Privacy protection and public goods: building a genetic database for health research in Newfoundland and Labrador

Patricia Kosseim,1 Daryl Pullman,2 Astrid Perrot-Daley,3 Kathy Hodgkinson,2 Catherine Street,3 Proton Rahman2

ABSTRACT

Objective To provide a legal and ethical analysis of some of the implementation challenges faced by the Population Therapeutics Research Group (PTRG) at Memorial University (Canada), in using genealogical information offered by individuals for its genetics research database.

Materials and methods This paper describes the unique historical and genetic characteristics of the Newfoundland and Labrador founder population, which gave rise to the opportunity for PTRG to build the Newfoundland Genealogy Database containing digitized records of all pre-confederation (1949) census records of the Newfoundland founder population. In addition to building the database, PTRG has developed the Heritability Analytics Infrastructure, a data management structure that stores genotype, phenotype, and pedigree information in a single database, and custom linkage software (KINNECT) to perform pedigree linkages on the genealogy database.

Discussion A newly adopted legal regimen in Newfoundland and Labrador is discussed. It incorporates health privacy legislation with a unique research ethics statute governing the composition and activities of research ethics boards and, for the first time in Canada, elevating the status of national research ethics guidelines into law. The discussion looks at this integration of legal and ethical principles which provides a flexible and seamless framework for balancing the privacy rights and welfare interests of individuals, families, and larger societies in the creation and use of research data infrastructures as public goods.

Conclusion The complementary legal and ethical frameworks that now coexist in Newfoundland and Labrador provide the legislative authority, ethical legitimacy, and practical flexibility needed to find a workable balance between privacy interests and public goods. Such an approach may also be instructive for other jurisdictions as they seek to construct and use biobanks and related research platforms for genetic research.

BACKGROUND AND SIGNIFICANCE

Newfoundland and Labrador is Canada’s youngest province, having joined the Canadian confederation only in 1949, with a current population of 514,500. According to historical geographers, 80–90% of its peoples can trace their ancestry to 20,000–30,000 settlers who made their way from Ireland and England in the 1700s and 1800s. Many of these early immigrants settled in small fishing villages along the rugged coast which were accessible only by boat. This settlement pattern and the concomitant isolation resulted in a number of genetic isolates in which a high degree of genetic homogeneity exists. Thus a number of common genetic conditions are more prevalent in Newfoundland and Labrador. For example, the incidence of juvenile type 1 diabetes mellitus (26 per 100,000) is one of the highest in the world, more than double that of admixed populations in the USA. Conditions such as colorectal cancer, certain cardiomyopathies, hereditary deafness, eye disease, psoriatic arthritis, and numerous others are more prevalent in the Newfoundland founder population. PTRG has used a variety of innovative means to populate its database, but has faced a number of privacy-related challenges with regard to the propriety of drawing upon certain sources to enrich that database. In particular, PTRG has not yet been able to confirm whether it can collect and use privately constructed family genealogies offered by individuals who are interested in furthering this work. The concern is that unless the individual who offers a genealogy has the consent of each member of the extended family, entering their genealogical information into the database might be a violation of the privacy rights of those family members, and thereby run afoul of the recently proclaimed Personal Health Information Act (PHIA) for Newfoundland and Labrador. While the issues raised here are directly relevant to the development of PTRG’s data management infrastructure for conducting genetic research in Newfoundland and Labrador, they are illustrative of some of the privacy issues that pertain to the development of biobanks generally when the privacy rights of individuals potentially conflict with those of family members who share some of that same personal information. They are also illustrative of some of the legal and ethical considerations that should be taken into account in the development and use of data management structures that may produce significant public goods.

OBJECTIVE

This paper provides a legal and ethical analysis of some of the implementation challenges encountered by the Population Therapeutics Research Group (PTRG) at Memorial University in Newfoundland and Labrador, Canada. Over several years PTRG has developed an innovative genetics research data management environment designed to facilitate genetic research on the Newfoundland
population. Newfoundland and Labrador is recognized as one of the world’s prime founder populations for conducting genetic research on a variety of monogenetic and complex conditions.3 4

The PTRG is a not-for-profit research team situated in the Faculty of Medicine at Memorial University. With primary funding from the Atlantic Canada Opportunities Agency, PTRG has been building a Newfoundland Genealogy Database (NGD) and has developed the Heritability Analytics Infrastructure (HAI), an innovative technological infrastructure capable of integrating genetic and genealogic information with drug information and health outcomes. Significant effort and resources have been expended to create a governance framework for the operation of this infrastructure.

MATERIALS AND METHODS

The development of the NGD has been a collaborative effort between PTRG, the Newfoundland and Labrador Statistics Agency, and the Canadian Century Research Infrastructure (CCRI).5 CCRI is a pan-Canadian initiative to create a national database of census records. The Atlantic arm of the project is led by the department of history at Memorial University. When the project was initiated in 2003, CCRI’s goal was to digitize a 5% sample of census records completed between 1911 and 1951. Access to census data is subject to the Canadian Statistics Act6 which stipulates the terms and conditions under which identifiable information extracted from census records can be released. In general, detailed census information that would allow identification of individuals is strictly controlled for 92 years from the time the census is completed, after which the secrecy provisions of the Statistics Act are lifted in respect of census information collected between 1910 and 2005.7 It is these provisions that limit both the amount of data the CCRI is able to digitize and the manner in which it can be used for research purposes. As Newfoundland did not join the Canadian confederation until 1949, the 92-year limitation does not apply to Newfoundland and Labrador census records collected before that date. Newfoundland’s pre-confederation census data are in the public domain and there are no restrictions on digitizing these.

It was this unique historical circumstance and resulting legislative flexibility that caught the attention of PTRG, prompting an inquiry to CCRI about the possibility of collaborating with that project to create the NGD. The aim was to digitize all of Newfoundland’s pre-confederation census data. Although CCRI’s primary intent was to facilitate historical research, PTRG understood the considerable potential for genetic research if census records were readily available in this format. Constructing family pedigrees is essential to effective genetic research, and literally hundreds of hours are invested in interviewing extended families, poring over church records, and using other creative means to establish family connections in the hope of identifying patterns of inheritance. Digitization of the census records to create the NGD would make it possible to build family pedigrees in a matter of hours.

For several years PTRG has worked closely with the Newfoundland Statistics Agency and the Memorial University arm of CCRI to complete the digitization of all pre-confederation census data. In order to facilitate genetic research, PTRG developed the HAI, a data management structure that stores genotype, phenotype and pedigree information in a single database, and custom linkage software (KINNECT) to perform pedigree linkages on the genealogy database8 (see figure 1).

The pedigree linkage and matching capabilities of the HAI have been validated against pedigrees developed in previous genetic studies, and this tool is now used to assist genetic researchers in a wide range of studies. However, its ongoing success is contingent upon the richness and completeness of the database. From the outset it was recognized that census records often contain erroneous data or are missing data that are essential to the construction of accurate pedigrees. PTRG continues to deal with these errors and gaps by supplementing the census data with other sources of published genealogical information, and has been granted permission by a number of major religious denominations to use church records. Church records often provide familial information from before the first census, helping to populate the database further back to the original founders. Another avenue to further enrich this data involves the use of StonePics, a unique project to photograph and index every cemetery headstone and monument in Newfoundland.9 Such markers often contain information such as maiden names that might not have been included in the census record.

While church records can assist in populating the database backwards beyond the initial census, there is an additional

---

**Figure 1** Visual depiction of the Heritability Analytics Infrastructure (HAI). CCRI, Canadian Century Research Infrastructure; NL, Newfoundland and Labrador; PTRG, Population Therapeutics Research Group; WHAP, Window-based Haplotype Association Program.
challenge in populating the database forward from 1949 to the present. Using the HAI for a genetic study generally requires that the proband is able to identify a known relative whose data was captured in a pre-confederation census. If the proband is not aware of any relative from that era a more traditional means of pedigree building is required until some link to the existing data is established. Some of this forward populating is accomplished by entering pedigree information from previously completed or ongoing genetic studies. Marriage records from vital statistics and church records can also supplement these more recent data when available.

Another means to supplement and correct existing data, and the focus of the analysis to follow, involves the use of family genealogies constructed by private individuals and voluntarily offered for inclusion into the digitized database. Genetic researchers often rely on such individuals when constructing pedigrees in the traditional manner to help fill gaps in a genealogy. It is not clear whether doing so would be a violation of the privacy rights of other living or deceased family members whose information was included in these genealogies, most of whom cannot be easily contacted, if at all, to provide consent. This specific issue raises the broader question of how individual, family and societal interests in privacy and access to information can be reconciled in the context of genetic research. In this paper, we offer a principled legal and ethical framework for balancing these competing interests.

DISCUSSION

We begin with a broad overview of the Canadian legislative landscape and move on to discuss more specifically recent relevant legislative developments in Newfoundland.

Canada has public-sector access and privacy legislation, federally and in every province and territory. Some Canadian jurisdictions also have private-sector privacy legislation, while some have privacy statutes specific to the health sector. Among the latter group of statutes, one will typically find an exemption permitting non-consensual collection, use and disclosure of personal information for health research purposes, subject to certain conditions. Some health privacy laws elaborate on these conditions more specifically than others; some expressly require research ethics board (REB) approval, while others go further to permit non-consensual collection, use and disclosure of personal information for health research purposes, subject to a lawful purpose and to collect such directly from the individual who is the subject of the information. Hence, individuals voluntarily offering their family genealogies to the PTRG for inclusion into the NGD for research purposes, would be providing valid consent under PHIA. What of other family members—both living and deceased—whose personal health information is included in those genealogies but who cannot practically or even possibly be contacted to provide consent or authorization?

PHIA exceptionally permits custodians to collect personal health information from a source other than the individual who is the subject of the information, and to use and disclose that information without consent for the purpose of carrying out a research project that has been approved by an REB appointed under the Health Research Ethics Authority Act (HREA). This unique legislation essentially creates a province-wide research ethics authority to ensure that all human health research is reviewed within the province and conducted ethically. The initial impetus for the legislation was to curtail the activities of outside genetic researchers who came to the province to conduct research without the knowledge of local authorities or health-care officials and without any accountability to the people of Newfoundland and Labrador.

Under HREA, health research involving human subjects cannot proceed without prior approval of a duly recognized REB established in conformity with the principles of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS). In exercising its review and approval powers under the Act, the REB shall apply the TCPS among other guidelines, where applicable. HREA is the first legislative scheme in Canada that incorporates the TCPS thereby elevating its status from national guidelines into law. Interestingly, however, HREA also provides for the possibility that the REB may, with the approval of the HREA, vary a standard or rule contained in the TCPS where the board considers it appropriate to do so.

The TCPS recognizes the importance of secondary use of identifiable personal information for research purposes and the collective social good that could come of it. However, it also recognizes and respects the privacy rights of individuals whose personal information is involved. Exceptionally, it allows researchers to make secondary use of personal information without consent only if the REB is satisfied that specific conditions are met.

Here we consider each of these conditions in turn as would an REB seized with the question of whether personal data originally collected for private genealogies could be included in the NGD without consent of all family members. It should be noted that the following analysis applies only to this initial entry of...
personal data into the NGD and does not consider whether individual consent might still be required for participation in a specific study that uses the NGD. This latter question would be dealt with by a duly constituted REB under PHIA on a case-by-case basis.

**Identifiable information is essential to the research**

In this case study, family genealogies offered by volunteers would almost invariably contain identifiable information. Researchers would essentially need identifiers for the purpose of populating the NGD, and linking genealogy with phenotype and genotype data through the HAI in order to enrich the meaningfulness of the information and better understand genetic patterns and predispositions. This being said, family genealogies prepared for personal purposes are likely to vary widely in the quantity and nature of the personal information included. They may include more or less information. Some may resemble simple pedigrees of names, relationships, dates of birth and death, while others may take a more narrative form with personalized accounts of the lives of each family member, and not all data may be objective, accurate, or relevant for research purposes. Hence, PTRG as data custodian would have to be selective in receiving these family genealogies and screen out (either by not accepting in the first place, or by immediately destroying thereafter) any personal information deemed non-essential for research purposes.

**Use of identifiable information without the participants’ consent is unlikely to adversely affect the welfare of individuals to whom the information relates**

The motivation of individuals who volunteer their family genealogies is often the hope of improving the health and welfare of their families—both present and future generations. Indeed, the possibility of identifying possible genetic links between individuals experiencing similar disease patterns and/or reactions to a particular course of treatment, could for instance, uncover important relationships that enhance understanding of adverse drug reactions and ultimately improve drug treatment.

But might the inclusion of family pedigrees adversely affect the welfare of family members who have not consented? In the case of deceased family members, it is important to acknowledge that privacy and reputational interests survive death. For instance, PHIA continues to protect the personal health information of deceased individuals up to 50 years after death, but allows their personal representative or nearest relative to exercise knowledge that such information exists, irrespective of its purpose. Information about already-known biological relationships may not be highly sensitive in the majority of cases. However, one can imagine rarer situations in which unknown paternal or adoption cases are uncovered. Should future contact be envisaged as a possibility for further research, there is a risk that unknown biological linkages might be inadvertently disclosed to some family members through others as information eventually makes its way through “the grapevine”. Psychological and/or social harms in such circumstances, though rare, could be traumatizing.

Hence, even with respect to family trees, the individual expectation of privacy is far from trivial. The REB must consider such risks in light of what the researchers intend to do with the information, whether future contact with individual family members is contemplated, and whether a plan is in place to mitigate the risk of social and psychological harms resulting from those relatively rare, but highly sensitive cases, where unknown paternity linkages may be inadvertently disclosed.

**The researchers will take appropriate measures to protect the privacy of individuals, and to safeguard the identifiable information**

The REB would also need to assess whether all reasonable safeguards were in place to maintain the security and confidentiality of personal information contained in the NGD, linkable and accessible through the HAI. This includes effective coding, de-identification, and encryption methods. PTRG has invested significant resources in creating appropriate governance structures and processes (including oversight mechanisms and sharing agreements) that clearly allocate responsibility and ensure accountability among all relevant actors, including PTRG staff and third-party researchers. Although this is a highly intricate and complex element to consider, it is not unique to the NGD. A full discussion of security safeguards in health research—physical, organizational, and technological—is beyond the purview of this paper.

**The researchers will comply with any known preferences previously expressed by individuals about any use of their information**

Although PTRG researchers are not likely to have had previous contact with family members and therefore, are not likely to know of expressed wishes, they should ask the individuals volunteering their family genealogies. Any known objection expressed by family members, either living or before death, should be respected by removing that member’s details from the data before entry in the NGD.

**It is impossible or impracticable to seek consent from individuals to whom the information relates**

Given the large number of individuals likely to be included in family genealogies dating back several generations, it is likely impracticable, if not impossible, to obtain consent from deceased family members and/or living members lost to contact. That being said, PTRG researchers could make other reasonable efforts through websites, local newspapers, radio announcements, phone-in hotlines, and other effective public dissemination vehicles to be open and transparent about the creation of the NGD and the HAI, and explain their intended purposes vis-a-vis both the general public and the appropriate regulator(s). Where contact with known family members can practicably be done to seek their authorization for inclusion in the NGD, the individual volunteering his or her family genealogy should be encouraged to establish preliminary contact with their family rather than have PTRG members make cold and unannounced calls.
The researchers have obtained any other necessary permission for secondary use of information for research purposes

In essence, this TCPS provision refers the REB back to the applicable legislative scheme in Newfoundland. This would encompass all other applicable provisions of HREA, including the principal investigator’s obligations to: obtain REB approval before implementing protocol changes; provide the REB with access to records for monitoring purposes; correct any deficiencies identified through REB monitoring; and submit to the REB a copy of the final report upon completion of the project(s).

Were PTRG to obtain REB approval under HREA in accordance with the principles of TCPS outlined above, it would seem they could then collect, use and disclose family genealogies without consent of family members who are deceased or lost to contact, although their privacy-related obligations would not end there. While PHIA allows exceptionally for the non-consensual collection, use and disclosure of personal information for research upon approval by a REB in accordance with HREA, it does not exclude these activities from the scope of the Act altogether. PHIA would continue to govern the activities of the PTRG as a custodian under the Act, as well as the personal health information it holds. Hence, apart from PHIA’s consent requirement from which HREA provides an exemption, other PHIA obligations to protect and secure personal health information, provide access rights thereto, notify in the event of breach, respond to complaints and submit to the review powers of the Information and Privacy Commissioner, persist as conditions for secondary research use of personal health information.

CONCLUSION

Having done its due diligence in reviewing all these conditions before approving the secondary use of family genealogies without consent for the purpose of creating and using the NGD, it is still open to the REB, with the approval of the HREA, to vary a standard or rule contained in the TCPS where the board considers it appropriate to do so in the context of a proposed research project. What are some unique circumstances that might lead the REB to consider doing so?

While Newfoundland’s unique founder population makes it an ideal place to conduct genetic research, it also results in an increased burden of disease for a variety of serious conditions. Arrhythmogenic right ventricular cardiomyopathy (ARVC) is one example of a particularly lethal genetic anomaly prevalent in Newfoundland which leads to sudden cardiac death. Fifty per cent of affected men are dead by the age of 40, and 80% by age 50. Research in Newfoundland has identified the gene responsible for this condition and testing is now available.20 Although the condition is medically untreatable, implantable defibrillators are a means of prophylactic intervention for affected individuals. However, researchers and clinicians are concerned that some branches of ascertained families may not yet have been identified. The NGD and the HAI can assist in constructing pedigrees rapidly to aid in identifying at-risk individuals. This provides a direct benefit to affected individuals, and also a broader public health benefit when those at risk of sudden cardiac arrest are no longer driving vehicles on the highways, or engaging in other activities that put the public at risk.

ARVC is a particularly poignant example of a situation in which the genetic health risks to individuals and the public in general are considered so significant as to out-weigh the relatively lower privacy risks associated with including family genealogical information in the NGD without the consent of all family members. Given the unique circumstances of Newfoundland and Labrador as a founder population, its increased prevalence of genetic risks, and higher incidence of related diseases, a proportionate approach that enables the creation of such a database for the public good may very well be justifiable. That said, each specific research project purporting to make subsequent use of the personal health information contained in the database would still have to be individually reviewed on its own merits by a duly authorized REB in the province, and its related consent and privacy concerns and other associated risks and benefits would have to be weighed accordingly. The beauty of PHIA is that it provides sufficient flexibility to do this, while HREA creates the necessary backstop to ensure proper accountability and oversight and to prevent potential abuses.

Although this discussion has focused on the NGD, the issues addressed are not unique to this particular research platform. Virtually every genomics research infrastructure, including biobanks, struggles to find a proportionate balance between protecting the privacy rights of individuals and family members whose data are stored in the repository, and allowing access to that data to promote broader public goods. The complementary legal and ethical frameworks that now coexist in Newfoundland and Labrador through PHIA, HREA, and TCPS provide the legislative authority, the ethical legitimacy, and the contextual flexibility needed to find a workable balance appropriate to particular circumstances. Such an approach may be instructive for other jurisdictions as they too grapple with similar challenges associated with emerging genomics research infrastructures.

The Newfoundland and Labrador model shows early promise as an innovative policy framework to be emulated: it reconciles legal and ethical principles through seamless integration; it enables a flexible and proportionate approach for balancing privacy interests and public goods; and, it establishes an effective governance framework that legitimizes decision-makers and holds them accountable to the population whose interests they are intended to protect.

Author footnote

The views in this paper are the author’s own personal views. They do not constitute a legal opinion nor represent the position of the OPC or any other data protection commissioner. The OPC, as a Federal Office, has no jurisdiction over the subject of this paper.

Contributors PK and DP are primary authors who conceived the idea for the paper and drafted the manuscript. AP-D, KH and CS reviewed the manuscript to ensure accuracy of the description of the Newfoundland Genealogy Database and Heritability Analytics Infrastructure as they were instrumentally involved in creating and testing the database. PR is the principal investigator for the project.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES

11. PHIA, (S.N.L. 2008 c. P-7.01) section 3.
13. PHIA, (S.N.L. 2008, c. P-7.01) subparagraph 5(1)(a), combined with paragraph 5(3).
19. PHIA, (S.N.L. 2008, c. P-7.01) paragraphs 10(1) and 7(e).