Commentary

New Models for Large Prospective Studies: Is There a Risk of Throwing Out the Baby With the Bathwater?

Michael B. Bracken*, Dean Baker, Jane A. Cauley, Christina Chambers, Jennifer Culhane, Dana Dabelea, Dorr Dearborn, Carolyn D. Drews-Botsch, Donald J. Dudley, Maureen Durkin, Barbara Entwisle, Louise Flick, Daniel Hale, Jane Holl, Melbourne Hovell, Mark Hudak, Nigel Paneth, Bonny Specker, Mari Wilhelm, and Sharon Wyatt

* Correspondence to Dr. Michael B. Bracken, Center for Perinatal, Pediatric and Environmental Epidemiology, Yale University School of Public Health, One Church Street, 6th Floor, New Haven, CT 06510 (e-mail: michael.bracken@yale.edu).

Initially submitted March 28, 2012; accepted for publication September 27, 2012.

Manolio et al. (Am J Epidemiol. 2012;175:859–866) proposed that large cohort studies adopt novel models using “temporary assessment centers” to enroll up to a million participants to answer research questions about rare diseases and “harmonize” clinical endpoints collected from administrative records. Extreme selection bias, we are told, will not harm internal validity, and “process expertise to maximize efficiency of high-throughput operations is as important as scientific rigor” (p. 861). In this article, we describe serious deficiencies in this model as applied to the United States. Key points include: 1) the need for more, not less, specification of disease endpoints; 2) the limited utility of data collected from existing administrative and clinical databases; and 3) the value of university-based centers in providing scientific expertise and achieving high recruitment and retention rates through community and healthcare provider engagement. Careful definition of sampling frames and high response rates are crucial to avoid bias and ensure inclusion of important subpopulations, especially the medically underserved. Prospective hypotheses are essential to refine study design, determine sample size, develop pertinent data collection protocols, and achieve alliances with participants and communities. It is premature to reject the strengths of large national cohort studies in favor of a new model for which evidence of efficiency is insufficient.

cohort studies; epidemiology; prospective studies; research design

Abbreviation: NCS, National Children’s Study.

Editor’s note: A response to this article appears on page 290, and a related letter appears on page 292.

In their commentary, Manolio et al. (1) discussed the importance of exploring new ideas for the design and conduct of large cohort studies. We agree that large prospective studies can be powerful instruments for making new discoveries of disease causation and acknowledge that the experiences with new research strategies being tested in other countries deserve close examination for their relevance to the particular challenges of US populations. However, several issues mentioned in the commentary warrant more critical examination. We categorize these into matters related to science and study design and issues of study conduct and the continuing successful engagement of large national cohorts.

SCIENTIFIC CONSIDERATIONS FOR NEW MODELS

First, although there is increasing interest and potential value in harmonizing data across different studies (2), harmonizing data from extant databases within a cohort study that used different clinical or administrative databases is fraught with difficulty. This undermines opportunities for
stringent harmonization (or standardization), and more flexible retrospective harmonization procedures have difficulty ensuring “inferential equivalence” to permit data synthesis (3, p. 262). Moreover, epidemiologists would caution that this approach has limited utility and risks bias that can confuse causal processes. Future important gains in understanding genetic and environmental causes of disease demand more precise specification of phenotype than is typically available in administrative or clinical databases. For example, advances in understanding the genetic etiology of adult macular degeneration required precise characterization of the wet versus dry adult macular degeneration phenotypes (4, 5); the etiology of cardiac ventricular septal defects varies for the muscular versus perimembranous phenotypes (6); and heritable as well as environmental determinants of bone mineral density vary according to bone site (7, 8).

Many of the examples provided by Manolio et al. come from countries with national health care systems, where harmonization is more feasible and standardized data on disease outcomes are readily accessible (although even in these systems, disease classification may be inadequate for research purposes). The lack of a universal health care system in the United States and the medical record system geared toward reimbursement, together with significant heterogeneity in the identification and measurement of disease outcomes across many healthcare systems, is a formidable if not lethal barrier to establishing a UK Biobank type of methodology in the United States for testing hypotheses about environmental exposures and health outcomes.

Second, environmental studies require sequential temporal monitoring of a geographic area and sometimes individual-level monitoring over time. Reliance on biologic samples alone to detect or quantify environmental exposure cannot evaluate many important exposures (e.g., ionizing and nonionizing radiation, sound, light, and other physical exposures) and is inadequate for others (e.g., air and water pollution) that require integrated exposure assessment over time (9–12). Blood samples, at best, may only represent a limited number of narrow exposure windows and do not always reliably reflect the external environment (13). Importantly, biologic samples cannot contribute to policy decisions that must be based on ambient external exposure levels that can be regulated to reduce risk of disease.

Third, the commentary suggests that selection bias and response rates as low as 1%–10% are concerns only for the estimation of disease frequency and not for investigations of risk associations. To the contrary, selection bias due to convenience sampling and low response and retention rates can compromise the internal validity of a study and cause consequential error in risk estimation of associations between exposures and health outcomes.

The risk of faulty inference is especially pronounced in the presence of interactions (14). A recent analysis of data from the Fragile Families Child Welfare Study, which is based on a probability sample, shows how and why previous results regarding gene-environment interaction and maternal depression were not generalizable. Mitchell et al. (15) examined 2 polymorphisms of the serotonin transporter gene 5-HTT and found that among mothers with no reactive alleles, level of education had no impact on depression, but when reactive alleles were present, high educational levels decreased the risk of depression and low educational levels increased risk. The authors concluded that sample selection could entirely determine whether any interaction was found, as well as the direction of the interaction (15). These and other examples contradict the statement in the commentary that low response and the “healthy volunteer effect” can lead to underestimation of disease prevalence and incidence, but [their] impact on relative risk estimates for environmental and genetic factors is generally not important (1, p. 862).

Fourth, large study sample sizes (on the order of 500,000 to 1 million participants) are proposed as being necessary for studying rare exposures and rare disease, but valid retrospective research designs already exist for these situations. However large the sample of a prospective study, it will still be too small to detect some rare events and highly inefficient because very large numbers in the cohort will be uninformative for any single research question. Rather than embarking on very costly cohort studies of rare disease, it would be a better investment of resources to improve the methodology of smaller but statistically more powerful retrospective research designs (16). The commentary did not make an economic or scientific case for prioritizing cohort studies with “hundreds of thousands of participants” (1, Table 3). The size of a study can only be scientifically determined in light of the scientific questions it is designed to answer. As the commentary gives no consideration to hypotheses, it provides no scientific way to decide the optimum size of any cohort.

Fifth, national cohort studies should use a probability sampling frame so that their results may be generalized to the national population as a whole (14, 17, 18). Few of the studies mentioned in the commentary allow for extrapolation of results to a national population. The US population-based studies cited by Manolio et al. are each located within 1 provider in a single geographic location (Marshfield Personalized Medicine Research Project, Marshfield, Wisconsin; Kaiser Permanente Research Program on Genes, Environment, and Health, Oakland, California; and Vanderbilt University’s DNA databank, Nashville, Tennessee). Convenience sampling used by studies such as the UK Biobank Study will inevitably lack appropriate representation of underserved populations, such as the poor, the less educated, immigrants, and ethnic minorities. This problem is accentuated in pregnancy cohort studies, such as the US National Children’s Study (NCS), in which social and ethnic factors associated with volunteering for research studies are especially prominent determinants of exposure and outcome (19). High-risk pregnancies that are of particular interest and have many of the outcomes of interest will be seriously underrepresented without careful implementation of probability sampling (20).

Sixth, Manolio et al. appear to prefer a hypothesis-free approach, a strategy that has worked quite well in large gene-hunting exercises. However, the structure of the genome is more amenable to hypothesis-free research than is investigation anchored in a constantly changing physical, social, and behavioral environment, where lack of hypotheses can lead
to wasteful research (e.g., sample sizes being too small or unnecessarily large) or where measured exposures are later determined to be inadequate to address subsequently developed hypotheses. Potential participants, community groups, and governmental funding agencies are unlikely to engage in research in which the only promise is to collect data for banking that will be used to address future and as yet undetermined research questions. Hence, for strategic, economic, and indeed ethical reasons, it is imperative that scientists who propose resource-intensive large cohort studies in which many future hypotheses cannot be anticipated still articulate a priori some important hypotheses that the protocol design can credibly address. The mandate for hypothesis-driven investigation must necessarily extend to epidemiologic-genetic and gene-environment research as well so that the environmental exposure can be adequately assessed. Such hypotheses would constitute the minimal “bang for the buck” that the taxpayers and scientists can expect the study to evaluate. We note that the UK Biobank project realized the importance of hypotheses about exposure-disease relations during protocol development (21).

STUDY MANAGEMENT AND CONDUCT OF LARGE PROSPECTIVE STUDIES

Seventh, the commentary (1) fails to emphasize the critical importance of continuous community engagement in the success of longitudinal cohort studies. Lessons learned from the Centers for Children’s Environmental Health and Disease Prevention Research (22, 23) and white papers developed specifically for planning the NCS (24) all emphasize the importance of a supportive community infrastructure. For example, Navajo tribal approval processes requires not only documented community support but also that the community and the culture be protected and receive benefit from any research conducted within the Nation (25). Cohorts established with the purpose of following participants for 21 years, and especially those in which investigators intend to collect biologic and genetic data, such as the NCS, must gain the support of their communities to sustain participant confidence needed for retention over decades. Other countries may not have to overcome the legacy of past breaches of trust by clinical and public health researchers, such as the Tuskegee syphilis study, the Willowbrook hepatitis research, and US government-sponsored radiation research on humans from the 1940s to the 1970s, which are fresh in the memories of communities and institutions today (26–28). Even the UK Biobank had difficulty engendering trust among the United Kingdom population, resulting in an initial 10% enrollment rate (29, 30) that appears to have dropped to 6% (500,000 of 9 million invitations sent) according to Manolio et al. (1).

Eighth, one important characteristic of an optimal cohort study not mentioned in the commentary by Manolio et al. is a high retention rate for follow-up visits. High retention rates require consistent engagement with the participants and their communities. Local fieldwork is often needed to track down the reasons for a failed contact. Although participants may leave the study at random, the majority leave because of systematic causes. This means that most dropouts will bias association studies. Moreover, the magnitude of bias could be great; for many pathologies, the subset at risk of dropping out are those at highest risk of disease. It is not accurate to assume that retention rates will be high in single-site studies carried out by health care organizations because it assumes that an individual will stay within that health care system. This is an assumption that has not been tested and is especially relevant in younger participants (like pregnant mothers) compared with older populations. Younger people tend to be more mobile in both their jobs and with their health care providers because of continuing rapid changes in managed care.

Although Manolio et al. (1) suggested that centralized invitations for appointments can maximize efficiencies in data collection, evidence is lacking. The impact on long-term retention of participants resulting from shifting from a multisite model with involvement of local institutions and investigators to a centralized model has never been rigorously tested. The Women’s Health Initiative, listed under the decentralized multisite design (1, Table 2), was largely organized by academic investigators and has already generated knowledge that has changed medical practice, reduced the incidence of invasive breast cancer, and may improve the health of postmenopausal women in other ways. It remains to be seen whether the nonacademically based hypothesis-free approach advocated by Manolio et al. will have similar measurable effects on public health (31–33).

Ninth, the commentary refers to the importance of including academic investigators in study leadership, and we agree. In contrast, the NCS, with which all the authors are associated, has largely excluded academic investigators from important scientific decisions. The study has cost hundreds of millions of dollars to date without articulation of a main study design that is hypothesis-driven or a protocol that is based on a sound scientific design. One proposal is that the NCS transition to a nongovernment organization to act as the coordinating center as has been done for some large cohort studies. The original 7 NCS Vanguard Centers were initially under this model, which was later terminated because of the large costs associated with contracting a coordinating center. In the case of the NCS, we estimate that about half of the $800 million spent so far has gone to nonacademic subcontractors; as yet, no economic analysis has been presented to document any gain in efficiencies. Centralized information management systems, instrument creation, and study protocol should be more cost effective, but this has no bearing on the need to accomplish participant recruitment and to maximize retention at the local level.

Tenth, meaningful inclusion of academic investigators can add value to a large study beyond the ability to provide scientific advice. Many have worked in their communities for decades and have a deep understanding of the organization and function of their local health care systems. For example, provision of antenatal care in the United States varies tremendously from region to region. Creating a representative pregnancy cohort has little chance of success without local knowledge of, and well-developed relationships with, involved health care providers and communities. Community engagement was emphasized (along with hypothesis development and population sampling) in developing the
Framingham Heart Study, one of the most successful research cohort studies (34). Eleventh, the experience with largely US government-organized and -directed cohort research, as exemplified by the NCS, is far from encouraging. Heavy reliance on non-academic consulting organizations in the place of university or academic medical center-based investigators, restriction of innovation by academic investigators, rapid and unexplained changes in study protocol, and inconsistent guidelines and timelines have characterized the NCS. Early and consistent input from NCS investigators with strong track records in community research, pregnancy and newborn recruitment, and epidemiologic methods would have led to a more productive and cost-effective result.

Twelfth, Manolio et al. (1) state that the UK Biobank costs were within budget ($100 million) and were substantially reduced by shifting to a centralized design. It is unclear whether other potentially more economical designs were considered. UK Biobank critics noted that case-control studies might have provided more power at a lower cost. Initial funding for the project was set at €45 million (35) and increased to €61 million in 2006 (36).

In summary, we suggest that the commentary authors’ conclusion that “large prospective studies are not simply small studies made large” (1, p. 864) leads them to ignore the value of lessons learned from the great body of prospective cohort studies of smaller size than the UK Biobank. The attention to detailed research design, hypothesis testing, internal and external validity, nurturing of community relationships, and care in participant recruitment and retention that are essential in smaller cohort studies are equally important in national cohorts. We argue that these “best practices” in community research are best entrusted to and managed by local investigators. Considerable budgetary savings may be achieved by putting more resources at the local level, thereby involving experienced investigators who preferably work under cooperative agreements, and fewer into central governmental program offices supported by an unwieldy array of nonacademic subcontractors with overhead costs possibly higher than those of academic centers. Successful models of these types of collaborative agreements are found in clinical consortia throughout the National Institutes of Health.

As medical and health researchers who have been active in conducting cohort studies, some with thousands of participants, we consider the approach recommended by Manolio et al. (1) to have narrow applicability. For scientific and public health goals that require serious assessment of the human environment matched with precision in phenotype definition, the model described in their commentary appears unworkable in the United States.

ACKNOWLEDGMENTS

Author affiliations: Center for Perinatal, Pediatric, and Environmental Epidemiology, School of Public Health and School of Medicine, Yale University, New Haven, Connecticut (Michael B. Bracken); Center for Occupational and Environmental Health, School of Medicine, University of California—Irvine, Irvine, California (Dean Baker); Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Jane A. Cauley); Department of Pediatrics, Division of Dysmorphology and Teratology, School of Medicine, University of California—San Diego, La Jolla, California (Christina Chambers); Department of Pediatrics, Division of Adolescent Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania (Jennifer Culhane); Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado (Dana Dabelea); Department of Environmental Health Sciences and Department of Pediatrics, Division of Pediatric Pulmonology, School of Medicine, Case Western Reserve University, Cleveland, Ohio (Dorr Dearborn); Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia (Carolyn D. Drews-Botsch); Department of Obstetrics and Gynecology, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Donald J. Dudley); Population Health Sciences and Pediatrics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin (Maureen Durkin); Carolina Population Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina (Barbara Entwisle); Department of Epidemiology, School of Public Health, Saint Louis University, St. Louis, Missouri (Louise Flick); Department of Pediatrics, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Daniel Hale); Institute for Healthcare Studies, Pediatrics and Preventive Medicine, General Academic Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Jane Holl); Health Promotion and Behavior Science, Center for Behavioral Epidemiology and Community Health (CBEACH), Graduate School of Public Health, San Diego State University, San Diego, California (Melbourne Hovell); Department of Pediatrics, Division of Neonatology, College of Medicine Jacksonville, University of Florida, Jacksonville, Florida (Mark Hudak); Department of Epidemiology and Biostatistics, College of Human Medicine, Michigan State University, East Lansing, Michigan (Nigel Paneth); Department of Pediatrics and Human Development, College of Human Medicine, Michigan State University, East Lansing, Michigan (Nigel Paneth); EA Martin Program in Human Nutrition, South Dakota State University, Brookings, South Dakota (Bonny Specker); Department of Nutritional Sciences, University of Arizona, Tucson, Arizona (Mari Wilhelm); and Department of Internal Medicine, Division of Hypertension, School of Medicine and School of Nursing, University of Mississippi Medical Center, Jackson, Mississippi (Sharon Wyatt).

Conflict of interest: none declared.

REFERENCES


