DOI: 10.1002/jcv2.12043

EDITORIAL

JCPP Advances ACAMH

The importance of selection bias in prospective birth cohort studies

Abstract

Sophisticated statistical analyses of data from large-scale prospective birth cohort studies combined with thoughtful study designs have advanced understanding about the causes, consequences and developmental course of child and adolescent mental health problems. Available large-scale prospective cohort studies, such as ALSPAC, MoBA, and TEDS have many noteworthy strengths, but they all suffer from non-random non-participation and attrition over time. Recent findings have highlighted that prospective birth cohort studies need to carefully consider the importance of selection bias.

INTRODUCTION

Psychiatric epidemiology was for a long time primarily represented by small clinical studies (Verhulst & Tiemeier, 2015). This has changed dramatically during the last three decades thanks to a growing number of large-scale prospective birth cohort studies (e.g., ALSPAC, MoBA, TEDS). Sophisticated statistical analyses of such longitudinal data combined with thoughtful study designs have generated findings that are advancing our understanding about causes, consequences and the developmental course of child and adolescent mental health problems. These prospective cohort studies also share one major challenge—participation rates have decreased substantially in recent decades as also willingness to participate in subsequent follow up assessments. This is a concern as self-selection into prospective cohort studies and loss to follow-up is likely to introduce selection bias.

With this brief context in mind, this editorial aims to (a) discuss the presence and implications of selection bias in prospective cohort studies, and (b) highlight key findings from the publication by Cadman (2021) in the October issue of JCPP Advances and discuss these in the context of selection bias.

PROSPECTIVE BIRTH COHORTS

JCPP Advances has in the first three issues (April, July, and October) published important study findings from prospective birth cohorts, including the Avon Longitudinal Study of Parents and Children (ALSPAC), Generation R (GenR), the Norwegian Mother, Father and Child Cohort Study (MoBa), and the Millennium Cohort Study (MCS). These internationally recognized cohort studies share several unique strengths. First, they all include large numbers of individuals (from about 10,000 and more), which is important to provide more precise estimates. Furthermore, large sample sizes are also needed to be able to identify small effects from genetic and environmental risk factors, as well as to study rare outcomes or important sub-groups. Second, the prospective birth cohorts have obtained data across several time-points—sometimes over decades, allowing for the study of individuals over time. Repeated measures from more than two waves of assessment (ideally with similar measures and raters) open up the possibility for trajectory modeling, which is a useful approach to examine how mental health problems change over crucial periods of development, thus providing opportunities to target time sensitive interventions and preventions. Third, the above-mentioned prospective birth cohorts have included multiple sources of information (e.g., self-reports, teacher-reports, parent-reports, cognitive data, and electronic health record data). This approach not only helps to reduce bias from shared method variance, but also offers possibilities to construct appropriate measures across development and to test for rater effects. Finally, several of the prospective birth cohorts are genetically-informative (e.g., Genome-wide data, Polygenetic risk scores, family-relationship/kinship data), which not only offers opportunities to understand genetic underpinnings of mental health conditions, but also possibilities to adjust for unmeasured confounding in studies focusing on potential environmental risk factors.

Despite these strengths, an important limitation in all these prospective cohort studies is non-participation, both at recruitment and during subsequent follow ups, which may introduce selection bias.

PRESENCE AND IMPLICATIONS OF SELECTION BIAS IN PROSPECTIVE BIRTH COHORTS

Non-participation at the baseline assessment or drop out from subsequent follow-up assessments (i.e., attrition) is clearly associated with a loss of statistical power. There is also clear evidence from prospective cohort studies that non-participation at baseline and follow-up is often non-random. For example, three previous studies from ALSPAC have demonstrated that attrition is associated with sex, socio-economic factors and polygenic risk

scores for psychopathology (Howe et al., 2013; Martin et al., 2016; Taylor et al., 2018).

It is well established that selection bias substantially influences prevalence estimates, most often towards decreasing prevalence estimates of child and adolescent mental health problems. In contrast, it has often been assumed that selection bias only has a limited impact on observed associations, in particular in genetic association studies (Munafo et al., 2018). Recent methodological review papers (Munafo et al., 2018), as well as studies using data from ALSPAC (Taylor et al., 2018) and MoBA (Biele et al., 2019) indicate that this assumption is problematic in many situations. It has been demonstrated that when both exposures and outcomes are associated with participation in a study, this can induce spurious associations between exposure and outcome, suitably conceptualized as collider bias (Munafo et al., 2018; Taylor et al., 2018).

This, of course, highlights the value of using appropriate strategies (e.g., create investment in the project from participants, incentives, appropriate length of the assessment protocol, new letters) to minimize non-participation and attrition in prospective cohort studies. It also highlights the importance of having detailed baseline data (e.g., electronic health record data, DNA) available on many participants when recruited into the study. The availability of rich baseline data offers the possibility to identify predictors of nonparticipation, which allows researchers to consider and acknowledge how likely non-participation is as a potential source of bias. Having access to detailed information at baseline also offers possibilities to use statistical approaches (e.g., inverse probability weighting, and multiple imputation) to reduce the impact of selection bias due to attrition (Munafo et al., 2018; Taylor et al., 2018). In situations where baseline data is lacking, replication may help, but only if the replication study is based on data from a cohort study that differs in the underlying predictors of non-participation. Using negative control exposures and/or outcomes has also been put forward as an approach to use when detailed baseline data is missing (Munafo et al., 2018; Taylor et al., 2018).

TRAJECTORIES OF DEPRESSIVE SYMPTOMS IN ALSPAC

Cadman (2021) in the October issue of JCPP Advances used data from ALSPAC to make an important contribution to field by exploring trajectories of depressive symptoms across adolescence in offspring of parents with and without maladaptive personality traits. Cadman (2021) used a binary measure of maladaptive maternal and paternal personality traits, while the outcome in children was measured via self-reported depressive symptoms from ages 11 to 24 years. Using multilevel growth curve models, the authors demonstrated that the offspring of mothers with high levels of maladaptive personality traits showed evidence of greater depressive symptoms throughout adolescence, while the evidence for associations with fathers' personality traits was weaker. The study outlines a thoughtful discussion of alternative explanations to the observed associations (e.g., common genetic influences, environmental mediation, reverse or reciprocal causality), which provide a useful guide for replication and extensions of the findings in future research.

Even more importantly, at least in the context of this editorial, the study by Cadman (2021) considered selection bias. First, they compared descriptive statistics for the full ALSPAC sample and the maternal and paternal study samples, which demonstrated that the study sample was of higher socioeconomic position than the full ALSPAC sample. Second, the trajectory modeling of depressive symptoms handled missing data using full information maximum likelihood estimation, which helped address selection bias due to attrition. Here it would have been informative if the authors also included sensitivity analyses using different levels of missingness. For example, the main analyses, which included individuals with at least one measurement of depressive symptoms, could have been supplemented by analyses using a more strict threshold for inclusion (e.g., 5 time-points required) all the way up to analyses based on only "complete cases." Third, Cadman (2021) conducted sensitivity analyses using inverse probability weighting to explore if results in the main analyses were robust to selection bias when considering missing exposure and covariate data. Similar results were obtained before and after introducing inverse probability weighting, which indicates that the main conclusions were largely robust to selection bias. Additional sensitivity analyses to explore the influence of selection bias could have involved multiple imputation and external replication. Triangulating findings across different approaches for missing data is probably a useful method to consider to advance the understanding of selection bias in studies using prospective cohort data.

CONCLUSIONS

In short, recent findings have highlighted that researchers using prospective birth cohort studies need to consider carefully the importance of selection bias. Addressing the impact of selection bias should help generate more accurate and replicable findings regarding the developmental course of child and adolescent mental health problems, and also how risk factors and outcomes associate with such problems.

KEYWORDS

ALSPAC, non-participation, prospective birth cohort studies, selection bias

ACKNOWLEDGMENTS

The author thanks Joanna Martin, PhD, of Cardiff University, as well as, Ebba Du Rietz, PhD, Ralf Kuja-Halkola, PhD, and Mark Taylor, PhD, all at Karolinska Institutet, for useful input on an earlier draft.

CONFLICT OF INTERESTS

Henrik Larsson reports receiving grants from Shire Pharmaceuticals; personal fees from and serving as a speaker for Shire Pharmaceuticals and Evolan Pharma AB outside the submitted work; and sponsorship for a conference on attention-deficit/hyperactivity disorder from Shire Pharmaceuticals outside the submitted work. He is Editor-in-Chief for JCPP *Advances*. [Corrections made on 22 June 2022, after first online publication: This Conflict of Interests statement has been updated in this version.]

Henrik Larsson[1,2]

[1]*School of Medical Sciences, Örebro University, Örebro, Sweden*
[2]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*

Correspondence
Henrik Larsson, Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden.
Email: henrik.larsson@ki.se

REFERENCES

Biele, G., Gustavson, K., Czajkowski, N. O., Nilsen, R. M., Reichborn-Kjennerud, T., Magnus, P. M., Stoltenberg, C., & Aase, H. (2019). Bias from self selection and loss to follow-up in prospective cohort studies. *European Journal of Epidemiology*, *34*(10), 927–938. https://doi.org/10.1007/s10654-019-00550-1

Cadman, T. (2021). Joint associations of parental personality traits and socioeconomic position with trajectories of offspring depression: Findings from up to 6925 families in a UK birth cohort. *JCPP Advances*, *1*(3).

Howe, L. D., Tilling, K., Galobardes, B., & Lawlor, D. A. (2013). Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology*, *24*(1), 1–9. https://doi.org/10.1097/EDE.0b013e31827623b1

Martin, J., Tilling, K., Hubbard, L., Stergiakouli, E., Thapar, A., Davey Smith, G., O'Donovan, M. C., Zammit, S. (2016). Association of genetic risk for schizophrenia with nonparticipation over time in a population-based cohort study. *American Journal of Epidemiology*, *183*(12), 1149–1158. https://doi.org/10.1093/aje/kww009

Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., & Davey Smith, G. (2018). Collider scope: when selection bias can substantially influence observed associations. *International Journal of Epidemiology*, *47*(1), 226–235. https://doi.org/10.1093/ije/dyx206

Taylor, A. E., Jones, H. J., Sallis, H., Euesden, J., Stergiakouli, E., Davies, N. M., Zammit, S., Lawlor, D. A., Munafò, M. R., Davey Smith, G., & Tilling, K. (2018). Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, *47*(4), 1207–1216. https://doi.org/10.1093/ije/dyy060

Verhulst, F. C., & Tiemeier, H. (2015). Epidemiology of child psychopathology: major milestones. *European Child & Adolescent Psychiatry*, *24*(6), 607–617. https://doi.org/10.1007/s00787-015-0681-9