

Linking Canadian Population Health Data: Maximizing the Potential of Cohort and Administrative Data

Dany Doiron, MPP,^{1,2} Parminder Raina, PhD,³ Isabel Fortier, PhD,^{1,2} on behalf of the *Linkage Between Cohorts and Health Care Utilization Data: Meeting of Canadian Stakeholders* workshop participants*

ABSTRACT

Linkage of data collected by large Canadian cohort studies with provincially managed administrative health databases can offer very interesting avenues for multidisciplinary and cost-effective health research in Canada. Successfully co-analyzing cohort data and administrative health data (AHD) can lead to research results capable of improving the health and well-being of Canadians and enhancing the delivery of health care services. However, such an endeavour will require strong coordination and long-term commitment between all stakeholders involved. The challenges and opportunities of a pan-Canadian cohort-to-AHD data linkage program have been considered by cohort study investigators and data custodians from each Canadian province. Stakeholders acknowledge the important public health benefits of establishing such a program and have established an action plan to move forward.

KEY WORDS: Medical record linkage; data collection; public health

La traduction du résumé se trouve à la fin de l'article.

Can J Public Health 2013;104(3):e258-e261.

The Canadian Longitudinal Study on Aging (CLSA)¹ and the Canadian Partnership for Tomorrow Project (CPTP)² are two important nationwide longitudinal cohort studies currently underway in Canada. The large number of subjects targeted for recruitment in these cohorts and the collection of repeated observations over time will allow investigators to identify the many factors influencing healthy aging and the complex etiological pathways of chronic diseases affecting an increasingly large number of Canadians. Information collected on lifestyle and behaviours, health outcomes, social and physical environments, and biological measures – including genetic information – will help establish two long-term, pan-Canadian research platforms, which will provide many opportunities for innovative health research. To complement information collected by each cohort, both CLSA and CPTP anticipate linking consenting participants' data to information collected in provincial administrative health databases (AHD) such as vital statistics, hospitalizations, physician billing, and prescription drugs. In an effort to begin addressing the challenges and opportunities related to a pan-Canadian linkage program, a workshop bringing together CLSA and CPTP investigators, and data custodians from each provincial ministry of health was held in Toronto in September 2011.*

Why link administrative and cohort data?

Implementing a data linkage strategy between administrative databases and population-based observational studies can provide

important contributions to the Canadian health research community. First, in the context of the CLSA, CPTP, and similar cohorts, such data linkage can provide a means of enriching study datasets with additional data not being collected directly from study participants, offer vital information on health outcomes of participants, and serve to validate self-reported information. Second, linkage of existing databases offers a cost-effective means to maximize the use of existing publicly funded data collections.⁴ Third, by combining a wide range of risk factors, disease endpoints, and relevant socio-economic and biological measurements at a population level, linkage lays the groundwork for multidisciplinary

Author Affiliations

1. Public Population Project in Genomics and Society, Montreal, QC
2. Research Institute of the McGill University Health Centre, Montreal, QC
3. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON

Correspondence: Dany Doiron, Research Institute of the McGill University Health Centre, 2155 rue Guy, Room 458, Montreal, QC H3H 2R9, Tel: 514-934-1934, ext. 71665, E-mail: ddoiron@p3g.org

Acknowledgements: The authors thank all data custodians and CLSA and CPTP investigators for participating in the workshop and providing invaluable feedback on this paper. Funds received from a Meetings, Planning and Dissemination grant from the Canadian Institutes of Health Research (Winter 2011 Competition) allowed us to organize the workshop of Canadian stakeholders. This work is also supported by Genome Canada, Genome Quebec, and the Canadian Partnership Against Cancer. Dr. Raina is the recipient of a Canada Research Chair in Geroscience and holds the Labarge Chair in Research and Knowledge Application for Optimal Aging at McMaster University.

Disclaimer: The interpretation and conclusions contained within this paper do not necessarily represent those of the various Ministries of Health or Data Protection Commissioners who participated in the *Linkage Between Cohorts and Health Care Utilization Data: Meeting of Canadian Stakeholders* workshop.

Conflict of Interest: None to declare.

* See Appendix A for list of workshop participants.

Box 1. What Is Data Linkage?

Data linkage is “the bringing together from two or more different sources, data that relates to the same individual, family, place or event”.³ When linking data at the individual level, a common identifier (or a combination of identifiers) such as a personal health number, date of birth, place of residence, or sex, is used to combine data related to the same person but found in separate databases. Data linkage has been used in a number of research fields but is an especially valuable tool for health research given the large amount of relevant information collected by institutions such as governments, hospitals, clinics, health authorities, and research groups that can then be matched to data collected directly from consenting individuals participating in health research.

health-research initiatives involving investigators from a wide number of fields such as public health, epidemiology, clinical epidemiology, genetics, pharmacology, economics, ethics, and policy. Such wide-ranging data collections allow the exploration of new hypotheses not foreseeable using independent datasets. Last, a coordinated pan-Canadian cohort-to-administrative linked database would establish legacy research infrastructures that will better equip the next generation of researchers across the country. Having foreseen such opportunities, the CLSA and the CPTP have included questions related to data linkage in their participant informed consent forms. Data custodians were also directly consulted in the development of these forms.

Building from experience: Existing data linkage initiatives

In a number of countries and jurisdictions around the world, integrating and linking different types of information from different sources is helping to create rich data repositories for research purposes. For example, the Western Australia Data Linkage System (WADLS), an international leader in data linkage, has now managed to link up to 40 years of data from over 30 population-based datasets (e.g., births, deaths, hospital inpatients, research surveys, electoral rolls) covering the 2 million inhabitants of Western Australia.³ Data stored, managed and linked at the WADLS have been used by over 400 studies, resulting in more than 250 journal publications and 35 graduate research degrees.³ The success of the WADLS program has also provided impetus for the establishment of the Population Health Research Network (PHRN; www.phrn.org.au), a linkage project which will establish data linkage units in all Australian states/territories and facilitate cross-jurisdictional data integration. Once fully established, the PHRN is expected to become the world’s largest population database supporting health research, policy, and planning.⁵ The Cohort of Norway (CONOR), a collaborative initiative between the Norwegian Institute of Public Health and four universities across the country, has made use of Norwegian unique identification numbers to link health survey data from consenting participants to administrative data (e.g., national health registries, prescription drugs, disease registers, census), and thereby help build a nationally representative multipurpose cohort.⁶ A number of data linkage initiatives also exist in Canada. The Manitoba Centre for Health Policy (MCHP; www.umanitoba.ca/centres/mchp/) is recognized as an international leader in providing access to, and making use of, linked administrative information for research purposes. Through various data-sharing agreements and collaborations, the MCHP now holds over 100 linkable datasets in domains ranging from administrative, survey and clinical health databases to justice and education databases. Other notable Canadian data linkage projects include pro-

grams led by the Institute for Clinical Evaluative Sciences (ICES; www.ices.on.ca) in Ontario, Population Data BC in British Columbia (www.popdata.bc.ca), and the Population Health Research Unit (PHRU; www.phru.dalhouse.ca) in Nova Scotia. At the federal level, Statistics Canada began linking National Population Health Survey (NPHS) data with administrative data in the early 1990s and is actively pursuing such work with initiatives like the Longitudinal Health and Administrative Data Initiative, a collaborative project with provincial and territorial ministries responsible for health care and public health.

Nationwide longitudinal cohorts and AHD linkage: Challenges and opportunities

Although we can build on the considerable work that has been achieved in data linkage in Canada and internationally, the longitudinal and pan-Canadian nature of the CLSA and CPTP makes linkage to AHD an unprecedented endeavour that will face extensive new challenges, as highlighted by data custodians at the Toronto workshop. For some provinces, the most significant challenge is in current legislative interpretation that administrative data cannot cross provincial jurisdictional boundaries, requiring that linkage take place province by province and prohibiting cross-provincial transfers of data. Moreover, some data custodians indicated limited resources and capacity to meet data access, extraction and linkage requests – secondary use data-access requests from researchers being generally met when, and if, there is time. Furthermore, the quality and completeness of AHD are generally not ideal, thus requiring data cleaning and validation work. Workshop participants also underlined that “longitudinal” linkage with cohort data (i.e., repeated or updated through time) has never been conducted on such a large scale in the past. Last, CLSA/CPTP-to-AHD linkage also poses unique challenges in that, in contrast to more traditional requests to link data to answer one-off research questions, it aims to establish a rich data repository that will allow investigators to answer a multitude of research questions over time.

As both the CLSA and CPTP aim to collect and link participant data for the next two to five decades, such a project will involve long-term and active engagement by provincial ministries. The unique and multijurisdictional nature of the project will require that: 1) some legislative impediments related to health-information access be addressed; 2) appropriate safeguards and data-sharing agreements be put in place; and 3) new linkage mechanisms be developed to allow interprovincial pan-Canadian research in a manner that protects participants and their personal information and respects applicable legislation. Innovative ethico-legal and information technology tools are helping address these conditions for a successful cohort-to-AHD linkage initiative. For example, conceptual frameworks such as risk-based de-identification⁷ and Privacy by Design^{8,9} can help cohort representatives and provincial data custodians establish appropriate measures which ensure the security and privacy of sensitive data. New protocols for data encryption and analysis of distributed databases can also offer solutions to the issue of sharing individual-level data across institutions and across jurisdictions.¹⁰⁻¹²

Toronto workshop participants generally agreed that the challenges are surmountable, and that the important public health benefits of an effective linkage process would indeed justify the efforts required to put it in place. A five-point action plan was therefore

Appendix A. List of Workshop Participants

	Last Name	First Name	Middle Initial	Institution	Province
1	Caetano	Patricia	A.	Manitoba Health	Manitoba
2	Deschênes	Mylène		P3G Consortium	Québec
3	Doiron	Dany		P3G Consortium/CPTP	Québec
4	Dufraîne	Elizabeth	E.	Alberta Health and Wellness	Alberta
5	Fay	Marina	A.	Health PEI	Prince Edward Island
6	Fortier	Isabel		P3G Consortium	Québec
7	Griffith	Lauren	E.	McMaster University	Ontario
8	Kirby	Emily		Centre of Genomics and Policy, McGill University	Québec
9	Kirkland	Susan		Dalhousie University	Nova Scotia
10	Kosseim	Patricia		Office of the Privacy Commissioner of Canada	Ontario
11	Lasheras	Maria	V.	Nova Scotia Department of Health & Wellness	Nova Scotia
12	Meagher	Nancy	L.	Population Data BC/UBC	British Columbia
13	Miller	Janice	E.	Canadian Institutes of Health Research	Ontario
14	Ndupuechi	Lauren		McMaster University-CLSA	Ontario
15	Norman	Chris		BC Ministry of Health	British Columbia
16	Paul	Carol		Ontario Ministry of Health and Long-Term Care	Ontario
17	Raina	Parminder		McMaster University	Ontario
18	Ringuette	Louise		Commission d'accès à l'information du Québec	Québec
19	Robson	Paula	J.	Alberta Health Services - Cancer Care	Alberta
20	Smallwood	Sara		NB Department of Health	New Brunswick
21	Spaull	Alison	M.	Canadian Partnership for Tomorrow Project, CPAC	Ontario
22	Spinelli	John	J.	BC Cancer Agency	British Columbia
23	Wickham	Sarah	M.	Newfoundland and Labrador Centre for Health Information (NLCHI)	Newfoundland and Labrador
24	Wolfson	Christina		McGill University	Québec
25	Zawati	Ma'n	H.	Canadian Partnership for Tomorrow, CPAC	Québec
Via teleconference					
	Last Name	First Name		Institution	Province
26	Beck	Patty		Saskatchewan Ministry of Health	Saskatchewan
27	Labine	Lisa		Manitoba Health	Manitoba
28	Lacasse	André-Anne		Commission d'accès à l'information du Québec	Québec
29	MacDonald	Don		Newfoundland & Labrador Centre for Health Information	Newfoundland and Labrador
30	Messer-Lepage	Jacqueline		Saskatchewan Ministry of Health	Saskatchewan
31	Murdock	Melissa		British Columbia Ministry of Health	British Columbia
32	Payette	Yves		CARTaGENE	Québec
33	Peters	Debbie		New Brunswick Department of Health	New Brunswick

established: 1) build strong collaborative relationships between stakeholders involved in data sharing (e.g., researchers, data custodians, and privacy commissioners); 2) identify an entity which could provide overall leadership as well as individual “champions” within each province; 3) find adequate and long-term resources and funding; 4) clarify data linkage and data-sharing models and develop a common framework within which the data linkage process takes place; and 5) develop a pilot project making use of a limited number of linked variables from participating provinces. Over the course of the next 3 to 5 years, this action plan will be implemented, as linked data become more and more useful and relevant over the course of the study follow-up phases.

CONCLUSION

Linkage of administrative databases to the in-depth measures collected by large longitudinal cohorts provides a unique opportunity for creating what Roos et al. (2004) have described as “information-rich environments” for the Canadian research community.¹³ Given the current economic climate, governments and funding agencies are looking for ways to implement cost-efficient improvements to the health and well-being of the population and to continue to provide essential health care services. Indeed, jurisdictions around the world are beginning to establish integrated health information systems and infrastructures that maximize the value of health and health-related data.⁵ In Canada, successfully combining and analyzing cohort data and provincially managed administrative data will require innovative approaches to data linkage and cross-jurisdictional data sharing. The meeting of Canadian stakeholders is a first step to moving this agenda forward. If successful, a CLSA/CPTP-to-AHD linkage could indeed lead to policy-

relevant research results capable of improving Canadian clinical care and public health, and would undoubtedly benefit other longitudinal research projects in the future by setting a precedent for similar cost-efficient and powerful analytical tools.

REFERENCES

1. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, et al. The Canadian Longitudinal Study on Aging (CLSA). *Can J Aging* 2009;28(3):221-29.
2. Borugian MJ, Robson P, Fortier I, Parker L, McLaughlin J, Knoppers BM, et al. The Canadian Partnership for Tomorrow Project: Building a pan-Canadian research platform for disease prevention. *CMAJ* 2010;182(11):1197-201.
3. Holman CD, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et al. A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev* 2008;32(4):766-77.
4. Roos LL, Brownell M, Lix L, Roos NP, Walld R, MacWilliam L. From health research to social research: Privacy, methods, approaches. *Soc Sci Med* 2008;66(1):117-29.
5. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health* 2011;32(1):91-108.
6. Naess O, Sogaard AJ, Arnesen E, Beckstrøm AC, Bjertness E, Engeland A, et al. Cohort profile: Cohort of Norway (CONOR). *Int J Epidemiol* 2008;37(3):481-85.
7. El Emam K. Risk-based de-identification of health data. *Security & Privacy, IEEE* 2010;8(3):64-67.
8. Cavoukian A. Privacy by Design. Report of the Information & Privacy Commissioner Ontario, Canada, 2012.
9. Pencarrick Hertzman C, Meagher N, McGrail KM. Privacy by Design at Population Data BC: A case study describing the technical, administrative, and physical controls for privacy-sensitive secondary use of personal information for research in the public interest. *J Am Med Inform Assoc* 2013;20(1):25-28.
10. Karr AF, Fulp WJ, Vera F, Young SS, Lin X, Reiter JP. Secure, privacy-preserving analysis of distributed databases. *Technometrics* 2007;49(3):335-45.
11. Wolfson M, Wallace SE, Masca N, Rowe G, Sheehan NA, Ferretti V, et al. DataSHIELD: Resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data. *Int J Epidemiol* 2010;39(5):1372-82.
12. El Emam K, Samet S, Hu J, Peyton L, Earle C, Jayaraman GC, et al. A protocol for the secure linking of registries for HPV surveillance. *PLoS ONE* 2012;7(7):e39915.

13. Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment. *Soc Sci Med* 2004;58(11):2231-41.

Received: November 22, 2012

Accepted: February 15, 2013

RÉSUMÉ

Le couplage des données recueillies par de grandes études de cohortes canadiennes avec des données hébergées dans les bases de données administratives provinciales portant sur la santé peut offrir des pistes très intéressantes pour la multidisciplinarité et la rentabilité de la recherche en santé au Canada. L'analyse concomitante de données provenant d'études de cohortes et de bases de données administratives de santé peut conduire à des résultats de recherche pouvant faire progresser la santé et le bien-être des Canadiens et améliorer la prestation des services de soins de santé. Cependant, une telle initiative nécessitera beaucoup de coordination et un engagement à long terme entre tous les acteurs impliqués. Les défis et les opportunités d'un programme pancanadien de couplage de données ont été considérés par les chercheurs d'études de cohortes et les dépositaires de données de chaque province canadienne. Les parties prenantes reconnaissent l'importante contribution en santé publique d'établir un tel programme et ont mis en place un plan d'action pour aller de l'avant.

MOTS CLÉS : couplage de dossiers médicaux; collecte de données; santé publique