Actes de conférences
Enquêtes longitudinales sociales et de santé
dans une perspective internationale

Conference proceedings
Longitudinal Social and Health Surveys in
an International Perspective

Developing a unified CIHR strategy for a Canadian:
National Cohort Initiative

Michael S. Kramer, MD and John Frank, MD

Planning within CIHR for one or more Canadian national cohort studies is now at a crossroads. When CIHR replaced MRC as Canada's national health research agency, the Canadian Lifelong Health Initiative (CLHI) became one of the cross-cutting initiatives that would supplement and complement the strategic research planned and funded through the 13 newly created Institutes. The CLHI was conceived as a large, nationwide cohort (i.e., prospective follow-up) study or combination of cohort studies that would provide new knowledge about the genetic, environmental, social, life-style, and behavioural determinants of key health and disease outcomes across the lifespan.

Initial discussions about the CLHI focused on two components: the determinants of successful aging (which evolved into the Institute of Aging's Canadian Longitudinal Study of Aging, or CLSA) and the etiology of common health outcomes of pregnancy and childhood (the Canadian National Birth Cohort, or CNBC), led by the Institutes of Population and Public Health, Genetics, and Human Development and Child and Youth Health. The CLSA started early and ambitiously to develop a detailed research protocol, with three principal investigators and 200 collaborators across the country. The protocol calls for a 50,000-member, population-based "tracking cohort," of whom 30,000 located near major medical centers would comprise an "intensive cohort" for closer scrutiny and more frequent, in-depth study of key exposures and outcomes. The protocol has been peer-reviewed by a panel of international experts and has received interim funding by CIHR's Governing Council for continued planning and pilot work.

The CNBC started later and has progressed more slowly. It has focused on quantitative traits such as height, weight, adiposity, blood pressure, blood lipids, IQ, and inattentiveness. The CNBC has held two design-oriented workshops, with several international participants. At one point, the CNBC considered a formal link to the CLSA as part of a single, multigenerational, family-based cohort study. This idea appealed to the Canadian genetics community, as well as the international experts who attended the workshops. Although concerns were raised about selection factors (e.g., the infeasibility of recruiting the older generation among recent immigrant groups and

the exclusion of children whose grandparents had died prematurely), adaptive designs were discussed that could at least partly circumvent these problems.

The multigenerational design was also enthusiastically endorsed by several internationally renowned population geneticists, epidemiologists, and biostatisticians who attended the annual meeting of the American Society of Human Genetics held in Toronto in the fall of 2004. Multigenerational cohorts would provide more genetic information "per subject" and, with repeated environmental and other exposure measures and biological samples, have more power than their sheer numbers might indicate. The 20-40 year age group, i.e., the middle generation, would be included in the study, but with less in-depth study than the older and younger generations.

While these meetings and discussions were proceeding on the aging and pregnancy/childhood fronts, members of the cancer research community have also been contemplating the possibility of a large cohort study focused on cancer, but also including other, more common chronic disease outcomes such as cardiovascular, neurologic, and musculoskeletal diseases. (The CLSA also incorporates the study of these outcomes.) The retrospective case-control study has served as the major methodologic approach to assess postulated etiologic determinants of most cancers, because the outcome is relatively rare and often occurs many years or decades following the exposures that cause it. But a prospective cohort approach is essential to identify preventable environmental and life-style factors (such as diet and physical activity) that are not routinely recorded at the time they occur and cannot be measured with sufficient validity and precision using a retrospective case-control design. Several large cancer cohort studies have been attempted elsewhere (e.g., EPIC in continental Europe and Biobank in the U.K.). But such studies have not incorporated such recent scientific advances as molecular characterization based on fresh tissue collection and genetic (e.g., haplotype mapping), genomic, and proteomic technologies that could improve screening and early detection and lead to individually-based therapy and prognosis. Indeed, the Alberta Cancer Board has already embarked on a cancer cohort study in that province, and Ontario has recently announced its intention to launch a similar study. Moreover, discussions have been held since the summer about Canada's possible contribution to a very large international cancer cohort study (John Potter's "last cohort"), which no individual country (and Canada in particular) is likely to be willing and able to mount on its own.

After several years of effort and several hundred thousand dollars of financial support, the time has clearly come to develop a single, unified strategy. CIHR and its potential funding partners (governmental and nongovernmental) must develop a scientifically rigorous, feasible, ethical, and cost-effective proposal that provides unique knowledge to improve the future health and health care of Canadians, and a wise investment for Canadian taxpayers and their government.

1. Goals of a unified cohort strategy

The overall goal of a unified strategy for a national cohort study is to identify modifiable factors that cause adverse health outcomes or promote beneficial ones among Canadian children and adults. The adverse outcomes causing the greatest mortality, morbidity, and disability among Canadian children are adverse pregnancy outcomes, asthma, obesity, injury, attention deficit disorder and hyperactivity, poor cognitive development, and conduct disorder. For older adults, the major conditions are cardiovascular disease (including stroke), cancer, obesity, type 2 diabetes, arthritis, injury, depression and other mental illness, dementia, disability, and social isolation. Some adverse health outcomes occur so infrequently in the population that a cohort study designed to investigate their risk or protective factors would require several hundred thousand participants. For children, diseases like autism, cerebral palsy, and most congenital anomalies require unrealistically large sample sizes to study using a cohort design. For adults, very large sample sizes are required to study all but the most common cancer sites, as well as rare chronic diseases such as multiple sclerosis. The beneficial outcomes to be studied for promoting factors (exposures) in children are good cognitive development, academic performance, and social adjustment with parents, siblings, and peers. Beneficial outcomes in older adults include health-related quality of life, productivity, mobility, independence, and social interaction.

The study of genetic and environmental determinants of quantitative traits (e.g., height, adiposity, fat distribution, IQ, insulin sensitivity, blood pressure, lipids) can be studied with much smaller cohorts than are necessary for dichotomous (present vs absent) outcomes. The role of gene-environment interactions in the etiology of such traits has not been a primary focus of previous cohort studies in either children or adults. But a long-term study of these traits may not excite the imagination of Canadian investigators, politicians, other potential funders, and the general public.

National cohort studies should provide new information that is pertinent to the health and health care of Canadians. The health outcomes studied may not be unique to Canadians, but the potential risk or protective factors studied should include such uniquely Canadian features such as cold climate, prolonged periods of exposure to indoor air, remote (rural and Northern) access to health care services, and – if logistically feasible and locally acceptable – Aboriginal social/cultural issues. It should also provide a long-term "population laboratory" for epidemiologists, biostatisticians, health service researchers, and population geneticists that will help attract and retain scientists of international stature and stimulate the training of new investigators in these domains.

The time, effort, and expense of designing and carrying out one or more national cohort studies in Canada are difficult to justify unless such an initiative fills a unique knowledge niche, i.e., unless the information it yields is both novel and useful. A "me too" approach that attempts to duplicate questions and approaches already addressed

in cohort studies in the U.S., U.K., Denmark, Norway, or Australia is unlikely to generate sufficient scientific interest, political will, or financial support.

Canada should capitalize on its proven strengths. These include a growing expertise in interdisciplinary collaboration; enthusiasm of the genetics, aging, cancer, chronic disease, child development, epidemiologic, health services, and environmental health research communities; interest in studying critical and sensitive time periods when transitions occur throughout the life course, including (perhaps) the preconceptional period; experience in measuring indoor and outdoor exposures associated with a cold climate; expertise in measuring the quality of maternal-child interaction and of day care facilities; and the capacity to link to existing administrative databases. For some exposures, it may be worth considering an embedded randomized controlled trial (RCT) component to avoid the otherwise inevitable confounding that would bias comparisons based on a purely observational design (e.g., comparisons of cognitive and behavioural outcomes among infants exposed to stimulating vs passive interactions with their mothers).

2. Size, "thickness", and costs

As shown in the accompanying Table, potential options for one or more Canadian national cohort studies depend on the choice (or relative contribution) between two contrasting approaches, each with its own philosophy, advantages, and disadvantages: a large, "thin" cohort vs a small, "thick" cohort. The size of the cohort refers to the sample size, i.e., the number of individual subjects followed in the cohort. "Large" cohorts are usually those of at least a hundred thousand, whereas "small" can be conceptualized as cohorts under 30,000 subjects. The "thinness" or "thickness" of the cohorts depends on the number of contacts with each subject and the amount of information collected at each contact. That information concerns exposure measures (i.e., risk factors), health outcomes, and other variables (potential confounding or modifying factors).

"Thin" cohort studies depend on infrequent, easy-to-obtain, and relatively inexpensive exposure measures, such as those that can be obtained on routine vital statistics, administrative records, DNA testing, single mailed questionnaires, or brief interviews and on outcomes that can also be obtained through administrative databases such as vital statistics, hospital discharge records, cancer registries, etc. "Thick" cohort studies generally require frequent and more expensive exposure measurements that combine environmental (including the physicochemical and psychosocial environments) and genetic measures. The outcomes may also be difficult and expensive to ascertain in "thick" cohorts, such as outcomes requiring biochemical, imaging, physiologic, or functional measures.

Measurement of exposures and outcomes in thin cohort studies can lead to substantial misclassification of exposure and/or outcome consequent and thus underestimation of

associations between exposure and outcomes. Perhaps more importantly, however, many of the exposures and outcomes of potential interest to a Canadian national cohort study simply cannot be obtained through infrequent, simple, inexpensive, and readily available data sources. "Thick" cohort studies may permit more pertinent measurements and less misclassification of exposures and outcomes, but the expense of obtaining those measurements often precludes large sample sizes. "Thick" cohort studies are therefore capable of detecting only large associations between exposures and outcomes.

The need for "one-off" vs repeated exposure measurements depends, of course, on the questions being asked. For example, associations between genetic mutations and specific diseases can be assessed easily and inexpensively. Analysis of single nucleotide polymorphisms (SNPs) is relatively inexpensive and requires only a single measurement, since the DNA sequence does not change over the subject's lifetime. Even entire genome scans may soon be affordable for large numbers of subjects. DNA samples are easily obtained through venipuncture or even buccal smears. On the other hand, detailed measures of the physicochemical environment and of individual behaviours are much more expensive, since the sampling and analysis of the air, soil, water, and diet requires measurement of many chemical entities. In addition, diet and physical activity change several times per day, over the days of the week, seasonally, and over longer periods of time, thus requiring multiple ascertainments of these exposures. Measuring the frequency and quality of family interactions also requires either invasive and expensive videotaping or frequent in-person interviews or observations.

Moreover, DNA can be frozen, requires only tiny quantities of blood or other nucleated cells, can be amplified as needed in the future, and analysis can be limited to diseased cases and a small fraction of controls (noncases). Many physical, chemical, and psychosocial exposures, however, are far more expensive to obtain, and for some of them (such as volatile air pollutants or interviewer-based observations of interpersonal interaction), storage for later analysis is difficult or even infeasible.

3. How can these goals be unified?

A cancer cohort study would require a very large sample size, at least 300,000 and perhaps as large as 1,000,000 to detect a statistically sufficient number of cases of individual nondermatologic cancers within a decade or so, even for such common sites as breast, colon/rectum, lung, lymphatic system, pancreas, ovary, and stomach. Because the CLSA focuses on functional measures of successful and unsuccessful aging, its sample size requirement is far more modest. The relevant study exposures (potential risk or protective factors) are likely to differ for these two studies, however. Achieving the goals of both studies would require that a smaller subcohort of a larger (cancer) cohort receive the in-depth measurements of the exposures and functional outcomes relevant to the CLSA. As indicated above, such an "intensive" subcohort of

approximately 30,000 is, in fact, already part of the CLSA's proposed design. If the CLSA study sample were a subcohort derived from Canada's contribution to a large, international cancer cohort study, the goals of both the aging and cancer/chronic disease cohorts could both be achieved.

For the pregnancy/child cohort, many of the key health outcomes of interest are relatively common, and it should be possible to identify separate genetic and environmental effects, as well as effects of gene-environment interactions, with a final sample size under 50,000 subjects. It can be argued, however, that the relevant exposures in children are highly outcome-specific. For example, pre- and periconceptional environmental and nutritional exposures may be particularly relevant to fertility, pregnancy, and early neurocognitive childhood outcomes; indoor air exposures are likely to be far more important for asthma; while maternal-child interactions are of key salience for attention deficit, hyperactivity, and conduct disorders. Given the specificity of exposures for specific outcomes, the expense of measuring the multiple relevant exposures at multiple potential time windows, and the high incidence of the outcomes, it may be both infeasible and highly inefficient and costly to measure all these exposures in all members of one large omnibus cohort. Several smaller purpose-built cohorts, perhaps as subcohorts under a larger umbrella cohort, may well be preferable. CIHR-IHDCYH's recently launched pregnancy/birth cohort study of asthma and allergy is an example of such a purpose-built cohort, although the interest and availability of numerous federal partners demonstrated for that study may be difficult or impossible to duplicate for other key questions concerning pregnant mothers and children.

As discussed earlier, adults and children could be combined in a multigenerational cohort study. An overall cohort initiative comprising 3 generations would be easier to "market" to government and other potential funding partners than separate studies of children and older adults. Such an approach would also facilitate access to grandparental DNA. A third advantage would be its potential to ascertain exposures over 3 generations to such key risk determinants as diet, physical activity, substance use, familial factors that may influence health and function, and social support.

A consensus is urgently needed before planning can proceed within and outside CIHR. The political will and financial commitment required to fund, design, and implement one or more large Canadian national cohort studies can be garnered only if the necessary effort and resources can be justified scientifically and in terms of likely impact on the future health and health care of Canadians. All are agreed that the large, long-term, stable funding required must be sought outside of CIHR's year-to-year budgetary allocation. Despite discussions over the last several years, further clarification of the following central questions is required before such funding is requested:

1) What is the unique scientific niche that can be filled by cohort studies in Canada, with immense geography and cultural diversity but limited population and financial resources?
2) Is a cohort (prospective follow-up) design required to answer the key scientific questions related to cancer, adult chronic diseases, successful aging, and important child health and development outcomes?
3) How can an overall national cohort initiative (or an "umbrella" study comprising several cohorts sharing a common infrastructure) be designed that addresses these diverse questions?
4) Should the healthy aging cohort (i.e., the CLSA) be nested within a larger cancer cohort of middle-aged and elderly adults?
5) What is the added value of linking the pregnancy/child cohort to the CLSA cohort through a multigenerational design beyond the availability of DNA across multiple generations, and is that added value worth the added cost?

Table. Cohort Study Attributes: Two Contrasting Approaches

Study Attribute	Large/Thin	Small/Thick
Types of Exposure Measure	simple, inexpensive	complex, expensive
Frequency of Exposure Measurement	infrequent	frequent
Types of Outcome Measure	simple, inexpensive	complex, expensive
Sample Size	high (≥100,000)	low (≤30,000)
Cost per Subject	low	high
Precision/Power	high	low
Exposure/Outcome Misclassification	high	low