;**77**:185–98.

29. Liang W, Chikritzhs T. Observational research on alcohol use and chronic disease outcome: new approaches to counter biases. *Sci World J* 2013;**2013**:1.

30. Corrao G, Rubbiati L, Bagnardi V, Zambon A, Poikolainen K. Alcohol and coronary heart disease: a meta-analysis. *Addiction* 2000;**95**:1505–23.

31. Roerecke M, Rehm J. Alcohol consumption, drinking patterns, and ischemic heart disease: a narrative review of meta-analyses and a systematic review and meta-analysis of the impact of heavy drinking occasions on risk for moderate drinkers. *BMC Med* 2014;**12**:182.

32. Haslam A, Prasad V. Multivitamins do not reduce cardiovascular disease and mortality and should not be taken for this purpose: how do we know that? *Circ Cardiovasc Qual Outcomes* 2018;**11**:e004886.

33. Monteith S, Glenn T, Geddes J, Whybrow PC, Bauer M. Big data for bipolar disorder. *Int J Bipolar Disord* 2016;**4**:10.

34. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. *Perspect Psychol Sci* 2016;**11**:702–12.

35. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999;**10**:37–48.

36. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;**45**:1866–86.

37. Corrao G, Bagnardi V, Zambon A, La Vecchia C. A meta-analysis of alcohol consumption and the risk of 15 diseases. *Prev Med* 2004;**38**:613–9.

38. Should protocols for observational research be registered? *Lancet* 2010;**375**:348. doi: 10.1016/S0140-6736(10)60148-1.

39. Loder E, Groves T, MacAuley D. Registration of observational studies. The next step towards research transparency. *BMJ* 2010;**340**:c950.

40. Boccia S, Rothman KJ, Panic N *et al*. Registration practices for observational studies on ClinicalTrials. gov indicated low adherence. *J Clin Epidemiol* 2016;**70**:176–82.

41. Lash TL, Vandenbroucke JP. Commentary: Should preregistration of epidemiologic study protocols become compulsory? Reflections and a counterproposal. *Epidemiology* 2012;**23**:184–8.

42. Dal-Ré R, Ioannidis JP, Bracken MB *et al*. Making prospective registration of observational research a reality. *Sci Transl Med* 2014;**6**:224cm1.