

Published in final edited form as:

Epidemiology. 2011 May ; 22(3): 446. doi:10.1097/EDE.0b013e3182117615.

Interviewer identity as exclusion restriction in epidemiology

Till Bärnighausen^{1,2}, Jacob Bor¹, Speciosa Wandira-Kazibwe³, and David Canning¹

¹Harvard School of Public Health, Boston, MA ²Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa ³Concave International, Kampala, Uganda.

To the Editor:

We read with great interest the thoughtful commentary by Geneletti et al¹ on two recent studies in this journal (one by Chaix² et al, the other by ourselves³), which correct estimation for selection effects. We agree with the conclusion of Geneletti and colleagues that modelling selection is “problem-specific, as well as dependent on assumptions made and the type of additional data available.” We would like to point out, however, that our problem-specific approach to using Heckman-type selection models should be widely applicable in epidemiology.

The performance of a Heckman-type model depends critically on the use of valid exclusion restrictions,⁴⁻⁵ i.e. variables that determine sample selection but do not independently affect the outcome of interest. Our innovation on the approach—to use the interviewer identity as an exclusion restriction—offers an opportunity to examine and control for selection on unobserved factors in many epidemiologic studies for several reasons.

1. Studies where interviewers act as agents of data collection, such as in surveys and surveillances, are a common source of data in epidemiology. Because epidemiologists are often closely involved in the data collection, they should have access to data on interviewer identity even in many of those cases where this information is not included in the routinely available datasets.
2. Interviewers differ in their experience, motivation and attitudes and thus have varying success contacting eligible individuals and eliciting consent from individuals they have contacted^{6,7}—i.e. interviewer identity determines sample selection. This hypothesis is testable.
3. Interviewer identity does not affect many of the variables of interest in epidemiology. While this hypothesis is usually not testable,⁴ an interviewer effect can often be ruled out on theoretical considerations. Interviewer identity cannot influence factors that are neither assessed by an interviewer nor affected in any way by interviewer contact (e.g., many factors measured in biological samples such as HIV status, haemoglobin levels, or the presence of a particular gene). While matching of interviewers to eligible individuals can introduce associations between interviewer identity and an outcome, as long as the matching criteria are known these associations can be easily controlled for in the analysis.

Address for correspondence: Till Bärnighausen, 665 Huntington Ave, Building 1, Room 1102, Boston MA 02115, (tel) 1.617.379.0372, (fax) 1.617.432.6733, tbaernig@hsph.harvard.edu.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Heckman-type selection models are well-established in economics, sociology and political science^{4,8,10} but rarely used in epidemiology. The recognition that epidemiologists often have at their disposal a highly plausible exclusion restriction to model the effect of selection on unobserved factors may increase the use of Heckman-type models, potentially leading to new insights into selection effects.

Acknowledgments

Funding: Supported by Grant 1R01-HD058482-01 from the National Institutes of Health/National Institute of Child Health and Human Development (NIH/NICHD), and the William F. Milton Fund, Harvard University (to T.B.), and by Grant 2008-2302 from the William and Flora Hewlett Foundation, Grant 5 P30 AG024409 from NIH/National Institute of Aging (NIA), and Grant 1R21AG032572-01 from NIH/NIA (to D.C.).

References

1. Geneletti S, Mason A, Best N. Adjusting for selection effects in epidemiologic studies: why sensitivity analysis is the only “solution”. *Epidemiology*. 2011; 22(1):36–9. [PubMed: 21150353]
2. Chaix B, Billaudeau N, Thomas F, et al. Neighborhood effects on health: correcting bias from neighborhood effects on participation. *Epidemiology*. 2011; 22(1):18–26. [PubMed: 21150351]
3. Bärnighausen T, Bor J, Wandira-Kazibwe S, et al. Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. *Epidemiology*. 2011; 22(1):27–35. [PubMed: 21150352]
4. Vella F. Estimating models with sample selection bias: a survey. *J Hum Res*. 1998; 33:127–169.
5. Heckman JJ. Sample selection bias as a specification error. *Econometrica*. 1979; 47:153–161.
6. Groves, RM.; Couper, MP. *Nonresponse in household interview surveys*. Wiley; New York: 1998.
7. Blohm M, Hox J, Koch A. The influence of interviewers’ contact behavior on the contact and cooperation rate in face-to-face household surveys. *International Journal of Public Opinion Research*. 2006; 19(1):97–111.
8. Winship C, Mare RD. Models for sample selection bias. *Annual Review of Sociology*. 1992; 18:327–350.
9. Puhani PA. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*. 14(1):53–68.
10. Dubin JA, Rivers D. Selection bias in linear regression, logit and probit models. *Sociol Methods Res*. 1989; 18:360–390.