

Longitudinal thinking

Increasing public and private sector cooperation to support long-term, longitudinal population studies is critical to advance translational research on complex diseases.

One of the most ambitious US research projects ever—the National Children’s Study on the effects of environmental influences on child health—will fully launch in the next 12 months. Although the project has been dogged by administrative missteps, wrangles over study design and poor allocation of budgetary resources, it promises to transform knowledge of how prenatal factors and early childhood experiences play a role in disease development in later life. It will also be an important testing ground for the potential of large-scale, multi-year epidemiological studies to facilitate elucidation of the role of hereditary, dietary, lifestyle and environmental factors in complex disease. Indeed, with the increasing availability of high-throughput technologies for monitoring many different types of omics data and the integration of information systems for sharing and accessing these with related health data, there has never been a better time to embark on such projects.

Legislation authorizing the Children’s Study was first enacted by the US Congress in 2000. The study was set up with the specific aim of examining the effects of environmental exposures on the growth, development and well-being of children and follows a long line of prospective longitudinal studies across the world yielding invaluable insights into risk factors for disease. Perhaps the archetype of these is the Framingham Heart Study, first established in 1948 in a Boston suburb, which has now amassed clinical data from more than 10,000 residents over three generations. The study’s data have been instrumental in identifying the role of factors, such as cigarette smoking, elevated cholesterol, hypertension, elevated triglycerides, sedentary lifestyle and diabetes, as causes of cardiovascular disease.

The Children’s Study aims to provide similar information about risk factors for illnesses, such as asthma, childhood cancers, neurodevelopmental disorders, obesity and type 2 diabetes, that are becoming increasingly common in children. Because these diseases have emerged comparatively recently, changes in human genetics are unlikely to account for their increasing prevalence. Thus, the design of sufficiently powered, long-term longitudinal studies to tease apart the gene-environment interactions behind these disorders is an important goal.

That is not to say such approaches are easy to conceive or implement. And the Children’s Study has suffered its fair share of problems, including a bloated consultation process (over 2,500 experts involved since the study’s inception), changes in strategic direction (a shift from a hypothesis-driven design to what is now ostensibly a data collection platform) and delays in recruiting study participants.

Despite these issues, there are still many reasons why the Children’s Study and other longitudinal, population-based, cohort studies like it deserve support—now more than ever. First, a host of patient monitoring mobile technologies are coming online that offer new possibilities for gathering phenotype information remotely. What’s more, current high-throughput analytical technologies, such as deep sequencing,

arrays, microfluidics and mass spectrometry, are enabling sample analysis on a hitherto unprecedented scale and breadth. Along with many of the phenotypic data gathered in traditional prospective studies, it is now possible to take urine, blood and even stool samples and catalog a host of molecular variables, including nuclear and mitochondrial DNA variants and copy number information; changes in DNA methylation, chromatin modification, coding and noncoding RNAs; and profiles of targeted sets of proteins, protein modifications and metabolites. At the same time, an appreciation that human ‘superorganisms’ comprise both host and microflora means that metagenomic and microbiome profiles can also be incorporated into study designs.

Advances in data storage, electronic health records and relational databases that can host, integrate and share data mean that information about subjects can be accessed by the research community more quickly and easily than ever before. Increasing awareness about reproducibility means that more complete records can be made of how data were generated and the types and version of instrumentation involved. In addition, there is an increasing appreciation of the importance of standardizing the patient consent process to ensure that data can be shared with researchers. Going forward, the introduction of new, more flexible forms of consent, like the Portable Legal Consent (*Nat. Biotechnol.* 30, 469, 2012), will also facilitate the reuse of interrogation of data by interested investigators.

Above all, industry is becoming increasingly aware of the importance of such large cohort studies in providing markers for use in trial design. In precompetitive collaborations like the Innovative Medicines Initiative (IMI) and The Foundation for the National Institutes of Health Biomarkers Consortium, companies are working with the public sector to identify biomarkers in patients with complex disease in the hope of breaking down conditions into molecularly defined subtypes and enabling the definition of presymptomatic individuals predisposed to disease.

Therapeutic interventions are likely to have a better chance of working earlier in a complex disease process than later. In the present situation, many of our current treatments fail because drugs are given to end-stage patients who manifest symptoms arising from disease processes that have become sufficiently dysregulated to result in a massive loss of function. In the field of Alzheimer’s, we are already seeing trials where a drug is being tested for its ability to forestall disease in cognitively healthy individuals (*Nat. Biotechnol.* 30, 731–732, 2012).

We are now on the cusp of a new era of digital medicine. Now is the time to start cataloging the molecular changes that happen during our lifetimes in longitudinal studies. The technology is ready. Industry is ready. And patients are ready. Even if our ability to unravel all of the risk factors is not quite ready, we will be providing the data foundation that will enable generations of researchers in the years to come to do just that.