Annals of Epidemiology xxx (2017) 1–10

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Multidisciplinary design and analytic approaches to advance prospective research on the multilevel determinants of child health

Sara B. Johnson PhD, MPH [a,b,*], Todd D. Little PhD [c], Katherine Masyn PhD [d], Paras D. Mehta PhD [e], Sharon R. Ghazarian PhD [f]

[a] Johns Hopkins School of Medicine, Baltimore, MD
[b] Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
[c] College of Education, Texas Tech University, Lubbock
[d] Georgia State University School of Public Health, Department of Epidemiology and Biostatistics, Atlanta
[e] University of Houston Department of Psychology, Houston, TX
[f] Johns Hopkins All Children's Hospital, St. Petersburg, FL

ARTICLE INFO

Article history:
Received 24 September 2016
Accepted 5 May 2017
Available online xxx

Keywords:
Precision medicine
Mixture model
Missing data
Longitudinal
Multilevel
Cohort studies
Child health
Study design

ABSTRACT

Purpose: Characterizing the determinants of child health and development over time, and identifying the mechanisms by which these determinants operate, is a research priority. The growth of precision medicine has increased awareness and refinement of conceptual frameworks, data management systems, and analytic methods for multilevel data. This article reviews key methodological challenges in cohort studies designed to investigate multilevel influences on child health and strategies to address them.

Methods: We review and summarize methodological challenges that could undermine prospective studies of the multilevel determinants of child health and ways to address them, borrowing approaches from the social and behavioral sciences.

Results: Nested data, variation in intervals of data collection and assessment, missing data, construct measurement across development and reporters, and unobserved population heterogeneity pose challenges in prospective multilevel cohort studies with children. We discuss innovations in missing data, innovations in person-oriented analyses, and innovations in multilevel modeling to address these challenges.

Conclusions: Study design and analytic approaches that facilitate the integration across multiple levels, and that account for changes in people and the multiple, dynamic, nested systems in which they participate over time, are crucial to fully realize the promise of precision medicine for children and adolescents.

Introduction

Characterizing the environmental determinants of child health and development over time, and identifying the multilevel mechanisms by which these determinants operate, is an area of intense scientific interest. The National Children's Study (NCS) was commissioned by Congress in 2000 to prospectively investigate the biological, physical, chemical, and social determinants of health in 100,000 children. In late 2014, after enrolling approximately 5000 patients, the National Institutes of Health closed the study on the recommendation of a National Academy of Sciences report [1]. The National Academy of Sciences report pointed to flaws in the study design and approach and emphasized the need for:"new study designs, informed by advances in technology and basic and applied research across multiple disciplines" [2]. These recommendations highlight the need to optimize the quality, feasibility, and reliability of data derived from cohort studies in ways that acknowledge the multiple dynamic influences on human health. In this article, we discuss historical challenges and new opportunities to accomplish this goal using interdisciplinary design and analytic approaches.

The growth of precision medicine has increased awareness and importance for use of conceptual frameworks, data management systems, and analytic methods for multilevel data [3, 4]. Precision

This work was supported by the All Children's Hospital Foundation.

* Corresponding author. Division of General Pediatrics & Adolescent Medicine, Johns Hopkins School of Medicine, 200 N. Wolfe St, Suite 2017, Baltimore, MD 21287. Tel.: 410-614-8437; fax: 410-502-5440.
E-mail address: sjohnson@jhu.edu (S.B. Johnson).

http://dx.doi.org/10.1016/j.annepidem.2017.05.008
1047-2797/

medicine refers to "the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease or their response to a specific treatment. Preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not" [3].

Precision medicine is inherently concerned with dividing heterogeneous groups of individuals into more homogeneous subpopulations. Sources of data that can inform this stratification process occur at multiple levels. Biological units of analysis (e.g., organ systems, cellular processes, proteome, metabolome, epigenome, and genome) are dynamically nested in individuals' social ecologies (e.g., policy environments, neighborhoods, schools, healthcare systems, families), and interactions among these levels drive individual- and population-level patterns of health outcomes [4–7]. To date, conceptualization of the precision medicine approach has largely been driven by molecular determinants of risk, particularly genomics and other "omics" (e.g., proteomics, metabolomics). Here, we aim to highlight opportunities to integrate well-established bioecological frameworks (e.g., Bronfenbrenner [5]) with omics-driven precision medicine frameworks [4]. Figure 1 illustrates key nested determinants of health and well-being relevant to precision medicine. Although the ability to describe phenomena that shape health risk and outcomes at multiple levels is burgeoning, meaningfully integrating these data remains a challenge [3]. Analytic methods that can integrate across multiple levels and account for changes in people as well as the multiple, dynamic, and nested systems in which they participate over time are crucial to fully realizing the promise of precision medicine for children and adolescents.

In this article, we summarize key methodological challenges that have arisen in efforts to advance multilevel prospective research in child health. We outline multidisciplinary study design and analytic methods that can help optimize the quality, reliability, and feasibility of cohort studies designed to investigate multilevel environmental influences on child health and development. Although these methods, approaches, and software tools have been used in the developmental and social/behavioral sciences, they are less common in the biomedical sciences. We use the Prospective Research in the Determinants of Child Health Trajectories (PREDICT) cohort study as an example of the type of study that can leverage these multidisciplinary approaches.

The PREDICT study

The Johns Hopkins All Children's Hospital PREDICT study investigates the multilevel determinants of child health beginning early in life, with a focus on obesity and developmental delay. Funding for PREDICT leveraged an institutional investment designed to facilitate scientific investigation related to precision medicine, including a pediatric biorepository, integrated medical management systems, and a center of resources for integrative biology (genomics, proteomics, metabolomics) research [8]. PREDICT enrolls pregnant women beginning in their 12th week of gestation from obstetrical care and follows them in pediatric primary care. Data are collected prospectively from the mother and child's medical records including growth, development, diagnoses, and developmental milestones. In addition, parents complete questionnaires about their own health, their family, neighborhood, and their child's health and development, behavior and functioning. Direct observations/assessments of parental functioning and child cognitive and socioemotional development are also conducted along with teachers' report on the child's development and academic performance. Administrative data and geocoded address data are used to characterize neighborhood resources and environments. Biospecimens (cord blood, blood/saliva/buccal cells) are collected during follow-up and stored in the biorepository for future analyses. PREDICT involves multiple levels of nesting, including the individual (e.g., parent and child survey, clinical and biomarker data), dyad (gestational, marital, and parent/child), family, clinic/practice, and neighborhood. Further, both chronological and developmental time is considered.

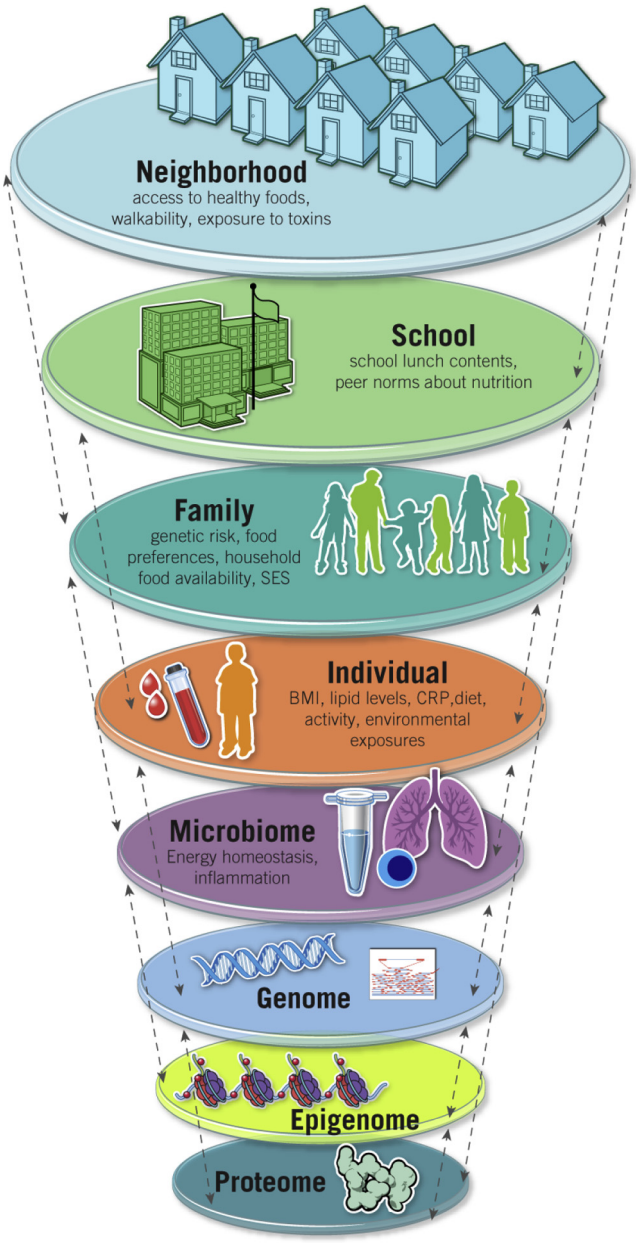

Fig. 1. Framework for pediatric precision medicine illustrating selected nested determinants of child health and well-being and levels of data commonly collected to facilitate targeted intervention and treatment activities. Arrows indicate interactions between levels. Note that not all possible arrows are included for readability.

Part 1: challenges in multilevel cohort studies

Cohort studies suffer from several potential threats to reliability and validity including multilevel data structures, missing data, and unobserved population heterogeneity; further, they require attention to variation in intervals of assessment, multiple reporters, and