13. Flanders WD, Boyle CA, Boring JR. Bias associated with differential hospitalization rates in incident case-control studies. *J Clin Epidemiol* 1989;42:395–401.
14. Schwartzbaum J, Ahlbom A, Feychting M. Berkson's bias reviewed. *Eur J Epidemiol* 2003;18:1109–12.
15. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
16. Pearl J. Causal diagrams for empirical research. *Biometrika* 1995;82:669–88.
17. VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am J Epidemiol* 2007;166:1096–104.
18. Berkson J. The statistical study of association between smoking and lung cancer. *Mayo Clin Proc* 1955;30:319–48.
19. Pearl R. Cancer and tuberculosis. *Am J Hyg* 1929;9:97–159.
20. Pearl R. Note on the associations of diseases. *Science* 1929;70:191.
21. Pearl R, Sutton AC, Howard WT. Experimental treatment of cancer with tuberculin. *Lancet* 1929;1:1078–80.
22. Robins JM. Data, design, and background knowledge in etiologic inference. *Epidemiology* 2001;12:313–20.
23. Bloemenkamp KW, Rosendaal FR, Buller HR, Helmerhorst FM, Colly LP, Vandenbroucke JP. Risk of venous thrombosis with use of current low-dose oral contraceptives is not explained by diagnostic suspicion and referral bias. *Arch Intern Med* 1999;159:65–70.
24. Feinstein AR. Methodologic problems and standards in case-control research. *J Chronic Dis* 1979;32:35–41.
25. Sadetzki S, Bensal D, Novikov I, Modan B. The limitations of using hospitalcontrols in cancer etiology—one more example for Berkson's bias. *Eur J Epidemiol* 2003;18:1127–31.
26. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14:300–06.
27. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
28. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
29. Westreich D. Berkson's bias, selection bias, and missing data. *Epidemiology* 2012;23:159–64.
30. Vandenbroucke JP, von Elm E, Altman DG *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology* 2007;18:805–35.
31. Knol MJ, Vandenbroucke JP, Scott P, Egger M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am J Epidemiol* 2008;168:1073–81.

Commentary: Three worlds collide: Berkson's bias, selection bias and collider bias

International Journal of Epidemiology, 2014, 521–524
doi: 10.1093/ije/dyu025
Advance Access Publication Date: 28 February 2014

Neil Pearce[1,2]* and Lorenzo Richiardi[3]

[1]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK and [2]Centre for Public Health Research, Massey University, Wellington, New Zealand and [3]Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Italy

*Corresponding author. Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. E-mail: neil.pearce@lshtm.ac.uk

Berkson's bias

In 1984 one of us (N.P.) was living on the beach in New Zealand. I was writing my PhD thesis, when I had a football ('soccer') injury which gave me a severe back problem for a week. After the first two days of lying around on my back listening to music and not being able to do much else, I woke up in the middle of the night with nausea and dizziness apparently caused by an inner ear infection. For the next five days I had severe back pain when I stood up, severe nausea and dizziness when I lay down and a mixture of the two when I sat in a chair. The purpose of recounting this sorry story is not to relate my medical history to readers of *IJE*, but rather because it is relevant to the story of Berkson's bias, which I had been studying at the time. I didn't get admitted to hospital, but it was a near thing, and it gave me firsthand experience of how 'persons with two or more diseases have a higher probability to be hospitalized than persons with only one disease—even if these results are independent'.[1] If I had been hospitalized and recruited for a study of inner ear infections, and if there had been enough other people like me, then we probably would have contributed to a false conclusion that

football injuries (which caused my back problem) were a cause of inner ear infections—this corresponds to the 'indirect' form of Berkson's bias,[2] as depicted in Figure 3a in the paper of Snoep *et al.*[1]

Berkson's bias (also termed 'Berkson's fallacy') is perhaps one of the best known, but least well understood, forms of bias. The paper by Snoep *et al*[1] clarifies what the bias is, why it sometimes matters, but why it usually doesn't. We will comment on three aspects of the paper: (i) the use of Directed Acyclic Graphs (DAGs); (ii) the components of Berkson's bias; and (iii) the likely strength and direction of such biases.

Directed Acyclic Graphs (DAGs)

The paper of Snoep *et al.* clearly illustrates the power and elegance of Directed Acyclic Graphs (DAGs). What we previously used to try to understand using words, probabilities and numerical examples, can now be explored much more elegantly using causal diagrams. This represents a real advance, and clarifies many aspects of Berkson's bias.

More generally, DAGs have clarified the previously murky relationship between selection bias and confounding. Traditionally, selection bias has been described as bias arising from inappropriate selection (or self-selection) of study subjects from the source population.[3] On one level this is clear enough, but the use of the word 'selection' has often led to the term being applied to inappropriate selection of a comparison group, thus leading to confusion as to whether phenomena such as the healthy worker effect are examples of selection bias[4] or of confounding,[5,6] The situation is further complicated because determinants of selection (e.g. age, gender, socioeconomic position) can effectively become confounders and be controlled for in the analysis, even if they were not confounders in the source population. The use of DAGs clarifies this, and distinguishes between biases resulting from (inappropriate) conditioning on common effects ('collider bias' or 'selection bias') and lack of conditioning on common causes of exposure and outcome (confounding).[6,7] The two phenomena can occur together, e.g. when we condition on a collider that is the effect of a cause of the outcome rather than being an effect of the outcome itself. Some would label this as selection bias,[6] others would consider it to also be a type of confounding.[8,9]

Thus, although the three terms are sometimes used almost interchangeably, collider bias is the more general phenomenon involving conditioning on common effects (although Hernan *et al.*[6] use the term 'selection bias' for this more general phenomenon); selection bias is then a particular type of collider bias in which the common effect is selection into the study; Berkson's bias is then a particular type of selection bias[10] in which selection of cases into the study depends on hospitalization, and the exposure is another disease, or a cause of another disease, which also results in hospitalization. It is unlikely that this would have been so easily clarified without the use of DAGs.

The components of Berkson's bias

In its bare essence, Berkson's bias can be seen as a biased estimation of the odds of exposure among the cases because exposed cases are identified with greater probability than non-exposed cases, when the hospitalization rate for the cases is less than 100% and the exposure is another disease, or a cause of another disease, which results in hospitalization. It is possible to illustrate with numerical examples the different steps involved in Berkson's bias. Let us start with the population reported in Table 5 of Berkson's paper[2] where the odds ratio is 1.0 (Table 1).

We now assume that the study compares hospitalized cases with general population controls (corresponding to Figure 1b of Snoep *et al.*[1]). We use the same probabilities of hospitalization for Disease 1 (the exposure—0.15), and Disease 2 (the cases—0.05) of the Berkson's paper. We also assume (differently from Berkson[2] and from Snoep *et al.*[1]) that the whole population has a prevalence of 0.2 and a hospitalization rate of 0.025 for any other disease than D1 and D2 (these different assumptions mean that our numbers are slightly different from those of Berkson[2] and Snoep *et al.*[1]). If the study compares hospitalized cases with general population controls sampled from non-cases with a sampling fraction of 10%, the corresponding findings are shown in Table 2. The estimated odds ratio is now

Table 1. Association in the general population as reported in Berkson[2]

	Exposed	Unexposed	Total
Cases	3000	97 000	100 000
Non-cases	297 000	9 603 000	9 900 000
Total	300 000	9 700 000	10 000 000

Odds ratio = 1.0.

Table 2. Association using hospitalized cases and general population controls

	Exposed	Unexposed	Total
Cases	590	5311	5901
Controls	29 700	960 300	990 000
Total	30 290	965 611	995 901

Odds ratio = 3.59.

Table 3. Association using hospitalized cases and controls from patients hospitalized for any disease, with a 0.2 population prevalence and a 0.025 probability of hospitalization for any disease other than D1 (exposure) or D2 (cases)

	Exposed	Unexposed	Total
Cases	590	5311	5901
Controls	45 812	48 015	93 827
Total	46 402	53 326	99 728

Odds ratio = 0.12.

Table 4. Association using controls hospitalized with a particular disease, with a 0.005 population prevalence and a 0.20 probability of hospitalization for the control disease

	Exposed	Unexposed	Total
Cases	590	5311	5901
Controls	480	9757	10 237
Total	1070	15 068	16 138

Odds ratio = 2.26.

3.59, because of higher exposure odds in the hospitalized cases (compared with all cases). This is caused by collider bias as shown in Figure 1b of Snoep *et al.*[1]

The corresponding findings from a study conducted among hospitalized patients in the same population are shown in Table 3. The bias is now in the opposite direction, because the increase in the exposure odds in the cases (compared with all cases) is more than offset by an even greater increase in the exposure odds in the non-cases (compared with the general population). This is again caused by the collider bias, as depicted in Figure 1a of Snoep *et al.*[1]

Table 4 shows an example more similar to that of Berkson,[2] in which one particular disease has been chosen for selection of controls and the control disease has a 0.20 probability of hospitalization and a prevalence of 0.005; the odds ratio is now 2.26, and the bias is now in the opposite direction to Table 3, because the hospitalization rate for the control disease is greater than the hospitalization rate for the case disease (and therefore the increase in the exposure is greater in the cases than in the controls).

Thus, when using general population controls, Berkson's bias will tend to produce elevated odds ratios (when the hospitalization rate for the case disease is less than 100%); when using hospital controls, the bias can be in either direction.

The strength and direction of bias

The paper by Snoep *et al.* not only clarifies these underlying mechanisms of Berkson's bias; it also provides estimates of its strength and direction in a variety of circumstances (see Snoep *et al.*, Table 3[1]). This reveals that, when Berkson constructed a scenario similar to Table 4 in which the rate of hospitalization in the control disease (refractive errors) was considerable higher than in the case disease (diabetes)—0.2 compared with 0.05—the odds ratios were strongly biased upwards (line 1 of Snoep *et al.*, table 3[1]); when the rate of hospitalization is lower in the control disease, the odds ratio is biased downwards (lines 4–5 in Table 3 of Snoep *et al.*[1]), whereas there is no bias when the rate of hospitalization is the same for the two diseases and patients who have both the case and the control disease are counted only as cases (lines 8–9, Table 3 of Snoep *et al.*[1]).

So Berkson's bias can certainly occur, but does it really matter? As Snoep *et al.*[1] note, Berkson's example was based on a hypothetical study involving the association between prevalent cases and another prevalent disease (the exposure). When exposure is not a direct reason for hospitalization in itself, only the indirect form of Berkson's bias is relevant—like the example about the back problem caused by a football injury (see above). This bias is largely attenuated by using incident cases and can be prevented completely by excluding cases that were hospitalized because of another disease. This can be seen from the example described above—many people with incident middle ear infections would have to be hospitalized for concurrent football injuries for material bias to occur. Thus, in many (or perhaps most) plausible situations, the bias will be extremely small, particularly if incident cases are used. Although it is theoretically interesting, in practice it has largely been 'much ado about nothing'.

Perhaps the main message here is that it is not sufficient merely to demonstrate that a bias could occur; it is necessary to also assess the likelihood that it will occur, and its likely strength and direction. Epidemiological studies are frequently criticized on the basis of the potential for information bias or residual confounding. In some instances these potential problems are real and important; in others they are trivial. Bayesian methods are becoming increasingly available to assess the likely strength and direction of such biases.[11]

Which brings us to the limitations of DAGs. Berkson's paper produced extreme results because it was based on prevalent cases, a situation which cannot be easily represented by DAGs. If we change from prevalent cases to incident cases, all of the DAGs in figures 1–6 in Snoep *et al.*[1] still look the same, but the biases have generally become trivial. This illustrates a more general problem of DAGs—they can show that a bias could occur, but do not provide estimates of its likely strength and direction. Without this, it is easy to succumb to 'analysis paralysis'

which stems from the fear of adjusting for a potential confounder (which could also be a collider in another path) because to do so might result in collider bias ('collider anxiety'). In some situations, collider bias may be comparable in size with uncontrolled confounding.[7] In others it will not, and the benefit from controlling confounding will far outweigh the effects of collider bias. It all depends.

Funding

The Centre for Public Health research is supported by a Programme Grant from the Health Research Council of New Zealand.

Acknowledgements

We thank Jan Vandenbroucke for his comments on the draft manuscript.

Conflict of interest: None declared.

References

1. Snoep JD, Morabia A, Hernandez-Diaz S, Hernan MA, Vandenbroucke JP. A structural approach to Berkson's fallacy: and a guide to a history of opinions about it. *Int J Epidemiol* 2014;43:515–21.
2. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bull* 1946;2:47–53. Reprinted *Int J Epidemiol* 2014;**43**:511–15.
3. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
4. Checkoway H, Pearce N, Crawford-Brown D. *Research Methods in Occupational Epidemiology*. New York: Oxford University Press, 1989.
5. Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology*. 2nd edn. New York: Oxford University Press, 2004.
6. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
7. Greenland S. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology* 2003;14: 300–06.
8. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash TL (eds). *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
9. Pizzi C, De Stavola B, Merletti F *et al*. Sample selection and validity of exposure-disease association estimates in cohort studies. *J Epidemiol Community Health* 2011;65: 407–11.
10. Westreich D. Berkson's bias, selection bias, and missing data. *Epidemiology* 2012;23:159–64.
11. Greenland S. Bayesian perspectives for epidemiologic research. III. Bias analysis via missing-data methods. *Int J Epidemiol* 2010;39:1116–16.

Commentary: Berkson's fallacy and missing data

International Journal of Epidemiology, 2014, 524–526
doi: 10.1093/ije/dyu023
Advance Access Publication Date: 28 February 2014

Daniel Westreich[1]* and Rhian M. Daniel[2]

[1]Department of Epidemiology, UNC-Chapel Hill, Chapel Hill, NC, USA and [2]Department of Medical Statistics and Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine, London, UK

*Corresponding author. Department of Epidemiology, CB 7435 McGavran-Greenberg Hall, Chapel Hill, NC 27599, USA. E-mail: djw@unc.edu

We thank Snoep *et al.* for their commentary[1] offering valuable perspective on Berkson's classic and important paper[2] and the controversies that followed. We wish here to mention what we feel is an important point arising from their discussion, and use this to highlight some interesting features of using directed acyclic graphs (DAGs) to discuss issues relating to selection bias and missing data.

The non-hospitalized subpopulation

An interesting point made by Snoep *et al.* is that, even in the simplest setting in which Berkson's fallacy may occur (Snoep *et al.* Figure 1b), where two independent diseases are found to be associated in hospitalized patients, 'one would not expect a spurious association between [two diseases] in a study restricted to non-hospitalized patients.' In Snoep Figure 2, a patient is not hospitalized ($H = 0$) if and only if she is not hospitalized for disease 1 ($H1 = 0$) and not hospitalized for disease 2 ($H2 = 0$); a patient is hospitalized ($H = 1$) if she is hospitalized for disease 1 ($H1 = 1$) *or* hospitalized for disease 2 ($H2 = 1$). Restricting to non-hospitalized patients therefore implies jointly conditioning