Published in final edited form as:

Addiction. 2015 January ; 110(1): 183–184. doi:10.1111/add.12708.

Sub-group analysis as a source of spurious findings: An illustration using new data on alcohol intake and coronary heart disease

Steven Bell^{1,*}, Mika Kivimäki¹, and G. David Batty^{1,2}

¹Research Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London, UK, WC1E 6BT

²Centre for Cognitive Aging and Cognitive Epidemiology, University of Edinburgh, UK

Ware and Munafò [1] discuss the important topic of chasing statistical significance and note that unplanned *post hoc* analyses do not equate to serious scientific misconduct. We agree, but want to emphasise that the practice can nonetheless be very misleading – particularly when the unplanned analyses become the focal point of a paper [2]. With ever increasing pressure to publish, the practice of taking well-established associations and examining them within strata of a third variable, also known as sub-group analyses, is becoming popular [3,4], yet the social or biological rationale for anticipating that the relationship will differ between sub-groups is rarely stated [5]. Ultimately, *post hoc* sub-group analyses can lead to spurious conclusions being drawn [6,7] that may mislead public health and clinical policy [8].

To illustrate this point, we present Cox proportional hazard analyses for the association between moderate alcohol intake and major coronary heart disease (CHD) events during 24 years of follow-up in the Whitehall II prospective cohort [9]. A total of 420 coronary events were observed during follow-up in 8517 men and women of mean age 45.1 years at baseline.

Taking the epidemiological literature together (and setting aside the limitations of this evidence base [10,11]), people who drink in moderation seem to experience a reduced CHD risk [12,13]. Consistent with this result, in new analyses of data in the Whitehall II study, study members who abstained from alcohol at baseline had a 1.47-fold (95% CI 1.17–1.84) increased risk of major CHD compared to those drinking within UK recommended weekly drinking limits (up to or less than 21 or 14 UK units per week for men and women, respectively). At this stage, we might wish to ascertain if there are any groups that do *not* benefit from moderate consumption. As Ware and Munafò [1] state, this may seem like a well-intentioned question if we wish to advise particular groups of people that the apparent beneficial effects of moderate drinking may not apply to them or, alternatively, an unplanned post-hoc analysis without clear justification.

*Corresponding author: steven.bell@ucl.ac.uk.

Conflict of interest: None to declare.

Here, we take the latter approach to an extreme by stratifying by Sign of the Zodiac [14,15]. As shown in Figure 1, if we interpret the estimates literally and do not test for heterogeneity of effect [6,7], the association between moderate drinking and major CHD seems only to exist for people born under the sign of the Capricorn and Cancer. For all others, there is no association between moderate drinking and CHD. Are there any plausible reasons for this? No. Do we think a campaign focusing the dangers of abstinence that specifically targets people born under these star signs should be rolled out? Again, no.

Even in the current academic environment where the requirement to publish is considerable [16], self-restraint is needed if we are to minimise these unhelpful and possibly even harmful findings infiltrating the literature. We agree with the suggestions proposed by Ware and Munafò [1] that sub-group analyses, when there is a rationale for anticipating the relationship to differ between groups, should be stated before testing, as is the case in randomised controlled trials with registered protocols [17] (there are also good examples of observational studies that require users to register their study and publish their protocol before they are provided with data [18]). If sub-group analyses are to be used to generate hypotheses (although one could argue that as an association was examined in the first place, a hypothesis was generated and subsequently tested from the outset [19]), then the procedures used in genetics, that is, correction for multiple testing and requirement for replication, may help avoid false conclusions [17].

Acknowledgments

Funding statement: SB is funded by the European Research Council (309337, PI: Britton, <http://www.ucl.ac.uk/alcohol-lifecourse>) and MK is supported by an ESRC professorial fellowship. The Whitehall II study is supported by grants from the Medical Research Council (K013351), British Heart Foundation (RG/07/008/23674), Stroke Association, National Heart Lung and Blood Institute (HL036310) and National Institute on Aging (AG13196 and AG034454). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. SB had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Ware JJ, Munafò MR. Significance chasing in research practice: causes, consequences and possible solutions. *Addiction: Advance Access*. 2014
2. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA J Am Med Assoc*. 2004; 291:2457–2465.
3. Canoy D, Cairns BJ, Balkwill A, Wright FL, Green J, et al. Body mass index and incident coronary heart disease in women: a population-based prospective study. *BMC Med*. 2013; 11:87. [PubMed: 23547896]
4. Ferrari P, Licaj I, Muller DC, Kragh Andersen P, Johansson M, et al. Lifetime alcohol use and overall and cause-specific mortality in the European Prospective Investigation into Cancer and nutrition (EPIC) study. *BMJ Open*. 2014; 4:e005245.
5. Feinstein AR. The Problem of Cogent Subgroups: A Clinicostatistical Tragedy. *J Clin Epidemiol*. 1998; 51:297–299. [PubMed: 9539884]
6. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet*. 2005; 365:176–186.
7. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med*. 2007; 357:2189–2194. [PubMed: 18032770]

8. Sleight P. Debate: Subgroup analyses in clinical trials: fun to look at - but don't believe them! *Curr Control Trials Cardiovasc Med*. 2000; 1:25–27. [PubMed: 11714402]
9. Marmot M, Brunner E. Cohort Profile: The Whitehall II study. *Int J Epidemiol*. 2005; 34:251–256. [PubMed: 15576467]
10. Fekjær HO. Alcohol—a universal preventive agent? A critical analysis. *Addiction*. 2013; 108:2051–2057. [PubMed: 23297738]
11. Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, et al. Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ*. 2014; 349
12. Ronksley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ*. 2011; 342:d671. [PubMed: 21343207]
13. Roerecke M, Rehm J. The cardioprotective association of average alcohol consumption and ischaemic heart disease: a systematic review and meta-analysis. *Addiction*. 2012; 107:1246–1260. [PubMed: 22229788]
14. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *The Lancet*. 1988; 332:349–360.
15. Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *J Clin Epidemiol*. 2006; 59:964–969. [PubMed: 16895820]
16. Anderson M, Ronning E, De Vries R, Martinson B. The Perverse Effects of Competition on Scientists' Work and Relationships. *Sci Eng Ethics*. 2007; 13:437–461. [PubMed: 18030595]
17. Kivimäki M, Singh-Manoux A, Ferrie JE, Batty GD. Post hoc decision-making in observational epidemiology—is there need for better research standards? *Int J Epidemiol*. 2013; 42:367–370. [PubMed: 23569177]
18. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, et al. Data Resource Profile: Cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). *Int J Epidemiol*. 2012; 41:1625–1638. [PubMed: 23220717]
19. Greenland S. Commentary: On “Quality in epidemiological research: should we be submitting papers before we have the results and submitting more hypothesis generating research?”. *Int J Epidemiol*. 2007; 36:944–945. [PubMed: 17954715]

Figure 1.

Hazard ratios (95% confidence intervals) for experiencing a major CHD event during 24 years of follow-up in abstainers compared to moderate drinkers by zodiac sign in the Whitehall II study (N=8517; adjusted for age, gender, socioeconomic position and smoking status).