Introduction

Low-cost computing power has made possible the storage and analysis of large quantities of health data. Individual electronic health records (EHRs) offer the potential for major improvements both in patient care and in the extent and quality of data available for population health research. The Canadian government has committed more than $1 billion in infrastructure funds for the so-called Canada Health Infoway, with a specific mandate to accelerate the development of EHR systems across Canada. Several provincial governments are actively supporting complementary initiatives. In addition to facilitating secondary uses of data that were gathered in the course of providing and financing health care, the ‘information revolution’ facilitates multiple, low-cost analyses of data gathered for research purposes. The proposed Canadian Lifelong Health Initiative (CLHI) relies on this capability. This longitudinal prospective research program would involve both an early birth cohort and an aging cohort. Health status information, information related to various social determinants of health, and biological materials would be collected and maintained as a "research platform" to support multiple individual research studies, over a period of decades. A larger scale study in the UK (Biobank) will collect baseline data and blood samples from 500,000 people aged 45–69 years, who will be followed for at least 10 years. Such projects offer unprecedented opportunities to study interactions between genetic variations and environmental variables, which can and should be defined in the broadest sense, and to provide an evidence base for policies and interventions to improve population health.

Somewhat belatedly, legislative and policy attention has focussed on the ethical and social implications of these new technological capabilities. In the United States, the main national response has been a health information privacy rule under the Health Insurance Portability and Accountability Act (HIPAA) of 1996. In Canada, legislation governing commercial users of health information and federally regulated industries – the Protection of Personal Information and Electronic Documents Act – which came fully into force in January 2004, complementing a variety of existing federal and provincial privacy statutes. However, this article is concerned not with the content and implementation of these requirements, but rather with a key omission in the underlying ethical and public policy rationale.

Privacy, and Beyond

At least in the Anglo-American countries, ethical analysis of research involving health data is highly individualistic. It tends to focus on potential encroachments on individual privacy; on preventing unauthorized or inappropriate access to data related to the health status of a particular, identifiable individual; and (when data are gathered specifically for research purposes) on the nature of the process by which consent is obtained from research participants. If confidentiality can be assured, then many ethical problems are considered solved, even if data were originally collected with no indication that they would subsequently be used for research purposes. For instance, the Tri-Council Policy Statement (TCPS) that provides guidelines for ethics review of federally funded research in Canada states that secondary
use of data “becomes of concern only when data can be linked to individuals.” Commentators often presume that a legitimate tradeoff exists between individuals’ wishes to control use of information they have provided and society-wide benefits in the form of improved health system performance or improved understanding of the determinants of health. When data are gathered specifically for research, ethical debate tends to focus on the question of whether and how the process of obtaining consent can take into account uses of data that might not be envisioned at the time the participant is originally recruited, and the circumstances under which renewed or amended consent should be sought. Somewhat paradoxically, this has the effect of setting a more demanding ethical standard when individuals are clearly informed that data are being gathered for research purposes than for research involving administrative and clinical data gathered in the course of providing health care or other services.

Ethics analysis of research uses of health data, whether or not they were gathered for research purposes, generally ignores harms or benefits that may result for individuals not because they are personally identified, but because of shared characteristics: they are part of a population whose interests and well-being are affected by the research or by its subsequent dissemination and use. A partial exception to this generalization involves literature on population genetics, where some authors acknowledge that potential harms to individuals, by virtue of their identification as members of a definable group or community (e.g., with high genetic vulnerability to a particular condition), should be considered when choices are made about such matters as how research is explained to participants and how consent is obtained. The issue of “group consent”, however, remains contentious.

As important as they are, privacy and confidentiality are only part of the ethics equation. Justice is one of the canonical “four principles” of bioethics, at least in its North American version. Crucially, justice includes distributive justice, and should (in the author’s view) incorporate special concern for populations disadvantaged by one or more of the interacting gradients of power and powerlessness (class, gender, education, (dis)ability, stigmatization) that characterize contemporary society.

The central argument of this article is that ethical analysis and institutional approval of population health research must be explicitly concerned with avoiding harms to subordinated or marginalized populations. Terminology is important here: ‘population’ is preferable to ‘community’ or ‘group,’ since each of these implies an element of self-recognition and internal cohesion. The descriptors ‘subordinated’ and ‘marginalized,’ although imperfect, are preferable to ‘vulnerable.’ Vulnerability connotes lack of agency and a need for paternalistic protection, and references to vulnerable populations in bioethics often address individual inability to provide fully informed consent for legal or other reasons, which is not the issue here. Suggestions for improvement are invited.

The argument made here was, in fact, noted but not elaborated upon in a recent compilation of case studies on secondary use of personal information published by the Canadian Institutes of Health Research (CIHR): “While the right to privacy is essentially recognized in law as an individual human right, collecting, using and linking personal information about many individuals, then analyzing that information and drawing generalizable conclusions from it may sometimes result in potential discriminatory harm to communities of individuals in certain circumstances.” So far, the problématique has been stated in abstract terms. Its importance can be illustrated with reference to three scenarios. These are distinctive in that none of them involves genetic information. Debate continues about whether the collection, storage and use of genetic information in population health research raise truly distinctive ethical issues (“genetic exceptionalism”). Thus, one can remain agnostic on this point while nevertheless acknowledging the existence and importance of a range of ethical issues that go beyond privacy.

**What’s Wrong with These Pictures?**

**Scenario 1** – A record linkage study is designed to document health status differences among workers in the health care sector, investigate predictors of long-term disability claims for musculoskeletal and mental health disorders, and identify characteristics of the health care workplace that are associated with increased risk of these disorders. The study links a database maintained by the benefits provider for health care workers with data on health service use obtained by the provincial Ministry of Health, and with workers’ compensation claims. Published results subsequently show that workers older than 45 are more likely than their younger, but otherwise comparable counterparts, to file long-term disability claims.

**Scenario 2** – Using either a pre-post or RCT design (the distinction is not important for purposes of this article), chart
audits are used in a program evaluation to assess the effect of introducing mobile mental health crisis teams in one or more mid-size cities on use of hospital emergency rooms, which are routine ‘last resort’ ports of call for people with certain psychiatric diagnoses. Because only anonymized data are used in the study, ethics review is expedited and individual consent to the research use of chart data is not required. Reductions in the outcome of interest are identified, but these are not statistically significant at conventional 95 percent confidence levels. Citing this negative result and its commitment to evidence-based service planning, the health authority responsible for funding priorities decides not to fund mobile crisis teams in the future, and to terminate the funding for existing teams.

Scenario 3 – Data from a national survey of health and household characteristics are used to assess the relations among children’s activities, obesity, and family characteristics. Published results identify a negative correlation between household income and obesity, and also identify growing up in a single-parent household as a risk factor for obesity in childhood and adolescence. The lead investigator explains the connection to the media: “The more convenient television or video game thing serves as a very good surrogate to a partner that keeps the kids safe and quiet and so forth.”

Each scenario raises important ethical issues that have nothing to do with privacy, and points to the importance of integrating concern for differentials in power and privilege into both the design and anticipated use of research studies. In the first scenario, employers might well use the results of the research as a basis for favouring younger employees in the recruitment process, for reasons that would be justifiable on the basis of ‘bottom-line’ considerations: limiting workers’ compensation premiums and costs for privately provided long-term disability insurance. This choice might also be impermissible under anti-discrimination statutes, but meeting a legal standard of proof with respect to discrimination, especially age discrimination, is always expensive and time-consuming. Since mental health disorders constitute one of the studied categories of disability claims, various considerations related to the stigmatization of people living with mental illness are also an issue.

The central argument of this article is that ethical analysis and institutional approval of population health research must be explicitly concerned with avoiding harms to subordinated or marginalized populations.

The second scenario demands considerations of how ‘negative’ results are defined and interpreted. Emergency room usage, although relevant to institutional and financial concerns, may not be the most meaningful or patient-centred indicator of therapeutic effectiveness. Perhaps the intervention’s apparent lack of effectiveness was due to exogenous factors beyond the control of researchers, such as the introduction of more demanding eligibility criteria for disability benefits or the repeal of rent controls, leading to a drastic increase in housing costs at the bottom end of the rental market. The value dimension involved in the choice of a threshold of statistical significance, and the potential for competing interpretations of ‘conservatism’ in drawing conclusions about cause and effect, has been extensively discussed in the context of environmental policy; it is also relevant to research on the effectiveness of health and social service interventions.

Unlike the first two, the third scenario involves data collected specifically for research purposes. It is nevertheless troubling, especially because it contains no hypothetical elements. Neither in the article reporting on the research nor in the interview given to Canada’s national public broadcaster did the researchers specifically acknowledge the time pressures on single parents, or how those pressures can be compounded by poverty or lack of adequate child care. In 1994, the year in which the data were gathered, 53 percent of all single parent households lived on incomes below Canada’s national Low-Income Cutoff; for households headed by single mothers, the overwhelming majority of such households, the figure was 56.4 percent. Income-related contextual factors may also be important in explaining the study findings: downtown neighbourhoods with narrow sidewalks and high levels of vehicle traffic, or high-rise buildings adjacent to multi-lane suburban arterial routes, do not offer the safe and convenient opportunities for outdoor physical activity available in manicured suburban culs-de-sac. Risk of criminal victimization may also be an issue in some neighbourhoods.

It is unrealistic to state – as do the authors of the most comprehensive study of consent requirements for population-based genetic research – that “the burden of consider-
ing group implications falls primarily on the participants themselves,\textsuperscript{25} and this position is of course inapplicable to studies involving secondary uses. Discriminatory or otherwise ethically impermissible use of published or unpublished results may be preventable, if at all, only at prohibitive cost to those adversely affected. The example of mental health services research provides an especially dramatic illustration of how discrepant life situations of researchers and the people they study\textsuperscript{26} may create a gap that confounds even the best intentioned and most reflective researchers. As Donna Haraway writes, “there is good reason to believe vision is better from below the brilliant space platforms of the powerful.”\textsuperscript{27}

\section*{Discussion – Policy Responses}

Considerations of equity or distributive justice suggest that ethics review should not be restricted to issues of privacy and confidentiality, although those remain crucially important. Rather, it should provide a deliberative context for considering a broader range of harms to all individuals comprising a subordinated or marginalized population, whether or not they are participants or data subjects\textsuperscript{28} in a particular study. It also needs to take into account potential benefits to the populations of greatest concern. For example, a study of the effects of user charges for prescription drugs in Quebec’s provincial drug insurance plan made use of the province’s administrative databases on health service utilization and prescriptions filled. It identified substantial negative impacts, including reduced use of essential drugs and an increase in the number of emergency department visits, among the poor and elderly beneficiaries of the plan.\textsuperscript{29} Both logistical considerations and response bias would have limited the value of a survey-based approach: “people who are willing to fill out a questionnaire and send it back are likely to be in better health than those who do not. If the researchers had used only the results of surveys to evaluate the impact of the new drug cost-sharing policy, they would have falsely concluded that the drug plan had a very minor impact on these groups of individuals.”\textsuperscript{30} Thus, it cannot be emphasized too strongly that the argument made here is not ‘anti-research.’ Rather, it is for a more inclusive approach to assessing both potential harms and potential benefits – an approach that, in the case of research on determinants of health at the population level, may have the additional advantage of generating richer, more policy-relevant research designs.

Guidelines for ethics review could be amended to specify create the necessary deliberative context. As noted, the TCPS now appears to foreclose such deliberation in many cases, so long as privacy and confidentiality are protected. Devising operational definitions of a subordinated or marginalized population is not easy, but any worthwhile definition would include several categories identified in the scenarios (workers injured or made ill by their employment, people with serious mental illness, low-income single parents) as well as, e.g., people living with HIV/AIDS. Alternatively, researchers might be asked to demonstrate the absence of economic or power differentials between themselves and research participants or data subjects sufficient to trigger a review that went beyond considerations of privacy and confidentiality.

Not enough research has been conducted on the operation of the research ethics boards (REBs) that are responsible for implementing the TCPS and its predecessor guidelines. The lead author of the most recent national study of this topic\textsuperscript{31} laments the lack of research on research ethics, and is also highly critical of many aspects of the ethics review process: “Canadian regulators are distracted and generally inattentive to the ethical challenges of health research”\textsuperscript{32} and “all the major actors (including research sponsors, institutions, and regulators) behave as if REB approval is all that there is the ethical conduct of research involving human subjects.”\textsuperscript{33} Until more is known about REBs and how they function, and until REBs are provided with adequate resources at least to carry out clearly specified existing duties,\textsuperscript{34} judgments of feasibility must be set aside even if its concluded that more expansive ethics review is desirable.

A proactive response to the issues raised here, which is more in keeping with the long time frame anticipated for research programs like CLHI, would actively seek out knowledge and experience from outside the research community, by empowering research participants, data subjects and similarly situated individuals. The Social Sciences and Human-
Representativeness will always be an issue, but the boards and advisory councils of agencies such as administrative databases and research platforms, and for management committees of institutions mandated to manage administrative databases and research platforms, and for the boards and advisory councils of agencies such as CIHR. Representativeness will always be an issue, but imperfect or even inconveniently strident representation is preferable to the silencing of important voices. Agencies like CIHR might award additional points for, or even require: involvement of members of affected populations in defining outcome variables; participatory action models that incorporate multiple perspectives into the design of research projects before protocols are finalized; and, for larger scale research projects, governance structures that specify and formalize the roles of members of vulnerable populations through advisory and steering committees. One or more university-based research centres could compile and maintain a national database of organizational templates for equity-oriented approaches to research design and governance.

This article has suggested, in necessarily generic terms, the need for creativity and innovation in responding to the ethical challenges presented by power imbalances and divergences of perspective between health researchers and members of subordinated or marginalized populations. Some researchers are likely to be sceptical of the need for adding to ethics review requirements that may already be considered burdensome. Others will recognize the proposals made here as part of the evolution of bioethics away from a primary focus on the clinic and the laboratory and toward a broader concern with the social context of health, illness and health research. Prudential considerations are also relevant, especially for funding agencies. Large scale population health research that links a variety of data sources can generate evidence in support of a range of preventive policies and interventions, but it also carries with it the possibility of large scale (and very expensive) failure. This could happen if public mistrust results either in active policy opposition or bias in recruitment of participants – for example, under-representation of precisely those populations whose health is placed most at risk by their social and economic situation.

As noted, the scenarios outlined in this article do not involve genetic information, although its collection is integral to the design of research platforms like Biobank and CLHI. Heightened public sensitivities about possible (mis)uses of health information, and the likelihood that those sensitivities will increase during the long time frame contemplated for such studies, mean that scientific researchers and funding agencies committed to such projects are well advised to ‘go the extra mile’ (or two, or three) to anticipate and address both matters of public perception and genuine ethical problems such as those identified here.

Ted Shrecker is a Research Associate at the Saskatchewan Population Health and Evaluation Research Unit, University of Saskatchewan, Saskatoon, Saskatchewan. The author acknowledges Bartha Maria Knoppers who encouraged the articulation of these ideas in article form. All views expressed are exclusively those of the author.

Notes


5. S.C. 2000, c. 5.

7. In this article, the term ‘research participant’ is used to describe individuals from whom data are gathered primarily or exclusively for research purposes, and ‘data subject’ describes individuals about whom data are gathered for non-research purposes in the course of providing or financing care but are subsequently used for research.


17. This is a condensed version of the description of an actual British Columbia study in *Case Studies, supra* note 15 at 120-123.

18. The author was employed on a similar study, which used a pre-post design to evaluate the introduction of a mental health crisis worker in a hospital emergency room in a mid-sized Ontario city, before relocating to another province; the study is still in progress.


25. Beskow et al., supra note 10 at 2319.


28. See supra note 7.


32. McDonald, ibid. at 7.

33. Ibid. at 9.

34. See McDonald et al., supra note 31 at 216-219.


37. The National Breast Cancer Coalition has been influential in setting research priorities in the United States, partly because it has established a science training program to enable advocates to contribute the priority-setting process more effectively: J.H. Platner et al., “The Partnership between Breast Cancer Advocates and Scientists” (2002) 39 Environmental and Molecular Mutagenesis 102.
