Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

DRAFT

Case Studies Involving Secondary Use of Personal Information in Health Research

DECEMBER 2001

Canada

Table of Contents

* N/A = Not available in this draft package, but will eventually be included in the final document.

Members of CIHR's Working Group on Case Studies

Dr. Sharon Buehler (Chair) is an Associate Professor of Epidemiology in the Division of Community Health, Faculty of Medicine at Memorial University of Newfoundland. She holds B.A. and M.A. degrees in biology and zoology. Her work on lymphoma in small communities on the northern peninsula of Newfoundland brought her to doctoral work in epidemiology at Memorial. Currently, with collaborators in oncology, she is assessing variation in the management of breast cancer and, with colleagues in family practice and business, looking at practice-based strategies to maintain continuity of primary care; she is part of the CIHR-funded Interdisciplinary Health Research Team on colorectal cancer based in Toronto and Newfoundland. She was director and is currently on the advisory committee of the Provincial Cancer Registry. She has served on the boards of the Canadian Society for Epidemiology and Biostatistics and the Canadian Cancer Society and was a member of the original Management Committee of the Canadian Breast Cancer Research Initiative. She currently sits on the CIHR Institute Advisory Board for Cancer. Her work in these groups and on her local ethics review board have fostered a keen interest in how research, and the involvement and protection of research participants, can be effectively communicated to the public and to policy makers.

Dr. Charlyn Black, Co-Director of the Manitoba Centre for Health Policy and Evaluation (MCHPE), is a CIHR Scientist and a nationally recognized health policy researcher. She plays a key role in working at the interface between research and policy and in ensuring that the Manitoba Centre's work is relevant to the policy process. Her research interests focus on applications of population-based information systems, uses of administrative data to assess and monitor quality, effectiveness and outcomes of medical care, and the development of data-driven information tools to inform and improve health care delivery. She serves on a number of influential committees, including the Federal/Provincial/Territorial Advisory Committee on Health Services (reporting to the Conference of Deputy Ministers of Health), the Canadian Population Health Initiative Council and the Steering Committee of the Western Canadian Waiting List Study. Since September 2000, she has taken a role as Senior Advisor to the President of CIHR and to the President and CEO of the Canadian Institute for Health Information (CIHI). Dr. Black received her medical degree from the University of Manitoba (1979) and her doctorate in health services research from the Johns Hopkins University (1990). As Associate Professor and Associate Head of the Department of Community Health Sciences, she is active in teaching at undergraduate, residency and graduate levels in the Faculty of Medicine at the University of Manitoba. She has practised medicine in a variety of primary care settings, including a core area clinic, a suburban private practice, and the far north.

Dr. Denis Cournoyer is an hematologist at the McGill University Health Center and an Associate Professor of Medicine, Oncology and Human Genetics at McGill University. He received his MD from the Université de Sherbrooke and trained in hematology at McGill. He then received training in molecular biology and molecular genetics in the Department of Biochemistry at McGill and in the Department of Molecular Genetics at Baylor College of Medicine in Houston. His research work relates to somatic gene therapy. He is Chair of the Research Ethics Office of the Montreal General Hospital and Director of the Research Ethics Office of the McGill University Health Centre. He is also a member of the Board of Directors of Héma-Québec, and of the Advisory Committee in Ethics of the Fonds de la Recherche en Santé du Québec.

Dr. Eric Holowaty is a cancer epidemiologist at Cancer Care Ontario, where he is also Director of the Cancer Surveillance Unit, which includes the Ontario Cancer Registry, the largest patient-specific population-based cancer registry in Canada. He is also an Associate Professor in the Department of Public Health Sciences at the University of Toronto, where his responsibilities include teaching as well as thesis supervision. His research interests include historical record linkage control studies, second primary cancers, health services research, cancer registration and quality control. He holds a number of research grants and contracts with Health Canada, the National Cancer Institute of Canada and the U.S. National Cancer Institute.

Dr. George Kephart is an Associate Professor and Clinical Research Scholar in the Department of Community Health and Epidemiology, Faculty of Medicine at Dalhousie University. He is also co-founder and Director of the Population Health Research Unit (PHRU), which maintains administrative health databases, and a variety of other databases, for health services and population health research. His research interests include health policy evaluation, access to health care services, and methodological issues in the use of administrative and longitudinal health data (including methods to protect privacy and confidentiality). Dr. Kephart's current research includes a study on the effect of user fees on prescription drug use in the Nova Scotia Senior's Pharmacare Program, a study on non-financial barriers to accessing health care services, and an evaluation of the validity of administrative health data for the study of diabetes. His research has been funded by the National Institutes of Health in the United States, the Medical Research Council of Canada (now CIHR), the Canadian Health Services Research Foundation, and the Canadian Population Health Initiative. Dr. Kephart holds an MS and a PhD from the University of Wisconsin-Madison in Sociology (Demography).

Dr. Malcolm Maclure is Senior Healthcare Epidemiologist in the Strategic Planning, Reports and Nursing Directorate of the British Columbia Ministry of Health. From 1991-2000, he was Manager of Statistical Analysis and Evaluation within BC Pharmacare. He is also Adjunct Associate Professor of Epidemiology at Harvard School of Public Health and Visiting Professor at the Department of Clinical Pharmacology and the Research Unit on General Practice at the University of Southern Denmark. He splits his time between Pharmacare-sponsored studies in British Columbia and, working by telecommunication, several projects at Harvard and in Denmark that use the case-crossover study design, a method he invented. Five years ago, he began studying evidence-based policy-making, reference pricing policies, how to build bridges between researchers and government decision-makers, and methods for using health databases for contacting of patients without violating privacy. This work culminated in a randomized drug policy trial in 1999. He has helped initiate other ongoing trials in BC with randomized delayed controls that are evaluating impacts of the Therapeutics Initiative's *Therapeutics Letter,* the Medical Services Commission's Guidelines and Protocols, physician-office access to PharmaNet, the Foundation for Medical Practice Education's Practice-Based Small-Group Learning Program, and several strategies for educating asthma patients in self-care.

Dr. Colin Soskolne, former Director of Graduate Training for the Department of Public Health Sciences and presently professor in the Epidemiology Program at the University of Alberta, received his PhD from the University of Pennsylvania in 1982. He has taught courses in Epidemiology, community medicine and occupational cancer Epidemiology at the University of Toronto (1982-1985) and, since 1985, at the University of Alberta. Dr. Soskolne's areas of research expertise are the human health consequences of global change, occupational cancer case-control studies, environmental cancer epidemiology, infectious diseases and professional ethics. He

established the Epidemiology Program at the U of A, and then developed the Department-wide Graduate Training Program. He has published more than 150 peer-reviewed papers, book chapters, books, editorials, letters, book reviews and an interactive video disk. He has held grants from international, federal and provincial agencies. Dr. Soskolne consults for the World Health Organization, The National Cancer Institute (Naples, Italy), and for the University of Pretoria, South Africa. His research on sulfuric acid led in 1991 to the I ARC designation of occupational exposures to strong-inorganic-acid mists containing sulfuric acid as a definitive human carcinogen. He has chaired and continues to serve on a number of professional committees, locally, nationally and internationally. Based on his work on a SSHRC-funded grant (1996-1999), Dr. Soskolne spent part of his sabbatical year (1998/99) with the WHO's European Centre for Environmentally & Health in Rome. There, he produced a Discussion Document calling on WHO to address concerns about the consequences of environmental degradation for public health. Dr. Soskolne's current research focus is to explore the role of epidemiology in linking indicators of health and well-being to ecological declines.

Dr. Robyn Tamblyn is an Associate Professor in the Departments of Medicine and Epidemiology and Biostatistics at McGill University, Faculty of Medicine. She also holds a position as Medical Scientist at the McGill University Health Center Research Institute, is a CIHR scientist and a McGill University William-Dawson scholar. She heads an FRSQ-funded team to study the relationships between medical training, practice and health outcome. She spearheaded a series of initiatives aimed at enhancing the early uptake of evidence into primary care practice, the medical office of the 21st Century projects – Phases I and II (MOXXI). More recently, she and her colleagues have obtained funding from the Canadian Foundation for Innovation to establish a novel provincial infrastructure for health care and research. The Quebec Integrated Health Care and Research Network will integrate data from the four academic university health centers and their extended primary care networks with the provincial health care database warehouse.

Dr. Ross Upshur received BA (Hons.) and MA degrees in philosophy before receiving his MD from McMaster University in 1986. After 7 years of rural primary care practice he returned to complete his MSc in epidemiology and fellowship training in Community Medicine at the University of Toronto. He is currently the director of the Primary Care Research Unit at the Sunnybrook Campus of the Sunnybrook and Women's College Health Sciences Centre. Dr. Upshur is a Research Scholar and Assistant Professor, Departments of Family and Community Medicine and Public Health Sciences at the University of Toronto. He holds a New Investigator Award from the Canadian Institutes of Health Research. He is a member of The Royal College of Physicians and Surgeons of Canada, The Joint Centre for Bioethics, University of Toronto and is an Associate Member of the Institute of Environment and Health at McMaster University and Adjunct Assistant Professor of Geography and Geology at McMaster University . His research interests include the concept of evidence in health care, medical epistemology, clinical reasoning, public health ethics, time series applications in health services research, empirical approaches in bioethics and environmental epidemiology.

Dr. Don Willison is Assistant Professor, Department of Clinical Epidemiology and Biostatistics, at McMaster University. He holds a Career Scholar award with the CIHR, and works out of the Centre for the Evaluation of Medicine at St. Joseph's Hospital in Hamilton, Ontario. Dr. Willison's training combines an undergraduate degree in pharmacy from the University of Toronto, a Master's degree in Design, Measurement, and Evaluation from McMaster University, and a Doctorate in Program

Evaluation from the Dept. of Health Policy and Management, Harvard School of Public Health. On the data privacy front, Dr. Willison has recently completed a survey of patients' views on use of information from their electronic health record for research purpose and is currently organizing a Delphi panel examining research and policy priorities on the use of personal health information for research. Dr. Willison is working with the Canadian Stroke Network to develop a cross-Canada consent-based patient registry combining clinical information from in-hospital encounters, self-reported health outcomes through follow-up telephone surveys, and provincial administrative records. He has been working with both the new Canadian Institutes for Health Research (CIHR) and Canadian Institute for Health Information (CIHI) on policies related to the secondary use of personal health information for research purposes to create an environment that will continue to allow the use of personal health information for research purposes while addressing the privacy and confidentiality concerns expressed by the public and advocacy groups.

***CIHR Staff*:**

Patricia Kosseim, Acting Director, Ethics Office

Matthew Furgiuele, Research Officer, Ethics Office

Preface

The Canadian Institutes of Health Research (CIHR) was created "*to excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system*" (*CIHR Act*, S.C. 2000 c. 6). CIHR's mission statement clearly emphasizes the importance, not only of advancing knowledge through health research, but of linking the results of health research to the relevant health needs of Canadians and our evolving health system.

The Federal Government funds research through CIHR because of the many benefits that the generation and use of new knowledge can bring to Canadians. Discovery leads the way toward progress, improved health, leading-edge innovation, new jobs and opportunities in a global knowledge-based economy. A commitment of funding in the order of $500 million to CIHR in 2001-2002 demonstrates Parliament's recognition of the importance of health research to Canadians.

The Institutes comprising CIHR span across disciplines, sectors and regions, in partnership and in collaboration with many others, both nationally and internationally, to address important health issues in areas such as:

- aboriginal people's health
- cancer
- circulatory and respiratory health
- gender and health
- genetics
- health services and policy research
- healthy aging
- human development and child and youth health
- infection and immunity
- neurosciences, mental health and addiction
- musculoskeletal health and arthritis
- nutrition, metabolism and diabetes
- population and public health

In particular, research on health services and policy, population and public health depends heavily on the ready availability of large volumes of existing data about people. Such data include health surveys; hospital, physician and laboratory records; provincial and federal billing and registration data; birth and death certificates; vital statistics; socio-demographic data; and, employment records. The data are analyzed for the purposes of:

- monitoring the health of the population;
- identifying populations at high risk of disease;
- determining the effectiveness of treatment;
- quantifying prognosis and survival;
- assessing the usefulness of preventive strategies, diagnostic tests and screening programs;
- influencing policy through studies on cost-effectiveness;
- supporting administrative functions; and,
- monitoring the adequacy of care.

While such research is of great social importance, Canadians also highly value their privacy. The right to privacy is in fact intimately connected with the right to respect for one's dignity, integrity and autonomy in a free and democratic society. This fundamental right is constitutionally enshrined in the *Canadian Charter of Rights and Freedoms*. It is also explicitly protected in Quebec's *Charter of Human Rights and Freedoms,* as well as the *Civil Code of Quebec*. The right to privacy lies at the very root of laws pertaining to data protection, statutory protection legislation, professional codes of conduct and other international and national ethics guidelines, including the *Tri-Council Policy Statement on Ethical Conduct of Research Involving Humans*. Though the right to privacy may seem well protected by this vast array of laws and ethical rules, the perceived threats to privacy as a result of recent technological advances, such as electronic data access, is of great concern to Canadians.

The recent debate over the new Federal *Personal information Protection and Electronic Documents Act* (S.C. 2000 c. 5) has brought into focus both the value Canadians place on world-class research needed to improve their health and health system, and the value they place on their individual right to privacy. In turn, provinces will also need to manage these values as they develop substantially similar provincial legislation before January 1, 2004. The practical challenge for policy makers and legislators, the research community, the health care sector and the broader public when developing, interpreting and applying legislation will be to reach an appropriate balance between the protection of personal information and access to such information in order to nurture and promote the values held by Canadians.

In this context, CIHR has created an Ad Hoc Working Group of researchers to develop a series of case studies. The objectives of these case studies are to foster dialogue:

- with those who draft policy/legislation and those responsible for interpreting it, by providing tangible illustrations of its practical application in the health research context;
- among researchers about how to comply with the spirit of the fair information principles and how to improve their information practices;
- with the privacy community and the broader public on the benefits and concrete realities of health research.

CIHR wishes to thank all the members of the Ad Hoc Working Group for their precious time and valuable contributions to this project, as well as their many colleagues who participated in the preparation of the case studies and supporting documentation.

Executive Summary

BACKGROUND

Ethics Guidelines

Health research has made critical contributions to improving health and the health system. Such research requires access to information which, in turn, raises concerns about privacy and confidentiality. Researchers have long recognized the importance of respecting the individual's rights to privacy and confidentiality in the context of health research. Internationally, researchers have been sensitized to privacy and confidentiality issues since the World Medical Association's *Declaration of Helsinki* (1964, as amended in 1975, 1983, 1989, 1996 and 2000). In the U.S., the seminal *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* (Washington, DC: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979) spawned a whole series of U.S. Federal Regulations governing all federally-funded research. In Canada, the former Medical Research Council published *Guidelines on Research Involving Human Subjects* which were later revised in 1987, and then replaced in 1998 by the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (Ottawa: Public Works and Government Services Canada, 1998) which covers all research funded by the three major federal granting agencies. In 1997, the Therapeutic Products Directorate of Health Canada adopted the *Good Clinical Practice Guidelines* of the International Conference on Harmonization of Technical Requirements for the Registration of Pharmaceuticals for Human Use; these international scientific and ethical guidelines set out the roles and responsibilities of sponsors and investigators conducting clinical trials that involve the participation of human subjects.

All of these fundamental documents require that researchers respect the data subjects' rights to privacy and confidentiality when designing and implementing research protocols. Rapid progress in technology and ever-increasing use of electronic records require that these guidelines not stay static, but rather, continue to evolve in order to ensure adequate protection. The further refinement and development of ethics guidelines is an important challenge that calls for the ongoing work and commitment of the health research community.

Research Ethics Boards

Research ethics boards (REBs) review research studies according to these ethics guidelines. In Canada, only those studies approved by REBs are eligible for funding by the three federal granting agencies and/or for regulatory approval under the *Food and Drugs Act* (R.S.C. c. F-27). Independent, multi-disciplinary REBs have been established at a local level in academic institutions across the country for several years now, in some cases, well over two decades. They embody a broad range of perspectives and an enormous wealth of hands-on experience in reviewing the ethical acceptability of research protocols. They are composed of at least two members with expertise in the area of the research under review; at least one member knowledgeable in ethics; at least one member knowledgeable in the relevant law, especially in bio-medical research; and at least one community member.

REBs have acquired specialized knowledge of the inherent complexity of research proposals involving various disciplines. REBs are well immersed in issues relating to both the protection of individual human subjects and the societal need for research. REBs review research protocols in accordance with fundamental principles of:

< respect for human dignity,
< respect for free and informed consent,
< respect for privacy and confidentiality,
< respect for justice and inclusiveness,
< balancing of harms and benefits,
< minimizing harm, and
< maximizing benefit.

REBs have unique experience applying these principles using a proportionate and flexible approach depending on the level of risk involved, and do so in a manner which seeks to achieve balance overall. REBs are specially placed to play both a review and educational role: they review research protocols with the aim of determining their ethical acceptability from the point of view of the research subject and they also provide an ongoing consultative and educational function for the research community.

The need to respect privacy and confidentiality in the current context of rapidly advancing technology and increasing use of electronic data poses particularly complex issues. A challenge for REBs will be to educate their members and acquire more specialized privacy expertise to deal with these issues as they arise. Also, an overriding challenge for the health research community as a whole will be to garner the necessary resources for REBs to carry out their responsibilities on a continuing basis and to establish a national system of REB accountability to secure greater public trust and confidence in the research enterprise. Finally, given the emergence of data protection laws, the relationship between REBs, privacy commissioners and other review bodies will have to be explored to ensure an oversight mechanism that is complementary, workable and feasible in practice.

Emerging Legislation

In recent years, Canadian legislation governing data access and privacy in health research has undergone (and continues to undergo) rapid change (See CIHR's *Compendium of Canadian Legislation respecting the Protection of Personal information in Health Research* (Ottawa: Public Works and Government Services Canada, 2000). Currently, the Canadian legislative landscape consists of a patchwork of laws with different requirements applying either at the provincial or federal level, in the private or public sector, to personal information generally or personal health information more specifically. Yet, large population studies often involve combining data from different provinces or even countries, meaning that a single research study or program may be subject to varying standards in different legal jurisdictions. Likewise, these research studies can be based on personal information derived from private and public sources, or be co-funded by public and private partners, and thereby be subject to both private sector and public sector legislation. Finally, examination of certain health determinants may require access to general personal information (eg. income level, education, work history) as well personal health information (eg. physician, laboratory, hospital records, registration and billing data, etc.), invoking both general and sectoral legislation. More than ever, there is a recognized need to harmonize legislative standards in order to afford similar protection across the country and to establish similar requirements for assuring that protection.

Most international standards are modeled after the *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data* developed by the Organization for Economic Cooperation and Development (OECD) in 1980. These guidelines have since been adapted by Canadian businesses, consumer groups and governments, under the auspices of the Canadian Standards

Association, reformulated into the *Model Code for the Protection of Personal information* CAN/CSA-Q830-96 (the "CSA" Code) and more recently incorporated as Schedule 1 of the federal *Personal information Protection and Electronic Documents Act* (S.C. 2000, c.5). The CSA Code is based on ten fair information principles:

- Accountability
- Identifying purposes
- Consent
- Limiting Collection
- Limiting Use, Disclosure and Retention
- Accuracy
- Safeguards
- Openness
- Individual Access
- Challenging Compliance

CASE STUDIES

In the fall of 2000, *Canadian Institutes of Health Research* (CIHR) created an Ad Hoc Working Group of health researchers to develop a series of case studies. The objective of this initiative, within the current ethical and legislative framework, is to foster dialogue:

- with those who draft policy/legislation and those responsible for interpreting it, by providing tangible illustrations of its practical application in the health research context;
- among researchers about how to comply with the spirit of the fair information principles and how to improve their information practices;
- with the privacy community and the broader public on the benefits and concrete realities of health research.

This collection of case studies will include 20 examples of actual research involving secondary use of data in Canada. They attempt to illustrate, in practice, what personal information researchers need, for what purpose, how they collect, use and disclose data, what retention practices are followed, what security safeguards are used and what oversight mechanisms are in place. These case studies, when reviewed in light of existing and proposed data protection laws, as well as the fair information principles enunciated in the CSA Code, raise interesting issues about the practical application of legislation to certain types of health research. Findings from a review of the case studies are outlined below:

The Value of Health Research

The case studies cover a range of research, including, as examples, studies that have:

< identified patterns of disease and certain adverse outcomes of medical treatments;
< tracked changes in payment policies for prescription medications; and,
< helped address perceived crises in the health care system.

These studies, and others like them, have contributed to our present level of understanding of the causes, patterns of expression and natural history of diseases, as well as the impact of ever improving strategies for diagnosis, treatment and prevention. The case studies illustrate that researchers critically depend on data to study large and unbiased samples that are representative

of larger populations. These samples, in turn, are required to generate meaningful conclusions regarding patterns of diseases, beneficial and adverse outcomes of therapies, and the effectiveness and economic efficiency of the health care system. Indeed, in the present climate of major public concern about the quality, effectiveness and cost of our rapidly changing health care delivery system, the need to support retrospective, epidemiological and health services research has become an urgent priority.

Secondary Use of Data

The case studies demonstrate that the ability to conduct health research to improve health and health care may depend heavily on large volumes of readily-accessible, existing data. Such data relates to peoples' health, their use of health care services, and their lives more generally. These may include information derived from: personal interviews; analyses of tissue samples; results of scientific tests; physician, hospital and laboratory records; birth and death certificates; billing claims; vital statistics; employee records; age; education; and socio-economic status. The case studies focus on examples of research using data that were originally collected for another purpose (secondary use of data). Existing data are often found to be extremely useful for identifying and understanding problems, as well as providing potential solutions. Possible secondary uses of data cannot always be anticipated during initial collection and therefore detailed and specific consent cannot always be obtained at that time. Yet, secondary uses of data can have great public benefit, provided the necessary confidentiality agreements are in place and appropriate safeguards are taken to protect the data against unauthorized disclosure.

Identifiability of *Data*

Researchers who study health services or health of overall populations rarely have any direct interest in knowing the specific identities of the people they study. Their focus is on aggregate trends. So, while personal information about identifiable individuals may be the source of data, this type of research is conducted with information that has either been made completely anonymous or has had as many identifiers removed as possible and replaced with encrypted codes. Indeed, many investigators conducting studies would not need any personal identifiers at all were it not for the need to adjust for important individual characteristics or to link data about individuals so as to construct histories over time. The possibility of linking de-identified data to other potentially identifying information (for example, birthdates, health insurance numbers) remains crucial. This is necessary in order to: study the relationship between certain health determinants and health status; group together individuals on the basis of common characteristics such as age or geographic location; or, track individuals over time in order to study the evolution of certain diseases after long latent periods or to assess their progress through the continuum of health care. Researchers ought to implement deliberate strategies that make it impossible (or at least extremely difficult) to determine the identity of an individual from the data they use. Current practice for anonymizing, de-identifying and linking personal information (whether carried out by the original data-holder before releasing the data for research purposes or by the researcher him or herself once in possession of the data) tends to vary significantly according to what is considered "identifiable". The ongoing challenge will be to reach agreement on what constitutes an appropriate degree of identifiability, a concept that will evolve over time. Approaches for de-identifying data need to achieve greater consistency to streamline efforts for meeting and continually improving best practices.

Security Safeguards

Traditional security safeguards include: organizational safeguards, such as limited personnel access, security clearance and employee confidentiality agreements; physical safeguards, such as

locked rooms, filing cabinets and facilities; and technological safeguards, such as special passwords and access codes. Further options for protecting personal information are increasing rapidly with advances in computing technology. A spectrum of solutions is emerging and techniques are increasingly available for limiting access to only the minimal data needed in the most general form possible, thereby ensuring confidentiality of the data while also retaining their usefulness for research purposes. The challenges now lie in: better disseminating information about existing security systems and processes; developing a set of minimum standards; ensuring greater consistency in the application of minimum standards; and, continually reviewing, updating and adapting those standards as technology evolves. There is clearly a need to identify best practices that are both cost-effective and sufficiently flexible to accommodate different research approaches.

Consent

In typical clinical research studies, researchers directly interact with patients in well-defined protocols and can provide them with the detailed information required for seeking informed consent. However, strict application of traditional consent procedures in health services and population health research raises problematic issues. Among the factors which often make seeking consent impracticable, impossible or self-defeating in these particular types of studies are: the sheer size of the populations studied; the wide range of relevant information examined; the age of the data; the significant number of persons who may have since relocated or died; the risk of introducing potential bias through the consent procedure itself thereby affecting the generalizability and validity of research results; the creation of even greater privacy risks by having to link otherwise de-identified data with nominal identifiers in order to communicate with individuals so as to seek their consent; the lack of any real opportunity for direct contact between the researcher and each individual in the study population; and, the practical difficulty of involving the original data holders to establish contact on researchers' behalf.

On the one hand, obtaining specific consent for all possible secondary uses of the information that often cannot be predicted at the time of collection is not feasible. On the other hand, obtaining unqualified, blanket consent for yet undefined future health research purposes is often empty and meaningless and may sometimes reduce rather than increase privacy protection. While it may be possible, over time, to expand consent mechanisms to make them more inclusive of future research uses in a manner that is both informed and meaningful, studies that rely on already-existing, historical or archival data - including samples - remain a challenge. The case studies demonstrate the need for constructive, creative and innovative ways of respecting peoples' right to know and to control how their information is used without necessarily having to obtain individual consent in writing from each and every one of them in each and every instance. The case studies also demonstrate the need to develop appropriate alternatives to the consent model, specifically designed for population health and health services research, taking into account the overall balance of risks and benefits both to individuals and society as a whole.

Retention and Destruction of Data

Many of the case studies were made possible by the existence of secure data archives containing historical records. These secure sites enable the linking of personal information about individuals in order to study important research questions. Just like research laboratories allow basic scientists to advance knowledge about disease, these data archives provide the necessary tools for population health and health services researchers to conduct important studies about human health and the health care system, (eg. studies on genetically modified foods, certain environmental exposures and hospital waiting lists.) These data archives also make it possible to identify potentially affected

patients in order to notify them about the risks of contracting fatal diseases or experiencing adverse effects that were unknown at the time of certain interventions (eg. risks of contracting Creutzfeldt-Jakob disease in human growth hormone trials, contracting HIV in hepatitis B trials or experiencing adverse effects from certain vaccines.) The automatic destruction of data and/or all possible identifiers upon the fulfillment of the original purpose for which the data were collected would prevent researchers from studying factors that may improve health and health care over time or to notify individuals and allay public fears when new problems emerge. Furthermore, the destruction of large databases would result in a huge waste of valuable public funds; having to re-create new data archives for each new research project would be completely impossible and/or entirely cost-prohibitive. In addition, researchers are often required to retain data for possible verification and auditing purposes, though these requirements tend to vary among sponsors and/or publishers. Creative means need to be further explored to secure the long-term existence of vitally important databases in the hands of trusted guardians subject to formal periodic audits and proper oversight.

Oversight Mechanisms

Health services and population health studies conducted in universities and affiliated institutions are typically reviewed by REBs to ensure compliance with fundamental ethical principles, including respect for privacy and confidentiality (see background, above). In cases involving secondary use of data or proposed data linkages, REBs consider, among other factors: the sensitivity of the information involved; the possibility of identifying particular individuals; the magnitude and probability of harm or stigma resulting from identification; the context in which the information was originally collected; the possibility of obtaining consent; the appropriateness of using alternative strategies for informing participants and/or consulting with representative members of the study group; as well as any legal provisions that may apply in the situation. In their review, REBs apply a proportionate approach in balancing risks and benefits and modulate their requirements accordingly. REBs are multi-disciplinary bodies, with specialized expertise and lay representation, close to the ground and sensitive to local needs and values. REBs play a critical role in ensuring the protection of individual privacy, within a larger ethical framework. Areas for further improvement include: strengthening privacy expertise and education of REB members particularly in light of rapidly evolving technology and emerging legislation; ensuring adequate resources for REBs to meet their mandate for continuing review, monitoring and periodic audits; increasing public accountability and transparency of REBs; and, further exploring the relationship between REBs, privacy commissioners and other oversight bodies. Indeed, in some provinces, legislation requires that privacy commissioners or special privacy committees designated by law also approve (or at least be notified of) the proposed research or data linkage. The challenge, therefore, will be to ensure complimentary forms of protection rather than redundant levels of unneeded bureaucracy.

Conclusion

In summary, the case studies assembled by CIHR provide examples of how researchers, using secondary data, attempt, through various ways, to comply with the spirit of the fair information principles contained in the CSA Model Code. They suggest the need to further develop creative, effective and innovative mechanisms for protecting privacy and confidentiality of data, as well as the need for ongoing discussion and continual improvement of best practices. The CIHR case studies further provide concrete illustrations of the importance for interpreting and applying privacy laws and policies in a flexible, feasible and workable manner in order to permit the valuable social benefits of health research to continue. The case studies suggest that the health research community should work actively with the privacy community, consumers and the general public to identify and implement strategies for balancing the right of individuals to have their personal

information protected and the need for researchers to access data so as to improve the health of Canadians and their health care system.

Values Which Health Researchers Believe in

To perform their work, some health researchers require access to data. They are granted access to the data they need on the understanding that they agree to adhere to rules concerning respect for the privacy and confidentiality of the data provided to them. As a group, and to ensure adherence to such rules, health researchers are expected to respect the individual's right to privacy.

Practice guidelines are the established norms for many specialized groups of health researchers. These guidelines are anchored in professional values. The values to which health researchers adhere include the protection of the public interest, respect for personal privacy, and objectivity and excellence in the conduct of research. They also address the need for the virtuous conduct of professionals engaged in health research. It is against these professional values that health researchers can be held accountable.

Like other scientists, health researchers uphold the values of free enquiry and the pursuit of knowledge. The goal of science, after all, is to explain and to predict observable phenomena. Health researchers thus not only pursue knowledge about health care and health systems, but they also uphold the value of improving the public's health through the application of scientific knowledge. Their research includes improving the efficiencies of what are very costly health systems to ensure that health services, and both the quality and quantity of life, are accessible to all Canadians. The large administrative data bases that have supported the delivery of services to Canadians for many years, have also facilitated the kinds of research that have contributed greatly to the health of Canadians. It is these administrative data bases that health researchers will often need to access.

Adherence to the values common to health researchers is ensured through professional training programs, professional self-regulation and institutional controls, including Research ethics boards. Health researchers engage in an ongoing review of procedures for ensuring that research being conducted in the public health interest is facilitated, taking into account the need to be responsive to changing technologies for the handling and transfer of data, as well as to social values and societal perceptions of the research enterprise. Owing to new technologies that make data access even more attainable, it behoves researchers to engage with privacy advocates to ensure that the public's interest continues to be well served.

Analysis of Case Studies

The processing and linking of a broad array of personal data for many purposes, including research, are generating considerable public concern. Consumer and professional groups have expressed the need for controls on the collection, use and disclosure of personal information. Issues surrounding the privacy and confidentiality of personal health information are particularly complex since personal health information is often regarded as highly sensitive. Patients and providers want assurances that such information will not be used inappropriately. Governments in Canada are moving to strengthen privacy protection through laws and regulations in step with the international community.

This said, the use of personal information is crucial in order for health researchers to examine issues of high public priority and to assist policy-makers in improving the health of Canadians and strengthening their health care system. Health research involves an integrated and multi-disciplinary approach. Health research embraces the study of disease at the biomedical level and specific treatment and preventive applications in well-controlled clinical contexts. At the more macroscopic level, health research also involves studying the health of populations and evaluating health services and policies, i.e.:

- studying geographic and temporal patterns of disease in the general population or certain sub-groups of the population;
- identifying and quantifying the risk of disease that may be attributed to a particular exposure, and
- developing and evaluating health strategies, treatments, services, programmes or policies.

Contrary to biomedical and clinical research that depend heavily on individual level data collected directly from research subjects through personal contact, a substantial proportion of population health research and health services research is conducted using large databases usually collected for other purposes. This is referred to as “secondary use” of existing data. Examples of data used for secondary purposes include data that were originally collected to:

- administer programs and services (e.g. physician and drug claims databases);
- manage employees, health professionals and hospitals (e.g. employee records and job descriptions);
- deliver clinical care (e.g. medical records); or,
- diagnose or identify a disease (e.g. blood, urine and tissue samples).

The secondary use of personal data for health research purposes can have broad social benefits. However, strict compliance with fair information principles aimed at protecting the privacy and confidentiality of data subjects (particularly those pertaining to identification of individuals, consent and retention of data) presents special challenges in the context of population health research and health services research.

For instance, it is often not practicable to obtain consent at the time the secondary use of the data is made, which could be several years after the original collection. Nor is it feasible to obtain, at the point of original collection, meaningful informed consent for all potential secondary uses that may eventually be made of the data since future purposes are often unknown and impossible to specify.

The requirement to destroy or completely anonymize all personal data once the specified purpose has been achieved also poses particular challenges for health researchers who are often obliged to retain data, sometimes up to twenty (20) years for auditing and verification of research results. Moreover, the necessary destruction of large and extremely valuable research databases each time a specific research project is completed entails a huge waste of scarce public resources and corresponding lost opportunity to improve health and health services.

The health research community has developed and adopted many security safeguards, accountability mechanisms and oversight mechanisms to protect the privacy and confidentiality of personal information. However, some of the more effective safeguards are not widely known or may be applied inconsistently. Also, there is significant variability in the degree of public accountability and oversight structures across legal jurisdictions (e.g. different provinces), sectors (e.g. private or public sectors) and institutional settings (e.g. government versus universities) where research is conducted.

There is a clear need to address public concerns regarding the secondary use of personal data for health research. Many questions need to be considered, for instance: where it is agreed that a given research proposal has social value, how can that research be done in a way that maximizes actual benefits and minimizes potential harms? What are best practices for safeguarding privacy and managing risks of unauthorised disclosures? How can these best practices be systematized? What oversight structures should be used to ensure the protection of research participants in an open, transparent and publicly accountable manner? What are the strengths and limitations of current legislation regarding the secondary use of personal information for health research purposes?

The 20 appended case studies* were developed to stimulate discussion and enhance mutual understanding among researchers, consumer and professional groups, the legal community and policy makers. They provide concrete examples of how secondary uses of personal data are answering questions that are of high priority to the general public. The case studies outline specific practices used by researchers and identify variations in the strategies they deploy to protect privacy and confidentiality of personal data. They also illustrate some of the special challenges associated with the interpretation and application of current legislation in the specific context of population health and health services research. Accordingly, the case studies highlight a number of ethical and legal issues that warrant further discussion and debate.

I. Why do Health Researchers Make Secondary Use of *Personal information*?

The attached case studies illustrate that researchers currently make secondary use of existing personal information to: 1) study patterns of diseases in the population; 2) identify causes of disease and their impact; 3) develop and evaluate preventive and therapeutic strategies, health services, programmes, and policies; and, generally, 4) assess data quality.

1. To study patterns of diseases in the population

Population health or health services researchers generally need to look at whole populations, or a

* Editor's note: the complete version of this document will contain 20 Case Studies. This initial version however contains only 8. For an explanation of the selection process please refer to appendix A.

representative sample of individuals to address questions about geographic or temporal patterns of disease, potential risks of disease or effectiveness of certain treatments, services or policies. Sometimes, all of the information needed to answer the research question is contained in databases created for other uses and the researcher does not need to contact individually the thousands of people involved to obtain any further information. The researcher does not need to know who the actual individuals are. It is information on the whole study population that is important to the researcher. However, an individual identifier is nonetheless necessary to link information about the same individual across databases (e.g. data on prescriptions with data on hospital admissions) or to link information on the same individual within a given database (e.g. to identify prescriptions written to the same patient in a prescription claims database). Usually, the identifier can be non-nominal, and once linkage is complete, the set of information about an individual can often be de-identified or completely anonymized such that re-identification is rendered extremely difficult, if not impossible.

- ***Rapid surveillance of cancer in neighbourhoods near point sources of pollution:*** The Ontario Property Assessment File (which identifies where people pay their taxes) can be used to link information on where Ontario residents live with the Ontario cancer registry and with Ontario death records. These linked data can then be used to learn if there are significantly higher risks of cancer or death for persons who live in certain areas. Because these data are already collected on an ongoing basis, assessment of risks can be done relatively quickly and easily to identify and divert potential risks. **[case #16]**

- ***Use of RFLP molecular epidemiology to find out how tuberculosis is spread among people infected with HIV:*** Researchers in Quebec used sputum samples stored in the Public Health Laboratory to assess how many cases of tuberculosis were due to new infections spreading between individuals, and how many were due to re-activation of previous infections. The weakening of the immune system due to HIV infection facilitates the spreading of new infections, including tuberculosis. Understanding the role of HIV is important to help control tuberculosis infections. The researchers did not need to know whose samples these were, nor did they require any other potentially identifying information. Contacting the persons who provided the samples for the purpose of obtaining consent would have been both intrusive to those providing the samples and logistically impossible. **[case #5]**

Other times, databases are used to identify or assemble potential research participants. In these cases, individuals identified as eligible participants are then contacted and asked whether they would agree to participate in a research study or to provide further information needed to answer a research question. Researchers may also use databases to identify potential controls or comparison populations.

- ***Patient outreach via Pharmanet:*** In British Columbia, a province-wide network of pharmacy computers was used to "flag" the records of persons taking five or more prescription drugs. Once this study population was selected, a one-line message was sent to the pharmacist to offer patients an educational service that could assist them in better managing their medications. The research aimed to see whether study patients who agreed to this additional instruction would be better at refilling their prescriptions at the right times, than were those control patients who did not receive similar instruction. **[case #17]**

- ***Ontario familial colon cancer registry:*** To learn more about the genetic causes of cancer, cancer patients identified through their provincial cancer registry are contacted to ask whether they, and their family members, would consent to provide more information through

questionnaires and tissue samples for possible inclusion into the family registry. If they are part of a cancer family and agree to be included in the registry, they may be asked to participate in a specific research project, conditional upon their informed consent for that project. **[case #15]**

2. To identify causes of disease and their impact

Secondary use of existing data is often required to conduct studies examining the causes of disease, or to determine whether persons who are exposed to a substance are at increased risk of adverse health effects. Many epidemiological studies investigate the health effects of exposures or events that occurred in the past (e.g. exposure to asbestos among shipyard workers, or exposure of gulf war veterans to depleted uranium). For both ethical and logistical reasons, researchers must often rely on retrospective studies that draw upon existing data designed for other purposes. Clearly, researchers cannot expose persons to a suspected environmental toxin to study its health effects. Instead, researchers need to study people who have already been exposed, and compare their overall health to that of control participants who have not been exposed.

- ***Identifying subgroups of the general population that may be susceptible to short-term increases in particulate air pollution: A time series study in Montreal, Quebec:*** Provincial death records were consulted to identify 140,939 Montreal residents who died between 1984 and 1993. Researchers asked the Régie de l'assurance-maladie du Québec to link these individuals' death records with information about their physician visits, hospitalisations and drug prescriptions. De-identified, linked information was then released to researchers who used this information to determine whether fluctuations in the levels of particulate matter in the air in Montreal (measured at Dorval Airport) may have contributed to cause of death in certain more vulnerable subgroups of the population. **[case #19]**.

- ***Cancer and other problems associated with breast implants:*** Surgical records were used to identify 25,000 women in Ontario and Quebec who had received implants for cosmetic reasons. Information about the women and their surgical procedures taken from physician and hospital records was linked by Statistics Canada with information on deaths or diagnosis of cancer to determine whether women who had had breast implants were at greater risk of particular cancers or of death than women in the general public. **[case #13]**

3. To develop and evaluate health strategies, treatments, services, programs and policies
Pre-existing databases are primary sources of information for monitoring and evaluating the performance and effectiveness of the healthcare system and developing new policies. Comprehensive databases are available in many provinces documenting physician visits, hospitalisation and prescription drug use. Studies using this information are critical for maintaining and improving the health care system and evaluating new health policies to ensure that they are effective.

- ***Seasonal Patterns of Winnipeg hospital use:*** There are repeated crises in emergency rooms and waits for hospital beds during flu season. This study examined patterns of hospital use, focusing on January to April over several years, to estimate the extent of this problem and suggest ways to avoid crises in the future. This was a descriptive study, and its validity and value depended on complete and accurate data on hospital use. Comprehensive data on hospital discharges, collected routinely for administrative purposes, were used for the study. The data were de-identified, and consent was not obtained. The

benefits of the study were substantial. It found that crises had resulted from flu outbreaks, that the increased need for hospital beds resulting from flu outbreaks could be predicted, and that there were ways to avoid the crises. **[case #2]**

- ***A Randomized Drug Policy Trial with Camouflaged Contacting of Patients.*** This study sought to evaluate the impact of a provincial drug plan policy to convert treatment of asthma patients from nebulizers to inhalers. Researchers obtained from the Ministry of Health de-identified, already linked data about affected patients, including data on prescriptions, hospitalisations, medical services and long-term care. In order for researchers to contact these patients to send a quality of life questionnaire, Pharmacare gave researchers a separate, camouflaged list of only names and addresses which included both affected (80%) and unaffected (20%) patients. Hence, when researchers used the list to mail out the questionnaire, they had no way of knowing the health status of those individuals. Only patients potentially affected by the new asthma policy and who agreed to self-declare and participate in the study by returning the questionnaire would thereby reveal their health status to the researchers. This camouflaged sampling protected patient privacy, while allowing researchers to obtain important information about their health so as to evaluate the impact of the Pharmacare policy. **[case #12]**

- ***A Randomized Controlled Trial of Call/Recall of Hard-to-Reach Women for Pap tests.*** This study set out to determine whether reminder letters sent to hard-to-reach women could increase their likelihood to undergo regular Pap tests and thereby decrease the rate of death from cervical cancer. In order to identify eligible women for this study, researchers needed to consult physician records, the provincial cancer registry and the provincial cytology registry. "Hard-to-reach" women who had never been screened or were seldom screened were sent a first letter and a subsequent follow up letter reminding them to go see their physician for a regular Pap test. The study results indicated that these hard-to-reach women were no more likely to come in for their Pap test after receiving a reminder letter and that the development of population-based programs and policies will likely require multiple approaches to successfully recruit these women for screening. **[case #10]**

4. To assess data quality

Researchers also study how to refine their methods to draw clearer and more accurate inferences from existing data. Data quality is an important issue in research on populations; research based on poor quality data is wasteful of resources and misleads policy-makers and the general public.

- ***Assessing the accuracy of the Nova Scotia Health Survey:*** Canada-wide or province-wide health surveys provide important information based on representative samples of the population. People who refuse to take part in such a survey or who cannot be reached for one reason or another often account for 25% or more of those selected to be in the sample. It is important to know whether non-respondents differ in any systematic way from those who do take part. Hence, in this study, data about physician visits, hospitalisations and drug prescriptions were linked with the sample of persons who were contacted to take part in the Nova Scotia survey. This permitted researchers to compare the health characteristics of persons who had taken part in the survey with those of persons who had not taken part. Based on this comparison, researchers were able to develop "correction" factors that now allow more accurate inference to be drawn from the survey. **[case # 3]**

II. Why is it Sometimes Impracticable to Obtain Consent Before Making Secondary Use of Existing Data for *Research* Purposes? What are some alternatives?

Researchers in the fields of population and public health, health services and health policy who depend heavily on access to large, pre-existing databases face special challenges when having to seek consent from individual data subjects. This situation differs significantly from the usual work of clinical researchers that depends on the prospective collection of new data from individual patients,
where the opportunity to seek consent is much more feasible in practice. There are several reasons for this difference.

1. Unforeseen opportunities at time of *collection* / Unfeasible conditions at time of *research*

A. Administrative databases collected for other purposes

It is often realized, much after the fact, that large databases routinely collected for another purpose provide a unique opportunity to answer new and important research questions about patterns of disease, risks of disease or the effectiveness of certain health services, treatments or policies. Fee-for-service claims data routinely collected under provincial health plans for the purpose of paying physicians and prescription drug claims data collected under provincial drug programs are examples of large, administrative health databases, the full value of which was not recognized when first collected. Physician claims data include information such as patient and physician identifiers, the date of service, the type of service performed, the primary diagnosis and the amount paid for the service. Prescription drug claims data typically include physician, pharmacy and patient identifiers, the type of drug, the quantity dispensed and the cost. Other related databases include: hospital discharge abstracts, patient files, health provider files, laboratory files and vital statistics data. Collectively, these databases describe nearly all of the publicly-funded healthcare services provided in each province.

Not until the 1980s did the potential value of these databases for research purposes become more widely apparent. These databases typically contain many years of comprehensive data maintained on an ongoing basis for accounting and auditing purposes. Such information would be virtually impossible for researchers to reconstruct for both logistical and financial reasons. These databases have thus become indispensable sources of data for health services research and planning in Canada. Many of the case studies illustrate the use of these databases, including:

- ***Use of anti-arrhythmia drugs in Saskatchewan:*** Some drugs used to correct irregular pumping of the heart can, in some people, cause dangerous changes in rhythm. Professional practice guidelines recommend what types of drugs should be prescribed to avoid this problem. In Saskatchewan, prescription drug data were linked with hospital records to determine the degree to which physicians in the province were following these practice guidelines and whether non-compliance with the practice guidelines was contributing to the incidence of rhythm problems. **[case #7]**

- ***The impact of having elderly and welfare patients in Quebec pay a greater share in the costs of their prescription drugs:*** This important study, the results of which were recently reported in the Journal of the American Medical Association (JAMA), examined the impact of a recent policy requiring patients to pay a greater share of their prescription drug costs under the provincial drug plan. The study showed that this cost sharing arrangement resulted in a significant reduction of drug use to control illness

among some patients most in need, leading to an increase in serious adverse effects. The study findings resulted in an immediate change to the provincial drug insurance policy allowing free access to drugs for elderly patients and those welfare participants unable to work due to illness. The change in policy came into force as law within 6 months of the final report of the study. **[case #11]**

All the valuable research uses that can be made of these data simply could not be anticipated at the time the data were collected. In addition, consent has never been explicitly sought for maintaining a clinical record from the patient encounter, or for the transfer of relevant subsets of this information for other clinical care (e.g. referral to a specialist). In such cases, consent for use of this information for administering the public health care system has been implied. Currently, even the attempt to contact the thousands of individuals involved in these large administrative databases to seek their informed consent to use their de-identified data for research purposes is unfeasible. At the time of use, many data subjects have relocated or died since the time of the original data collection. Moreover, the time and costs required to try to reach remaining subjects can be prohibitive. This reality is reflected in most provincial Freedom of Information legislation that allows public bodies to disclose personal information for research or statistical purposes without consent under strict conditions that may vary from province to province.

B. Patient registries developed for statistical and research purposes

Other examples of systematic and ongoing collection of large amounts of data that are extremely useful for research purposes are patient registries. Such registries collect personal information on patients afflicted with a particular disease (e.g. cancer) or patients who received specific health interventions (e.g. cardiovascular surgery or immunizations). These databases are often collected specifically for public health and/or research purposes. In some cases, use of the personal information for research purposes is expressly permitted by the enabling statute that creates the registry (*e.g. The Cancer Act*, R.S.O. 1990, c. C.1, section 7(1)). In other cases, consent to use personal information for research purposes is obtained from the patients themselves.

Obtaining consent for use of personal information for research purposes presents special challenges in the case of registries. Registries are not created for a single research study. Rather, they provide critical infrastructure to support a broad array of studies that meet important information needs in our healthcare system. They are used to, among other things, assess patterns of disease, determine risks of disease, develop and evaluate health policies, assess the safety and effectiveness of therapies, measure the quality of care being provided and provide direct information to healthcare providers and patients to enhance the quality of care. As mentioned in section I (1)(A) above, consent could be sought, either at the time data are collected for inclusion into the registry or when used for specific research studies. However, special challenges arise at either stage.

At the time the data are collected, it may be difficult to provide specific information on the potential uses of the data. In other words, while consent could be obtained, it may not necessarily be meaningful informed consent. Low participation rates could also diminish the utility and value of the data in describing attributes of the general population. Further, data elements in some registries may pertain to multiple persons (e.g. familial registries containing personal information about individuals may also reveal related personal information about his or her family members).

At the time the research is conducted, the study can be described in more specific detail and consent obtained at that point would arguably be much more informed and meaningful. However,

having to contact all potential research subjects in the registry may prove to be costly, time-consuming and difficult for some of the reasons elaborated above. That is, many subjects may be hard to trace years after collection due to relocation or even death. This being said, some registries have implemented a regular follow-up program precisely to maintain a complete and accurate record of changes over time. It may be that individuals are regularly contacted to check blood pressure and serum lipids or to answer questions about quality of life and adverse health events. In the case of those registries, establishing contact for research purposes may be significantly less problematic. Some examples of research using patient registries include the following:

- ***Second cancers following treatment for non-Hodgkin's lymphoma*:** Data from the Ontario Cancer Registry were used to study whether the type of chemotherapy and radiation therapy used to treat non-Hodgkin's lymphoma (NHL) was associated with increased risk for subsequent cancers. The study was able to demonstrate that patients with NHL continued to be at significantly elevated risk of second cancers for up to two decades following the diagnosis. This persistently elevated risk has important implications for the medical surveillance of these patients. **[case #14]**

- ***Ontario familial colon cancer registry:*** This computer-based research registry will collect personal data on individuals with a genetic predisposition to colorectal cancer (CRC), their families, as well as unaffected individuals. This registry will facilitate studies on how inherited and external risk factors affect the risk of CRC. Such research will help us better understand whether genetic screening would be useful and what preventive interventions and/or different treatment regimens might reduce the high risk of developing cancer and the risk of resulting death. **[case #15]**

- ***National diabetes surveillance system***: Currently under development, this national diabetes surveillance system will use physician and prescription claims data from provinces, territories and aboriginal communities to describe trends and geographic differences in the care received by people suffering from diabetes. The surveillance system will be an important resource for monitoring the quality of diabetes care across the country and for developing more effective strategies of prevention and planning. **[case #4]**

- ***National immunization registry***: The development of a national immunization registry will bring together immunization records from across the country that are currently kept sometimes in doctors' offices, sometimes in public health clinics and sometimes in hospitals. The registry can provide parents with a more reliable and permanent record of their children's vaccinations; allow physicians to send more prompt reminders when immunizations are due; permit health planners to better assess the extent to which the Canadian population is vaccinated for purposes of developing immunization programs; and, provide researchers with much needed data to study how immunization can help reduce illness and prevent death from certain diseases. **[case #8]**

2. Inability to study important health research questions

The value of many types of population health or health services studies depends primarily on their ability to accurately describe or compare the characteristics of populations or groups. Examples include studies that aim to measure the distribution and frequency of risk factors and health problems in a population, to measure the health care needs of a population and to evaluate the

actual effects of treatments in a population. Data suitable for such studies must either cover the entire population of interest or else constitute a representative sample of the population.

Data collected in the context of clinical trials are of limited value for describing the effects of a treatment in a population and are inappropriate for describing the characteristics of a population*. Clinical trials are typically designed to assess cause-and-effect relationships in a select group of patients. These study patients must meet strict criteria for inclusion in the research and therefore are usually not representative of the general population. Clinical trials assess the effectiveness of a treatment or intervention in certain well-defined circumstances. However, there is growing recognition of the need to understand how well treatments *actually* work in real-world settings. Data collected from surveys, and large amounts of data collected systematically and routinely for administrative uses (e.g. claims databases, registries, and employment records) are usually much better suited to address the real-world effectiveness of treatments, since they provide a more representative sample of the general population.

The potential benefits of descriptive studies and studies designed to assess real-world effectiveness of preventative and therapeutic strategies can sometimes be significantly undermined by efforts to obtain consent for secondary uses of data. In many cases, the likelihood of being able to contact and obtain consent from the thousands – sometimes millions -- of individuals in many large databases is extremely small. Even if it were feasible in terms of economic costs and logistics, a requirement for individual consent could result in a systematic exclusion of persons who could not be reached or who declined to participate. It is known that persons who can be located and who have the time and inclination to participate in studies are often systematically different with respect to their age, sex, health and other characteristics than the general population. As a result, the requirement of obtaining consent to use secondary data can result in biased findings and undermine the potential benefits of the research.

Exorbitant costs, practical difficulties and biased findings that may result from a consent requirement do not, in and of themselves, constitute adequate justification for using data without consent. However, these factors do emphasize that the impact of requiring consent for secondary use of data can be substantial for some types of studies, even to the degree of defeating the potential benefits of a study. It would seem only reasonable then, that the loss of potential benefits resulting from the consent requirement be weighed against the risk of harm resulting from possible

---

* It is important here to discuss the notion of 'validity':
Validity refers to the degree to which the inference drawn from a study, especially generalizations extending beyond the study sample, are warranted when account is taken of the study methods, the representativeness of the study sample, and the nature of the population from which it is drawn. Two varieties of study validity are distinguished:

1. Internal validity: The two groups being compared are selected in such a manner that the observed differences between them on the dependent variable under study [i.e. the outcome of interest] may, apart from sampling error, be attributed only to the hypothesized effect under investigation.
2. External validity (generalizability): A study is externally valid or generalizable if it can produce unbiased inferences regarding a target population beyond the subjects in the study. This aspect of validity is only meaningful with regard to a specified external target population. For example, the results of a study conducted using only white male subjects might or might not be generalizable to all human males (the target population consisting of all human males). It is not generalizable to females (the target population consisting of all people). The evaluation of generalizability usually involves much more subject-matter judgment than internal validity.

The above is adapted from John M. Last, *A Dictionary of Epidemiology*, 4th ed. New York: Oxford University Press, 2001.

infringement on privacy and confidentiality.

- ***The Computerization of Medical Practices for the Enhancement of Therapeutic Effectiveness:*** The purpose of this research is to evaluate the usefulness of electronic medical records (EMRs) of 42,000 patients in primary care offices in the Hamilton-Niagara area and to study the benefits, risks and costs of these EMRs for physicians and their patients. A rigorous, comprehensive analysis of the use of EMRs is necessary to allow researchers to determine whether this health care technology could help improve patient care and physician practices. **[case #1]**

- ***Studying the Health of Health Care Workers***: This study examines, over a fifteen-year period, the impact of workplace characteristics on specific causes of morbidity (musculoskeletal and mental health disorders) among health care workers in B.C.'s acute health care sector. As health care workers represent a significant portion of the workforce in Canada, healthy workers are considered essential to the delivery of care in our health system. A comprehensive study of occupational risks facing this large population over a long period of time is essential to help researchers and policy-makers develop methods of early detection and intervention to prevent or minimize the functional limitations and disabilities that force employees to leave the workforce. **[case #20]**

Moreover, some studies are aimed specifically at examining the characteristics of certain sub-groups of the population who are non-responsive, hard-to-reach or generally non-compliant with recommended treatment. For instance, researchers sometimes need to understand whether those who do respond to requests to participate in research studies differ, in any systematic way, from those who do not respond to attempted contact by researchers. Being able to characterize non-respondents allows researchers to gauge the possible bias in their study results and draw more accurate inferences therefrom. Also, researchers sometimes set out to examine the specific characteristics of hard-to-reach populations who do not comply with recommended treatment. Better insight into the common characteristics of this particular sub-group of the population is necessary in order to help improve health services and policies targeted specifically to them.

Requiring researchers to obtain consent from non-respondent, hard-to-reach or generally non-compliant individuals is often highly impracticable given the very nature of the group itself, or may be self-defeating given the very purpose of the study. Two specific examples follow:

- ***Assessing the accuracy of the Nova Scotia Health Survey***: In this study, researchers set out to determine whether the health characteristics of non-respondents to the N.S. Health Survey differed in any systematic way from the health characteristics of persons who did respond. Based on this comparison, researchers were able to develop "correction" factors that now allow more accurate information to be drawn from the survey results. Requiring researchers to individually contact all non-respondents to obtain their consent before undertaking this comparison would have been practically impossible since many of the non-respondents were non-respondents precisely because they could not be reached in the first place. **[case #3]**

- ***A randomized controlled trial of call/recall of hard-to-reach women for Pap tests:*** This study set out to determine whether reminder letters sent to hard-to-reach women would increase their likelihood to undergo regular Pap tests. "Hard-to-reach" women who had never been screened or were seldom screened were sent a first letter and a

subsequent follow up letter by the research team reminding them to go see their physician for a regular Pap test. Had the physicians contacted these women in advance to request their consent to be contacted by the researchers, this would have made it impossible to determine whether any increase in the frequency of Pap tests among these women was due to the specific strategy under investigation (*i.e.* the reminder and follow up letters) or the preliminary contact by physicians to obtain consent. Hence, in this case, preliminary contact by physicians to obtain consent would have confounded the results and completely defeated the purpose of the study. **[case #10]**

3. Additional risks to *privacy* and *confidentiality*

There are situations where privacy protections, when applied in the research context, can actually result in unintended adverse consequences. For instance, in some cases, the very process of having to contact data subjects to obtain their consent to use de-identified data for research purposes could pose greater privacy and confidentiality risks for the data subjects themselves, sometimes outweighing the potential benefits of the study. When only de-identified data are released by the original data holder to the researcher, the researcher has no direct way of knowing - nor any desire to know - who the individual data subjects are. While it may still be possible to re-identify persons indirectly using other information about them (e.g. age, sex and a prescription refill history), it is difficult to do in a database of hundreds of thousands of persons and completely unnecessary for the research purposes. Requiring researchers to obtain individual consent before making secondary use of de-identified data places researchers in a particularly difficult situation for two reasons.

Researchers must either rely on the original data holder to obtain prior authorization from the individual data subjects or researchers must seek access to identifying information in order to themselves contact individuals directly. As regards the first option, many data holders do not have the time, interest or resources to obtain this prior authorization on behalf of researchers; their engagement in the consent process may be virtually impossible to obtain in practice or less than optimal given their resource constraints or priorities. As regards the second option, researchers are placed in the odd situation of having to request access to identifying information from the original data holder that would otherwise not be needed to conduct the study. The release of direct identifiers in situations where they are not necessary to fulfill the research purpose defies the fair information principle of limiting collection and actually places the privacy and confidentiality of the data in greater jeopardy than would otherwise be the case.

Another unintended consequence of privacy protection arises in some cases where researchers are in direct contact with individual data subjects to request their participation in an anonymous questionnaire or their donation of an anonymous tissue sample for research purposes. The stringent requirement for written consent may, rather than afford greater protection, actually pose greater risk to the privacy and confidentiality of personal information. That is, individuals who would otherwise be willing to consent verbally to anonymous participation in research may be much more hesitant to sign a written consent form where their name becomes associated with the study. The inclusion of their name in circumstances where, but for the consent requirement, it would be completely unnecessary to fulfill the research purpose, naturally increases the risk of unauthorized disclosure to third parties.

- ***Use of RFLP molecular epidemiology to find out how tuberculosis is spread among people infected with HIV:*** In this study, the Department of Health linked results of a genetic analysis of tuberculosis (TB) bacteria (grown from sputum samples stored at the

Public Health Laboratory) with basic demographic data such as age, sex and residence. The linked, but de-identified, information was provided to researchers who then proceeded to study patterns in the spread of TB and identify contributing factors. If researchers were to seek prior consent for conducting this study, they would have required names and addresses in order to contact the individuals concerned. Releasing identifying information to researchers, that they would not otherwise require but for the consent requirement, would have, in effect, posed greater risks to individual privacy and confidentiality. **[case #5]**

- ***HIV seroprevalence among women undergoing abortion in Montreal*:** Better understanding of the incidence of HIV infection and risk factors can help guide educational programs and policies aimed at preventing infection. In this study, women undergoing therapeutic abortions who had to have a blood test as part of their standard treatment were asked to donate part of the blood sample for anonymous HIV testing and to fill out an anonymous questionnaire about certain risk factors. A computer-generated scrambled code was used simply to link the results of the participants' blood test with their answers to the questionnaire. There was no possibility of identifying the individual women and once the links were made, even the scrambled code numbers were destroyed. In most legal jurisdictions where this international study was conducted, verbal consent sufficed for the purposes of recruiting participants. Under Quebec law, however, consent has to be given in writing. As a result, Quebec women who otherwise would have been willing to participate in the study on a completely anonymous basis refused to do so because they did not want to sign their name on a written consent form which would document their participation in an HIV study. Ironically, therefore, a legal requirement intended to provide greater protection to data subjects, in this case, actually increased the perceived and actual risks to individual privacy. **[case #6]**

4. Need for rapid response to a potentially urgent public health threat

Often, use of existing datasets can meet a critical need for rapid or timely access to information. Whereas, obtaining individual consent for the use of the dataset or the prospective collection and analysis of new data could take years, or even decades. Examples of relatively urgent health threats include concerns about the safety of a drug (*e.g.* cisapride andor "fen-phen"), the potential risks posed by an exposure (*e.g.* blood transfusions potentially infected with HIV or Hepatitis C; E-coli infections caused by contaminated drinking water), and the adequacy and sufficiency of health services at times of crises (*e.g.* hospital emergency services, delays for cardiac surgery). Secondary use of existing data can, in such circumstances, provide pertinent information for rapid and effective risk assessment.

- ***Seasonal Patterns of Winnipeg Hospital Use**:* This study was undertaken in response to concerns about bed shortages and emergency room overloads in Winnipeg hospitals during certain periods of the year. Similar problems were occurring in other provinces. By using existing hospital data routinely collected for administrative purposes, researchers provided timely findings illustrating certain patterns and services of use, specifically, an increase in the need for hospital beds each year due to flu. The findings allowed policy-makers to predict and avoid annual crises the following year by adopting
  new policies for strengthening the influenza immunization program in Manitoba and better managing the use of hospital beds. These successful policies and programs have since been introduced in other provinces as well. **[case #2]**

- ***Rapid surveillance of cancer in neighbourhoods near point sources of pollution**:* A

rapid computerized surveillance system capable of assessing the relationship between residential proximity to a potential source of pollution and the subsequent incidence of cancer can provide timely and reliable evidence to communities, alerting them if significant hazards exist, or reassuring them if no association is found. Notable examples of community concerns include residential proximity to nuclear reactors, metal smelters and foundries, chemical contamination of drinking water and industrial pollution in general.**[case #16]**

5. Alternatives to Individual Consent

For reasons elaborated above, obtaining individual consent in the context of large studies on population health and/or health services can sometimes be impracticable to obtain. However, this does not prevent researchers from seeking alternative means for providing individuals with the opportunity to become engaged and/or opt-out of the process and creative ways of better understanding what might be their perspective. Below are two examples of alternative means employed by researchers when individual consent was impracticable to obtain.

- ***Cancer and other problems associated with breast implants:*** Following the recommendations of the REB that reviewed the research proposal, the researchers in this case initiated a general information program that publicized the study aims and methods at professional meetings, through women's interest groups and in lay and scientific periodicals and newspapers. Informational pamphlets were also distributed to 35 000 physician offices across Canada for display in patient reception areas. A toll-free, bi-lingual hotline was set up in order to provide more detailed information and to allow women to opt out of the research. **[case #13]**

- ***Rapid surveillance of cancer in neighbourhoods near point sources of pollution:*** In this particular case study, it will be nearly impossible to obtain informed consent given the sample size of 100 000 people, as well as the long latency period for potentially adverse health effects and the likelihood of relocation or death. Nevertheless, the researchers will undertake a priori qualitative studies and interview focus groups representative of individuals, interest groups, government agencies, other stakeholders and cancer patients in order to better elucidate community concerns and interests prior to commencing the study. [**case #16**]

III. Access to Existing Data, Security Safeguards and Review/Oversight Mechanisms – What is the Current State of Affairs? What are some of the Challenges Facing Health Researchers?

The following case studies reveal a significant degree of variation in: 1) the conditions required for accessing existing data; 2) the deployment of security safeguards aimed at protecting personal information; and, 3) the review/oversight mechanisms in place for approving the collection, use and disclosure of data for research purposes. The variations can be found at the level of researchers, data custodians, institutions, sectors and/or jurisdictions. These can be due to differences in the practices currently adopted and/or requirements legally imposed. This variability highlights the challenges that health researchers often face in seeking secondary access to data for research purposes.

The case studies demonstrate an urgent need to harmonize general practices and legal requirements. At the same time, the case studies highlight an equally important need to open dialogue, build public trust and explore flexible and creative approaches for dealing with the unique ethical challenges specific to particular research studies.

1. Access to Existing Data

The case studies illustrate how access to existing data for research purposes can involve many different stewards or custodians. These include hospitals, public health clinics and laboratories, physicians' offices, research centers, pharmacies, employers, specially created organizations with statutory responsibility for data (e.g. cancer registries, Canadian Institute for Health Information), and federal/provincial/municipal government departments and agencies (e.g. Health, Environment, Revenue, Statistics).

Health research frequently requires linking different data maintained by several different custodians; this can result in considerable complexity. For example, in the study examining the association between breast implants and cancer (case # 13), researchers needed access to patient files, including surgical records. In Quebec, researchers could seek access to this personal information from public hospitals subject to prior approval by the Director of Professional Services of each institution in accordance with sections 59 and 125 of the *Act respecting Access to documents held by public bodies and the Protection of Personal information* (R.S.Q. c. A- 2.1) and section 19.1 of the *Act respecting health services and social services* (R.S.Q. c S-4.2). Whereas, in Ontario, researchers could only request access to this personal information from physicians' offices in accordance with the criteria specified under section 34(b) of the regulations under the *Ontario Medicine Act* (O.Reg. 856/93). Patient identifiers obtained from these records were then sent to Statistics Canada that, in turn, linked the data with information on the incidence of cancer and mortality. In order to do so, prior authorization had to be obtained from each provincial cancer registry and each provincial registrar of vital statistics that regularly maintain these cancer and death records and routinely transmit them to Statistics Canada in compliance with the federal *Statistics Act* (R.S.C. c. S-19).

In some cases, the same data might be maintained by one or several custodians. For instance, in some provinces (*e.g*. Saskatchewan, Alberta and Quebec), access to administrative data (*e.g.* physician and drug claims data) can only be sought from provincial governments directly. Whereas, in other provinces, such data are also maintained by research centers with a specific mandate to manage the data for research purposes. Examples include the Centre for Health Services and Policy Research (CHSPR) at the University of British Columbia, the Manitoba Centre for Health Policy and Evaluation (MCHPE) at the University of Manitoba, the Institute for Clinical and Evaluative Sciences (ICES) in Ontario, and the Population Health Research Unit (PHRU) at Dalhousie University in Nova Scotia. An advantage of these centres is that they can streamline the access and linkage of data from different custodians in a way that enhances security. For example, all four centres manage data provided by several different custodians (e.g. hospitals and different government departments), house the data in secure environments and have the internal capacity to link this data and remove all identifiers before releasing the data to researchers.

Before allowing researchers to gain access to personal information, some data custodians may either choose - or be required – to enter into detailed research agreements with the researcher. The essential purpose of the research agreement is to obtain a legal commitment on the part of the researcher to maintain the privacy and confidentiality of the data released. The specific terms of such agreements can vary significantly, as can sanctions for non-compliance.

2. *Security* Safeguards

The case studies also illustrate the variety of safeguards and strategies used by the research community to secure the privacy and confidentiality of personal information. Approaches vary among physical, technological and organizational safeguards.

A. Technological Safeguards

In many of the case studies, technological safeguards were used to limit or control researcher access to identifiable data. Researchers used denominalized or non-nominal data where direct identifiers, such as names and addresses, were removed and replaced with scrambled code numbers or encrypted identifiers. In virtually all these cases, the data were denominalized by the original data custodian or independent third party so that researchers did not have access to any names and addresses – nor did they have any reason or desire to access these. In denominalized or non-nominal data, each individual can still be indirectly identified by creating a unique profile or record, based on other information in that record. Variables at high risk of indirectly identifying a unique record, when used in combination, include: date of birth, sex, ethnic origin, or presence of a relatively rare health condition. Other variables that carry the potential for indirectly identifying an individual on their own include: health insurance number and social insurance number. Re-identification may occur when this information is then combined with another data source that contains this information in addition to directly identifying information. However, the ease with which re-identification is possible can vary with the number of variables in the record and the capacity of those variables to identify a unique individual. For example, date of birth carries a relatively high risk of identifying an individual, year of birth a lesser risk, age somewhat less, and age category even less. Thus any record, apart from aggregated data, carries some risk of indirect identification, which must be estimated at the time of release of that information. This underlines the importance of the principle of collecting only as much information as is necessary to answer the study question, of user agreements limiting the additional uses of the information obtained, and the use of safeguards applied to information, even if direct identifiers have been removed.

Some of the cases illustrate unique and innovative technological approaches for masking the identity of research subjects from the researchers. For example, in the randomized drug policy trial designed to evaluate a Nebulizer-to-Inhaler conversion program (case #12), researchers used a camouflage technique to blind themselves to the health status of eligible research subjects. More specifically, the study involved sending a questionnaire to all patients potentially affected by the new policy. In order to produce a mailing list without violating privacy and confidentiality, researchers obtained from the provincial drug plan, names and addresses of patients potentially affected by the policy (80%), mixed in with names and addresses of patients not affected by the policy (20%). The questionnaire was mailed out to all of the people on the list, and only those who were affected by the policy and who agreed to respond returned the questionnaire to the research team. This camouflage technique allowed the research team to target the relevant study population without knowing who they were until they had consented to participate in the policy trial.

In some cases, however, researchers required nominal information for the purpose of identifying eligible research subjects who could then be contacted in order to seek their consent to be enrolled in the study. In other cases, researchers required nominal information to link data about individuals in cases where the linkage could not be conducted by the data custodians themselves nor any suitable third party. In such situations, various technological means can be used to protect

identifiable data while in the possession of researchers, including: individual authentication of users through unique log-on I.D.s; regular review of audit logs to detect any inappropriate access to sensitive information; special protection for remote electronic access and external communications; virus-checking programs and disaster recovery safeguards such as regular back-ups; and removal of all direct identifiers at the earliest possible opportunity in the study.

B. Physical Safeguards

The case studies also demonstrate a range of physical security measures that can be used to protect data holdings. These include housing servers and computers that contain protected data in secure settings that are physically inaccessible to all but legitimate staff for legitimate purposes. Architectural designs have been used to preclude public access to areas of research space where sensitive data are housed. Automatically locking doors and other security measures such as routine monitoring by a surveillance system have also been used to provide physical security to protect sensitive data. Special physical security measures can also protect data from hazards such as floods and fires.

C. Organizational Safeguards

Many of the research centers have implemented specific processes to make their organizations "privacy sensitive". Approaches to achieve this are variable, but may include: commitment to privacy and continued emphasis of its importance by management; development of privacy programs and implementation of security policies and procedures; regular staff training and education programs for newcomers and continuing employees; appointment of privacy officers and creation of security and confidentiality committees; development of regular self-audits and external privacy reviews. While such measures could be achieved in nearly any research setting, small research teams may lack sufficient financial and staff resources to implement them; therefore, an important and essential safeguard is that most organizations require their research staff to sign oaths of confidentiality as an essential condition for employment and can use violation of privacy as grounds for dismissal.

3. Review/Oversight Mechanisms

All of the case studies were subject to prior review and approval. The oversight mechanisms range from: research ethics boards, peer review panels, internal privacy committees of data custodians, special privacy committees created by statute and privacy commissioners. These oversight mechanisms vary in terms of ethical expertise, scientific expertise and privacy expertise; they also vary in their degree of openness, transparency and public accountability and their ability to monitor compliance and audit practices over time. Some case studies were reviewed and approved by several oversight bodies depending on where the research was
conducted. The sheer multiplicity of reviews, the associated time delays, the resources required to prepare several applications according to varying criteria, the unpredictability of outcomes and the potential for contradictory outcomes pose particular challenges for researchers.

In all but a few of the cases, studies were reviewed by institutional research ethics boards (REBs) in accordance with the ethical guidelines set out in the *Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans*. These are multi-disciplinary boards, composed of at least two members with expertise in the area of research under review; at least one member

knowledgeable in ethics; at least one member knowledgeable in law; and at least one community member. REB review is mandatory for research funded by the federal granting agencies and generally required for all research conducted under the auspices of universities, affiliated teaching hospitals and research centers. REBs review research protocols according to a coherent ethical framework, which includes principles for the respect for privacy and confidentiality.

Peer review for scientific merit was also required for many of the case studies. The potential benefits of research depend on the attainment of high scientific and scholarly standards – poorly designed or executed studies do not yield reliable information, are wasteful of public resources and, to the extent they involve human participants, undermine human dignity and worth. Evaluation of scientific merit and the qualification of investigators are rigorously carried out by expert peer review panels of federal and provincial research granting agencies as a fundamental condition for funding.

A number of data custodians also required some form of internal privacy or confidentiality review before releasing data to researchers. For example, the British Columbia Ministry of Health and B.C.'s PharmaNet require prior review and approval by their internal Data Access Committees before releasing personal information to researchers. Saskatchewan Health requires review and approval by its Cross-Agency Study Committee for the case study involving linkage of databases. The Nova Scotia Department of Health and the Nova Scotia Population Health Research Unit similarly requires internal privacy review and approval before releasing data.

In some jurisdictions, specialized bodies created by statute are charged with overseeing the collection, use and disclosure of personal health information by data trustees. For example, the Manitoba Health Information Privacy Committee established under section 59 of Manitoba's *Personal health information Act* (S.M. 1997, c. P-33.5) is responsible for reviewing requests for access to personal health information for research (and other) purposes according to specific criteria set out in the Act.

Finally, in other jurisdictions, such as Quebec, requests for access to personal information for research or statistical purposes must be first reviewed and approved by the Commission d'accès à l'information in accordance with the same specific criteria set out in both the *Act respecting Access to Documents Held by Public Bodies and the Protection of Personal information* (R.S.Q. c. A-2.1, s. 125), as well as the *Act respecting the Protection of Personal information in the Private Sector* (R.S.Q. c. P-39.1, s. 21). Where such requests are made by professionals for study, teaching or research purposes in the context of hospitals or other health care institutions, the statutory discretion of review and approval is delegated to the Director of Professional Services or the Executive Director of the institution (*Act respecting Health Services and Social Services*, R.S.Q. c. S-4.2, s. 19.1).

Conclusion

This analysis has addressed a number of issues that arise from the attached case studies. It has attempted to highlight a variety of strategies that researchers use to protect the confidentiality of research subjects and ensure ethical use of secondary data. However, it is only a start. The objective of the Working Group on Case Studies has been to provide concrete examples that will enhance the ability of researchers, policy makers, consumers and privacy advocates to engage in meaningful and constructive dialogue on the appropriate use of secondary data. The purpose of this analysis is to stimulate more informed discussion about the underlying values at stake.

CASE STUDY # 2

TITLE
Seasonal Patterns of Winnipeg Hospital Use, completed in 1999 by the Manitoba Centre for Health Policy and Evaluation, University of Manitoba, and conducted under contract with Manitoba Health.

RATIONALE
There are repeated crisis periods in emergency rooms and waits for hospital beds during flu season. The researchers wanted to estimate the extent of the problem and suggest ways to avoid these crises in the future.

PURPOSE
The purpose of this study was to examine patterns of use of Winnipeg hospitals focusing on January to April, over several years.

POTENTIAL BENEFITS
If the effects of flu on the healthcare system could be predicted, then it might be possible to better manage the system to deal with outbreaks. In addition, the study could serve to strengthen the drive towards an influenza vaccination program. Given that this study has already been performed, we can observe the following: researchers found that each year there is a small increase in the need for hospital beds because of flu. They also found that these needs can be predicted and that there are ways to avoid the annual crises. This study led to a stronger influenza immunization program in Manitoba. It also identified strategies to better manage use of hospital beds. In the first year after policy changes were made, the annual crises in Manitoba were avoided. Since then, largely as a result of the study findings and the reduced pressures in Manitoba, other provinces that also experience these flu-related problems, including British Columbia and Ontario, followed suit by introducing similar programs.

METHOD
The health information used in this study was not collected specifically for research purposes; it was originally collected by Manitoba Health for routine administrative purposes. However, because of its utility for research, a secure database consisting of only de-identified data has been created by the University of Manitoba for research purposes. Manitoba Health, as original trustees of the data, removes or alters all names and health insurance numbers prior to releasing the data to the University so that individual persons cannot be identified.

Included in the research database at the University of Manitoba are de-identified hospital discharge records for the Province of Manitoba, as well as a de-identified population registry file. For the purpose of the study, the hospital discharge data were linked to the population registry file using scrambled numbers so that only Winnipeg residents could be studied. This de-identified, linked data made it possible to study the number and type of cases in Winnipeg hospitals over time and to determine what proportion of these cases were in fact related to flu.

This specially-linked set of data was created for this research project alone. When the study ended, this specially-linked set of data was destroyed. The computer program that did the linkages was

maintained, however, for the purposes of verifying the research results.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

The study posed a very low informational risk to Manitobans because the focus was on large groups (*i.e.* Winnipeg residents). The researchers did not look at any individual person's health information. Rather, their intention was to look at groups of individuals in order to study overall patterns of use of hospital beds. All information which could identify an individual person was removed or altered by the original trustee (Manitoba Health) in such a way that researchers could use de-identified information to track people's hospitalizations during flu season, but they could not actually identify any of the individuals. Researchers also undertook not to publish results in any manner which might permit certain individuals to be recognized.

In general, several further precautions are taken to protect the security and confidentiality of the research database while it is being used. The research database is maintained in a secure data laboratory at the University of Manitoba. All of the information which is in this special database is kept in an unlinked format. Information is linked using scrambled numbers only for specific projects that have received ethical approval for the linkage. The database is maintained over the long term as an important resource for research.

The scientists who maintain this database use a number of safeguards to protect the information. Access to data is carefully monitored. All staff sign an oath of confidentiality. If they break this oath, they lose their jobs. In the fifteen years, since the creation of this database, there has never been a breach of security. The University of Manitoba also conducts regular security and privacy audits.

LEGAL/ETHICAL ISSUES

This study used potentially sensitive data for the entire population of Manitoba. Persons were not asked individually to consent to letting their health information be used for this research, given the sheer size of the population, the real practical difficulty of contacting individuals to obtain consent, the very low informational risks involved and the important social value and potential benefits of the research.

The study was reviewed and approved by the Research ethics board of the Faculty of Medicine of the University of Manitoba. The Health Information Privacy Committee was also informed about the study and its methods. This Privacy Committee was set up by Manitoba Health to make sure that researchers, like other health data users, follow the principles of the *Manitoba Personal health information Act*. This *Privacy* Committee considers the informational risks of any research in relation to the benefits that might result from the research (see Terms of Reference of the Privacy Committee and its *Guidelines for Assessment of Intrusion into Privacy for Research Purposes*, in Appendix G).

Upon completion of this research study, only the linked data were destroyed. The actual database was not destroyed because of its enormous potential in helping to answer a number of other important health research questions.

If strict interpretations of the privacy law prevented health researchers from making secondary use of this existing administrative data, they could not have carried out this study and would not have been able to describe how flu affects hospitals in Manitoba. New policies could not have been developed to prevent these yearly crises. This research project provided results that clearly served

the public interest, while posing very minimal risks to the persons whose data were studied. Likewise, future health research designed to answer equally important questions is made possible through the maintenance of this valuable database.

PROPOSED PRACTICES

This case study suggests that large repositories of de-identified data can be set up strictly for the conduct of research. Such repositories can have rigorous procedures to protect security and confidentiality but, at the same time, provide researchers with the opportunities to use the data to conduct projects that serve the public interest.

Case Study # 3

TITLE

Assessing the accuracy of the Nova Scotia Health Survey, funded by the Faculty of Graduate Studies and the Population Health Research Unit at Dalhousie University. (Funding for the Nova Scotia Health Survey was from the National Health Research and Development Program and the Nova Scotia Department of Health.)

RATIONALE

Health surveys are increasingly being used to measure health status, guide the allocation of health resources, measure how well the health system performs and, in general help us better understand what determines health. Examples of health surveys are the National Population Health Survey that was conducted by Statistics Canada in 1994, 1996, 1998 and 2000, and the Canadian Heart Health Surveys conducted across the country from 1986 to 1991.

Surveys are used to describe the characteristics of populations (e.g. the proportion of persons who have a disease), but not every person in the population can be surveyed. Instead, researchers use a sample of the population that is large enough to represent the whole population. Normally, these are “random” samples; anyone in the population could potentially be picked to be in the sample. The accuracy of a survey depends on how well the sample represents the population. In a good sample, the percentage of persons in different age groups, the ratio of men and women and the proportion of people living in different areas should be the same in the sample and in the whole population.

While good methods are used to pick random samples of potential participants in health surveys, many of the people selected may not want to participate. Others cannot be located, either because they have moved to another location, or because the researchers are simply unable to make contact with them. Non-respondents often amount to more than 25% of the people selected to be in the sample. If non-respondents are different than those who respond, it can bias the overall results of the health survey. For example, if non-respondents were less healthy than those who do did respond, researchers would underestimate the extent of health problems in the population. This bias is called “non-response bias”, and is among the most important threats to the accuracy of the information obtained from surveys

PURPOSE

The purpose of this research study was to measure the direction and amount of non-response bias in the Nova Scotia Health Survey (NSHS), and develop adjustment factors to correct for the bias. More generally, the purpose was to understand how the accuracy of health surveys is affected by non-response bias.

POTENTIAL BENEFITS

The adjustment factors created by this study can be used in the future to obtain more accurate information from the NSHS about the general health status of Nova Scotians and what health services they use. Little is known about how non-response biases health surveys. Very few

studies have been done on non-response bias in health surveys, either in Canada or other countries. The opportunity for data linkage with the NSHS provided a rare and unique opportunity to obtain valuable knowledge on non-response bias likely to be found in similar health surveys.

METHOD

The NSHS was conducted in 1995 by the Nova Scotia Department of Health and researchers at Dalhousie University. The sample of those to be contacted and asked to participate in the survey was selected, at random, from a database of all Nova Scotians who were registered with Medicare (the "registration file"). The registration file contained peoples' names, addresses, age, sex and health insurance numbers. Those sampled were sent a letter from the Nova Scotia Department of Health explaining the survey, and telling them that a public health nurse would contact them. The public health nurses were unable to locate or contact 15% of the sample; 23% of the sample were contacted by a nurse but declined to participate; and only 61% participated in the survey.

For the purpose of this study (i.e. assessing the accuracy of the NSHS), the health insurance number for each person in the sample (respondents and non-respondents) was used to link to other provincial health databases and obtain summary data on the use of health services and health characteristics. More specifically, this study linked to the hospital discharge abstract database (which contains data on all hospital stays) and the physician claims database (which contains data on almost all visits to a physician). The summary information obtained from the linked files summarized each subject's use of health services (e.g. number of doctor visits, number of hospital visits and the total number of days in hospital) and general health characteristics (e.g. evidence of heart disease or high blood pressure based on diagnostic codes on physician claims or hospital discharge abstracts).

For the analysis, the researchers described how respondents differed from non-respondents in terms of age, sex, region of residence, use of hospital and physician services, and evidence of a selected group of health conditions. After this comparison, the researchers measured the degree of non-response bias and developed "correction factors" that can be used to reduce the non-response bias and obtain more accurate information from the NSHS.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

The potential confidentiality risk is that persons who did not want to participate in the survey might be identified, along with information about their use of health services and their health characteristics. There might also be concerns that information on non-respondents could be retained.

However, because of the design of the study protocol and the review process, it was very unlikely that non-respondents could be identified, that the data could find its way into the hands of unauthorized third parties, or that data on non-respondents would be retained once the study was completed.

To conduct this study, the researchers designed a protocol that included a variety of security measures. The study was able to link health information without the researchers having any access to names, addresses or actual health numbers that could be used to directly identify individuals. The protocol worked as follows:

1. The custodian for the NSHS (i.e. the researcher who originally conducted the NSHS) was asked to prepare a data file containing the NSHS study identifier, and the respondent/non-respondent status of each person in the sample. The custodian obtained this information from an electronic

file that did not contain names or addresses. The custodian maintains names and addresses for NSHS respondents, but they are stored in locked filing cabinets in a secure area separate from the electronic data files.

2. The custodian sent the file containing the NSHS study identifier and the respondent/non-respondent status to the Nova Scotia Department of Health. Using their copy of the NSHS sample list, the Department of Health added and encrypted the health card number, and sent the file to the Population Health Research Unit (PHRU) at Dalhousie University. The PHRU manages provincial health care data for research purposes, and all the databases they manage contain health card numbers that are encrypted using the same procedure. PHRU does not have access to actual health card numbers, names or addresses. Since the Department of Health holds the encryption algorithm, their approval and assistance is required for all linkages of data to the PHRU databases.
3. PHRU linked the summary health care data to the NPHS file, removed the encrypted identifier, and made the file available to the researcher for analysis. The file did not contain either NSHS identifiers or encrypted health card numbers, and included only the variables from the NSHS which were needed for the study, and the summary variables on health services use and health characteristics.
4. Before obtaining access to the data, the researchers were required to sign a confidentiality agreement, and an agreement specifying that sensitive data would not be retained following the study.
5. Once the study was completed, detailed information on non-responders was no longer needed. The file made available to the researcher was returned to PHRU, and all individual level study data in possession of the researcher was destroyed. Only aggregate results summarizing the differences between respondents and non-respondents were retained. However, all computer programs used to create the files and do the analysis were retained should they be needed in the future for verification purposes.

A research proposal, which included a detailed description of the above-mentioned study protocol, was reviewed and approved by the Dalhousie Faculty of Medicine Ethics Review Committee, the Nova Scotia Department of Health, and the population Health Research Unit.

LEGAL/ETHICAL ISSUES

Based on the study design, it was not possible to contact subjects to ask study subjects, especially non-respondents, for consent. Many of the non-respondents were people who could not be contacted after repeated attempts, or for whom current contact information was not available. Moreover, contacting study subjects to ask for consent would have required researchers, or at least a third party, to access identifiable information. The study protocol was specifically designed to prevent researchers from accessing any identifiable information, so as to maintain anonymity throughout the study.

Evidence suggests that if consent for this study had been sought, most of the non-respondents would have agreed to participate, even though they had declined to participate in the survey. In many surveys, persons who are contacted but decline to participate in the survey, are asked if they would answer a few basic questions anyway (e.g whether they smoke and their education level) so that the researchers can learn something about the health characteristics of non-respondents. A clear majority of non-respondents agree to do so.

Data linkage was used to obtain health information about non-respondents, some of whom may have declined to participate in the survey because they did not want to divulge personal information about their health in the first place. Clearly, this study traded some loss of privacy for a public

benefit – knowledge of non-response bias. However, this research study did not collect personal information on non-respondents for inclusion in a database. The loss in privacy resulting from this study was limited in scope and duration as much as possible.

The study protocol provided strong protection of confidentiality. The informational risks to the non-respondents in this case were virtually non-existent. No identifying information was used in the study and once the correction factors were developed, all non-respondent information was completely removed from the research database. In this specific case, the harms and benefits were weighed and considered by the Dalhousie Faculty of Medicine Research Ethics Review Committee, the Nova Scotia Department of Health and the Population Health Research Unit. It was reasonably felt by multiple reviewers that the benefits of proceeding with the study outweighed any inconvenience or harm that might result to non-respondents.

PROPOSED PRACTICES

This study illustrates strategies to link data from different sources while maintaining confidentiality. Only anonymous identifiers (encrypted health numbers) were used in this study, and the Population Health Research Unit at Dalhousie University, that did the linkage, manages health care data for research purposes. In all of the data files at PHRU, identifying information such as names and addresses have been removed. The only identifier is the encrypted health number. The encryption method is held by the provincial health department, and any data linkage requires their permission and assistance. The data is located on a secure computer. Access is limited to a few persons, and all access is carefully monitored.

This study also shows the value to requiring different types of approval. For this study, review and approval was required from the health department, PHRU and a university ethics committee. Multiple levels of accountability help to guard against risks to privacy and confidentiality.

Case Study # 5

TITLE

Use of RFLP molecular epidemiology to find out how tuberculosis is spread among people infected with HIV*, sponsored by the National Health Research and Development Program of Health Canada and conducted by academic researchers affiliated with a Montreal University and its teaching hospitals. The laboratory aspect of the study was conducted in the Public Health Laboratory of Quebec. The data linkage was performed by the Department of Public Health. The investigators performed the data analysis at one of the teaching hospitals. The study was conducted between 1996 and 1999.

RATIONALE

Tuberculosis (TB) has again become a problem in certain parts of Canada. People who are HIV positive are more likely to get infections, including TB. It would seem, therefore, that the presence of HIV in the population might contribute to the spread of TB. If researchers could determine whether in fact HIV infection contributes to the spread of TB, this information may be able to assist policy makers in designing strategies to decrease TB spread in the general population.

PURPOSE

The researchers wanted to learn how various factors like HIV infection affect the spread of TB in the population.

POTENTIAL BENEFITS

A better understanding of how TB spreads could help develop better ways of controlling TB infection in the general population.

METHOD

The Public Health Laboratory in Quebec is a reference laboratory for the testing of blood and tissue samples. Doctors who want to find out whether their patients have a certain infection send the patients samples to this lab. This is the laboratory that did the analysis for this study using RLFP molecular epidemiology to examine TB bacteria. The samples of TB bacteria to be studied were already stored at the Public Health Laboratory and identifiable (i.e. linked with identifying information). There is mandatory reporting of TB in Quebec, but these cases had already been reported.

RFLP molecular epidemiology is a laboratory method that uses parts of chromosomes to identify similar sets of genes. Using this technique, it is possible to look at the TB bacteria in or grown from different samples of sputum (coughed-up material from the lungs) and to determine whether the bacteria which caused one case of TB was likely related to that which caused another case. Usually, TB in adults is caused by reactivation of an old infection. In those cases, the TB

* RFLP stands for Restriction Fragment Length Polymorphism. This describes a technique in which DNA is isolated and cut with a restriction enzyme. The DNA so cut is then separated onto a gel. If a given trait or gene fragment appears as different bands in a given population, it is said to be polymorphic (having more than one form). This technique thus enables researchers to map genes or to follow their passage from one generation to the next.

bacteria of various cases are generally unrelated. However, individuals with a compromised immune system (for instance persons who are HIV positive) more likely become ill from a recent external exposure to the TB bacteria. In those cases, the TB bacteria may be genetically related. Hence, the genetic analysis of the bacteria can provide information on how the disease spreads.

The analysis of each sputum sample by the Public Health Laboratory in Quebec was provided to the Department of Health. Using the personal identifiers (names, date of birth) provided with the results of RFLP analysis, the Department of Health linked these lab results to non identifying demographic information stored within the Department, such as age, sex and residence. Once the Department of Health completed this linkage, all names were removed from the linked information. The linked, but unidentifiable information, was then provided to researchers so that they could study how the TB disease actually spreads.

The researchers did not seek consent to conduct this study. Many of the sputum samples were sent to the lab years ago. Since then, many of the patients had relocated and could not be easily contacted. Even if they could be contacted, the researchers would need to know the names of the patients in order to trace them and seek their consent. Yet, the study was precisely designed in such a way as to avoid releasing identifying information to the researchers. It was the Department of Health that was responsible for linking the lab results and the individual demographic data and then removing the identifiers before releasing the linked information to the researchers. This way, the researchers would not be able to identify individuals with the information that was provided to them.

The investigators had their study reviewed and approved by the Research ethics board of their teaching hospital. The committee considered that the privacy and confidentiality of the research subjects would be well protected and that the potential benefits far outweighed any possible harm. However, the REB did have concerns about the Quebec laws that might apply.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

The main possible risk was breach of confidentiality. In theory, there could be concern that a researcher might be able to find out about the health information of a particular person. However, the Public Health Laboratory provided information without names so that the researchers could not do so.

LEGAL/ETHICAL ISSUES

The Quebec Civil Code requires written consent for the use of a part of the body for research.

> **Art. 22.** *A part of the body, whether an organ, tissue or other substance, removed from a person as part of the care he receives may, with his consent or that of the person qualified to give consent for him, be used for purposes of research.*
>
> **Art. 24.** *Consent to care not required by a person's state of health, to the alienation of a part of a person's body, or to an experiment shall be given in writing. It may be withdrawn at any time, even verbally.*

One important issue for the researchers and for the REB was whether bacteria grown from sputum samples of TB patients were *'part of their body'* so that using the samples would require written consent from the subjects.

PROPOSED PRACTICES

The Research ethics board consulted various experts. Some experts felt that the bacteria grown from the sputum samples should not be considered part of the body, because the samples were something grown, developed after the TB patients gave their sputum sample. This research material was thus considered as not being part of the body of the patients. Thus, article 22 of the Civil Code was not applicable to this project, and written consent for the patients was considered not required. The Quebec Ministry of Health was told of the expert opinion but Ministry officials did not officially reply to the researchers.

The REB's interpretation of the law was sufficiently flexible in this case to allow the research to be carried out as planned. However, this is an example of how legislation intended to protect research subjects might inadvertently interfere with important and legitimate research projects that have already integrated as part of their design, adequate ways of protecting research subjects by removing any possible way of identifying them.

Case Study # 6

Please note: tissue and formalities around informed consent (i.e written consent vs implied consent) raise more issues than data in other forms. We are working on those issues, and will have a more complete analysis in the final version.

TITLE

HIV seroprevalence among women undergoing abortion in Montreal, conducted from 1993 to 2000 by academic researchers affiliated with a Montreal University and its teaching hospitals; sponsored by the Laboratory Centre for Disease Control of Health Canada.

RATIONALE

There is a continuing need to carefully monitor the frequency of HIV infection in the population. In part, this can be done by examining the frequency of infection in smaller groups that can be more easily accessed and tested than the general population at large.

PURPOSE

The purpose of this study was to survey and test women undergoing therapeutic abortions in order to assess the frequency of HIV infection in this particular group.

POTENTIAL BENEFITS

Current information about the incidence of HIV infection and certain practices can help inform and better guide educational programs and policies aimed at preventing infection. Moreover, such information may help indicate how certain risk factors for infection may have changed. If public educational programs or policies are not working or if the importance of certain risk factors for infection have changed over time, strategies can be adapted to help better prevent HIV infection.

METHOD

One of the groups of patients who necessarily have to have a blood test before they are treated are women undergoing therapeutic abortions. Researchers approached these women in the clinic of a large teaching hospital. Researchers sought their consent to participate in the study. Those who agreed to participate were asked to fill out questionnaires about certain risk factors for HIV infection. Also, with their permission, leftover blood from the blood test these women would be undergoing in any event as part of their treatment was used to test for HIV infection. For each participant, a computer generated a specific scrambled code identifying the blood sample for the HIV serology and the answers to the questionnaire. Once the results of the HIV serologies were linked to the corresponding questionnaire, the computer-generated code was removed. This way, it was not possible to identify the research subjects, even if one had used the same computer program to try to retrace the scrambled codes. The linked information for each person was thus completely anonymized so that the researchers could look at risk factors and determine the incidence of HIV infection but could not identify any of the research subjects.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

This study started before the Quebec Civil Code was reformed in 1994. The Research ethics board of the teaching hospital worried that the results of the HIV test might be traced back to individual women in the study. But they agreed that replacing names with code numbers for linkage purposes would effectively prevent this risk.

There was also the risk that, sometime in the future, the women might have to declare in an insurance contract or job application that they had once been tested for HIV and that this might cause them prejudice even though the test was conducted for research purposes and even though the women could never know the results of the test. The REB required that this risk be disclosed to the women at the beginning of the study.

LEGAL/ETHICAL ISSUES

Before 1994, women were informed about the study both verbally and through posters in the clinic waiting area. Every woman was given a clear choice to decline to be part of the study. Those who did agree to participate, did so through verbal consent. Almost all of the women approached - 99.6%. - agreed to be in the study.

Since 1994, the revised Quebec Civil Code required that consent to donate blood or tissue for research purposes be given in writing.

> **Art. 22.** *A part of the body, whether an organ, tissue or other substance, removed from a person as part of the care he receives may, with his consent or that of the person qualified to give consent for him, be used for purposes of research.*
>
> **Art. 24.** *Consent to care not required by a person's state of health, to the alienation of a part of a person's body, or to an experiment shall be given in writing. It may be withdrawn at any time, even verbally.*

Since this change in the law, the REB required that written consent be obtained before women could participate in the research study. With this new requirement for written consent, the participation rate dropped from 99.6% to only 90%. Women who would have been willing to participate under conditions of complete anonymity, now hesitated to sign their name on a written document that would record the fact that they actually had participated in a HIV study. Ironically therefore, a legal rule that was designed to better protect research subjects (ie. the requirement for written consent) in this case afforded research subjects with less protection and actually put them at greater risk of being personally identified.

The frequency of HIV was very low in the study population (about 0.2%). However, the investigators and sponsor of the study expressed concern that women more likely to test positive for HIV might have been even less willing to sign a written consent form and therefore, might have declined to participate in the study altogether. If this was indeed the case, the requirement for written consent was jeopardizing the scientific validity of the study because the researchers may not have had a completely accurate picture of the frequency of HIV in this population of women, let alone the general population.

Moreover, this study was but one arm of a broader national initiative. It exemplifies how lack of harmonious standards across the country can potentially impact the generalizability of research which is designed to be national in scope.

PROPOSED PRACTICES

In this case, the requirement for a written informed consent actually turned out to be in conflict with the women's right to privacy and confidentiality. The women would have preferred not to sign a document that they felt might be seen by others and might link them with a HIV study. In the specific circumstances of this case, verbal consent might have been more desirable for all concerned, but after 1994, was not permitted under Quebec law.

Case Study #12

TITLE

*A Randomized Drug Policy Trial with Camouflaged Contacting of Patients, s*ponsored by the Canadian Health Services *Research* Foundation, Health Transition Fund, British Columbia Pharmacare and the US National Institute on Aging. The study occurred between 1999 and 2001, and was conducted at the UBC Faculty of Pharmaceutical Sciences.

RATIONALE

In 1999, British Columbia Pharmacare, the publicly funded drug insurance program within the provincial Ministry of Health, introduced a Nebulizer-to-Inhaler Conversion policy. Under this policy, Pharmacare stopped covering the costs of respiratory medications that used nebulizers to deliver the drug to the lungs and encouraged doctors to switch patients to the same medications delivered using metered-dose inhaler devices. If the doctor requested, patients with special clinical needs were granted exemptions to the policy.

PURPOSE

To measure the intended and unintended impacts of Pharmacare's Nebulizer-to-Inhaler Conversion Program on health care utilisation and quality of life.

POTENTIAL BENEFITS

Patients on nebulized medications funded by drug benefit programs in other provinces that are considering implementing a similar program may benefit if unintended, avoidable impacts of the policy are found in some subgroups. Findings might enable better versions of the policy to be designed elsewhere.

METHOD

To evaluate the intended effects and possible unintended impacts of the policy, Pharmacare agreed to grant 10% of physicians in the province an optional 6-month exemption from the new policy. Thus the policy was implemented with a randomized control group, like a scientific experiment, although the policy itself was not considered experimental.

The policy was not considered experimental because Pharmacare had already decided, after consultation with clinical experts, that the policy should be applied to all Pharmacare clients, with exemptions granted only to certain types of patients. The purpose of the optional six-month delay was simply to evaluate the impact of that policy.

With the approval of the Data Access Committee of the B.C. Ministry of Health, the principal investigator at UBC and co-investigator at Harvard were given anonymous data on the health care utilization (prescription drugs, medical services, hospitalizations, long term care) and mortality (from 1997 through mid 2000) of all Pharmacare clients affected or potentially affected by the policy. The data were already linked by personal health numbers (PHNs) at the Ministry of Health and then replaced by unique study numbers before the data were released to the researchers for study and comparison.

Furthermore, to measure any change in quality of life of patients immediately affected by the policy compared with those in the control group, questionnaires were sent to all patients potentially affected by the policy. To assemble a mailing list for this purpose without violating the privacy of patients' Pharmacare data, the Ministry of Health produced a "camouflaged" list of patients as follows. The scrambled PHNs of all patients potentially affected by the policy were combined with scrambled PHNs of a random sample of Pharmacare clients who were not affected by the policy. When the resulting list of scrambled PHNs was unscrambled and converted to names, addresses and telephone numbers by the Ministry of Health's Client Registry, the health status of each patient remained unknown to Pharmacare or the researchers.

To complete the camouflaging, a general survey applicable to any Pharmacare client was also included in the mailing. Covering letters from both Pharmacare and the principal investigator explained that the Pharmacare computer had selected the patient from either a random list or a special list of patients, and therefore their health status was unknown to Pharmacare and the researchers. Any patient who did not wish to complete the questionnaire could merely decline to respond. The only way that researchers would learn anything about the health status of patients on the mailing list was if those patients voluntarily completed and returned the anonymous questionnaire. The questionnaires were mailed to one of the researchers. A majority of patients agreed to have their quality of life linked with health care utilization data and supplied their PHNs for this purpose. When analyses showed no difference in quality of life between treatment and control groups, the researchers decided such linkage of data was not needed after all. However, if linkage had been done, it would have been done using the PHNs supplied by the participants. The patient's health status and whether they participated in the study remained unknown to Pharmacare.

The initial grants ended in March 2001. However, additional funds to compare alternative control groups were obtained by colleagues at Harvard University from the US National Institute on Aging. This justified extending the duration of data retention and providing the Harvard investigators with these anonymous data in which the scrambled personal health numbers were replaced by study IDs.

The study was approved by the UBC ethics committee, which has approved camouflaged contacting for several other studies. The privacy branch of the Ministry of Health also approved of camouflaged contacting.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

There was no risk to patients of their health status becoming known without their permission. However, some patients were at risk of *feeling* that the privacy of their health data had been violated, even if it had not, given that the data remained anonymous unless the individuals agreed to complete and return the quality of life questionnaires.

In order to be most effective, camouflaging should aim to protect the privacy of targeted patients, while limiting the number of untargeted patients who need to be contacted. For instance, if the targeted patients constitute only 1% of the total population, then 99% of the random sample would be camouflaged. 99% camouflage would be very inefficient. In this study, the sample was 80% targeted and 20% camouflage. This proportion successfully preserved the privacy of the targeted people's data, yet greatly reduced the number of people to be contacted.

LEGAL/ETHICAL ISSUES

Participation in the delayed control group

Participation in the delayed control group was optional for physicians. They could withdraw from the control group by complying with the policy 6 months sooner than they were required to. They could also contact Pharmacare or UBC and ask to have their patients withdrawn from the analysis. (Only one out of approximately 600 control physicians asked for his patients to be excluded.) In this way, the physicians did give their consent, according to a negative consent model. That is to say, they were informed of the study and given the opportunity to withdraw from the group at any time.

Moreover, physicians were invited to inform their patients about the study. However, the informed consent of patients was not sought before including them in the delayed control group. This was justifiable because governments do not generally seek informed consent from a population before implementing a new policy. The only novelty about Pharmacare's policy change was that it involved designed delays that allowed for a more rigorous evaluation between those patients to whom the policy applied immediately and those patients to whom the policy applied six months later.

Contacting of patients for the quality-of-life questionnaires

When a Ministry responds to a citizen's request for aggregate data, individual health records are "used" by the Ministry computer without the permission of those patients. When a camouflaged mailing list is produced, the same records are "used" by the same computer. Therefore, if the first use is considered routinely permissible, the latter should likewise be possible, i.e., it should be permissible to "use" patients' health records to create a camouflaged mailing list so that patients may be contacted to seek their permission to respond to a questionnaire. What constitutes an adequate degree of camouflage is an important question. Does 20% camouflage provide sufficient protection of privacy? Should it be 50% or 80% in more sensitive circumstances?

PROPOSED PRACTICES

Privacy legislation should permit the anonymous use of health care data and camouflaged contacting of patients for health care evaluation research.

Evaluations of policies with well-designed delays (control groups) should not be held to higher standards of data access than evaluations of policies with undesigned delays. Some people initially assume that, because there is a randomized control group, this necessarily means the policy evaluation automatically should be held to the same standards of data protection that normally apply to randomized clinical trials of experimental treatments. After careful consideration of the circumstances, people generally agree that well-designed delays do not affect data privacy concerns. Therefore, whatever standards of data access apply to ordinary studies of policy impacts should also apply to randomized policy trials.

Case Study #13

TITLE
Cancer and Other Health Problems Associated With Breast Implants.
Conducted by Cancer Care Ontario (Ontario) and Laval University (Quebec) 1995-present
Sponsor: Health Canada

RATIONALE
The use of silicone gel-filled breast implants was stopped in 1992 because of reports of health problems. Many Canadian women remain concerned about the long-term health effects of these medical devices. This study was developed to look at these long-term effects.

PURPOSE
The purpose of this long-term follow-up study is to identify harmful health effects about a large number of women who received breast implants for cosmetic reasons over the period of 1975 to 1989, compared with a group of women who had other kinds of cosmetic surgery.

POTENTIAL BENEFITS
This study may or may not find that there are certain health problems associated with breast implants. Either finding will be important knowledge for women and their doctors.

METHOD
Most breast implantation in Canada was done in Ontario and Quebec. The investigators reviewed surgical records in these two provinces to identify approximately 25,000 women who received breast implants for cosmetic reasons. In Quebec, these records were reviewed in public hospitals where the surgery was performed, and in private clinics, with prior authorization from the Quebec Access to Information Commission and the Directors of Professional/Medical Services in each institution. In Ontario, the records of plastic surgeons were reviewed in their private practices, following written authorization from each surgeon. In both provinces, the research teams pledged that there would be no direct contact with the research participants and that all sensitive information would be kept confidential. Trained health record technicians were employed by each investigator to abstract personal information from these records. This information included patient names, birth dates, and health numbers , as well as relevant surgical and medical details. These pieces of information were entered directly into a database file using laptop computers. The file was protected by a password known only to the individual abstractor. On a daily basis, these records were transmitted by modem from the laptop being used at the sites where the records were abstracted to central database files in the two study centres in Quebec City and Toronto. All of the records were password-protected and encrypted before they were sent so that unauthorized access could not identify specific persons. And after successful completion of transmission, these patient records were deleted from the laptops. At the study centers, all patient records were separated into two files, one holding only patient identifiers, and the other only surgical and medical details, with a sequential study number serving as the common key.

The women who were in this study had to be identified by going from hospital to hospital and clinic to clinic. This method was necessary because there is no central registry of this information. The investigators did not contact the women. Early pilot work indicated that three-quarters of the

participants were no longer residing at the address recorded in the surgical records. It would not have been practical for either the investigators, or the surgeons, to contact these subjects for informed consent.

Only the patient identifiers needed for record linkage (e.g. names, addresses, birthdates,outcome dates, if known) were sent to Statistics Canada; no information about the surgery itself was sent. This allowed Statistics Canada to undertake accurate linkage with any cancer registrations or death notices that occurred after the surgery. Statistics Canada holds the databases for newly diagnosed cancers and for all deaths in Canada. When this linkage is finished, the cancers and death records linked to the identifiers of the surgical records will be sent back to the investigators with the permission of each provincial cancer registry and each provincial registrar of vital statistics. These provincial agencies are the primary data custodians and retain the right to review all proposals for disclosure, even though they routinely transmit these cancer and death records to Statistics Canada, in compliance with the Federal Statistics Act.

In the central database at each study centre, only two people can look at personal information - the study coordinator and the person who provides the technical help for the database. Both of these staff are directly supervised by the Principal Investigators and are in the employ of CCO and a Laval University-affiliated hospital research center. The first analysis of the data will be done later in 2001 with the information de-identified and pooled from both study centres. The pooled file will have sequential study numbers for each patient and coded numbers for surgeons so that they cannot be identified. The deidentified (no names, complete dates, addresses, incl. Postal codes, phone nos.), participant-specific file will reside at Health Canada, in accordance with the Research Contracts. The Principal Investigators will oversee the analysis, which will initially consist of basic tabulations, and person-years statistical analysis, and then progress to multivariate analysis, including Poisson regression and proportional hazards regression.

Prior to commencement of this study, external peer review was undertaken by three experienced scientists. Ethics review was undertaken at a university-based health sciences Research ethics board (REB) in each province. The REBs accepted the impracticality of collecting informed consent from the participants but recommended that a general educational program publicizing the study be developed along with a toll-free telephone hotline for further information. The REB also recommended that all women who do not wish their records to be included in the study should be given the opportunity to opt-out of the research by telephoning the hotline.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

The primary hazard to women participating in this research relates to harmful disclosure of sensitive, personal information. Information relating to breast implants may be so sensitive that women may choose not to inform spouses or other family members, or other health professionals.

Safeguards taken to minimize the risk of harmful disclosure include personnel practices, database security measures and physical security measures. All staff employed to abstract surgical records were trained health record technicians who swore an oath of confidentiality in relation to this study. This oath reminds all staff that any unauthorized use or disclosure of sensitive information is grounds for immediate dismissal. The physical measures taken included the transportation of laptop
computers and other materials in locked cases and the location of the central study office in secure quarters behind locked doors. In terms of database measures, personal identifiers were separated from surgical or medical information, and passwords used to gain access to the identifying information were known only to two staff in each of the two research centres in Toronto and Quebec

City.

LEGAL/ETHICAL ISSUES

Within the Province of Quebec, the existence of privacy legislation regulating access to the hospital records of interest required a formal application to the Provincial Commission on Access to Information, primarily because a complete listing of relevant surgeries existed in the central Med-Echo database held by the Government of Quebec. This database was used to identify all potentially eligible participants and the location of their surgical records. Subsequently, permission to review and abstract from the original surgical records was obtained from the Director of Professional/Medical Services at each public hospital.

In Ontario, the required information was not indexed in a central government database, or even at the province's hospitals. The only feasible means of identifying eligible participants was through review of the surgical records held in each plastic surgeon's office. In Ontario, a personal health information statute did not yet exist regulating access to records held by private practitioners However, under the *Ontario Medicine Act*, a new regulation indicated that physicians had the discretion to decide whether to disclose sensitive patient information in instances where (1)consent would be extremely difficult, (2)the research purpose was valid and (3)adequate safeguards would be taken by the researchers to protect confidentiality.

PROPOSED PRACTICES

In order to receive the cooperation of women, surgeons and hospitals, a general informational program was initiated, publicizing the study aims and methodology at professional meetings, with women's interest groups and through lay and scientific periodicals and newspapers. Informational pamphlets were distributed to approximately 35,000 licensed physicians across Canada, for display in the patient reception areas of their offices . A bilingual hot-line was established, with a dedicated toll-free phone number, in order to provide more specific information to interested women. For women who asked that their records not be included, their records were removed from the database. Of approximately 25,000 women recruited for this study, there were approximately 20 who asked that their records be removed (<1 in 1000).

For individual plastic surgeons in private practice who had eligible subjects, the letter requesting access to records included a summary protocol and a list of the relevant data elements to be abstracted. The investigators pledged no further disclosure of patient information than what was described in the protocol. Any future research projects would require additional consents from the surgeons, as well as REB approval. The identifiers will remain in a secure location until approximately 20 years of follow-up can be completed for most of the subjects.

Case Study #16

TITLE

Rapid Surveillance of Cancer in Neighbourhoods Near Point Sources of Pollution.
Conducted by Cancer Care Ontario, in collaboration with the Durham Regional Health Department.
Funded by Health Canada and the Canadian Nuclear Safety Commission. 2001-2002.

RATIONALE

There are many documented examples of community concerns about potential health hazards associated with residential proximity to point sources of pollution. Notable examples include proximity to nuclear reactors, to metal smelters and foundries, chemical contamination of drinking water, and industrial pollution in general. Existing health outcome reports based on cross-sectional mortality or morbidity statistics have a number of flaws, most notably inability to control for residential mobility, inaccurate residence information, inability to control for known risk factors, and inability to take into account the long latency from initial exposure to associated outcomes, at least for chronic diseases such as cancer. Communities are distrustful of existing statistical reports and desire better surveillance systems to alert them if significant hazards exist, or reassure them if existing health outcomes are unlikely to be associated with the exposures of concern.

PURPOSE

The purpose of this research is to design, develop and test a computerized surveillance system to rapidly assess the relationship between residential proximity to real or perceived point sources of pollution, and subsequent risk of cancer. This new system will be pilot tested in the Durham Region of Ontario, where the question about the relation between neighbourhood proximity to the Pickering Nuclear Reactor and subsequent risk of cancer will be addressed.

POTENTIAL BENEFITS

The pilot study may or may not find that there are certain cancers associated with living close to a perceived point source of pollution. Either finding will be important knowledge for residents and stakeholder groups, particularly if the knowledge can be generated quickly.

METHOD

A longitudinal cohort design is most appropriate for periodically estimating the risk of chronic diseases, such as cancer, among humans who are exposed to real or perceived external exposures. The choice of an historical, rather than prospective, cohort (i.e., a group of persons with something in common) shortens the start-up time of the surveillance program, which is a major issue given the long lag that usually exists between exposure and outcome for most chronic diseases, including cancer (i.e. 10-30 years).

In order to overcome the problems of incompleteness, inaccuracy and imprecision of residential information on existing health records (e.g. CIHI hospitalization abstracts; death records), the Provincial Property Assessment File will be utilized to identify, with a high level of accuracy and completeness, physical location of the usual place of residence of each inhabitant of Ontario. This file is prepared on an annual basis within each municipality, and eventually forwarded to the Ontario Ministry of Revenue where it is amalgamated to form a province-wide file. Completeness of this file

is almost as high as the Canadian census, and it has sufficient identifiers to permit fast, accurate and cost-effective linkage to Ontario-wide health files, including the Ontario Cancer Registry and the Ontario Mortality Database. As well, this file exists back as far as the early 1980s. This permits identification of the place of residence of Ontarians 15 to 20 years ago. And this ensures that sufficient latency will have occurred in order for researchers to conduct a sensible assessment of the relationship between residential exposure to pollution and subsequent risk of chronic disease. Under the Ontario *Freedom of Information and Protection of Privacy Act*, a research application for access to these files, including personal identifiers necessary for record linkage (e.g. names, complete birthdates), will be sent to the Ministry of Revenue. The investigators have previously received copies of these files for certain years for similar research purposes, and they do not anticipate much difficulty with this new application.

For the first phase of this project, a feasibility study will be undertaken in the Durham Region of Ontario, which has a population of approximately 0.5 million inhabitants, and includes the Pickering Nuclear Reactor, a notable site of concern to many of the residents of the Region.

In order for such a surveillance system to be effective, it must be both rapid and accurate. Given the need for speed, and cost efficiency, the use of existing computerized files and record linkage techniques would seem the most viable option for developing a surveillance system that represents a significant enhancement over traditional cross-sectional studies.

It is proposed that the feasibility study be undertaken at Cancer Care Ontario's (CCO's) Provincial Office, which is the home of the Ontario Cancer Registry (OCR). An electronic copy of the Ontario Mortality Database also resides at CCO, to facilitate its cancer research. Additionally, expertise in linking computerized files in a secure fashion exists at CCO. And, again, CCO has considerable experience working with the Ontario Property Assessment File in relation to specific cancer research projects.

Because a unique personal identifier does not exist for all persons on these files, it will be necessary to use whatever discriminating personal information is available, including surnames, given names, birth dates, ages, sex, and residence. Probabilistic techniques are used in order to determine whether pairs of records likely describe the same person, or different people. Probabilistic techniques are necessary because of incompleteness, errors and truncations in many of these variables. Additionally, computerized techniques of record linkage minimize drastically the amount of human viewing that otherwise would be required to accurately link records across several files.

An additional scientific enhancement in this project will be to take into account the possible confounding effects of known risk factors (e.g. smoking) and socio-demographic factors (e.g. poverty) as they may also affect the risk of various cancers. This information will be available at the neighbourhood level, from community surveys and the national censuses. This neighbourhood level information is publicly available and may either be purchased from Statistics Canada, or accessed via the University of Toronto Data Liberation Initiative.

Ultimately, if the feasibility study proves successful, then a proposal will be prepared to design, develop and maintain an on-going surveillance system that will permit the linkage and rapid assessment of cancer risk for all residents in the province, back to the early 1980s. As is currently done with the OCR, these files will be maintained, linked and summarized on a dedicated computer, with password protection and other security measures consistent with CCO's UNIX environment and ORACLE relational databases. Only four named permanent staff at CCO will have access to the linked records of named individuals in this surveillance system.

The protocol for the feasibility study, which is still under development, will be submitted to the Office of Research Studies at the University of Toronto for ethics review. Further, this protocol will also be reviewed for scientific merit by three external reviewers with known expertise in this area of geographic surveillance.

Finally, all results emanating from the system will be carefully reviewed prior to publication or wider dissemination in order to be assured of no residual risk of disclosure. Eventually, dissemination will be undertaken through multiple vehicles at the national, provincial and local levels.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

The most notable risk relates to the risk of harming individuals if sensitive medical information about them is disclosed without their consent. This information may be embarrassing, distressing or possibly even discriminatory. Additionally, there is some risk that publicity about the surveillance system and ensuing findings may be harmful or discriminatory to the group of residents living in the area of concern. Worries may exist about falling or stagnant property values or low community morale.

While it is likely that better information about the possible association between health and point sources of pollution will be reassuring to communities, or at least lead to better remedial action, the risk of harming communities should not be discounted. While individual informed consent will be virtually impossible to solicit, given the large sample size necessary (e.g. approximately 100,000 individuals) and the long latency of historical residential information that is required, it will be beneficial to do a priori qualitative studies, utilizing focus groups representing individual citizens, interest groups, government agencies, other stakeholders and cancer patients, to better elucidate community concerns and interests. These focus groups will be interviewed prior to initiating the pilot study. And they will be facilitated by an independent behavioural scientist who will be identified through a competitive RFA process.

LEGAL/ETHICAL ISSUES

Access to the required files falls under several statutes in Ontario, the most notable being the *Cancer Act of Ontario*, which regulates access to the Ontario Cancer Registry, and the *Freedom of Information and Protection of Privacy Act*, which regulates access to all files held in the custody of the provincial government, including the Provincial Property Assessment File and the Ontario Mortality Database.

Under the *Cancer Act*, Cancer Care Ontario has a statutory mandate to create the Ontario Cancer Registry and to maintain it for research purposes. The Cancer Act states clearly that sensitive information about cancer patients must be kept confidential and can only be used for statistical purposes or for epidemiologic and medical research. Applications for access to the OCR for the purpose of record linkage research studies require a detailed protocol and review and approval by an appropriately constituted research ethics body. The work of linking a research file to the OCR is conducted at CCO, by only two employees. Generally, files that are released back to researchers, whether internal or external to CCO, have been deidentified in order to protect confidentiality, but permit researchers to complete their analysis.

Under the Ontario *Freedom of Information and Protection of Privacy Act* (FIPPA), disclosure of sensitive information without individual consent is permitted for research purposes so long as there

is evidence that the research cannot be undertaken without this disclosure and that adequate safeguards are taken by the researchers to protect the confidentiality of this information. A formal application for access to these files is necessary which must be reviewed and approved within the appropriate Ministry of the Ontario government. For example, access to the Provincial Property Assessment File falls under the Ministry of Revenue, whereas access to the Ontario Mortality Database falls under the Ministry of Community and Commercial Relations. In the event that the pilot study is successful, then a full fledged on-going surveillance system will require formal agreements between CCO and the Government of Ontario to facilitate the regular transfer of necessary files and on-going surveillance activities.

PROPOSED PRACTICES

It is likely that explicit agreements will eventually be necessary between CCO and the Government of Ontario in order to permit the proposed surveillance system to operate.

These agreements, which will have to be consistent with current and new legislation (e.g. the proposed Ontario *Health Information Privacy Act*), will more explicitly identify necessary data elements, agreed-upon formats, periodicity of reporting and security standards.

Further, it will be necessary to develop methods of reporting the findings of this surveillance system utilizing graphical tools that facilitate understanding, while minimizing any residual risk of disclosure. New mapping tools, including appropriate statistical smoothing techniques and anonymizing methods, will be necessary.

Case Study # 17

TITLE

*Patient Outreach via PharmaNet, s*ponsored by the Health Transition Fund and British Columbia Pharmacare. The study was conducted throughout BC in 1999-2001.

RATIONALE

In 1995, British Columbia Pharmacare, the BC College of Pharmacists, the BC Pharmacy Association and pharmacies throughout BC, introduced PharmaNet, a province-wide network among pharmacy computers enabling on-line access to a unique-to-patient electronic record of the patient's past 14 months of filled prescriptions. PharmaNet permits real-time adjudication of Pharmacare benefits and automated warnings of drug interactions.

Patients treated with multiple medications are at greater risk of experiencing an undesirable therapeutic outcome resulting from the complexity of their drug regimen. Adverse therapeutic outcomes such as drug interactions, allergic reactions, and accidental falls in the elderly, to name only a few, result in substantive physical and economic repercussions for the patient, his or her community, and the health care system.

PURPOSE

To measure the impact on patient care resulting from the addition of a flag to patients' unique electronic record in PharmaNet indicating to a dispensing pharmacist that the patient is a candidate for educational intervention designed to improve medication compliance.

POTENTIAL BENEFITS

Patient outcomes may improve if individualized instruction in managing complex medication regimens could be targeted to patients in need of such service. Flagging PharmaNet records can be done anonymously, which enables centralized health care databases to be used to improve patient care without violating privacy.

METHOD

A form of automatic notification was tested in the Patient Outreach Project. A one-line message was created by the central computer and flagged in the record of any patient receiving five or more concurrent medications. The flag was visible only to pharmacists in pharmacies participating in the study. A geographically stratified sample of pharmacies (n=110) was randomly assigned to experimental and control groups. When pharmacists saw the flag, they informed patients that, based on the number of medications they currently receive, they were eligible for an educational service that could assist them in managing their medications. Patients were invited to participate and asked to sign a consent form. Patients recruited at experimental sites received an individualized educational session. Patients were recruited at control pharmacies using an identical protocol including signing a consent form, but received only the customary care provided by their pharmacist.

The researchers never saw personal identifiers in the flagging process because identifiers were not needed to flag the records. The statistical analyses of changes in patients drug use were done

using data extracted from PharmaNet only for patients who gave their written consent, as this is an essential condition for getting access to that data. The project involved no linkage with other data sets. The impact measure was refill compliance of prescribed medications in the two comparison groups. It was expected that refill compliance would be improved in patients in the experimental pharmacies more than in the control pharmacies. Final analyses were not yet complete at the time of writing this case study.

The drug dispensing data from PharmaNet will be retained for approximately one year past the end of the funding for the project to enable additional analyses for publication in a scientific journal.

The PharmaNet Data Access Committee, the University of British Columbia Human Research Ethics Committee, and the privacy branch of the Ministry of Health approved the study.

POTENTIAL RISKS AND INTENDED SAFEGUARDS

There was no risk to patients of their health information becoming more widely available than it already was. Pharmacists already have access to the patient record, which has optional password protection. No information from other sources was added to PharmaNet. What was added was a new algorithm for processing the data to create a new type of message in the record.

The flagging process was designed so that no human handling of data was required between the activation of the algorithm and the pharmacist seeing the flag in the patient's record.

LEGAL/ETHICAL ISSUES

When a Ministry responds to a citizen's request for aggregate data, individual health records are "used" by the Ministry computer without the consent of those patients. Likewise, this project "uses" patients' data without their consent, but does not disclose any identifiable information to anyone who does not already have the authorization to see it. If this project were deemed impermissible, this might also mean that the current practice of periodically updating PharmaNet's existing automatic drug-interaction warning system would likewise be considered impermissible. Clearly, this would have significant repercussions on patients' health.

Informed consent was obtained from patients for the use of their identifiable data to analyze study outcomes; it was not obtained before anonymous flagging. Patients were informed they could withdraw from the study at any time.

PROPOSED PRACTICES

Privacy legislation should permit the anonymous use of health care data for flagging electronic records for patient care and for recruiting patients for health care evaluation.

Appendix A: Method Followed in the Preparation of the Case Studies

Drawing upon its strength as Canada's leading federal granting agency in the area of health research, CIHR gathered together a group of researchers from various disciplines and from across the country to propose, review and discuss actual case studies involving secondary use of data. No hypothetical cases have been used in this exercise; all of the cases retained in this document are real.

The objective of this initiative is to demonstrate in concrete terms for specific research studies: their purposes; the rationale for undertaking them; the potential health benefits they can lead to or have led to; what information is required at what level of identifiability and for what purpose; how the information required is collected, used and disclosed; how consent is obtained and in what form; where consent is not obtained, an explanation of why not; what security measures are implemented to protect the data; how long data is retained, for what purposes and under what conditions. These case studies will hopefully provide opportunity for non-health researchers to better understand the inherent complexities of research, as well as an opportunity for health researchers to exchange best practices.

In order to fulfill its objective, CIHR established a Working Group in the Fall of 2000. Members included clinical and non-clinical researchers in epidemiology, health services and health policy. Meetings, electronic discussions and telephone conferences ensued over the next several months to determine the range of case studies that would be canvassed and develop a process to assist in identifying and articulating the privacy issues underlying the case studies. A template was designed as a model for developing the case studies. Members of the Working Group then prepared a series of case studies based either on their own experience or that of their colleagues who agreed to participate in the process. A call for papers was also made to the Canadian Association for Population Therapeutics.

The draft case studies were then reviewed from legal and ethical perspectives. Volunteer lawyers and ethicists were recruited from across the country to participate in this exercise, including academics, federal and provincial government lawyers, and individuals in private practice. The volunteers were from Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba and British Columbia.

Each individual case author was teamed up with a lawyer or ethicist to provide opportunity for more fulsome discussion of the relevant privacy issues raised in each case study. Through this exercise, the researcher-lawyer teams were able to discuss questions from legal and ethical perspectives and bring out more factual detail to enrich the study description and improve the case study as an analytical and discussion tool. This preliminary review allowed the researchers to tell a fuller story so that privacy advocates, policy makers, legislators and other interested stake-holders may better understand the kind of work they do. The preliminary review also proved to be insightful for the researchers themselves as they re-considered and re-evaluated the facts and issues in each case and proposed some possible practices.

The Working Group then analyzed these case studies in light of the fair information principles. The analysis discusses many practical issues that arise when applying these general principles to health research studies that depend largely on secondary use of data. The analysis identifies the many challenges that lie ahead and have yet to be more fully debated among health researchers, policy-makers, consumers and privacy advocates in an effort to elucidate and refine information principles in the specific context of health research.

All of the Case Studies and appendices are useful and relevant, having been chosen by the assembled working group. Given that some of the cases raise more complicated issues than others (for instance the ones dealing with use of tissue) and given the voluntary nature of participation in the working group, it became clear that some of the cases required more work to provide the necessary details to hold informed discussions about the research performed. The entire working group agreed that rather than wait for all case studies to reach the same level of completion, it was better to put together a representative yet incomplete package which responded to the urgent need, and pressures of all stakeholders instead of delaying the process further.

Appendix B: Glossary of Terms

Aggregate Data: This type of data may take two forms: aggregate or micro aggregate. Aggregate data is data where, within each data sub-element the data have been averaged or grouped into ranges, and only the averages or ranges reported, not revealing the identity of the data subjects.

Micro-aggregate data is data with small randomly assembled clusters of cases averaged, in effect generating a set of pseudo-cases that represent the real population.[1]

Anonymous: This type of data has been permanently stripped of all identifiers such that the information has no reasonable potential for any organization or person to identify a specific individual.[2]

Best Practices: A standard for the design, conduct, performance, monitoring, auditing, recording, analyses, and reporting of research that provides assurance that the data and reported results are credible and accurate, and that the rights, integrity, and confidentiality of trial subjects are protected.[3]

Camouflaged: This refers to a technique which 'scrambles' the identifiers of a given target research group with a control group so that when researchers or third parties are examining the data, the health status of the specific individuals remains unknown. That is, whether a specific person belongs to the targeted group or to the control group, is not ascertainable by the researchers or the third party involved in the camouflaging.

Coded Data: Data for which personal identifiers are removed and secreted but which are still potentially traceable via a matching code.[4]

Confidentiality: Exists when information is communicated in the context of a special relationship (such as doctor-patient, lawyer-client, etc.) where the information is intended to be held in confidence or kept secret.[5]

Collection: The act of gathering, acquiring, or obtaining personal information from any source, including third parties, by any means.[6]

---

[1] Alexander M. Walker, "Generic Data" *Pharmacoepidemiology and Drug Safety* 4, 265-267 (1995).
[2] Canadian Institutes of Health Research (CIHR), *CIHR consultation session On draft recommendations for the interpretation and application of the Protection of Personal information and Electronic Documents Act in health research.* Ottawa, Public Works and Government Services Canada, 5 (2001)
[3] Adapted from Health Canada, Therapeutic Products Directorate Guidelines, *Good Clinical Practice: Consolidated Guideline*. Ottawa, Public Works and Government Services Canada, 5 (1997)
[4] William W. Lowrance, *Privacy and Health Research* A Report to the U.S. Secretary of Health and Human Services. Washington: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Vii. (1997)
[5] Adapted from *Black's Law Dictionary* 5th edition, 269-70 (1979)
[6] Canadian Standards Association (CSA), *CSA Model Code for the Protection of Personal information*

De-identified: This refers to personal information from which identifiers have been removed, such that it is difficult or nearly impossible to identify the specific individual whom it is about.

Disclosure: This refers to making *personal information* available to others outside the organization.[7]

Identifiable: This refers to information which may identify either directly or indirectly, a specific individual; or which may be manipulated by a reasonably foreseeable method to identify a specific individual, or which may be *linked* by a reasonably foreseeable method with other accessible information to identify a specific individual.[8]

Identifier: This refers to a piece of information which may, by itself or when *linked* with others lead to the identification of a specific individual. Examples of direct *identifiers* are names, addresses and telephone numbers. Examples of other *identifiers* include: postal code, date of birth, provincial health insurance number, social insurance number, other dates (i.e. death, diagnosis), sex, local *identifier* (i.e. hospital or physician billing number), ethnic group, occupation, age, etc.[9]

Informed consent: Consent is informed when it is given by a person who understands the purpose and nature of the study, what participation in the study requires the person to do and to risk, and what benefits are intended to result from the study.[10]

It also refers to the dialogue, information sharing and general process through which prospective subjects choose to participate in *research* involving themselves.[11]

Linkage: The bringing together of two or more separately recorded pieces of information concerning a particular individual or family.[12]

***Nominal/ Nominative*:** This refers to personal information which includes direct identifiers (i.e. name, address, telephone number).[13]

***Non-nominal (De-nominalized)*:** This refers to personal information from which direct identifiers have been removed (i.e. names, addresses, telephone numbers).[14]

---

[7] *CSA Model Code*.
[8] CIHR *Consultation session*, (2001) 4-5
[9] Adapted from NHS Executive, *The Caldicott Committee: Report on the Review of patient-identifiable information – December 1997.* Appendix 7 -- Patient-*identifiable* information.
[10] Council for International Organizations of Medical Sciences (CIOMS), *International Guidelines for Ethical Review of Epidemiological Studies*, 11-2 (1991)
[11] Tri-Council Policy Statement (TCPS), *Ethical Conduct for Research Involving Humans*, Public Works and Government Services Canada, 2.1 (1998)
[12] Howard B Newcombe, et al., "Automatic linkage of vital records." *Science,* 130:954-9, (1959) 87:420
[13] Adapted from *CIOMS,* 17 (1991)
[14] Adapted from *Ibid.*

Personal Information: This refers to information about an identifiable individual, and is broader than Personal health information which refers to information concerning a living or deceased individual:

physical or mental health; health services provided; the donation by the individual of any bodily part or any bodily substance or information derived from the testing or examination of a body part or bodily substance; information that is collected in the course of providing health services to the individual; or, information that is collected incidentally to the provision of health care.[15]

Privacy: This refers to the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others.[16]

Research: This refers to a class of activities designed to develop or contribute to generalizable knowledge. Generalizable knowledge consists of theories, principles or relationships, or the accumulation of information on which they are based, that can be corroborated by accepted scientific methods of observation and inference.[17]

Security: This consists of a number of measures that organizations implement to protect information and systems. It includes efforts not only to maintain the confidentiality of information, but also to ensure the integrity and availability of that information and the information systems used to access it.[18]

Use: This refers to the treatment and handling of personal information within an organization.[19]

[15] *Personal information Protection and Electronic Documents Act*, (S.C. 2000 c.5)
[16] Alan F. Westin, *Privacy and Freedom*, 7 (1967)
[17] Council for International Organizations of Medical Sciences (CIOMS), *International Ethical Guidelines for Biomedical Research Involving Human Subjects*, 11 (1993).
[18] National *Research* Council (USA), *National Research Council, Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure, Computer Science and Telecommunications Board, For the Record: Protecting Electronic Health Information*. 1-1 (1997).
[19] *CSA Model Code*.

Appendix C: Useful *Links**

American Health Information Management Association (AHIMA):
http://www.ahima.org/

Caldicott Committee Report on Patient *Identifiable* Health Information in the National Health Service in Great Britain
http://www.doh.gov.uk/confiden/crep.htm

Canadian Institutes of Health Research (CIHR) *Personal information Protection and Electronic Documents Act Questions and Answers for Health* Researchers, 2001
http://www.cihr.ca/about_cihr/ethics/intro_qa_e.shtml

Canadian Institutes of Health Research (CIHR), *A Compendium of Canadian Legislation Respecting the Protection of Personal information in Health Research, 2000*
http://www.cihr.ca/about_cihr/ethics/compendium_e.pdf

Canadian Institutes of Health Research (CIHR) *Personal health information: Balancing Access and Privacy in Health Research*, Summary, Recommendations & Follow Up, 2000.
http://www.cihr.ca/about_cihr/ethics/personal_health/personal_health_e.shtml

Canadian Institutes of Health Research, Natural Sciences and Engineering *Research* Council of Canada, Social Sciences and Humanities *Research* Council of Canada, *Tri-Council Policy Statement Ethical Conduct for Research Involving Humans, 1998*
http://www.nserc.ca/programs/ethics/english/ethics-e.pdf

Canadian Institutes of Health Information (CIHI), *Privacy and Confidentiality Guidelines on Health Information at CIHI: Principles and Policies for the Protection of Health Information*,
http://www.cihi.ca/weare/pcsmain.shtml

CMA Health Information *Privacy* Code:
http://www.cma.ca/inside/policybase/1998/09-16.htm

Council of International Organizations of Medical Sciences (CIOMS) guidelines (both for epidemiology and for *research* involving humans):
http://www.cioms.ch/frame_menu_texts_of_guidelines.htm

Department of Health and Human Services (HHS) (USA), bibliography concerning Confidentiality of Electronic Health Data:
http://aspe.os.dhhs.gov/datacncl/privbibl.htm

Health Privacy Project:
www.healthprivacy.org

Epidemiology for the Uninitiated:
http://www.bmj.com/collections/epidem/epid.shtml

MRC UK *Personal information* in Health *Research*:

* Note: these links are accurate and functional as of November 8th 2001.

http://www.mrc.ac.uk/PDFs/PIMR.pdf
executive summary:
http://www.mrc.ac.uk/PDFs/PIMR_summary.pdf

Nuremberg Code:
http://ohsr.od.nih.gov/nuremberg.php3

OECD List of *Privacy* and Data Protection Authorities in member states:
http://cs3-hq.oecd.org/scripts/pwv3/privcontacts.htm

Privacy Laws & Business: Data Protection & Privacy Information Worldwide
http://www.privacylaws.com/

Saskatchewan consultation on protection of personal health information:
http://www.gov.sk.ca/health/phiq/response

World Medical Association (WMA), Declaration of Helsinki:
http://www.wma.net/e/policy/17-c_e.html

APPENDIX D: SELECT BIBLIOGRAPHY

Canadian Institutes of Health Research (CIHR) *Personal Information Protection and Electronic Documents Act Questions and Answers for Health Researchers*, Ottawa: Public Works and Government Services Canada, 2001.

Canadian Institutes of Health Research (CIHR) *A Compendium of Canadian Legislation Respecting the Protection of Personal Information in Health Research*, Ottawa: Public Works and Government Services Canada, 2000.

Canadian Institutes of Health Research (CIHR) *Personal Health Information: Balancing Access and Privacy in Health Research*, Summary, Recommendations & Follow-up, Ottawa: CIHR, 2000.

Canadian Institutes of Health Research, Natural Science and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada, *Tri-Council Policy Statement Ethical Conduct for Research Involving Humans*, Ottawa: Public Works and Government Services Canada, 1998.

Carman D. and Britten N. Confidentiality of medical records: the patient's perspective. British Journal of General Practice 1995; 45: 485-488.

Coughlin SS, Soskolne CL, Weed DL, Buck G, Snider DE, McTyre RB, American College of Epidemiology Ethics Guidelines, Annals of Epidemiology 10(8); 487-497, 2000.

Coughlin Steven S, Soskolne Colin L and Goodman, Kenneth W, Case Studies in Public Health Ethics. Washington: American Public Health Association, 2000.

Gordis L. and Gold E. Privacy, confidentiality and the use of medical records in research. Science, 1980; 207: 153-156.

Hodges et al. Legal issues concerning electronic health information : privacy, quality and liability JAMA 1999; 282: 1466-1471.

Jacobsen S. et al. Potential effect of authorization bias on medical record research. Mayo Clinic Proceedings 1999; 74: 330-338.

Lako C.J. Privacy protection and population-based research. New England Journal of Medicine 2000; 342

McCarthy D et al. Medical records and privacy : empirical effects of legislation. Health Services Research 1999; 34: 417-425.

Medical Research Council UK, Personal Information in Medical Research. Medical Research Council, London: October 2000.

Melton, L. Joseph III. The threat to medical records research. The New England Journal of Medicine 1997; 337:1 1466-1470.

National Health & Medical Research Council, Medical Research Ethics Committee. Report on Ethics in Epidemiological Research. Australian Government Publishing Service, Canberra: 1985.

Office of Management and Budget, Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology. Prepared by Subcommittee on Disclosure Limitation Methodology, Federal Committee on Statistical Methodology. May 1994.

Patterson, ICM. Consent to cancer registration – an unnecessary burden. British Medical Journal 2001; 322: 1130.

Traynor, Ben. Medical research threatened by patient consent. The Guardian, Thursday, april 12, 2001.

Veatch R and Pitt J. The myth of presumed consent in new organ procurement strategies. Transplantation Proceedings 1995; 27: 1888-1892.

Willison D. Health services research and personal health information. CMAJ 1998; 159: 1378-1380.

Yawn B et al. The impact of requiring patient authorization for use of data in medical records research. The Journal of Family Practice. 1998; 361-365.