OPEN

Using Linkage to Electronic Primary Care Records to Evaluate Recruitment and Nonresponse Bias in The Avon Longitudinal Study of Parents and Children

To the Editor:

Participation is often incomplete in observational research—because of initial failure to recruit or loss to follow-up—resulting in loss of statistical power and possible bias. Selective participation often leads to biased prevalence estimates, but appears to be less important in relation to estimates of exposure–outcome associations, though exceptions have been reported.[1–7]

Linkage to routine health data offers a means to examine the extent of these biases by providing data on participants and nonparticipants in a study. Here, within appropriate permissions, we have linked those eligible to take part in The Avon Longitudinal Study of Parents and Children (ALSPAC), a birth cohort, to the General Practice Research Database, an anonymized database of primary care records of

around 5 million UK residents. The linked sample comprised 749 individuals eligible to take part in ALSPAC (4% of the ALSPAC-eligible population). Of these, 519 (69%) had enrolled (originally been recruited) in ALSPAC, of whom 348 (67%) had participated at 10 years of age and 223 (43%) at 17 years of age. Further details about the sample are given in the eAppendix (http://links.lww.com/EDE/A898).

We used the linked resource to define the cumulative incidence (by ages 11 and 19) of six outcomes: any mental illness, any respiratory illness (excluding asthma), asthma and/or allergies, ever smoked (up to 19 only), been pregnant (19 only), and been classified as "child at risk" (eAppendix; http://links.lww.com/EDE/A898). Data on exposures and potential confounders were obtained from ALSPAC questionnaires administered during pregnancy and early infancy.

For each outcome, we calculated the cumulative incidence among those eligible to take part in ALSPAC, those who had enrolled and those who participated at 10 and 17 years of age. We compared these using ratios of cumulative incidences. We calculated odds ratios (ORs) for four exposure–outcome associations (three examined separately at 10 and 17 years of age) among all enrolled subjects, and compared them—using relative odds ratios (RORs)—to ORs from individuals participating at 10 and 17 years of age. Confidence intervals were constructed around the logarithm of the ratios of cumulative incidence and RORs using a nonparametric bootstrap method.

The cumulative incidence of each of the outcomes was similar among enrolled and eligible subjects (eTable 4; http://links.lww.com/EDE/A898). However, participants were less likely than all enrolled subjects to have ever smoked, been pregnant or been "at risk" (eTable 5; http://links.lww.com/EDE/A898). For all the exposure–outcome associations, the ORs among participants at 17 were substantially different from those among all enrolled subjects (Table), although confidence intervals were inevitably wide given the small numbers with linked data available. The RORs for participants at 10 years of age were all

closer to unity. Adjusting for other factors predictive of nonparticipation made little difference to these RORs. For three of the outcomes, there was evidence for an interaction between the exposure and outcome with respect to participation at 17 years of age, resulting in the outcome being missing not at random in one exposure group but not the other (eTable 6; http://links.lww.com/EDE/A898). These interactions were not seen at 10 years of age.

In conclusion, we have shown that bias due to loss to follow-up in some exposure–outcome associations may be substantial, although this study was small and the uncertainty in our estimates of bias consequently quite large. Our study demonstrates that linkage between observational studies and electronic patient records can provide useful information on nonresponders. This utility is likely to increase as the proportion of individuals with linked data increases. Comparative information available via linkage could be combined with, or used in place of, observational data. In addition, linked data could provide important information on factors predictive of nonresponse; such information can be incorporated into statistical analyses, thus reducing bias[8] and potentially increasing the precision of estimates.

Rosie Cornish
School of Social and Community Medicine
University of Bristol
Bristol, United Kingdom
rosie.cornish@bristol.ac.uk

Kate Tilling
School of Social and Community Medicine
University of Bristol
Bristol, United Kingdom
MRC Integrative Epidemiology Unit
School of Social and Community Medicine
University of Bristol
Bristol, United Kingdom

TABLE. Crude and Adjusted Relative Odds Ratios (RORs) for Different Exposure–Outcome Associations Measured up to 11 Years and up to 19 Years Among Subjects Enrolled in the Avon Longitudinal Study of Parents and Children Also Appearing in the General Practice Research Database

Exposure–Outcome Association	Crude OR (95% CI)		Crude ROR (95% CI)	Adjusted[a] OR (95% CI)		Adjusted[a] ROR (95% CI)
Before 11 Age of Years	All Enrolled Subjects[b]	Participated at 10 Years of Age		All Enrolled Subjects[b]	Participated at 10 Years of Age	
Social class (low vs. high) and mental illness[c]	1.48 (0.51, 4.26)	1.30 (0.35, 4.84)	0.88 (0.34, 2.28)	1.39 (0.45, 4.31)	1.00 (0.24, 4.15)	0.66 (0.25, 1.76)
Smoking in first trimester (smoking vs. nonsmoking) and respiratory illness	1.36 (0.80, 2.30)	1.68 (0.88, 3.18)	1.23 (0.83, 1.84)	1.28 (0.73, 2.23)	1.75 (0.89, 3.44)	1.37 (0.91, 2.05)
Breastfed <6 months and asthma/allergies	1.07 (0.65, 1.74)	1.01 (0.57, 1.78)	0.94 (0.70, 1.27)	1.16 (0.68, 1.96)	1.07 (0.58, 1.97)	0.92 (0.66, 1.28)
Before 19 Years of Age	All Enrolled Subjects[b]	Participated at 17 Years of Age		All Enrolled Subjects[b]	Participated at 17 Years of Age	
Social class (low vs. high) and mental illness	0.65 (0.30, 1.56)	1.89 (0.58, 6.24)	2.92 (1.13, 7.54)	0.78 (0.30, 1.99)	1.99 (0.53, 7.50)	2.55 (0.68, 9.53)
Smoking in first trimester (smoking vs. nonsmoking) and respiratory illness	1.88 (1.03, 3.41)	5.09 (1.62, 16.00)	2.71 (0.89, 8.25)	1.68 (0.90, 3.14)	4.53 (1.41, 14.56)	2.70 (0.82, 8.84)
Breastfed <6 months and asthma/allergies	1.13 (0.67, 1.91)	0.68 (0.28, 1.65)	0.60 (0.29, 1.26)	1.13 (0.64, 1.97)	0.60 (0.23, 1.56)	0.53 (0.23, 1.22)
Sex (female vs. male) and smoking[c]	2.51 (1.33, 4.73)	0.89 (0.32, 2.51)	0.36 (0.14, 0.88)	2.72 (1.43, 5.20)	0.96 (0.32, 2.84)	0.35 (0.13, 0.97)

[a]Adjusted for maternal education, parity as well as sex, breastfeeding, and smoking in first trimester when these were not the exposure variables.
[b]All enrolled subjects with complete information on baseline covariates.
[c]Parity was not adjusted for in this analysis because of sparse data.
CI indicates confidence interval.

Andy Boyd

John Macleod

School of Social and Community Medicine
University of Bristol
Bristol, United Kingdom

Tjeerd Van Staa

Health e-Research Centre
University of Manchester
Manchester, United Kingdom
Utrecht Institute for
Pharmaceutical Sciences
Utrecht University
Utrecht, The Netherlands

REFERENCES

1. Wolke D, Waylen A, Samara M, et al. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *Br J Psychiatry.* 2009;195:249–256.
2. Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology.* 2013;24:1–9.
3. Ferrie JE, Kivimäki M, Singh-Manoux A, et al. Non-response to baseline, non-response to follow-up and mortality in the Whitehall II cohort. *Int J Epidemiol.* 2009;38:831–837.
4. Greene N, Greenland S, Olsen J, Nohr EA. Estimating bias from loss to follow-up in the Danish National Birth Cohort. *Epidemiology.* 2011;22:815–822.
5. Bjertness E, Sagatun A, Green K, Lien L, Søgaard AJ, Selmer R. Response rates and selection problems, with emphasis on mental health variables and DNA sampling, in large population-based, cross-sectional and longitudinal studies of adolescents in Norway. *BMC Public Health.* 2010;10:602.
6. Tin Tin S, Woodward A, Ameratunga S. Estimating bias from loss to follow-up in a prospective cohort study of bicycle crash injuries. *Inj Prev.* 2014;20:322–329.
7. Osler M, Kriegbaum M, Christensen U, Lund R, Nybo Andersen AM. Loss to follow up did not bias associations between early life factors and adult depression. *J Clin Epidemiol.* 2008;61:958–963.
8. Gorman E, Leyland AH, McCartney G, et al. Assessing the representativeness of population-sampled health surveys through linkage to administrative data on alcohol-related outcomes. *Am J Epidemiol.* 2014;180:941–948.

Simulation Comparison of the Quality Effects and Random Effects Methods of Meta-analysis

To the Editor:

The problem with the conventional random effects model of meta-analysis is