DISCUSSION AND DEBATE

Creating Clinical Cohorts: Challenges Encountered in Two Canadian Provinces

Constituer des cohortes cliniques : défis rencontrés dans deux provinces canadiennes

ESTHER S. SHOEMAKER, PHD
Bruyère Research Institute
University of Ottawa
Ottawa Hospital Research Institute
Institute of Clinical and Evaluative Sciences
Ottawa, ON

MARISSA L. BECKER, MD
Centre for Global Public Health
University of Manitoba
Winnipeg, MB

CLARE E. LIDDY, MD
Bruyère Research Institute
Department of Family Medicine
University of Ottawa
Ottawa, ON

LEIGH M. MCCLARTY, MSC
Centre for Global Public Health
University of Manitoba
Winnipeg, MB

SHABNAM ASGHARI, MD, PHD
Department of Family Medicine
Centre for Rural Health Studies
Memorial University of Newfoundland
St. John's, NL

JILLIAN HURD, MSC
Department of Family Medicine
Centre for Rural Health Studies
Memorial University of Newfoundland
St. John's, NL

SEAN B. ROURKE, PHD
Li Ka Shing Knowledge Institute
St. Michael's Hospital
Department of Psychiatry
University of Toronto
Toronto, ON

SOURADET Y. SHAW, PHD
Centre for Global Public Health
University of Manitoba
Winnipeg, MB

CHRISTINE BIBEAU
C.T. Lamont Primary Health Care Research Centre
Bruyère Research Institute
Ottawa, ON

RON ROSENES
C.T. Lamont Primary Health Care Research Centre
Bruyère Research Institute
Ottawa, ON

Creating Clinical Cohorts: Challenges Encountered in Two Canadian Provinces

PHILIP LUNDRIGAN
C.T. Lamont Primary Health Care Research Centre
Bruyère Research Institute
Ottawa, ON

LOIS CROWE
C.T. Lamont Primary Health Care Research Centre
Bruyère Research Institute
Ottawa, ON

LAURIE IRELAND, MD
Nine Circles Community Health Centre
Winnipeg, MB

CARLA LOEPPKY, PHD
Manitoba Health, Seniors and Active Living
Government of Manitoba
Winnipeg, MB

CLAIRE E. KENDALL, MD, PHD
Bruyère Research Institute
University of Ottawa
Institute of Clinical and Evaluative Sciences
Li Ka Shing Knowledge Institute
Ottawa, ON

Abstract

In 2013, the Living with HIV (LHIV) Innovation team established clinical cohorts of people living with HIV in Manitoba and Newfoundland and Labrador, and they linked the data to provincial health administrative databases. Access to these data enabled researchers to conduct studies across provincial borders; contribute to a national dialogue on HIV health system performance; and give recommendations for evidence-based healthcare, health policy and public health. However, research funding is episodic; maintaining cohorts requires stable funding. We support the establishment of a cross-jurisdictional approach to facilitate streamlined data collection and linkage without interruption and to allow for meaningful analysis in order to inform national policies.

Résumé

En 2013, l'équipe d'innovation Vivre avec le VIH constituait des cohortes cliniques de personnes vivant avec le VIH au Manitoba et à Terre-Neuve-et-Labrador. L'équipe a couplé ses données aux bases de données administratives provinciales. L'accès à ces données a permis aux chercheurs de mener des études transprovinciales, de contribuer au dialogue national sur le rendement du système de santé quant au VIH et de formuler des recommandations pour des politiques de santé, des soins et des mesures de santé publique éclairés par les données probantes. Toutefois, le financement de la recherche est sporadique, alors que la continuité d'une cohorte dépend de la stabilité du financement. Nous appuyons la mise en place d'une approche pangouvernementale afin de simplifier la cueillette et le couplage des données sans interruption et de permettre des analyses approfondies pour éclairer les politiques nationales.

To make evidence-informed policy and public health decisions, stakeholders, including federal, provincial and regional policy makers, practitioners and community organizations, require high-quality information and access to valid and reliable data to measure the impact of their decisions (Plsek and Greenhalgh 2001). In Canada, although there is a growing imperative to measure HIV health system performance (Johnston et al. 2015) and strengthen existing data management and infrastructure (Law et al. 2007), significant gaps remain with respect to comprehensive, population-level data that could guide resource allocation and healthcare delivery, including prevention, treatment and care for people living with HIV (PLWH). Despite the universality of coverage for medically necessary services delivered in hospitals or by physicians in Canada and the fact that the federal government transfers funding to the provinces and territories to support such coverage, the majority of these services are paid for directly by the provinces, and therefore, the majority of health administrative data are housed provincially, with varying degrees of completion and comprehensiveness (Widdifield et al. 2013). As the Council of Canadian Academies states, "in Canada ... those who need access to data must navigate a 'complex environment of heterogeneous entities,' often including numerous data custodians, privacy offices and research ethics boards, whose collective governance and operational practices fall short of constituting a well-defined and coherent system" (Council of Canadian Academies 2015).

Over the past five years, a number of research-funded initiatives focusing on different areas of health, such as the Canadian Longitudinal Study on Aging and the Canadian Partnership for Tomorrow Project, have sought to develop systems or platforms to link data collected through large cohort studies across different Canadian provinces to administrative health data (Doiron et al. 2013; Dummer et al. 2018). They are working to devise access protocols that respect jurisdictional requirements while supporting rigorous and effective harmonization practices (Fortier et al. 2017), but these are still nascent and are not yet in a position to provide researchers with access to multi-province data. It is thus heartening to see the Canadian Institutes of Health Research (CIHR) award $39 million over seven years to the Pan-Canadian Real-World Health Data Network (Smith et al. 2018) to create the Strategy for Patient-Oriented Research (SPOR) National Data Platform that can "receive multi-jurisdictional service requests and enable a single portal of access" to provincial and territorial health data, and "provide leadership and a forum to harmonize data access requirements and processes across jurisdictions" (CIHR 2017, 2019a), a sum which has recently been increased to $81 million by the federal government (CIHR 2019b).

In the absence of such a single portal of access, however, researchers wishing to establish cohorts in more than one province/territory that are linked with administrative health data must still work one jurisdiction at a time, which is more costly and less efficient. This paper describes how we used research funding to create linked cohorts in two provinces to address data gaps, opportunities and challenges encountered in doing so and sustainability of bolstering national capacity around the HIV health system performance by using this approach.

Cohorts as a Response for Measuring Health System Performance

Measurement of the HIV cascade, a framework that incorporates testing for HIV diagnosis, presentation to and engagement in care and successful treatment to reach full virologic suppression, is one example of health system performance and an urgent public health priority (Levi et al. 2016). To fully characterize the HIV cascade and implement actionable policies, it is essential to know who is accessing care, who is providing care and the extent to which care is optimized (Rice et al. 2018). However, there is no national consensus about whether administrative, clinical cohort or electronic health data sources best capture cascade data; who holds accountability regarding data integrity and validity; and who is responsible for resourcing comprehensive data collection and analysis for sustainability and monitoring/evaluation (National Collaborating Centre for Infectious Diseases 2017).

The Public Health Agency of Canada collates and reports provincial and territorial public health data on HIV-relevant variables, including age, sex, ethnicity, country of birth, geographic location, transmission risk groups and HIV viral loads (Public Health Agency of Canada 2017). Although critical to the HIV cascade picture, these data may not reflect the broader health complexities and health services needs of PLWH (Johnston et al. 2015). Administratively collected data, on the other hand, are limited to routinely collected health services data, with varied linkage to social and other data across jurisdictions, and these do not include important patient characteristics or health system indicators to facilitate policy and provider responses (Tu et al. 2014). This gap has led research teams in Ontario (Ontario HIV Treatment Network 2018), British Columbia and Quebec (Canadian Observational Cohort Collaboration [CANOC] 2008; Klein et al. 2010; Loutfy et al. 2017) to create provincial clinical cohorts of PLWH by comprehensively linking public health, administrative and cohort data.

Until recently, comparable cohorts did not exist in Manitoba or Newfoundland and Labrador. As part of a five-year CIHR–funded program of research, "Advancing Primary Health Care for Persons Living with HIV in Canada" (LHIV), and building on our Ontario experience, new clinical cohorts of PLWH were created in Manitoba and Newfoundland and Labrador and linked to provincial administrative databases. This linkage created a repository of population-level clinical, administrative and public health data that could facilitate comparisons across provinces and contribute to a national dialogue on the HIV health system performance. In addition to cascade measures, these data can be used to assess the proportion of patients using primary or specialized healthcare services given provincial geographic disparities and to provide insight into how services are being used and how they can be optimized.

TABLE 1. Summary of information obtained through clinical cohorts from Manitoba and Newfoundland and Labrador*

	Manitoba (*N* = 871) *n* (%)		**Newfoundland and Labrador (*N* = 251) *n* (%)**
Sex			
Male	620 (71.2)		189 (75.3)
Female	251 (28.8)		61 (24.3)
Other/missing			1 (0.4)
Age at diagnosis			
<25	126 (14.5)		34 (13.5)
25–34	308 (35.4)		78 (31.1)
35–44	235 (26.9)		55 (21.9)
45–54	136 (15.6)		36 (14.3)
≥55	66 (7.6)		14 (5.7)
Missing	0		34 (13.5)
Ethnicity			
Caucasian	373 (42.8)		180 (71.7)
Indigenous	358 (41.1)		–
African Caribbean Black	94 (10.8)		13 (5.2)
Asian	31 (3.6)		–
Other/missing	15 (1.7)		58 (23.1)

	Manitoba Regional Health Authorities		**Newfoundland and Labrador Health Regions**	
	Region of residence	***n* (%)**	**Region of residence**	***n* (%)**
	Southern	35 (4.0)	Central	21 (8.4)
	Prairie Mountain	33 (3.8)	Western	18 (7.2)
	Winnipeg + Churchill	00 (80.4)	Eastern	202 (80.5)
	Interlake-Eastern	48 (5.5)	Grenfell-Labrador	≤6
	Northern	33 (3.8)	Unknown	≤6
	Unknown	22 (2.5)		
Primary care access		621 (71.3)		232 (92.4)
Co-infections/co-morbidities				
Hepatitis C		152 (17.5)		6 (2.4)
Asthma/COPD		123 (14.1)		29 (11.6)
Hypertension		104 (11.9)		48 (19.1)
Type II diabetes		92 (10.6)		17 (6.8)
Ischemic heart disease		29 (3.3)		7 (2.8)
Receiving antiretroviral treatment		823 (94.5)		81 (32.3)
Suppressed viral load (<200 copies/ml)		128 (14.7)		66 (26.3)

* data from June 2017

Creation of LHIV Provincial Clinical Cohorts

Recruitment for Manitoba's clinical cohort began in October 2013. We sought consent during clinical encounters at two Winnipeg-based sites of the Manitoba HIV Program, the primary provider of HIV care in the province. Clinical data are regularly collected manually from charts or electronic medical records. Anonymized and de-identified clinical data are linked to provincial administrative health databases housed within a division of Manitoba Health, Seniors and Active Living. Manitoba is well positioned to undertake large population-based studies given its strong infrastructure of linkable, population-based administrative health databases (University of Manitoba 2017). Currently, data from 871 PLWH living in Manitoba are included in the cohort, representing the first comprehensive source of health data among PLWH in Manitoba. This cohort will facilitate more sophisticated epidemiological analyses that can inform HIV care programming and provincial policy, for example, by establishing which populations are currently underserved.

Similarly, a cohort of PLWH was developed in Newfoundland and Labrador in September 2013. Because the majority of PLWH are referred to and receive care at the St. John's nurse-practitioner–led HIV clinic, a clinical cohort was developed similar to the Manitoba cohort. Further, an iterative approach was used to extract variables from three different databases – provincial laboratory data, HIV clinic data and administrative health data – using validated algorithms (Nosyk et al. 2013). The extraction and compilation of these databases were approved and completed by a trusted third party, the Newfoundland and Labrador Centre for Health Information. The cohort currently includes data from 251 PLWH, including demographics, laboratory tests, hospitalization visits, physician claims, mortality, cancer and co-morbidity information, pregnancy status, smoking status, country of origin and medication use. With cohort data access, researchers in Newfoundland and Labrador are able to identify the number of PLWH in the province and how many are accessing care at the HIV clinic. Table 1 provides preliminary demographic and clinical data acquired from these cohorts.

Challenges to Cohort Creation

The development of these cohorts was not without challenges. In Manitoba, some potential participants expressed apprehension regarding the kinds of data collected and how they will be used and concerns regarding confidentiality. The idea of using provincial administrative health data for research is a new concept to many individuals, and the study staff spent considerable time explaining the processes through which the study is able to link clinical and administrative data while maintaining anonymity and confidentiality. Potential participants were assured that they would never be contacted by the research team outside of the clinic setting to maintain confidentiality and that access to the master list containing study identification numbers and names was highly secured and restricted to two members of the research team. At the institutional level, it is challenging to integrate the cohort enrolment protocols into established clinic operations. In the context of an over-burdened healthcare system, it is difficult to ask busy providers to incorporate additional procedures related to research studies into their routine encounters with clients. The Manitoba LHIV study team engages regularly with providers to emphasize the benefits of the cohort for their own practice and for Manitoba HIV Program's ability to meet the needs of its clients. Furthermore, because enrolment protocols are clinic-based, the clinical cohort under-represents PLWH who are sub-optimally engaged in care, thereby limiting the generalizability of cohort findings.

In Newfoundland and Labrador, there are risks to confidentiality among its predominantly rural population (Statistics Canada 2011), and a number of steps and ethical considerations had to be considered to ensure the cohort remained anonymous. Newfoundland and Labrador developed a data governance model that united researchers, data custodians, clinicians, patients, trainees and data experts, who met regularly to discuss any challenges and to develop mitigation strategies (Asghari et al. 2019). Inconsistencies across databases (provincial laboratory, clinic and health administrative data) posed another challenge, as PLWH were not necessarily identified by all three sources. Continuous data quality is a challenge for any clinical cohort; administrators and policy makers need to adopt the latest coding standards and diagnostic systems and ensure coders and physicians are trained to use them correctly (Nicholls et al. 2017). Some provincial databases were transitioning to an electronic medical system during cohort development, and the data of patients whose files had not yet been transferred may have been missed. Thus, the cohort was developed to be an interactive database that is retrospectively updated bi-annually with new PLWH data.

Our processes of stakeholder engagement, cohort development, data linkage and cross-system alignment of variable and outcome definitions have been undertaken entirely using research funds, as the research questions answered by the LHIV team could not have been addressed before the creation of clinical cohorts in these two provinces and their linkage with established administrative data. Leveraging research funds was beneficial during cohort development because it provided the flexibility of using innovative approaches guided by scientific evidence, but such funding is time-limited and therefore cannot ensure the sustainability of the cohorts.

Conclusion

Research funding has allowed the LHIV research team to develop clinical cohorts of PLWH in Manitoba and Newfoundland and Labrador, and it has created platforms for linking these cohorts to provincial health administrative databases. The established clinical cohorts will, for the first time, allow us to use rich individual-level clinical data to understand population-level healthcare delivery to PLWH in these provinces. One aim of the LHIV team was to compare HIV epidemics across provinces, but the unique provincial challenges experienced with respect to data collection and linkage, as well as the complexity of and time required to pool cohort data of multiple provinces into one repository, mean comparisons will have to be conducted at the aggregate level. Further, for these cohorts to be developed and maintained, a more stable and sustainable source of funding is required than research funding, which is time-limited and episodic. LHIV funding ends in 2020, and it is unclear how the linked cohorts we have established will be maintained thereafter. Without meaningful multi-stakeholder investment in data collection, linkage and analysis, system-level measures of comprehensive HIV performance measurement in Canada are unlikely in our foreseeable future (Low-beer et al. 2018). We urge stakeholders, including researchers, public health and stewards of administrative data at the provincial and national levels, to collaborate across silos, collect data in Canadian jurisdictions that are generally under-represented and commit to action on infrastructure such as the SPOR National Data Platform. This will facilitate rigorous harmonization across studies and will inform the direction of the evidence-based health system for the care of PLWH and other chronic conditions.

Correspondence may be directed to: Claire Kendall, Bruyère Research Institute, 43 Bruyere St., Annex E, Ottawa, ON K1N 5C7; e-mail: ckendall@uottawa.ca.

References

Asghari, S., S. Boyd, J. Blackmore, O. Hurley, J. Knight, D. Kelly et al. 2019. *Developing a Comprehensive Database with Sensitive Health Information: A Profile of People Living with HIV in Newfoundland and Labrador, Canada*. Vancouver, BC, Canada. In STI & HIV 2019 World Congress.

Canadian Institutes of Health Research (CIHR). 2017. *SPOR National Data Platform*. Retrieved: doi.org/10.1002/ejoc.201200111. <https://www.researchnet-recherchenet.ca/rnr16/vwOpprtntyDtls.do?prog=2794&view=currentOpps&org=CIHR&type=EXACT&resultCount=25&sort=program&next=1&all=1&masterList=true>.

Canadian Institutes of Health Research (CIHR). 2019a. *Funding Decisions Database – CIHR*. Retrieved April 25, 2019. <http://webapps.cihr-irsc.gc.ca/decisions/p/project_details.html?applId=385842&lang=en>.

Canadian Institutes of Health Research (CIHR). 2019b. *Minister of Health Announces $81M Initiative to Increase Access to Health Research Data – Canada.ca*. Retrieved April 25, 2019. <https://www.canada.ca/en/institutes-health-research/news/2019/04/minister-of-health-announces-81m-initiative-to-increase-access-to-health-research-data.html>.

CANOC. 2008. *CANOC – Canadian Observational Cohort Collaboration*. Retrieved January 6, 2018. <http://www.canoc.ca/>.

Council of Canadian Academies. 2015. Accessing Health and Health-Related Data in Canada. Ottawa, ON: Author.

Doiron, D., P. Raina and I. Fortier. 2013. "Linking Canadian Population Health Data: Maximizing the Potential." *Canadian Journal of Public Health* 104(3): 258–61.

Dummer, T.J.B., P. Awadalla, C. Boileau, C. Craig, I. Fortier, V. Goel et al. 2018. "The Canadian Partnership for Tomorrow Project: A Pan-Canadian Platform for Research on Chronic Disease Prevention." *The Canadian Medical Association Journal* 190(23): E710–17. doi.org/10.1503/cmaj.170292.

Fortier, I., P. Raina, E.R. Van den Heuvel, L.E. Griffith, C. Craig, M. Saliba et al. 2017. "Maelstrom Research Guidelines for Rigorous Retrospective Data Harmonization." *International Journal of Epidemiology* 46(1): 103–05. doi.org/10.1093/IJE/DYW075.

Johnston, S., C. Kendall, M. Hogel, M. Mclaren and C. Liddy. 2015. "Measures of Quality of Care for People with HIV: A Scoping Review of Performance Indicators for Primary Care." *PLoS One* 10(9): e0136757. doi.org/10.1371/journal.pone.0136757.

Klein, M.B., S. Saeed, H. Yang, J. Cohen, B. Conway, C. Cooper et al. 2010. "Cohort Profile: The Canadian HIV–Hepatitis C Co-Infection Cohort Study." *International Journal of Epidemiology* 39(5): 1162–69. doi.org/10.1093/ije/dyp297.

Law, S., C. Flood, D. Gagnon, and on behalf of the Listening for Direction III partners. 2007. *Listening for Direction III: National Consultation on Health Services and Policy Issues*. Ottawa, ON. Retrieved July 23, 2018. <https://www.cfhi-fcass.ca/sf-docs/default-source/listening-for-direction/LfDIII-FINAL_ENG.pdf?sfvrsn=1a1494d5_0>.

Levi, J., A. Raymond, A. Pozniak, P. Vernazza, P. Kohler and A. Hill. 2016. "Can the UNAIDS 90-90-90 Target be Achieved? A Systematic Analysis of National HIV Treatment Cascades." *British Medical Journal Global Health* 1(2): e000010. doi.org/10.1136/bmjgh-2015-000010.

Loutfy, M., A. de Pokomandy, V.L. Kennedy, A. Carter, N. O'Brien, K. Proulx-Boucher et al. 2017. "Cohort Profile: The Canadian HIV Women's Sexual and Reproductive Health Cohort Study (CHIWOS)." *PLoS One* 12(9): e0184708. doi.org/10.1371/journal.pone.0184708.

Low-beer, D., M. Mahy, F. Renaud and T. Calleja. 2018. "Sustainable Monitoring and Surveillance Systems to Improve HIV Programs: Review." *Journal of Medical Internet Research Public Health and Surveillance* 4(2): e3. doi.org/10.2196/publichealth.8173.

National Collaborating Centre for Infectious Diseases. 2017. *Reflections on the Development of the HIV Cascade Measures*. Winnipeg, MB.

Nicholls, S.G., S.M. Langan and E.I. Benchimol. 2017. "Routinely Collected Data: The Importance of High-Quality Diagnostic Coding to Research." *Canadian Medical Association Journal* 189(33): E1054–55. doi.org/10.1503/cmaj.170807.

Nosyk, B., G. Colley, B. Yip, K. Chan, K. Heath, V.D. Lima et al. 2013. "Application and Validation of Case-Finding Algorithms for Identifying Individuals with Human Immunodeficiency Virus from Administrative Data in British Columbia, Canada." *PLoS One* 8(1): e54416. doi.org/10.1371/journal.pone.0054416.

Ontario HIV Treatment Network. 2018. *OHTN Cohort Study (OCS)*. Retrieved February 1, 2018. <http://www.ohtn.on.ca/ohtn-cohort-study/>.

Plsek, P.E. and T. Greenhalgh. 2001. "Complexity Science: The Challenge of Complexity in Health Care." *British Medical Journal* 323(7313): 625–28.

Public Health Agency of Canada. 2017. *Can We Eliminate HIV? Canada Communicable Disease Report* 43(12). doi.org/10.14745/ccdr.v43i12a01.

Rice, B., T. Sanchez, S. Baral, P. Mee, K. Sabin, J.M. Garcia-Calleja and J. Hargreaves. 2018. "Know Your Epidemic, Strengthen Your Response: Developing a New HIV Surveillance Architecture to Guide HIV Resource Allocation and Target Decisions." *Journal of Medical Internet Research Public Health and Surveillance* 4(1): e18. doi.org/10.2196/publichealth.9386.

Smith, M., M. Schull, K. McGrail, A. Katz, B. Diverty, T. McDonald et al. 2018. "Building a Pan-Canadian Real World Health Data Network." *International Journal of Population Data Science* 3(2): doi.org/10.23889/ijpds.v3i2.550.

Statistics Canada. 2011. *Population, Urban and Rural, by Province and Territory (Newfoundland and Labrador)*. Retrieved October 23, 2017. <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo62b-eng.htm>.

Tu, K., T.F. Mitiku, N.M. Ivers, H. Guo, H. Lu, L. Jaakkimainen et al. 2014. "Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD)." *The American Journal of Managed Care* 20(1): e15–e21.

University of Manitoba. 2017. *Manitoba Population Research Data Repository Data List*. Retrieved January 23, 2018. <http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/datalist.html>.

Widdifield, J., S. Bernatsky, J.M. Paterson, K. Tu, R. Ng, J.C. Thorne et al. 2013. "Accuracy of Canadian Health Administrative Databases in Identifying Patients with Rheumatoid Arthritis: A Validation Study Using the Medical Records of Rheumatologists." *Arthritis Care & Research* 65(10): n/a-n/a. doi.org/10.1002/acr.22031.