

Trends in Collection, Use and Disclosure of Personal Information in Contemporary Health Research: Challenges for Research Governance

Donald J. Willison

Background: Changes in the Nature and Use of Personal Information for Health Research

Health research encompasses a heterogeneous set of research activities. This paper focuses on challenges that arise in the governance of observational research which is usually carried out without any direct contact between the researcher and the individuals being studied. Two broad areas of health research are heavily dependent on access to a wide range of existing *person-level* health information:

1. Public health, occupational health and safety, and the non-medical determinants of health and disease. The latter examines the relationship between health and lifestyle, environmental, and socioeconomic factors including income and education. Epidemiology is the foundation of much of this type of research. Research in this domain links both health and non-health information, such as occupation, education, and lifestyle information.
2. Health policy, health services research, and program evaluation examine the health care system and the effects of different policies and methods of health care delivery on the quality and efficiency of care pro-

vided. This type of research is informed by a wide variety of disciplines, including: economics, health policy, political sciences, sociology, anthropology, medicine, and epidemiology.

Most health research requires person-level data, chiefly to increase precision in analysis. For example, when trying to determine the effect of exposure to an environmental toxin in a neighbourhood, with person-level data one can better examine the causal relationship by “controlling for” or holding constant known personal factors such as age and sex of the individual that relate to the outcome of interest. Similarly, when evaluating a policy to increase co-payments for prescription drugs, it is prudent to examine across different income brackets the impact of that policy on the tendency to discontinue medications. In some cases, if using aggregate rather than individual-level data, it is possible to come to spurious conclusions about the effect of exposure (whether to a policy or an environmental toxin) on health outcomes.¹ Also, individual-level data are required to link information from disparate databases. This linkage creates the ability for researchers to answer a much broader set of questions about the determinants of health, but it also raises major privacy concerns when these activities are being conducted without individual consent. Although data may be stripped of direct personal identifiers, the resultant records are often so rich in



information that the residual risk of disclosure of identity through indirect means is sufficiently high that the data must be treated as if they were identifiable. In fact, with as little information as date-of-birth, sex, and full postal code, the majority of individuals in a particular region may be re-identified by linking with census tract information.²

Trends in Data Collection, Use, Storage, and Disclosure

Twenty years ago, only a handful of research centres across North America had the capacity to manipulate and link large data sets, and most government and other data repositories were used only for claims adjudication. Medical records were all paper-based. Advances in the capacity of computers and the internet to store, manipulate and disseminate large amounts of data have changed dramatically the nature of collection, use and disclosure of personal information in contemporary health research. These advances have spawned two parallel developments: the planning and development of large disseminated health information networks that will serve multiple purposes beyond those for direct clinical care; and the proliferation of decentralized holdings of personal data.

In Canada, the United States, much of the European Union, Australia, and New Zealand, major efforts are underway to computerize patient records across health care settings, with the ability to share and link information from the records of physicians, diagnostic facilities, and health care institutions. In addition to their primary use for direct patient care and claims adjudication, it is intended that these records will be used for quality and risk management, disease surveillance, research, and education of students in the health care professions. In Canada, health infohighway plans include common information architecture across all provinces and territories, to facilitate information sharing across jurisdictions.³ While the mechanisms by which health researchers will gain access to these infostructures have not yet been determined, considerable attention has been given to developing a consolidated pan-Canadian framework for managing privacy and confidentiality of health information.⁴

The same technologic advances that have spurred ambitious plans for massive health infohighway projects have also advanced personal computers to the point where they now have greater computing power than the mainframe computers of two decades ago. As a result, most universities no longer maintain centralized facilities to manage the computing needs of researchers. This shifts responsibility for the man-

agement of large amounts of personal data onto individual researchers and their staff, who are often ill-trained in issues of privacy, confidentiality and data security matters.

Over the past decade, a third development in data use for research has emerged, involving the prospective collection of large amounts of data to serve as broad platforms for health research, including future yet-to-be-defined research questions. These data repositories – registries and biobanks – have been developed either as by-products of existing health information collection (e.g. separate data holdings on demographics and concurrent medical conditions of patients visiting speciality clinics) or separate collections intended specifically for research (e.g. the development of multi-site disease-specific or treatment-specific registries and biobanks). While some have developed specifically in recognition of the limitations of existing clinical and administrative data records for research, many collections have evolved over time from modest beginnings (e.g. lists of names of patients with specific conditions or leftover laboratory samples).

The introduction of new data protection laws in Canada and elsewhere is causing the research community to take a closer look at its data collection and management practices. While, in general, the laws provide considerable scope for self-governance on the part of the research community, there are several challenges associated with each type of research activity. Some of these are explored below.

Emerging Governance Challenges and Possible Ways Forward

For this paper, governance will be considered chiefly in the context of ensuring that the collection, use, and management of data for research purposes are consistent with the “fair information practices” that are at the foundation of legislation in Canada and most Western industrialized countries.⁵

Each of these three developments – the design of large health infostructures, the proliferation and decentralization of a large number of smaller personal data holdings, and the development of large multi-centred registries and biobanks for future research purposes – poses particular challenges around research governance. In the first case, the chief challenge comes in regulating access to these data in the context of fuzzy boundaries in the purposes for which the data are used and the multiple roles that health researchers may hold. In the second case, the chief challenge is one of ensuring



responsible data management in a decentralized, largely unsupervised context. In the second and third scenarios, a pivotal question is “Where, ultimately, does responsibility lie for stewardship of the data?” Finally, across all scenarios is the question of the circumstances under which this information may be used for research purposes without consent. What systems need to be in place to garner the trust of the public in the responsible use of their personal information for health research?

Challenge 1 – Fuzzy Boundaries

A challenge that currently exists with research conducted using information collected for clinical care is one of “fuzzy boundaries” in both the purposes for which the data are used and in the roles of the individual accessing the data. For example, it is increasingly difficult to distinguish between research and quality improvement.⁶ Yet, historically, these two uses of data have been treated very differently. While consent has been implied for quality improvement and risk management purposes, such is not the case for research and other secondary uses. Further, studies using personal data for quality improvement without consent have largely been carried out without any formal ethics review whereas equivalent studies under the label “research” are subject to review by a research ethics board (REB). There have been numerous attempts to distinguish between the two activities, for example:

- By original purpose – to improve or assess service delivery (quality improvement) vs. contributing to a growing body of generalizable knowledge (research).⁷
- By intended uptake, where quality improvement is characterized by rapid turn-around of data for implementation of change.⁸
- By level of risk to individuals, be they patients or staff, where quality improvement involves minimal risk.⁹
- By size of reform or innovation being introduced.¹⁰

However, in practice these distinctions are largely artificial. A particular evaluation may have multiple purposes, the purpose may change, or generalizable knowledge may come

out of work that was initially intended only for assessment of local service delivery. Chart review and interviews are common to both quality improvement and health services research. Each has inherent risks to individuals such as breach of confidentiality through record review and distress induced in interview participants.¹¹

Once electronic health records become more popular, it will be much easier to implement interventions designed to change physicians’ behaviour – e.g. to influence test-ordering or prescribing behaviour. Under these conditions, the blur between clinical care, quality improvement and research activities will increase further. In the extreme, all patients become research subjects at one point in time or another, simply through their participation in the health care system. While some researchers may regard this as utopian, advocates of human subject protections in research see this as a critical erosion of existing protections.¹² Indeed, O’Neill cites the increased complexity

in the care-giving environment as one reason to consider alternatives to the current consent process.¹³

Failure to obtain ethics review for quality improvement activities has sometimes led to problems when people have attempted to publish their quality improvement work only to discover that the editor requires evidence of prior REB approval before the manuscript may be published or, having successfully published the material, they have been called to task for failure to submit their material to a REB. In the United States, concern has been expressed over the chilling effect that the requirement of formal REB review has had on quality improvement activities.¹⁴

Given that the risks to patients are associated with particular activities, regardless of whether the label is quality improvement or research, it is the institution’s responsibility to have some sort of ethics review of both activities, proportionate to the risk to patients, to determine necessary patient protections and whether or not individual consent for participation is required. Whether review of quality improvement activities occurs through REBs or through a parallel ethics review, the process will need to be sufficiently resourced as

“Advances in the capacity of computers and the internet to store, manipulate and disseminate large amounts of data have changed dramatically the nature of collection, use and disclosure of personal information in contemporary health research.”



to allow timely review of projects, so as not to chill quality improvement activities.

Challenge 2 – Ensuring Adequate Data Management in a Decentralized Research Environment

Universities and hospitals are now only beginning to recognize the need to come to grips with the challenge of ensuring that the hundreds of research databases within their organizations are adequately safeguarded from unauthorized disclosure and from corruption. This includes physical, technical, and procedural safeguards – not just against theft of equipment but of the data itself. This author is unaware of any formal audits in the public domain on security measures in university settings. However, apart from research institutes which are repositories for massive amounts of data, attention to basic practices re: password-protected files, locked filing cabinets, and securing computers with personal information from research studies are generally quite relaxed.

In general business, most disclosure breaches occur when authorized data users either inadvertently or deliberately disclose information to outsiders.¹⁵ The extent to which this occurs in the research setting is unclear. However, in academic research, despite increased emphasis on intellectual property, it is still common for data or biological samples to be freely shared among researchers conducting similar work in other centres or even with industry. Data protection legislation across Canada generally prohibits disclosure of personal health information without individual consent. For example, under the new Ontario legislation, any researcher who has received personal health information from a health information custodian is prohibited from further disclosure of the information to others without the data subject's approval, except as required or permitted by law, with limited exceptions such as for quality improvement purposes.¹⁶ Similar provisions exist in the Tri-Council Policy Statement¹⁷ and the CIHR privacy best practice guidelines.¹⁸

An equally important concern is what happens to the data once the project has concluded? Data protection laws generally specify that personal data be destroyed or rendered irreversibly anonymized once the purpose for collecting the data has been accomplished. Such practice is an anathema to researchers, who prefer to keep data for future analyses. While this may still be required under conditions of a user agreement with the original data custodian, where the data has been collected *de novo* by the researcher, there is great hesitancy to destroy the data.

All this suggests a much more active role be taken on the part of the institutions within which researchers conduct their research. This could include: developing standard operating procedures (SOPs) for safeguarding data, and providing educational and technical support to facilitate uptake procedures. It could even include periodic audits of compliance with SOPs. As for completed studies, one alternative would be for the institution to take responsibility for the custody of the data at the end of the study, where release to the original investigators (or to others) occurs only after REB approval.

Challenge 3 – Locus of Responsibility for Data Stewardship

Suggesting that institutions take a more active role in ensuring responsible data management raises a contentious question: What are the limits of accountability of the institution, as opposed to the principal investigator, not only for safeguarding data but, ultimately, for ensuring responsible use of data – particularly when there is some breach of fair information practices? A case in point is that of the anthropologist, Ryk Ward, who, in the 1980s while on faculty at the University of British Columbia, collected blood samples of the Nuu-chah-nulth people of British Columbia to study genetic markers for an unusual form of arthritis. He then went on to use the samples for research other than that for which he obtained consent.¹⁹ During the period of dispute over use of the blood samples, Ward moved twice to positions in other universities in the United States and the United Kingdom, taking with him the disputed samples in each case.²⁰

As collections become larger in size and research questions become more open-ended — e.g. with the registries and biobanks that may collect and analyse data for decades — the question of who may collect and manage the holdings becomes even more pressing. Should an individual investigator be ultimately responsible or the institution within which she works – be it a university, hospital or foundation? If the founding principal investigator moves on to another institution, should the registry or biobank remain with the institution or move with the researcher? Further, when registries are established as platforms for future unknown research, there should be an independent oversight body that reviews the scope of work being done over time by the researchers and to publicly declare their judgments as to how well they are exercising their stewardship of the data.²¹ This is particularly important with very large holdings, as they are highly public and their long-term viability is largely



dependent upon ongoing trust of those contributing data to the data/biobank. Perhaps the most developed model is that of UK Biobank, which uses language of “stewardship” rather than “custodianship” or “ownership”, and for which an independent Ethics and Governance Council has been established.²²

Challenge 4 – Consent and the Collection, Use, and Disclosures of Health Data

The role of consent is particularly contentious in both secondary use of information collected originally for clinical care purposes and in the creation of large research registries. In both cases, it is frequently argued by researchers that obtaining consent is impracticable and that ensuing selection biases would invalidate the usefulness of the data collected.²³

Current laws frame consent for research purposes in the context of finite intervention trials. This is a poor fit with the types of observational and longitudinal research described in this paper, particularly when conducting longitudinal studies in which potential future research applications of the data may not yet be known at the time of collection.²⁴ O’Neill has argued that the complexity of the modern care environment, the vulnerability of acutely ill individuals at the time of approach, and challenges of excess information to be conveyed render true informed consent a heroic or impossible cognitive feat, even among individuals with mature faculties.²⁵

Recognizing the challenges associated with obtaining individual consent, some researchers have argued for the waiver of consent for whole classes of research²⁶ while others have argued the need to recognize the value of some form of broad consent²⁷ The CIHR has created guidelines that help REBs in interpreting when it may be impracticable – or even inappropriate – to obtain consent.²⁸ The burden of proof is considerably higher in prospective collections, when there is contact with the individual whose information is to be used. The decision whether or not to waive consent, and the breadth of consent required will likely require case-by-case adjudication by REBs.

Challenge 5: Trust and Accountability in the Use of Personal Information for Health Research

Regardless of the place of individual consent in these research activities, for continued access to personal information for research, the research community must ensure that systems are in place that will engender public trust that

their personal information is being used responsibly. In particular:

- Support is needed to train researchers and research ethics board members in how to bring data collection and management practices for research in line with fair information use practices. The new CIHR privacy practice guidelines²⁹ provide an important first-step in providing a harmonized response to new legislation. However, there is still a need for ongoing oversight and support on the part of individual institutions to ensure that intended practice translates into actual practice. Initially, funding will need to come through infrastructure support on the part of the institution, or perhaps the granting councils directly. However, just as grant budgets include support for programming staff, it would be reasonable if grants were to cover costs associated with managing data security.
- As discussed above, the differing oversight requirements of quality of care research vs. in-house quality improvement activities need to be reconciled. To preserve public trust, ethical oversight should be applied to research *and* quality improvement, proportionate to the risks involved to human participants.
- The research community needs a shift in culture with regard to data. In particular, researchers need to shift their thinking:
 - from ownership to stewardship, where data are held in trust by the researcher;
 - from data as mere objects for manipulation to something that has meaning for individuals, the misuse of which could lead to harm to those individuals and loss of public trust.
- Trust requires some form of external accountability. O’Neill calls for “intelligent accountability” that does not damage professional performance and that provides institutions with ample self-governance to perform their task.³⁰

Finally, in planning for future systems for collection, use, and disclosure of personal health information, most of the consultation has been with those who will make use of the information, either for their primary purpose of providing patient care or for the growing list of secondary uses, including research. In Canada, there has been relatively little input from the public. Yet, with such fundamental changes in the



use of personal information, it is imperative that those who would govern the use and disclosure of such information engage in meaningful dialogue with the public to get direction on the broad parameters under which these uses of personal information will take place – particularly, the role of consent for uses of this information and the governance structures that the public will find trustworthy. For what types of research is express (opt-in) consent essential? Under what circumstances would a broad consent be acceptable? When may consent be waived entirely, with or without an opportunity to opt-out? Do adequate checks and balances in the governance of uses of personal information for health research mitigate the need for express consent of individuals? How can citizens' interests best be represented in an ongoing fashion? These questions all have profound implications for the governance of research requiring the use of personal information. There are many technical challenges in gauging the values of the public. Indeed, this may raise more questions than it resolves. However, without meaningful public engagement on these issues, public trust could be lost, resulting in greater restrictions on use of that information.³¹

Donald J. Willison, Assistant Professor, Department of Clinical Epidemiology & Biostatistics, McMaster University Centre for Evaluation of Medicines, St. Joseph's Healthcare, Hamilton, Ontario.

1. Hal Morgenstern. "Uses of Ecologic Analysis in Epidemiologic Research" (1982) 72 *American Journal of Public Health* 1336.
2. Latanya Sweeney, "Weaving Technology and Policy Together to Maintain Confidentiality" (1997) 25 *Journal of Law Medicine & Ethics* 98; Latanya Sweeney "k-Anonymity: A Model for Protecting Privacy" (2002) 10 *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* 557.
3. Canada Health Infoway, *EHRs blueprint: an interoperable EHR framework* (July 2003), online: <<http://www.canadahealthinfoway.ca/>>.
4. Preliminary draft of the pan-Canadian health information privacy and confidentiality framework for stakeholder consultation. (Ottawa, 2004).
5. CSA., CAN/CSA-Q830-96, *Model Code for the Protection of Personal Information: A National Standard*

of Canada. (Ontario: Canadian Standards Association, 1996). To paraphrase, an organization is compliant with the fair information practices if it:

- Is accountable for personal information in its custody.
- Identifies, in advance, the purposes for which information is collected.
- Obtains consent – to collect, use or disclose the information.
- Limits collection of personal information to that necessary to accomplish the purpose
- Limits use, disclosure & retention to the purposes for which the information was collected.
- Ensures information is accurate, complete and current for its intended purposes
- Employs adequate safeguards – from unauthorized data use, disclosure or from corruption
- Is open about its information use, policies and practices
- Allows individuals individual access to information about them, and to challenge its accuracy and completeness, and amend as required; and
- Allows individuals to challenge compliance with these practices.

6. National Health & Medical Research Council, "When Does Quality Assurance in Health Care Require Independent Ethical Review?", online: (Canberra: Commonwealth of Australia, 2003) <<http://www.nhmrc.gov.au/publications/pdf/e46.pdf>> [NHMRC]; J. Lynn, "When Does Quality Improvement Count as Research? Human Subject Protection and Theories of Knowledge" (2004) 13 *Quality and Safety in Health Care* 67; L. Doyal. "Preserving moral quality in research, audit, and quality improvement" (2004) 13 *Quality and Safety in Health Care* 11; Wilfred E. Thurston, Ardene Robinson Vollman & Michael M. Burgess. "Ethical review of Health Promotion Program Evaluation Proposals" (2003) 4 *Health Promotion Practice* 45. A similar fuzzy boundary problem occurs in the area of public health. A case in point is the outbreak of severe acute respiratory syndrome in the Toronto area in 2003, wherein access to personal data for infection control purposes was permitted without ethics review but any research on the outbreak required REB review.
7. Alberta Research Ethics Community Consensus Initiative, *Draft Recommendations for Ethics Screening and Review of Research Program Evaluation, and Quality Assurance or Quality Improvement* (2004),



8. Lynn, *supra* note 6.
9. Doyal, *supra* note 6; *Ibid*.
10. Lynn, *supra* note 6.
11. NHMRC; Lynn; Thurston; *supra* note 6.
12. Beverley Woodward, "Challenges to Human Subject Protections in US Medical Research" (1999) 282 *Journal of the American Medical Association* 1947.
13. Baroness Onora O'Neill, "Informed Consent and Genetic Information" (2001) 32 *Studies in History & Philosophy of Biological & Biomedical Science* 689.
14. Thurston, *supra* note 6.
15. Michael Pastore, "Internal Threats Justify Increase in Security Spending" 19 June (2001), online: <http://www.clickz.com/stats/big_picture/applications/print.php/787251>.
16. Bill 31, *Health Information Protection Act, 2001*, 1st Sess, 38th Leg., Ontario 2004, s. 4 (6) (d) (given Royal Assent Thursday, May 20, 2004).
17. Tri-Council, *Tri-Council Policy Statement on the Ethical Conduct of Research Involving Humans (with updates of May 2000 and September 2002)* (Ottawa: Tri-Council, 1998), section 36, article 3.2.
18. Canadian Institutes of Health Research Privacy Advisory Committee. *Guidelines for protecting privacy and confidentiality in the design, conduct and evaluation of health research: best practices*. (Consultation draft, April 2004), online: <<http://www.cihr-irsc.gc.ca/e/about/22085.shtml>>.
19. David Wiwchar, "Bad Blood" *Ha-Shilth-Sa - a First Nations Newspaper* (S, 7-9 October 2000); K. Kleiner, "Blood Feud" (2000) *New Scientist* 7; and Rex Dalton, "Tribe Blasts Exploitation of Blood Samples" (2002) 420 *Nature* 111.
20. Under most data protection legislation, this new use of data (in this case, blood samples) would be unacceptable without individual consent for the new uses, unless the research ethics board affiliated with the researcher's institution deemed otherwise. Criteria for waiver of consent include:
 - The research cannot be achieved without personal information
 - It is impracticable to obtain consent
 - Adequate safeguards are in place to protect the information, and
 - The REB has weighed the public interest in research vs. the public interest in protecting individuals' privacy.
 Until recently, there has been little guidance for REBs for the determination of impracticability of obtaining consent. The CIHR has recently developed practice guidelines to assist in this regard.
21. REBs may approve individual research projects, but have generally not taken on this kind of oversight function of data holdings.
22. UK Biobank Ethics and Governance Framework: Summary of Comments on Version 1.0, (2004), online: <<http://www.ukbiobank.ac.uk/ethics.htm>>.
23. Jack V. Tu, *et al.*, "Impracticability of Informed Consent in the Registry of the Canadian Stroke Network" (2004) 350 *New England Journal of Medicine* 14.14; S.J. Jacobsen, *et al.*, "Potential Effect of Authorization Bias on Medical Record Research" (1999) 74 *Mayo Clinic Proceedings* 330.
24. Timothy Caulfield, Ross E.G. Upshur & Abdallah Daar, "DNA Databanks and Consent: A Suggested Policy Option Involving an Authorization model" (2003) 4 *BMC Medical Ethics* 1.
25. *Supra* note 13.
26. *Supra* note 23; L. Joseph Melton, "The Threat to Medical Records Research" (1997) 337 *New England Journal of Medicine* 1466.
27. *Supra* notes 24; Ross E. Upshur, B. Morin & V. Goel, "The Privacy Paradox: Laying Orwell's Ghost to Rest" (2001) 165 *Canadian Medical Association Journal* 307.
28. NHMRC, *supra* note 6.
29. *Ibid*.
30. Baroness Onora O'Neill, "Intelligent Trust and Intelligent Accountability" (Presented by Yale Divinity School's Center for Faith and Culture, September 17, 2004), online: <<http://www.yeale.edu/divinity/video/o'neill.htm>>. O'Neill contrasts "intelligent accountability" with "managerial accountability" that focuses on controlling performance through setting targets, measuring success through achievement of targets, and setting sanctions for failure. These targets are proxy indicators for more complex activities that lack readily-measurable targets. While purported to be cheap, objective, and transparent, they may be none of the above in complex circumstances and run the risk of institutions responding to improve the indicator rather than the underlying activity of interest.
31. Donald J. Willison, "Privacy and Secondary Use of Data for Health Research: Experience in Canada and Suggested Directions Forward" (2003) 8 *Journal of Health Services Research and Policy* 17.

