Early Human Development 90 (2014) 769–780

Contents lists available at ScienceDirect

Early Human Development

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

Approaches for drawing causal inferences from epidemiological birth cohorts: A review

Rebecca C. Richmond [a], Aleef Al-Amin [b], George Davey Smith [a,1], Caroline L. Relton [c,d,*]

[a] Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK
[b] University of Bristol Medical School, University of Bristol, Bristol, UK
[c] Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK
[d] Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK

ARTICLE INFO

Keywords:
Birth cohort
Causal inference
Consortia
DOHaD
Epidemiology
Epigenetics
Life course
Metabolomics
Omics

ABSTRACT

Large-scale population-based birth cohorts, which recruit women during pregnancy or at birth and follow up their offspring through infancy and into childhood and adolescence, provide the opportunity to monitor and model early life exposures in relation to developmental characteristics and later life outcomes. However, due to confounding and other limitations, identification of causal risk factors has proved challenging and published findings are often not reproducible. A suite of methods has been developed in recent years to minimise problems afflicting observational epidemiology, to strengthen causal inference and to provide greater insights into modifiable intra-uterine and early life risk factors. The aim of this review is to describe these causal inference methods and to suggest how they may be applied in the context of birth cohorts and extended along with the development of birth cohort consortia and expansion of "omic" technologies.

Contents

* Corresponding author at: Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. Tel.: +44 117 3314028; fax: +44 117 3310123.
E-mail addresses: rebecca.richmond@bristol.ac.uk (R.C. Richmond), aleefalamin@gmail.com (A. Al-Amin), kz.davey-smith@bristol.ac.uk (G. Davey Smith), caroline.relton@bristol.ac.uk (C.L. Relton).
[1] Tel.: +44 117 3310099; fax: +44 117 3310123.

http://dx.doi.org/10.1016/j.earlhumdev.2014.08.023
0378-3782/

1. Introduction

Large-scale population-based birth cohorts recruit women during pregnancy or at birth over a defined time period and follow up their offspring through infancy and into childhood and adolescence. The longitudinal design of these cohorts is a key feature, providing the opportunity to monitor and model early life exposures in relation to developmental characteristics and later life outcomes, with prospective data collected at repeat follow-ups. Data are often collected on both parents and offspring and include information on demographic, socio-economic and lifestyle characteristics and environmental exposures obtained from questionnaires, clinic data for assessing health and development, and data from biological samples. Some cohorts have been designed as multipurpose resources, whilst others focus on specific health or exposure-related research questions. The size of birth cohorts varies considerably, from a few hundred individuals to over 100, 000 in countries where population-based record linkage is possible.

A major focus of such studies is exposure to risk factors during early life developmental periods which can have important consequences for health and disease. The "Developmental Origins of Health and Disease" (DOHaD) hypothesis outlines how the risk of chronic disease in adult life is initially induced through biological programming of the foetus or infant in response to early environmental signals [1,2]. These responses include molecular, hormonal, metabolic or physiological changes which may have negative impacts on later health. Of particular interest is the data captured on maternal exposures acting during pregnancy, driven by the notion that the intra-uterine environment is a critical period for influencing offspring development and programming events [3,4].

Studies have reported associations between foetal growth, maternal nutrition, exposure to drugs, pollutants and hormones in-utero and a whole host of perinatal and later life offspring traits. The influence of postnatal factors has also been explored, including early life growth [5] and breastfeeding [6]. Of particular value are historical birth cohorts which can be used to study the influence of early life exposures on later disease [7]. As well as DOHaD, other aspects of research within lifecourse epidemiology may be investigated within the context of a birth cohort [8,9] and details of these can be found elsewhere [10].

An attractive feature of birth cohorts is the ability to obtain information on other family members, not only the mothers of the offspring, but sometimes fathers, siblings and grandparents. Family-based sampling can facilitate inter-generational studies of the influence of parental characteristics on a range of offspring outcomes and may aid in disentangling the genetic determinants of disease from environmental risk factors [11].

Increasingly, birth cohorts collect and store biosamples from their participants, which can be used to obtain genetic, epigenetic and metabolic profiles, and to measure biomarkers of environmental exposures such as smoking and pollutants. Biosampling allows the exploration of how social and environmental factors leave biological imprints, independent of or in combination with genetic background. The 'omics' revolution [12] offers the potential to explore putative mechanisms by which specific exposures convey disease risk, whereby identified molecules provide robust biomarkers of early life exposure or may act as intermediates in pathways between exposure and risk of later outcomes.

In addition to the wealth of data collected, longitudinal birth cohorts can offer more to observational epidemiology than other study designs because they allow for prospective time-ordering of the associations of interest i.e. with exposures preceding outcomes, which is useful for establishing causality. However, a key limitation to causal inference in epidemiological birth cohorts is potential confounding, leading to spurious observational associations [13,14]. Distinguishing causality from correlation is essential to identify key early life modifiable causes of ill health and disease and to uncover new mechanistic pathways for therapeutic intervention. A suite of methods has been developed in the last decade to minimise problems afflicting observational epidemiology and to strengthen causal inference. The aim of this review is to describe the causal inference methods that have been used to provide greater insights into modifiable intra-uterine and early life risk factors in the context of large epidemiological birth cohorts and to suggest how we may improve methodological approaches, especially in relation to the expansion of "omics" technologies.

2. Challenges of establishing causality in birth cohorts

Key problems of observational epidemiology which limit its ability to establish causal effects include: 1) reverse causation—where the outcome of interest affects the exposure; 2) confounding—the presence of common causes of the risk factor of interest and the outcome; 3) selection bias—when the study participants are selected in a manner that biases the effect estimate in an association; and 4) measurement error in the exposure, confounding factors or outcome. The characteristics of birth cohorts are such that some of these problems can be minimised. For example, their prospective study design means that there is no biased retrospective assessment and the likelihood of reverse causation is reduced due to the time-ordering of the exposure-outcome associations. These studies also allow for repeated measures to be taken at different time points and appropriate analytical techniques may be used to account for missing data, reducing the role of measurement error and selection bias [15,16].

Observational epidemiology undertaken in the context of a birth cohort generally relies on the assumption that confounding characteristics have been identified and measured with little or no error. However, confounders may be inadequately measured (residual confounding) or there may be unobserved factors (unmeasured confounding) [17] which can lead to spurious associations and conclusions about intra-uterine and early life risk factors [18,19]. Inconsistent findings between cohorts and randomised controlled trials (RCTs) highlight the methodological challenges in establishing robust causal links [13,20]. For example, in observational studies maternal vitamin C intake has been found to be associated with higher birth weight in the offspring [21]. However, large RCTs where pregnant women have been randomised to vitamin C supplements [22–24] have found no benefit of supplementation on birth weight. These conflicting findings are likely due to confounding in the observational association, as mothers with higher vitamin C intake tend to have lower rates of smoking and are from a higher socio-economic background, which influence birth weight [25].

Other limitations introduced by the very nature of birth cohorts include the long time gap between outcomes and exposures, increasing the likelihood of confounding. Another implication of this time gap is the relevance of early life exposures experienced when the birth cohorts were established to contemporary cohorts. Finally, given the high correlation between maternal exposures and behaviours in pregnancy with those postnatally it is often difficult to tease apart intra-uterine from postnatal effects [26].

3. Classic epidemiological approaches for drawing causal inferences

Data collected on parents, offspring and other family members in epidemiological birth cohorts may be integrated in a suite of methods which minimise problems of confounding, strengthen causal inference and provide greater insights into modifiable early life risk factors. The strength of evidence obtained from these methods can be placed between observational associations and RCTs in the hierarchy of evidence for clinical guideline production. Table 1 includes a selection of large, well-established cohorts and the data available in these cohorts which may permit the application of the causal inference methods described in this review. Table 2 outlines each of the main causal inference methods, with examples and linked schematic diagrams in Fig. 1.

4. Randomised controlled trials

Well-conducted, large RCTs , where study participants are randomly allocated to a treatment to avoid potential confounding between