Journal of Psychiatric Research 116 (2019) 189–190

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Letter to the editor: When does selection generate bias in clinical samples?

To the editors

The article in the previous issue of this journal entitled "Does Attrition Affect Estimates of Association: A Longitudinal Study" (Saiepoura et al., 2019) provides rich results from a large sample and across a follow-up time of up to 27 years. The authors summarize: "The findings of this paper suggest that differential loss to follow-up, even if it applies to both the dependent and independent variables and is associated with the rate of loss to follow-up, rarely significantly affects estimates of association." And "that loss to follow-up may produce misleading findings only in circumstances where loss to follow-up is combined with a number of other sources of bias." These conclusions ignore what has long been known methodologically on bias due to selection in clinical samples. Besides, they don't consider a pitfall in using statistical tests, and the choice of analyzed associations needs to be challenged against the authors' claims on generalization ("rarely … " and "only … ").

First, loss to follow-up is just one instance for bias due to selection to occur. The general problem is that a parameter of interest, here an association between an **X** and a **Y**, $\mathbf{A_{XY}}$ (e.g. **X** = pregnancy, **Y** = mental disorder; measured as ln (odds ratio)), might differ between the population observed (typically filtered through self-selection into treatment and various inclusion criteria) and the population of interest (e.g. the population with all individuals having a disease irrespective of help-seeking). Bias due to selection (or lack of "external validity" or "non-transportability") can arise at different stages of investigation: 1. sampling (who is selected, e.g. hospitalization; or self-selects through help-seeking?), 2. participation (who among the selected participates?) and 3. completing a follow-up assessment (who among the participants remains in the study?) (Lash et al., 2009). Saiepoura et al. (2019) only assess bias due to selection at stage 3. This is fine for addressing this stage, but it has to be noted that selection already has had two chances to occur. Bias that emerges at stage 1 carries over to 2. and 3., since participants and completers are restricted to those that have been sampled at all (e.g. if a clinical sample includes only severe depression cases, participation and completion will still not add mild cases to the sample).

At whatever stage, bias occurs if selection/participation/completion (shortly **S**; **S** = 0 unselected, 1 = selected) is related to *both* **X** and **Y**. This had firstly been noted in medical research as "Berkson's bias" in the context of disturbed comorbidity findings in hospitals. Berkson (1946) wrote: "It appears […] that it is hazardous to apply in a hospital population the method of the fourfold table analysis for an inquiry into the correlation of diseases."

Today, this bias is graphically illustrated as "conditioning on a collider" in causal diagrams. Whenever selection is a *common consequence* (graphically a "collider") of both **X** and **Y** (or caused by or causes a common consequence), the estimation of $\mathbf{A_{XY}}$ is biased if investigated only among the selected (**S** = 1) (Elwert and Winship, 2014; Pearl, 2009). Conditioning on **S** (here through sampling, what resembles adjusting for **S** in a regression model) causes an association between **X** and **Y**. Specifically, if there is truly no association ($\mathbf{A_{XY}} = 0$), a spurious association is generated. Under presence of an association ($\mathbf{A_{XY}} \neq 0$), the association is biased through the spurious additional association (See Fig. 1).

In both cases, the direction of bias depends on the *signs* of the relations with **S**. If $\mathbf{A_{XS}}$ and $\mathbf{A_{YS}}$ are both positive or both negative, $\mathbf{A_{XY}}$ is *underestimated*. If the signs of $\mathbf{A_{XS}}$ and $\mathbf{A_{YS}}$ are different, *overestimation* occurs.

Next, these theoretical findings apply to *populations*. Sampling introduces random error. If **S** is truly related with **X** and with **Y**, and $\mathbf{A_{XY}}$ is truly different between **S** = 0 and **S** = 1 (in the target population = all individuals of interest with **S** = 0 or **S** = **1**), selection **S** might be found to be "significantly" (at the usual two-sided 05 level) related to both **X** and **Y**, but $\mathbf{A_{XY}}$ might not be found to be significantly different between **S** = 0 and **S** = 1. This might happen just because the statistical power is smaller for interactions (**S** x $\mathbf{A_{XY}}$) than for associations. The comparison of p-values (or even binary test decisions) between differently powered analyses is just one instance of many misunderstandings and misuses of statistical tests (Greenland et al., 2016). Rather than relying mechanically on statistical tests, medical researchers should assess bias due to selection with substantive considerations.

Another line of thought besides selection as common consequence of **X** and **Y** is reflecting *on determinants of* $\boldsymbol{A_{XY}}$. If, for instance, a comorbidity measure $\mathbf{A_{XY}}$ (e.g. ln odds ratio between depression and anxiety disorder) is assumed to decrease with increasing social support, one expects a smaller $\mathbf{A_{XY}}$ in a population with higher average social support. Now, if social support is positively associated with hospitalization, *downward bias* in estimating $\mathbf{A_{XY}}$ is expected in a convenience sample of hospitalized patients.

Another set of issues relates to the authors' choice of associations and whether this choice allows concluding that bias due to loss at follow-up constitutes "rarely" a problem. Actually, the analysis is restricted to the impact of loss at follow-up on *baseline* associations. We doubt a wide scope of the related findings for two reasons. First, clinicians are rather interested in associations with (or effects on) course of disease or treatment response. Second, dropout has no practical impact on estimating a baseline association just because no follow-up information is required for this. Although the convenience of the authors' analysis is obvious in that it does not require assumptions on true mechanism of how patients get lost, further investigations should focus on *prospective* associations. As here, data from a large sample with a large number of follow-up completers should be used to minimize random error, but to *define a population*. Then dropout can be simulated by imputing it into the data as if caused by baseline (e.g. symptom severity) and course variables (e.g. symptom change) and according to various scenarios (e.g. small and large effects of **X** and **Y** on **S**). Besides, selection should already be investigated at the first stage of sampling

https://doi.org/10.1016/j.jpsychires.2019.02.010
Received 21 January 2019; Received in revised form 28 January 2019; Accepted 20 February 2019

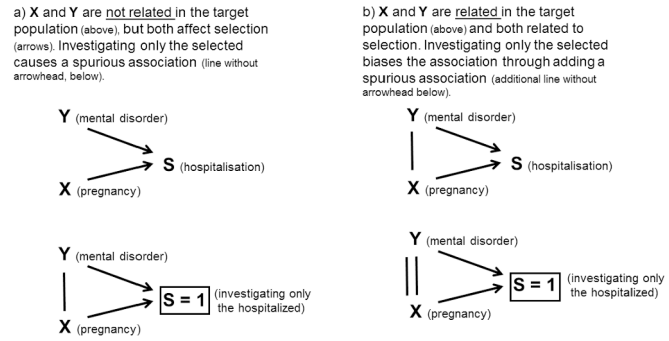

Fig. 1. a) Causal diagram for unrelated **X** and **Y**: conditioning on selection (**S**) generates a spurious association. b) Causal diagram for associated **X** and **Y**: conditioning on selection (**S**) generates an additional association.

where probably most bias yet creeps into the data. Lastly, *sensititvity analysis* should be used. This method computes how an estimated association changes under assumptions on a particular selection mechanism that might have generated the data (Lash et al., 2009).

Statement on competing interests

The authors have no competing interests.

References

Berkson, J., 1946. Limitations of the application of fourfold table Analysis to hospital data. Biom. Bull. 2, 47–53. https://doi.org/10.2307/3002000.

Elwert, W., Winship, C., 2014. Endogenous selection bias: the problem of conditioning on a collider variable. Annu. Rev. Sociol. 40, 31–53. https://doi.org/10.1146/annurev-soc-071913-043455.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., Altman, D.G., 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur. J. Epidemiol. 31, 337–350. https://doi.org/10.1007/s10654-016-0149-3.

Lash, T.L., Fox, M.P., Fink, A.K., 2009. Selection bias. In: Lash, T.L., Fox, M.P., Fink, A.K. (Eds.), Applying Quantitative Bias Analysis to Epidemiological Data. Springer, New York, pp. 43–55.

Pearl, J., 2009. Causality. Models, Reasoning and Inference, second ed. Cambridge University Press.

Saiepoura, N., Najman, J.M., Ware, R., Baker, P., Clavarino, A., Williams, G.M., 2019. Does attrition affect estimates of association: a longitudinal study. J. Psychiatr. Res. 110, 127–142.

Michael Höfler*
Clinical Psychology and Behavioral Neuroscience, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany
E-mail address: michael.hoefler@tu-dresden.de.

Sebastian Trautmann
Medical School Hamburg, Department of Psychology, Germany

* Corresponding author. Klinische Psychologie und Psychotherapie, TU Dresden, Chemnitzer Str. 46, D–01187, Dresden, Germany.