**Bias in matched case-control studies: DAGs are not enough**

Neil Pearce

Department of Medical Statistics and Centre for Global NCDs
London School of Hygiene and Tropical Medicine
United Kingdom

Address for correspondence:
Professor Neil Pearce
Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT
United Kingdom
E-mail: neil.pearce@lshtm.ac.uk
Website: http://www.lshtm.ac.uk

December 2017

There is perhaps no other procedure in epidemiology which is so apparently simple, but so practically and theoretically complex as matching in case-control studies. The paper by Mansournia et al [1] in this issue of the Journal is the most recent in a long line of papers which explore and clarify the methodological issues involved, including a previous paper by the same first author [2]. The issues and solutions discussed are not original, but Mansournia et al nevertheless make a major contribution in systematically and comprehensively addressing the key issues associated with matching in case-control studies. Unfortunately, it seems that each generation of epidemiologists needs to learn these concepts all over again.

Mansournia et al make extensive reference to a paper which I recently published in the British Medical Journal[3], in which I wrote that matching can 'introduce confounding by the matching factors…'. They note that the bias introduced by matching is actually selection bias, not confounding. Thus, unadjusted estimates in matched case-control studies involve a combination of confounding by the matching variable (already present in the source population) and selection bias by the matching variable (introduced by the selective study recruitment, including the matching). Both biases can be removed by adjusting for the matching variable, just like other 'confounder' adjustments. Thus, other authors have referred to 'selection confounding'[4 5]), and my own rather lazy use of language (and the space constraints of a short BMJ piece), led me to describe matching bias as confounding, rather than selection bias, an error which I am pleased to take the opportunity to correct here.

The original draft of my paper included a Directed Acyclic Graph (DAG) which showed 'that the bias introduced by matching was in fact selection bias (see figure 1 – the labelling has been changed to make it consistent with that of Mansournia et al), but unfortunately this was not included in the published version of the paper. One of the frustrating things I have experienced in teaching matched case-control studies is that methodological issues difficult to explain in words, or with numerical examples, can easily be addressed in DAGs, but this requires the audience to be familiar with those methods.

However, it is also striking that DAGs are not useful for all of the problems of bias in matched case-control studies (or many other epidemiological problems), and sometimes are not useful at all. This mirrors more general recent debates about the strengths and limitations of DAGs[6-13]. I would therefore like to take the opportunity to explore these issues further with regards to matched case-control studies, and the five main issues discussed by Mansournia et al. In this context, it is of interest that Mansournia et al use DAGs to address only some of these issues, and use statistical theory and numerical examples to address the others.

*Bias introduced by case-control matching is an intentional selection bias*

As noted above, matching in case-control studies introduces selection bias, since selection into the study (S) creates a backdoor pathway from E to C to S to D. This can only be removed by controlling for the matching factor (C). Figure 1 illustrates the power of DAGs to address these issues. The figure shows both confounding in the source population (the direct arrows from C to E and D), and the selection bias introduced by the matching (E-C-S-D). Both can be removed by control for the matching factor (C). However, the DAG does not

tell us the likely strength and direction of the two biases; the confounding could be in any direction (in table 1 it is strongly negative[1 3]) whereas the matching bias is always towards the null, but you cannot tell this directly from this DAG.

A related striking finding from Mansournia et al (see Figure 2 in their paper) is that there is no net bias after matching if the exposure has no effect on the disease (i.e. there is no direct arrow from E to D). In this situation, the confounding in the source population (E-C-D) is exactly cancelled out by the selection bias introduced by the matching (E-C-S-D)[2]. The reasons for this are not immediately clear, and are certainly not established by the DAG (why should these two biases be in opposite directions and of the same magnitude?), which is presumably why Mansournia et al use other evidence and arguments to establish this point.

*Adjustment of matching variables should account for both the actual matching protocol and further confounding effects*

Adjustment for the matching variable blocks both the backdoor path through selection (E-C-S-D), and the 'standard' confounding backdoor path (E-C-D). The DAG makes it clear that controlling for the matching factor (e.g. age), blocks both of these paths, but it does not make it clear that a different level of precision may be required in the two adjustments. For example, if there is strong confounding by age, then five-year age-groups (or smaller) may be required to adjust for confounding, even if the matching (and resulting selection bias) has only been done in ten-year age-groups. In general, control for selection bias should involve at least as much precision as was involved in the original matching [14], although in practice such rigorous precision may not always be required[3]; for example, if controls have been

matched on exact age (to the day), control for age may still be quite adequate if five-year age-groups are used. This could be represented in a DAG (David Richardson, personal communication) by having two versions of 'C' in the DAG (the coarse version C*, and the finer version C), but this solution relies on external knowledge to draw the correct DAG (of course, this is true of all DAGs). A related issue is that DAGs are non-parametric, and as Mansournia et al note, cannot easily address situations where adjusting for age as a continuous variable does not adequately control for age, because precise age matching creates a discontinuous 'saw-tooth' age-disease association[1].

*Identically matched sets should be collapsed together*

This was the main point of my BMJ paper, i.e. that a pair-matched design does not necessarily require a pair-matched analysis. Once again, this is not readily apparent from the DAG, and relies on statistical theory[1]. This can be represented in the DAG by having a coarse (C*) and a finer(C) version of the matching factor in the DAG (David Richardson, personal communication).

*Case-control matching on a non-confounder associated with disease may lead to selection bias*

If there were no direct arrow from C to D, then there is no confounding, but selection bias is introduced by the matching process, creating the backdoor path through S. This is an example where DAGs can help clarify an important methodological issue.

*Matching may lead to overadjustment, thus harming precision or creating uncorrectable bias*

If the matching factor has little or no association with the disease (i.e. there is no arrow, or just a weak arrow, from C to D), but is strongly associated with exposure, and is also strongly associated with selection (because of matching), the result of matching is to create a strong backdoor pathway through the selection/matching process. In extreme situations, blocking this backdoor pathway results in a loss of precision, because there are very few instances where the cases or controls are discordant with respect to exposure. Once again, the DAG cannot show this easily, since it is not (easily) possible to indicate the strengths of the various associations.

On the other hand, DAGs can clearly show us why certain variables are causally inappropriate for adjustment (such as an intermediate variable, or a variable affected by both exposure and disease)[15 16] They thus identify variables which should be avoided for matching, since matching on them will cause both the unadjusted and matching-adjusted analyses to be biased[1].

*Conclusions*

So where does this leave us? I seem to spend half my time advocating the increased use of DAGs, and the other half warning about the dangers of their overuse[13]. DAGs have clarified a number of key concepts in epidemiology, including the difference between biases resulting from (inappropriate) conditioning on common effects (collider bias) and lack of conditioning on common causes of the exposure and outcome (confounding)[5 15-17] . They clarify that confounding occurs not only when the exposure and disease are both affected by another variable (C), but also under less intuitive conditions.

Traditional definitions of confounding involve three criteria (C is predictive of disease, associated with exposure, and not affected by exposure or disease), which are traditionally

treated as necessary but not sufficient for C to be a confounder. I always wondered what these strange situations where a variable could satisfy all three criteria, but still not be a confounder. This was clarified (for me, at least), by the advent of DAGs, which revealed the existence of a type of collider bias called 'M bias'[16], in which C satisfied the traditional criteria but was not a confounder at all: adjustment for it instead produced confounding! This occurs when there are variables which are causes of exposure and C but not disease, and independent variables which are causes of C and disease but not exposure. Adjustment for C then opens a path between the two types of variables, thus unblocking a backdoor pathway from exposure to disease[16].

Unhelpfully, Hernan et al gave the name of 'selection bias' to all forms of collider bias[5], even those in which no selection of participants from the source population has occurred (e.g. M bias can occur even when all of the source population is included in the study). For causal inference, it can be argued that 'collider bias' is the more general concept, whereas selection bias is a particular type of collider bias where the common effect that is conditioned on is selection into the study[15 18].

More generally, as illustrated here, not all problems in epidemiology can be solved by DAGs, nor can DAGs provide a practical perspective on how likely particular biases are, or how strong they may be [9 13 18] Thus, I completely agree with the conclusion of Mansournia et al that it is 'highly misguided if not destructive to ignore the practical difficulties of locating and recruiting valid population control groups while attempting to avoid theoretical biases that are likely to be minor'. A related example of this is 'collider anxiety', where potential confounders are not adjusted for because there is a hypothetical risk that their control

might create M-bias [18], even though in many plausible scenarios this bias is likely to be very small[16].

My own experience is that although there are many problems where drawing the right DAG may help solve a problem, there are others where you need to solve the problem first before you can draw a sensible DAG, which in turn depends on deep subject matter knowledge[8 11 12]. The same issues apply here, in which some of the key issues in matched case-control studies require deep understanding of statistical theory[1], and not all can be easily addressed with DAGs. As with all mathematical and statistical methods, DAGs are an aid to thought, not a substitute for it[9].

**Acknowledgements**

This work was funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 668954.
The Centre for Global NCDs is supported by the Wellcome Trust Institutional Strategic Support Fund, 097834/Z/11/B. I would like to thank Sander Greenland, Deborah Lawlor, Mohammad Mansournia, and David Richardson for their comments on the draft manuscript.

## References


1. Mansournia MA, Jewell NP, Greenland S. Case-control matching: effects, misconceptions, and recommendations. *European Journal of Epidemiology* 2017:in press.
2. Mansournia MA, Hernan MA, Greenland S. Matched designs and causal diagrams. *International Journal of Epidemiology* 2013;42(3):860-69.
3. Pearce N. Analysis of matched case-control studies. *British Medical Journal* 2016;352:i969.
4. Greenland S, Lash TL. Chapter 19: Bias analysis. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008:345-80.
5. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15(5):615-25.
6. Aalen OO, Roysland K, Gran JM, Kouyos R, Lange T. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Statistical Methods in Medical Research* 2016;25(5):2294-314.
7. Daniel RM, De Stavola BL, Vansteelandt S. Commentary: The formal approach to quantitative causal inference in epidemiology: misguided or misrepresented? *International Journal of Epidemiology* 2016;45(6):1817-29.
8. Greenland S. Overthrowing the tyranny of null hypotheses hidden in causal diagrams. In: Dechter R, Geffner H, Halpern JY, editors. *Heuristics, probabilities and causality: a tribute to Judea Pearl*: College Press, 2010:365-82.
9. Greenland S. For and Against Methodologies: Some Perspectives on Recent Causal and Statistical Inference Debates. *European Journal of Epidemiology* 2017;32(1):3-20.
10. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *European Journal of Epidemiology* 2015;30(10):1101-10.
11. Krieger N, Smith GD. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *International Journal of Epidemiology* 2016;45(6):1787-808.
12. Krieger N, Smith GD. Response: FACEing reality: productive tensions between our epidemiological questions, methods and mission. *International Journal of Epidemiology* 2016;45(6):1852-65.
13. Pearce N, Lawlor DA. Causal inference-so much more than statistics. *International Journal of Epidemiology* 2016;45(6):1895-903.
14. Greenland S. Chapter 16: Applications of stratified analysis methods. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
15. Glymour MM, Greenland S. Chapter 12: Causal diagrams. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008.
16. Greenland S. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology* 2003;14(3):300-06.
17. Cole SR, Platt RW, Schisterman EF, Chu HT, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology* 2010;39(2):417-20.
18. Pearce N, Richiardi L. Commentary: Three worlds collide: Berkson's bias, selection bias and collider bias. *International Journal of Epidemiology* 2014;43(2):521-24.

**Figure 1: Directed Acyclic Graph (DAG) for a matched case-control study**