The importance of cohort research starting early in life to understanding child health

Nigel Paneth[a] *and Catherine Monk*[b]

Purpose of review
The current review addresses the importance of the prospective cohort design in large, unselected populations starting early in life for understanding the origins of childhood health disorders.

Recent findings
Cohort studies originating in healthy populations have contributed to great advances in health, especially in cardiovascular diseases, but have only recently been applied systematically to study the origins of childhood disorders. Several large population-based pregnancy and/or birth cohorts have been developed in different parts of the world, and these are beginning to contribute to better understanding of the underlying causes of rare but important childhood disorders, such as autism. The environmental influences on child health outcomes (ECHO) Program is distinct in leveraging and building upon 84 *existing* cohorts to prospectively investigate the role of early-life exposures and underlying biological mechanisms in childhood health and disease, specifically perinatal conditions, obesity, neurodevelopmental disorders, asthma and related pulmonary disorders as well as optimum child health. ECHO is expected to comprise approximately 50 000 children. It is the first US study of this size and scope since the US Collaborative Perinatal Project of 1959–1966.

Summary
The ECHO project represents a new approach to cohort studies in childhood, efficiently making use of extant cohorts while adding new data collection elements that should permit novel insights into the underlying causes of several important pediatric conditions.

Keywords
birth cohort studies, disease epidemiology, environmental factors

INTRODUCTION

The study of disease, not surprisingly, has historically focused largely on the patient, the individual already affected by disease. In the search for causes, this approach was suitable for diseases whose onset closely followed the inciting exposure, as with many infectious diseases. But for chronic diseases, whose origins lie in exposures and behaviors operating years or decades before the onset of disease, study of the patient is unlikely to fully reveal the underlying causes.

Medicine was slow to come around to the idea that to fully investigate the origins of chronic diseases, or more precisely, diseases whose origins are remote in time from the onset of clinically recognizable symptoms, it is necessary to initiate studies in the prediseased population and track exposures as they happen, so as to observe the evolution of disease. That understanding was not to come to full fruition until nearly the second half of the 20th century. Cohort studies of disease that tracked healthy individuals over time until the onset of disease were extremely rare before the founding of the Framingham heart study in 1948.

It is perhaps not surprising that such studies were not mounted earlier, inasmuch as the size of the assembled study population and the duration of time required to study the evolution of disease are formidable obstacles, requiring substantial investment of resources. The essential ingredients for a

[a]Departments of Epidemiology & Biostatistics, and Pediatrics & Human Development, Michigan State University, College of Human Medicine, East Lansing, Michigan and [b]Departments of Psychiatry & Obstetrics & Gynecology, Columbia University Medical Center, New York State Psychiatric Institute, New York, New York, USA

Correspondence to Nigel Paneth, MD, MPH, Departments of Epidemiology & Biostatistics, and Pediatrics & Human Development, Michigan State University, College of Human Medicine, 909 Fee Road, East Lansing, MI 48824, USA. Tel: +1 517 884 3961/+1 517 290 5062; e-mail: paneth@msu.edu

Curr Opin Pediatr 2018, 30:292–296
DOI:10.1097/MOP.0000000000000596

KEY POINTS

- The essential ingredients for a successful cohort study of the origins of any disease are a large enough sample size to provide enough cases of the disease of interest, time for disease progression, and a cohort that reflects the population to which the findings are expected to apply.
- Longitudinal cohort studies that track healthy individuals over time until the onset of disease, such as the US Framingham Heart Study, were transformative in securely establishing, as key risk factors for cardiovascular disease, elevated blood pressure, abnormalities of cholesterol and its fractions, cigarette smoking and glucose intolerance.
- Pediatric cohort studies initiated in pregnancy have identified, for example, the benign nature of most uncomplicated febrile seizures, leading to reductions in anticonvulsant use, and showing that cerebral palsy is not solely a result of brain injury from the birth process.
- The ECHO program is distinct in leveraging and building upon 84 *existing* cohorts to comprise approximately 50 000 children and prospectively investigate the role of early-life exposures and underlying biological mechanisms in childhood health and disease, specifically perinatal conditions, obesity, neurodevelopmental disorders, asthma and related pulmonary disorders as well as optimum child health.
- In line with newly promulgated NIH policy, data from ECHO will be publically available closely following collection and upload to central repositories thus ensuring that ECHO will be a rich source of useful, actionable information about child health for decades to come.

successful cohort study of the origins of any disease are first, the commitment to assembling a population sufficiently large so that enough instances of the disease under investigation will appear and so that any important differences in risk of disease between groups ranging in hypothesized causal factors are unlikely to be due to chance. The second commitment is to have sufficient patience, in the face of publish-or-perish tenure committees in universities, to wait the requisite number of years before disease emerges and relevant findings can be published.

But that is not all that is required. All successful cohort studies of the origins of disease are embedded in well defined populations from which results can be generalized. These need not be statistically ideal probability samples of the general population. Although many of the best known large cohort studies are from defined geographical regions, several important cohort studies have been mounted in medical professionals, such as the Nurse's Health Study [1], or the British Doctor's Study [2], or other professional or industry groupings, such as civil servants [3]. But cohort studies should not be based on participants selected in ways that make generalization risky – volunteers, for example, or – participants identified in medical settings, with some important exceptions, such as prenatal care settings that cover most pregnant women. Cohort studies need clear hypotheses in advance, as well as an a-priori plan to collect the appropriate exposure information, whether through biospecimens, questionnaires or clinical assessments, and to prospectively define the features of the diseases or conditions whose cause and preventability motivates the study. Without specification of the condition or conditions of interest, power calculations are not possible; and without specification of the exposures of interest, one cannot obtain the key ingredient in prospective cohort studies, which is real-time assessment of exposures of interest. In studies starting with extant disease, (i.e. case–control studies) exposure information can only be ascertained from what happened to be recorded earlier or remembered later. The prospective cohort study allows the investigators to carefully measure the hypothesized exposures long before their effects are manifest in the disease state.

NOTABLE SUCCESSES OF PROSPECTIVE STUDIES

The establishment of the Framingham Heart Study in 1948 [4], a prospective cohort study of some 5000 generally healthy citizens of Framingham, Massachusetts, USA with follow-up that continues to this day, was the first of a large number of such studies across the United States and overseas. The contributions of these studies to public health cannot be overstated. This cohort study securely established the key risk factors for the West's leading cause of death, cardiovascular disease, as elevated blood pressure (BP), abnormalities of cholesterol and its fractions, cigarette smoking and glucose intolerance. Primary prevention of heart disease through reductions in smoking and adoption of less atherogenic diets, and secondary prevention through management of high BP, diabetes and elevated cholesterol levels has been the major factor behind the gratifying reduction in United States death rates from ischemic heart disease and stroke, which are now at levels 75% lower than when the Framingham Study was initiated.

The British Doctor's study (cited above) was the first prospective study to clearly delineate the risks of cigarette smoking for lung cancer and heart disease and established a durable tradition of using

medical professionals as cohorts to be followed, such as the contemporary Nurses' Health Study. Medical professionals provide efficiencies in cohort assembly and follow-up, as do civil servants, who have contributed to several important new forms of understanding. The British Whitehall Study [3] demonstrated the important role of social stratification, income and job stress on risk of cardiovascular disease. A cohort study initiated in Taiwanese civil servants [5] demonstrated that the small fraction of the population who were carriers of the Hepatitis B antigen accounted for nearly all cases of primary liver cancer, then one of the commonest cancers in Taiwan, Japan and China, and set the stage for primary prevention of this cancer via immunization of newborns against Hepatitis B.

ORIGINS OF THE COHORT STUDY STARTING AT BIRTH OR IN PREGNANCY

The first of the British National Birth Cohort studies was mounted in 1946 [6], sampling 1 week of births in March of that year in England and Wales. Initially designed to study the medical and other costs of having a baby and the distribution of maternity services in the nation [7], subsets of the 1946 cohort (which numbered nearly 14 000 at birth) have been followed throughout life. This study was followed by two additional British birth cohorts, using the same 1-week-of-births sampling frame in 1958 [8] and 1970 [9].

The focus of the 1946 study was more social than biological, examining economic circumstances of parents in relation to a range of perinatal and child health outcomes particularly growth, school performance and cognitive functioning [10]. The latter cohorts focused on the impact of pregnancy medical conditions, largely based on information collected at birth. As the cohorts have aged, they have increasingly been used to study the early origins of adult disorders such as cardiovascular disease and cancer [11].

In the United States at about the same time as the second British National Birth Cohort study, two large cohort studies intended to study child health were established that had an advantage over the British birth cohorts in that they were initiated in pregnancy, but the disadvantage that they were selected from the prenatal clinics of large medical centers. The Child Health and Development Study (CHES) [12] enrolled some 10 000 pregnancies in the Kaiser Permanente System in California, and the National Collaborative Perinatal Project (NCPP) enrolled some 50 000 pregnancies in 12 major urban medical centers, largely in the North East United States [13]. Both were initiated in 1959. CHES is not diverse; its population is nearly entirely white. The NCPP, by contrast, had many African-American and Latino participants. CHES, taking advantage of its single location, has been the source of a large number of spinoff studies, particularly focusing on adult disease [14]. The NCPP followed children to age 7 years, and its most notable contributions were in child neurology. The benign nature of most uncomplicated febrile seizures was demonstrated by the NCPP [15], leading to a sharp decrease in the number of affected children treated with anticonvulsants, some of which had been demonstrated to impair learning capacity. The NCPP also showed that the cause of cerebral palsy (CP) was more complicated than initially thought [16]. Contrary to popular belief, the majority of children with CP had not been brain injured by the birth process, nor had they suffered from birth asphyxia.

CONTEMPORARY LARGE PREGNANCY COHORTS

Studies of the major sources of childhood morbidity require large sample sizes because important causes of childhood disability and handicap such as autism (1–2% of the population); severe intellectual disability (0.5% of the population) and CP (about 0.3% of live births). Birth defects are individually rare, and very early preterm birth is found in fewer than 1% of births. Such health outcomes are too few to be studied in cohorts of the size of Framingham. For this reason, recent child health studies have been much larger than most adult cohort studies of the past.

The Danish National Birth Cohort (DNBC) and the Norwegian Mother and Child Cohort study (MOBA) are the largest of several European efforts to mount meaningfully scaled cohort studies beginning in pregnancy. Each enrolled about 100 000 relatively (but not completely) unselected pregnancies from routine prenatal care settings and followed them primarily in a passive manner, making use of linkage to the extensive population-based disease registries in both countries. Data from the two cohorts are now being combined to study CP, using record linkage to national registries, in a cohort of 200 000 births [17▪]. Because the two national studies required surveying in the national language, many recent immigrants were not included. Among important results emerging from MOBA and DNBC are the findings that periconceptional folate supplementation may prevent autism [18], that mobile phone use in pregnancy is unrelated to latter child development [19], and that low maternal iodine intake is associated with delayed development [20▪].

A PEDIATRIC COHORT FOR 21st CENTURY RESEARCH AND PREVENTION

In late 2015, when the National Institutes of Health (NIH) announced the establishment of the 7-year environmental influences on child health outcomes (ECHO) program (https://www.nih.gov/echo) through various Funding Announcement Opportunities (FOAs), existing pediatric cohorts were envisioned as a central component. The overall aim was to prospectively investigate the role of early-life exposures and underlying biological mechanisms in childhood health and disease, and the approach was novel – to form a consortium of established pediatric research cohorts to leverage already existing longitudinal data so that the collective efforts would, scientifically, have far more impact than the isolated studies on their own. The cohorts were, however, given the opportunity to expand as well as to extend their period of follow-up.

The ECHO Pediatric Cohort plan envisioned supporting 'multiple, synergistic, longitudinal studies using existing study populations, called cohorts, to investigate environmental exposures – including physical, chemical, biological, social, behavioral, natural and built environments – on child health and development'. The exposure period was established as from before conception to age 5 years, with an explicit inclusion of the prenatal period, and in some cases, the preconceptional period. The outcomes of interest were identified as the following disorders of high public health impact: upper and lower airway disease, principally asthma; obesity; prenatal, perinatal and postnatal problems (e.g. low birthweight) and neurodevelopmental disorders. A fifth outcome, positive health, was added later.

The Pediatric Centers and their cohorts were charged with investigating these outcomes using two parallel approaches: the development of new aims specific to each pediatric cohort for the local extant cohorts and their expansions; and the development of an ECHO-wide uniform data collection protocol to which all cohorts would adhere through new data collection and/or harmonization of data already collected. One year after ECHO was launched in the fall of 2016, the Pediatric Cohort component consists of awards to 35 centers that have 74 principal investigators (many centers have multiple principal investigators) and who follow 84 cohorts, with some centers aggregating as many as 10 different cohorts. The ultimate sample size, including new recruitments to expand extant cohorts, is expected to reach about 50 000.

In the first year, cohorts have launched their individual cohort-specific research projects (ranging from analyses of breast milk for environmental toxins to task-based neuro-imaging to self-report indicators of positive health) while simultaneously developing the ECHO-wide protocol and related scientific policies (such as for publications and biospecimen utilization). The sample size, geographic, racial/ethnic diversity, the range of exposures and the commitment to new methodology for data collection that characterize the ECHO Pediatric Cohorts are such that ECHO is poised to significantly add to information generated by earlier birth cohort studies: 36 states are represented in ECHO as well as two territories and districts (Washington, District Of Columbia and Puerto Rico) and ECHO includes rural, urban and suburban populations.

Numerous known or suspected categories of contributors to children's health outcomes are being or will be measured, including nutrition, physical activity, environmental toxicants, psychosocial stressors and an array of omic technologies, with a particular focus on exposures that are remediable or susceptible to change, so that discoveries may be used to lead to improved child health. It is anticipated that the depth of measurement and the large sample size will permit meaningful studies of gene–environment interaction. Other goals that are expected to be reached include

(1) Clear clinical characterization of incident cases and preclinical stages of disease and the frequency of these disorders in the population.
(2) Assessment of the influence of multilevel environmental exposures on the risk of childhood health outcomes.
(3) Identification of the measurable normal human variation (of key exposures and key outcomes) in the population and across the lifespan studied.
(4) A degree of generalizability of findings to the US population not found in single cohort studies.
(5) Moreover, in line with newly promulgated NIH policy, data from ECHO will be publically available closely following collection and upload to central repositories (6–12 months, including results from bioassays).

Another key, as well as novel, aspect of the ECHO Pediatric Cohorts comes from the program-wide structure that integrates observational research with a clinical trials network. Specifically, as part of the ECHO program, established clinical investigators and their support teams working in academic research settings in Institutional Development Awards Program States (this NIH program provides resources to 23 states and Puerto Rico with historically low levels of NIH funding) will conduct

pediatric clinical trials research, with priority given to ECHO's outcome areas. This integration is consistent with recent initiatives in so-called 'translational epidemiology' [21,22], which inverts medicine's traditional bench-to-bedside information flow from bedside (i.e. from population health observations)-to-bench (i.e. to clinical trials and basic science research).

With ECHO's integrated structure, cohort observations of contributors to poor health trajectories, as well as protections from them, will suggest interventions that can efficiently be tested in clinical trials with two useful outcomes: first, underserved and rural children will benefit from potentially useful and new health interventions and second, the rigor and methodology of the Randomized Control Trial can be leveraged to rigorously test whether removal of observed risk factors identified in ECHO leads to changes in health outcomes. In addition, since the Pediatric Cohorts will test, using the many disciplines of its investigators, an array of biomarkers and biological pathways for exposure effects on health (including, for example epigenetic and microbiome work) it is likely that findings will spark laboratory research into fundamental mechanisms.

CONCLUSION

Pregnancy and birth cohorts are expensive undertakings, justified, ultimately, if they produce scientific knowledge that can advance child health. The ECHO program of research uses extant cohorts, harmonizes data already collected across cohorts to the extent possible, encourages cohort investigators to collaborate in special studies of interest, and plans to collect new data in expanded cohorts in uniform ways. These efficiencies in study design, combined with the broad and diverse set of scientific interests found among study investigators and the plans to make the data available to all serious scientific researchers suggest that ECHO will be a rich source of useful, actionable information about child health for decades to come.

Acknowledgements

We would like to thank the PIs and co-PIs of our ECHO projects in New York, Puerto Rico and Michigan: Cristiane Duarte, Glorisa Canino, Jonathan Posner, John Vena, Ronald Wapner, Kelly Hunt, Douglas Ruden, Michael Elliott, Glenn Copeland and Charles Barone.

Financial support and sponsorship

The current work was supported by the National Institutes of Health, grants: UG3OD023285, UG30D023328 and UG3OD023316.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:
- ■ of special interest
- ■■ of outstanding interest

1. Belanger CF, Hennekens CH, Rosner B, *et al.* The nurses' health study. Am J Nurs 1978; 78:1039–1040.
2. Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. Br Med J 1964; 5395:1399–1410.
3. Marmot M, Brunner E. Cohort profile: the Whitehall II study. Int J Epidemiol 2005; 34:251–256.
4. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham Study. Am J Pub Health 1951; 41:279–281.
5. Beasley RP. Hepatitis B virus. The major etiology of hepatocellular carcinoma. Cancer 1988; 61:1942–1956.
6. Wadsworth M, Kuh D, Richards M, *et al.* Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). Int J Epidemiol 2006; 35:49–54.
7. Douglas JWB. Social and economic problems of child-bearing; a proposed inquiry. Mother Child 1946; 16:160–163.
8. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol 2006; 35:34–41.
9. Elliott J, Shepherd P. Cohort profile: 1970 British Birth Cohort (BCS70). Int J Epidemiol 2006; 35:836–843.
10. Douglas JWB. The home and the school. London: MacGibbon and Kee; 1964.
11. Hardy R, Wadsworth MEJ, Langenberg C, *et al.* Birth weight, childhood growth and blood pressure at 43 years in a British birth cohort. Int J Epidemiol 2004; 33:121–129.
12. Van den Berg BJ, Yerushalmy J. The relationship of the rate of intrauterine growth of infants of low birth weight to mortality, morbidity, and congenital anomalies. J Pediatr 1966; 69:531–545.
13. Klebanoff ML. The collaborative perinatal project: a 50-year retrospective. Paed Perinat Epidemiol 2009; 23:2–8.
14. Cohn BA, Cirillo PM, Hopper BR, *et al.* Third trimester estrogens and maternal breast cancer: prospective evidence. J Clin Endocrinol Metab 2017; 102:3739–3748.
15. Nelson KB, Ellenberg JH. Predictors of epilepsy in children who have experienced febrile seizures. N Engl J Med 1976; 295:1029–1033.
16. Nelson KB, Ellenberg JH. Antecedents of cerebral palsy. Multivariate analysis of risk. N Engl J Med 1986; 315:81–86.
17. ■ Tollånes MC, Strandberg-Larsen K, Forthun I, *et al.* Cohort profile: cerebral palsy in the Norwegian and Danish birth cohorts (MOBAND-CP). BMJ Open 2016; 6:e012777.

The article describes the creation of the MOBAND-CP cohort, a pregnancy cohort of more than 210 000 children, among whom 438 children with cerebral palsy (CP) were located in the Danish and Norwegian CP registries. Biological samples and survey data were collected in pregnancy and can now be examined for a wide range of potential causes in the largest prospective cohort study of CP ever mounted.

18. Surén P, Roth C, Bresnahan M, *et al.* Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. JAMA 2013; 309:570–577.
19. Divan HA, Kheifets L, Olsen J, *et al.* Prenatal cell phone use and developmental milestone delays among infants. Scand J Work Environ Health 2011; 37:341–348.
20. ■ Abel MH, Caspersen IH, Meltzer HM, *et al.* Suboptimal maternal iodine intake is associated with impaired child neurodevelopment at 3 years of age in the Norwegian mother and child cohort study. J Nutr 2017; 147: 1314–1324.

The study of nearly 50 000 mother–child pairs recruited during pregnancy found that low maternal iodine intake in pregnancy, assessed by a validated food-frequency questionnaire, was associated with child language delay, externalizing and internalizing behavior problems and fine motor skills. The importance of this observation is underlined by the finding, from NHANES data, that iodine levels in women of childbearing age have dropped significantly in the United States in recent years.

21. Khoury MJ, Gwinn M, Ioannidis JP. The emergence of translational epidemiology: from scientific discovery to population health impact. Am J Epidemiol 2010; 172:517–524.
22. Weissman MM, Brown AS, Talati A. Translational epidemiology in psychiatry: linking population to clinical and basic sciences. Arch Gen Psychiatry 2011; 68:600–608.