

The SPOR-Canadian Data Platform: a national initiative to facilitate data rich multi-jurisdictional research

Lindsey Todd Dahl^{1*}, Alan Katz², Kimberlyn McGrail³, Brent Diverty⁴, Jean-Francois Ethier⁵, Frank Gavin⁶, James Ted McDonald⁷, P. Alison Paprica⁸, Michael Schull⁹, Jennifer D Walker¹⁰, and Juliana Wu¹¹

Submission History

Submitted:	18/06/2020
Accepted:	24/09/2020
Published:	09/11/2020

¹Manitoba Centre for Health Policy (MCHP), Rady Faculty of Health Sciences, Winnipeg, Manitoba R3E 3P5

²University of Manitoba, Departments of Community Health Sciences and Family Medicine; Director, Manitoba Centre for Health Policy (MCHP), Rady Faculty of Health Sciences, Winnipeg, Manitoba R3E 3P5

³Centre for Health Services and Policy Research, School of Population and Public Health, Vancouver, British Columbia V6T 1Z3

⁴Vice President, Programs Division, Canadian Institute for Health Information, Ottawa, Ontario K2A 4H6

⁵Associate professor, GRIIS, Université de Sherbrooke, Sherbrooke, Quebec J1K 2R1; Scientist, Centre de Recherche sur le vieillissement, 1036 Rue Belvédère S, Sherbrooke, Quebec J1H 4C4

⁶Public Advisory Council, Health Data Research Network Canada, Toronto, Ontario M4S 1M4

⁷Director, New Brunswick Institute for Research, Data and Training; Professor of Economics, University of New Brunswick, Fredericton, New Brunswick E3B 5A3

⁸Executive Advisor and Affiliate Scientist, Institute for Clinical Evaluative Sciences (ICES), 2075 Bayview Ave, Toronto, Ontario M4N 3M5

⁹CEO, Institute for Clinical Evaluative Sciences (ICES), 2075 Bayview Ave, Toronto, Ontario M4N 3M5; Senior Scientist, Evaluative Clinical Sciences, Trauma, Emergency & Critical Care Research Program, Sunnybrook Research Institute, 2075 Bayview Ave, Toronto, Ontario M4N 3M5; Professor, University of Toronto, Institute for Health Policy Management and Evaluation, 155 College Street, Suite 425, Toronto, Ontario M5T 3M6

¹⁰Indigenous Lead, Institute for Clinical Evaluative Sciences (ICES), 2075 Bayview Ave, Toronto, Ontario M4N 3M5; Canada Research Chair in Indigenous Health, School of Rural and Northern Health, Laurentian University, Sudbury Ontario P3E 2C6

¹¹Manager, Corporate Data Request Program, Canadian Institute for Health Information (CIHI), Toronto, Ontario M2P 2B7, jwu@cihi.ca

Abstract

Administrative health data is recognized for its value for conducting population-based research that has contributed to numerous improvements in health. In Canada, each province and territory is responsible for administering its own publicly funded health care program, which has resulted in multiple sets of administrative health data. Challenges to using these data within each of these jurisdictions have been identified, which are further amplified when the research involves more than one jurisdiction. The benefits to conducting multi-jurisdictional studies has been recognized by the Canadian Institutes of Health Research (CIHR), which issued a call in 2017 for proposals that address the challenges. The grant led to the creation of Health Data Research Network Canada (HDRN), with a vision is to establish a distributed network that facilitates and accelerates multi-jurisdictional research in Canada. HDRN received funding for seven years that will be used to support the objectives and activities of an initiative called the Strategy for Patient-Oriented Research Canadian Data Platform (SPOR-CDP). In this paper, we describe the challenges that researchers face while using, or considering using, administrative health data to conduct multi-jurisdictional research and the various ways that the SPOR-CDP will attempt to address them. Our objective is to assist other groups facing similar challenges associated with undertaking multi-jurisdictional research.

Keywords

administrative data; cross-jurisdictional; data platform; health research; population

*Corresponding Author:

Email Address: lindsey_dahl@cpe.umanitoba.ca (Lindsey Todd Dahl)



Introduction

The aims of health research are to advance knowledge and influence decision making in order to improve the health of individuals and populations. While there is no universal definition of health, modern interpretations commonly refer to the interactions among the physical, mental, and spiritual states of individuals as they exist within their social and environmental contexts [1]. In addition, the relative value or importance placed on each of these states can vary considerably from individual to individual as well as over time within individuals, illustrating the dynamic nature of how we perceive health [2], and the importance of engaging the public and patients as active participants in the research process. Given the multi-dimensional aspects underpinning the concept of health, it is not surprising that health research draws upon a variety of experimental and observational approaches.

One approach to health research that has contributed significantly to health improvements over the past 30 years focuses on the use of population-level data routinely collected when individuals interact with the health system, which contain individual-level information on healthcare services use (e.g., hospitalizations, prescription drug dispensations) and demographics [3, 4]. Using these data for multi-jurisdictional research, defined as research that accesses data from more than one jurisdiction, can overcome some of the limitations of single jurisdiction studies (e.g., studying rare conditions, or comparing health care system outcomes), but it does not come without challenges [5–8]. Addressing the challenges is acknowledged as a necessary step towards facilitating this type of research as various countries are developing the infrastructure to overcome them [7–11].

In Canada, the Health Data Research Network Canada (HDRN) received funding in 2018 to develop the Strategy for Patient Oriented Research Canadian Data Platform (SPOR-CDP) that will address the challenges of multi-jurisdictional research in Canada. The purpose of this paper is to describe the challenges multi-jurisdictional research that are faced by countries around the world, and how the SPOR-CDP initiative will address them in Canada with a specific focus on the activities during its first year of existence.

Administrative data in Canada and its application to research

A wealth of administrative data exists in Canada that is collected from the publicly-funded universal health care insurance legislated under the Canada Health Act [12], as well as from additional services provided at the discretion of jurisdictions [13]. While these data are primarily collected through the administration of health insurance programs [14], they possess many attractive features in the application of health research [15]. First, a central component of the administrative data is the health insurance registries, which are maintained in each of Canada's thirteen provinces or territories (hereinafter referred to as jurisdictions). These registries contain demographic information on virtually all residents in the provincial or territorial health systems which provides denominators for population-based studies [15]. Those individuals whose health information is not captured in

the provincial or territorial data include members of the Royal Canadian Mounted Police, Canadian Armed Forces, federal prison inmates, short-term temporary foreign workers and those who have received Indigenous health services provided by the federal government [14].

Second, an important element of the registries is the personal identifying numbers assigned to each resident that may be scrambled or encrypted to protect privacy and serve as a mechanism to link data across various domains providing contextual information about individuals (such as data from the social and education sectors, clinical trials, longitudinal cohorts, and electronic medical records) or geographical areas in addition to the health administrative data [16, 17]. These linkages provide an efficient way to capture population-based information on the 'real-world' conditions in which patients live; this is in contrast to purpose-built data collection, such as large-scale longitudinal studies, surveys, and censuses that involve more investment of time and resources [18, 19].

Third, administrative data are routinely collected allowing an individual's experience to be followed over time. This simplifies issues related to measurement and loss to follow-up that researchers contend with while conducting studies with primary data collection [20].

These features enable research using administrative data to be conducted for a variety of purposes, such as studying diseases of low prevalence, including disease prediction and surveillance, measuring the safety and effectiveness of an intervention, exploring outcomes from domains outside of health (e.g., education, social), and measuring exposure effects over time [15, 21–23]. While the advantages associated with administrative data make them an attractive option for conducting health and social research, administrative data are not without challenges, which are amplified when conducting research across multiple jurisdictions.

Challenges associated with multi-jurisdictional research

Theoretically, administrative data provide an opportunity to conduct research on a national scale or to make comparisons across jurisdictions. However, in practice this is not straightforward. In Canada, the Canadian Constitution ensures that each province and territory has a significant level of independence for operationalising and administering its own health programs [12]. In addition, federally funded and administered health services also exist for several populations, including Indigenous Peoples, notably registered First Nations people and Inuit. Sometimes, these services are administered through local Indigenous governance, which adds additional complexity to multi-jurisdictional work. This has essentially resulted in multiple distinct sets of administrative health data and led to differences in the availability and structure of data and in the processes involved in accessing data from each jurisdiction. Legal and policy barriers often create obstacles to sharing or pooling these data across jurisdictions. These differences and barriers pose a challenge to researchers and ultimately result in many not choosing to take on multi-jurisdictional research studies despite their potential benefit in academic and policy arenas [4, 24].

A recent paper explored challenges associated with cross-centre or multi-jurisdictional research [5]. The challenges identified, consistent with previous literature [5, 8, 24], can be generalized into three categories: (i) data access challenges, (ii) analytical challenges (including data organization and comparability), and (iii) culture of academia and data governance challenges.

Data access challenges

Challenges associated with gaining access to administrative data represent a major barrier for researchers. Administrative data are usually de-identified when used for research purposes meaning that any identifying or potentially identifying information about an individual has been removed in a way that minimizes the likelihood that an individual's identity can be determined by any reasonably foreseeable method. While these studies are considered low risk to individuals whom the data represents by ethics boards, researchers must meet a set of regulatory and governance requirements to ensure appropriate use of data, including protection of the confidentiality and privacy of the information and demonstration of the potential for public value. These regulations are in place to ensure that the data are appropriately safeguarded, but have also been criticized for being too restrictive, with lengthy and unpredictable approval times [8, 25]. There is also inconsistent application of regulations between institutions and of privacy requirements, even within a single jurisdiction. Data centres within some provinces have developed processes to facilitate access in their own jurisdiction; however, there is no harmonizing of these processes within and across jurisdictions. Therefore, researchers interested in conducting multi-jurisdictional studies using administrative data must navigate separate requirements and processes for each jurisdiction that will be providing data, which is cumbersome and further delays the time to access the data.

Conditions of data access usually require researchers to describe the procedures in place to meet jurisdictional requirements obviating the need for individual consent, including ensuring the privacy and confidentiality of the data and, where appropriate, respecting principles of Indigenous data governance [26–29]. This requires an understanding of the information contained in the datasets used in their research and limiting linkages to other data to ensure that there is no risk of re-identification [15, 30]. For Indigenous data, it also requires relationships, reciprocity and respect for diverse Indigenous data governance, and linkages that may be in place in different jurisdictions [31]. Understanding the data is also necessary to know whether the data are adequate for answering the specified research question. However, since researchers are typically not responsible for the collection of the data, and since research is not the primary purpose for collecting the data, understanding what information is contained in datasets and how it is collected or recorded poses a challenge, especially in the instance of multi-jurisdictional research. Many of these data intricacies are recorded in metadata; however, some jurisdictions do not have metadata information to accompany their datasets, and there is no standardized approach to generating and reporting metadata for the jurisdictions that do. Therefore, it is difficult for

researchers to know if (a) specific variables of interest are available in certain jurisdictions or (b) whether a variable label means the same thing across jurisdictions. Even with access to metadata, administrative data are complex and their use often requires an in-depth understanding of the data (including data provenance) and the systems that they are generated from, which often comes from experience.

Analytical challenges

The challenge of analytics stems from the way the data are organized and that has implications for designing and conducting multi-jurisdictional research using administrative data. Each jurisdiction is responsible for the administration of its health program, so there is no requirement that the data be collected or stored in a standard way across jurisdictions (and sometimes even within jurisdictions). This leads to differences in the way data are structured and variables are defined and in the availability of data across jurisdictions, all of which make it difficult to make appropriate comparisons [5]. For example, some jurisdictions limit the number of diagnoses collected in physician billing data which results in “missing” diagnoses, or payment policies might influence physicians to record more of some diagnoses than others; and differences in the number of digits collected for physician billing claims between provinces mean analysts need to develop creative ways to ensure consistency in outcomes [24]. Furthermore, there may be changes in the data quality or to the existence of data elements over time, which makes it difficult to provide and compare disease estimates (for example) from year to year [8, 32].

Regulatory requirements for data to remain in their original jurisdiction pose a practical challenge to analyses involving multi-jurisdictional data by prohibiting the pooling of data in many cases and impose distributed analyses [24]. However, preparing comparable datasets in each jurisdiction may not be possible, or running them in a distributed mode may have analytical implications, such as diminished power. All of these barriers and challenges impact the pace of the research and the amount of resources that are required as a considerable amount of time is spent preparing the data, especially if study personnel do not have adequate experience and familiarity with the data sets [24].

Culture of academia/data governance challenges

The competitive culture and funding model of the health research enterprise does not incentivize research that explicitly addresses these challenges [33]. In highly competitive research, funding environments researchers may be reluctant to take the risk of proposing research that is more complex. The added costs associated with addressing the challenges above, and their similarity to infrastructure may also discourage researchers for fear of requiring more funding than would seem acceptable by review panels. The lack of specific mechanisms and incentives, such as dedicated funding to support multi-jurisdictional research, may be seen as a barrier to undertaking this type of research.

Addressing the challenges in Canada

The opportunity

The Pan-Canadian Real-World Health Data Network (PRHDN) was established in 2015 as an informal collaborative network aimed at increasing multi-jurisdictional research in Canada [34]. The idea of such a network in Canada began from discussions between three different data-holding organizations (Canadian Institute of Health Information, Manitoba Centre for Health Policy, and ICES, formerly the Institute for Clinical Evaluative Sciences) around the possibility of more coordinated pan-Canadian data-intensive health research. At that time, the links between provincial, national, and territorial data-holding organizations were relatively weak, with limited one-off collaborative projects versus a comprehensive approach toward multi-jurisdictional research based on population-wide administrative data holdings. The idea and benefits of a network that facilitated multi-jurisdictional research in Canada were presented at research conferences and to stakeholders, which helped to shape PRHDN's objectives and implementation plan.

In December 2017, recognizing the potential benefit of conducting multi-jurisdictional research using administrative data and as part of the Canadian Strategy for Patient-Oriented Research (SPOR) [35], the Canadian Institutes of Health Research (CIHR) issued an open call for research proposals to create a platform – the SPOR Canadian Data Platform (SPOR-CDP) – in order to facilitate access to routinely collected data and accelerate multi-jurisdictional research, with a focus on patient-oriented research. The early work in setting up the PRHDN, and the establishment of processes to agree and act on priorities, positioned the PRHDN well to develop a proposal in response to the CIHR call. In 2018, CIHR announced a seven-year, \$81.35 million CAD grant (\$39 million from CIHR, the remaining from matching partner contributions) to the PRHDN to establish the SPOR-CDP.

Upon receipt of grant funding, PRHDN incorporated as a not-for-profit organization and rebranded as Health

Data Research Network Canada (HDRN) in January 2020. HDRN currently consists of 18 organizations from across Canada, including SPOR Support Units, which assist in the facilitation of patient-oriented research [36], and provincial and national data centres, which act as stewards of the data held within their jurisdiction [37]. HDRN is governed by an independent Board of Directors and Scientific Director/Chief Executive Officer and supported by an Executive Committee. Representatives from the 18 organizations populate working groups and teams, which meet monthly or more frequently to coordinate the work of the network and discuss issues and opportunities.

The SPOR-CDP

The overall vision of the SPOR-CDP is to become a distributed network that facilitates and accelerates multi-jurisdictional research in Canada and will focus on the seven objectives (Box 1) identified by HDRN to achieve this vision [38]. There are numerous activities to be undertaken for each objective in order for the full vision of the SPOR-CDP to be realized. Table 1 connects the objectives and their activities to each of the three general categories of challenges described above. While there has been much discussion and collaboration around appropriately carrying out the activities dedicated to each objective, the most visible or tangible activities during the first year apply to the first two listed objectives: creating a data access support system that helps navigate multi-jurisdictional requests and harmonizing and validating definitions for important chronic diseases and other key analytic variables.

A major accomplishment during the first year has been the creation and public launch of an online portal called the Data Access Support Hub (DASH) [39]. The DASH serves as the central location that researchers may go to for information and/or support pertaining to multi-jurisdictional research using administrative data in Canada. In an effort to address the challenges associated with accessing data, DASH

Box 1. SPOR-CDP objectives aimed at facilitating and accelerating multi-jurisdictional research

1. Create a data access support system that helps navigate multi-jurisdiction requests
2. Harmonize and validate definitions for important chronic diseases and other key analytic variables
3. Continue to expand the sources and types of data and linkages available through PRHDN organizations, including linkage to clinical and social data
4. Develop the technology infrastructure required to improve the data access request process as well as the documentation, storage, and re-use of algorithms and existing data
5. Create supports for advanced analytics and infrastructure for data collection and analysis
6. Establish strong partnerships with patients and the public and with Indigenous communities
7. Build strong governance and enable national coordination

Abbreviations: PRHDN, Pan-Canadian Real-World Health Data Network; SPOR-CDP, Strategy for Patient Oriented Research Canadian Data Platform.

Table 1: SPOR-CDP action during the first year that address the challenges of multi-jurisdictional research using administrative data

Challenge category	Specific challenge	SPOR-CDP objective	SPOR-CDP action
Data Access	<ul style="list-style-type: none"> • Restrictive policies • Lengthy, inconsistent approval processes • Different administrative processes • Lack of harmonization in data sharing laws across jurisdictions • Limited capacity to share and use data across jurisdictional boundaries • Differences in data provider requirements to obtain data access across jurisdictions • Multiple data provider involvement 	<ol style="list-style-type: none"> 1. Create a data access support system that helps navigate multi-jurisdiction requests 	<ul style="list-style-type: none"> • Central intake, coordination and support to researchers via DASH • Access Processes Inventory (DASH)
Analytic (Analysis/Availability/Data Management)	<ul style="list-style-type: none"> • No standardized definitions across jurisdiction • Inconsistencies among variables and indicators making it difficult to compare across jurisdictions • Data heterogeneity across jurisdictions • Absence of metadata and standards • Differences in data availability across provinces • Inability to make comparisons across jurisdictions - comparable data to create similar patient cohorts and measurements • Technical infrastructure to allow sharing to occur • Data compatibility for combined analyses • Data cannot be aggregated directly from multiple jurisdictions • Differences in data structure • Time required for data preparation • Coding differences • Restrictive and different data formats across jurisdictions • Changes in data quality over time • Some data not retained over time 	<ol style="list-style-type: none"> 2. Harmonize and validate definitions for important chronic diseases and other key analytic variables 3. Continue to expand the sources and types of data and linkages available through HDRN organizations, including linkage to clinical and social data 4. Develop the technology infrastructure required to improve the data access request process as well as the documentation, storage, and re-use of algorithms and existing data 5. Create supports for advanced analytics and infrastructure for data collection and analysis 	<ul style="list-style-type: none"> • Algorithm Inventory (DASH) • New Algorithm Development • Data Holdings Inventory (DASH) • Metadata Standards



Table 1: (continued)

Challenge category	Specific challenge	SPOR-CDP objective	SPOR-CDP action
Culture	<ul style="list-style-type: none"> • Disagreement on data uses • Achieving good balance between the need for individual privacy and the public good • Lack of trust and reciprocity • Nurturing appropriate socio-technical systems to support data-intensive science • Absence of guidelines on ownership and copyright • Perceived lack of information on cross-centre working in general and knowing people of similar research interests • Motivated by attracting new monies for organizations in order to raise their profile • Actively facilitating sharing • Academic institutional forces • Incentivised to publish papers • Disincentivised to share data • Lack of dedicated funding for cross-centre working • Dedicated funding and incentivising data custodians to share information • Lack of resources 	<p>6. Establish strong partnerships with patients and the public and with Indigenous communities</p> <p>7. Build strong governance and enable national coordination</p>	Initiation of public engagement, Indigenous engagement, and a focus on stakeholder relations

Abbreviations: DASH, Data Access Support Hub; HDRN, Health Data Research Network Canada.

provides a one-stop concierge service that helps researchers address the approval requirements for each jurisdiction.

Investigators initiate the process by submitting a central intake form, which provides an overview of their intended study and is reviewed for eligibility and feasibility by DASH staff associated with each intended study site/data centre. Consultations between the investigators and the data centres are coordinated by central DASH staff to ensure comprehensive and efficient information sharing, as opposed to investigators needing to identify their own contacts and navigate the access processes in each data centre. Local DASH staff prepare a cost estimate for the data access and analytic services for each data centre in the applicable jurisdictions. While the DASH coordinating and support services are currently free, the local data centre specific cost estimates for data extraction and/or analytic services apply for data access and use. Next, the investigator submits a data access request form to DASH, which coordinates the data access requirements, including facilitating ethical approvals, provider approvals, and other data access agreements. Currently, there are multi-jurisdictional studies underway utilizing the DASH services described above, which will serve as opportunities to review and improve the DASH processes. As data centres work together to support investigators through DASH, opportunities

exist to learn from each other's practices, and align aspects of the data access processes where feasible and desirable. Ultimately, the aim is to establish a streamlined and automated process for handling multi-jurisdictional data access requests.

The DASH site also includes three interactive inventories that provide valuable information that may be used by researchers and the general public. First, the data access processes in each jurisdiction have been mapped and compiled in an inventory that gives an overview of the requirements and enables site visitors to compare and contrast requirements between jurisdictions. A big advantage is that the terminology used to describe the processes has been harmonized so the same word will have the same meaning when used in different jurisdictions. Second, a searchable inventory of all the available data holdings across Canada is available. Presently, there are over 380 health, social, survey, and registry datasets listed in the DASH inventory. Third, there is a list of algorithms for identifying health conditions, health services use outcomes, and determinants of health that have been either validated or assessed for feasibility in more than one jurisdiction. The algorithms included were obtained from a systematic search of the published and grey literature, which will be revisited twice a year as a way to ensure the inventory remains current. Presently, there are over 120 algorithms in the inventory.

These three inventories contribute to meeting the first and second SPOR-CDP objectives as they provide valuable assets for researchers who are in the planning or exploratory stages of their research projects. These inventories will grow over time as more jurisdictions provide information about their data holding or expand their data holdings, and new algorithms for high priority conditions or measures are developed.

Analytic support will be addressed through activities related to SPOR-CDP objectives two through five, which involve data harmonization and validation of analytic variables, expanding the availability of data sources and linkages, developing the technological infrastructure required to improve the documentation, storage, and re-use of algorithms and existing data, and creating supports for advanced analytics and infrastructure for data collection and analysis. While the first year activities have focused primarily on laying the foundation for multi-jurisdictional data access, the SPOR-CDP is also committed to providing supports that will enhance distributed analyses. First, the SPOR-CDP aims to develop metadata for all of the available data holdings across the participating centres, which will benefit researchers and analysts as they use the data. The online inventory of data holdings is a starting point but this needs to be expanded to include comprehensive information on the contents of data sets. The current inventory of data holdings contains mostly free text information, but the addition of standardized metadata will enable better evaluation of comparability across jurisdictions and lay the foundation for more automation of processes including distributed analyses. Finally, the SPOR-CDP will continue to increase the availability of data sources and types of linkages, such as social and electronic medical records data, which will open new doors for more multi-jurisdictional research for researchers not only in health but also in other fields of study.

Addressing the practical and technical challenges will undoubtedly make it easier to do multi-jurisdictional research. However, in order for more of this type of research to occur, a shift in the culture will also need to accompany these improvements. Exploitation of the rich data held in the multiple jurisdictions and the significant potential value of the natural experiments they represent will take time. Most researchers are not aware of the opportunities or advantages offered in multi-jurisdictional research using administrative data. In addition, the collaborative relationships necessary to fully exploit the richness of the data will initially require deliberate and thoughtful dialogue in consideration of both real and perceived barriers to data access, while meeting jurisdictional legislative requirements and local established policies and procedures. The SPOR-CDP objectives and related activities are not explicitly aimed at changing the culture; however, there is an underlying focus on fostering collaborative relationships and partnerships across the country, which may nudge a change more organically.

In addition to incorporating and establishing a formal Board of Directors, HDRN has created a SPOR-CDP Advisory Committee, created capacity at each of our centers to respond to researcher requests (through DASH), and is developing new partnerships with important national health data initiatives in Canada and internationally, such as the Canadian Partnership for Tomorrow's Health [40], and Health Data Research-UK [9]. Important progress has also been made

in other priority areas such as engagement with patients and public, and with Indigenous communities. This engagement will build on existing relationships and data governance agreements at HDRN organizations. A Public Advisory Council has been established, which aims to engage with patients and public groups who are and are not directly associated with other SPOR entities in a continuous, substantial, and meaningful way. The development of the SPOR-CDP has also prioritized Indigenous data governance so that Indigenous-led research is supported and the use of Indigenous-identified data is respected. This has involved compiling the diverse Indigenous data governance requirements across jurisdictions and developing relationships with Indigenous partners. This will enable the SPOR-CDP to support the diverse Indigenous governance and health services organizations in their vision for harmonized Indigenous-led health research across multiple jurisdictions.

As the SPOR-CDP progresses and continues to evolve, the meaningful relationships with individuals, organizations, and communities will require purposeful strategic investments of time and energy to build the necessary trust. The relationships that are developed are a necessary component of multi-jurisdictional research, which may also catalyze a cultural shift where this type of research is more appropriately acknowledged and recognized.

Conclusion

There is enormous potential for multi-jurisdictional research using administrative data to continue to provide novel insights that improve the health and well-being of populations and individuals. There are many challenges that have been identified that must be overcome in order to move this potential into reality. Recent funding to support the development of the SPOR-CDP will facilitate processes for conducting administrative health data research across Canada and eventually international research. The activities pertaining to the ambitious SPOR-CDP objectives will progress over time and many will be used to inform other research. This paper documents some of the activities initiated during the first year that address some, but not nearly all, of the challenges of multi-jurisdictional research. The particular activities thus far have focused primarily on important first steps, such as improving data access processes and describing existing data. While the work in these areas will continue, new activities will also begin that will move the SPOR-CDP closer to establishing itself as a comprehensive solution to the challenges that face multi-jurisdictional research in Canada. Careful documentation of the activities and approaches undertaken in the development of the SPOR-CDP will be a valuable resource for other countries facing similar challenges. Revisiting the table provided in this paper each year can provide a simple way to observe and monitor the SPOR-CDP's progress from accepting to overcoming the challenge.

Acknowledgements

This work was supported by a grant from the Canadian Institutes of Health Research (FRN NDP-160882).

Statement on conflicts of interest

The authors declare there are no conflicts of interest

Ethics statement

As this is a commentary paper, ethical approval is not required.

References

- Svalastog AL, Donev D, Kristoffersen NJ, Gajović S. Concepts and definitions of health and health-related values in the knowledge landscapes of the digital society. *Croat Med J*. 2017;58(6):431–5. Available from: <https://dx.doi.org/10.3325/cmj.2017.58.431>
- Bircher J. Towards a dynamic definition of health and disease. *Med Heal Care Philos*. 2005;8(3):335–41. Available from: <http://dx.doi.org/10.1007/s11019-005-0538-y>
- Lucyk K, Lu M, Sajobi T, Quan H. Administrative health data in Canada: Lessons from history. *BMC Med Inform Decis Mak [Internet]*. 2015;15(1):1–6. Available from: <http://dx.doi.org/10.1186/s12911-015-0196-9>
- Martens PJ. Straw into gold: Lessons learned (and still being learned) at the Manitoba Centre for Health Policy. *Healthc Policy*. 2011;6(SPEC. ISSUE):44–54. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5319572/>
- Jones K, Heys S, Daniels H, Ford D. Exploring barriers and solutions in advancing cross-centre population data science. *Int J Popul Data Sci*. 2019;4(1). Available from: <https://dx.doi.org/10.23889/ijpds.v4i1.1109>
- Palamuthusingam D, Johnson DW, Hawley C, Pascoe E, Fahim M. Health data linkage research in Australia remains challenging. *Intern Med J*. 2019;49(4):539–44. Available from: <http://dx.doi.org/10.1111/imj.14244>
- Flack F, Smith M. Access to linked data: challenges and progress. *Intern Med J*. 2019;49(10):1339–40. Available from: <https://dx.doi.org/10.1111/imj.14455>
- Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health*. 2014;14(1):1–9. Available from: <https://dx.doi.org/10.1186/1471-2458-14-1144>
- Health Data Research UK [Internet]. [cited 2020 Jun 9]. Available from: <https://www.hdruk.ac.uk/>
- Flack F, Smith M. The population health research network - Population data centre profile. *Int J Popul Data Sci*. 2019;4(2). Available from: <https://doi.org/10.23889/ijpds.v4i2.1130>
- O'Reilly D, Bateson O, McGreevy G, Snoddy C, Power T. Administrative Data Research Northern Ireland (ADR NI). *Int J Popul Data Sci*. 2020;4(2):1–9. Available from: <https://doi.org/10.23889/ijpds.v4i2.1148>
- Canada Health Act. R.S.C. 1985 p. c. C-6.
- Lewis S, Donaldson C, Mitton C, Currie G. The future of health care in Canada. *BMJ*. 2001;323(7318):926–9. Available from: <http://dx.doi.org/10.1136/bmj.323.7318.926>
- Quan H, Smith M, Bartlett-Esquilant G, Johansen H, Tu K, Lix L. Mining Administrative Health Databases to Advance Medical Science: Geographical Considerations and Untapped Potential in Canada. *Can J Cardiol*. 2012;28(2):152–4. Available from: <https://doi.org/10.1016/j.cjca.2012.01.005>
- 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. Available from: <http://dx.doi.org/10.1146/annurev-publhealth-031210-100700>
- Guttmann A. The SPOR Canadian Data Platform: opportunity for multi-provincial research. *Can Med Assoc J*. 2019;191(40):E1091–2. Available from: <http://dx.doi.org/10.1503/cmaj.191040>
- Mazzali C, Duca P. Use of administrative data in healthcare research. *Intern Emerg Med [Internet]*. 2015;10(4):517–24. Available from: <http://dx.doi.org/10.1007/s11739-015-1213-9>
- Lyons RA, Ford D V., Moore L, Rodgers SE. Use of data linkage to measure the population health effect of non-health-care interventions. *Lancet*. 2014;383(9927):1517–9. Available from: [https://dx.doi.org/10.1016/S0140-6736\(13\)61750-X](https://dx.doi.org/10.1016/S0140-6736(13)61750-X)
- Orr J, Smith M, Burchill C, Katz A, Fransoo R. Outcomes of an investment in administrative data infrastructure: An example of capacity building at the Manitoba Centre for Health Policy. *Can J Public Heal*. 2016;107(4–5):e480–1. Available from: <http://dx.doi.org/10.17269/cjph.107.5659>
- Roos LL, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: Comparisons with primary data collection. *J Chronic Dis*. 1987;40(1):41–9. Available from: [http://dx.doi.org/10.1016/0021-9681\(87\)90095-6](http://dx.doi.org/10.1016/0021-9681(87)90095-6)
- Daneman N, Chateau D, Dahl M, Zhang J, Fisher A, Sketris IS, et al. Fluoroquinolone use for uncomplicated urinary tract infections in women: a retrospective cohort study. *Clin Microbiol Infect [Internet]*. 2020 May;26(5):613–8. Available from: <https://doi.org/10.1016/j.cmi.2019.10.016>
- Bernatsky S, Lix L, Hanly JG, Hudson M, Badley E, Peschken C, et al. Surveillance of systemic autoimmune rheumatic diseases using administrative data. *Rheumatol Int*. 2011;31(4):549–54. Available from: <http://dx.doi.org/10.1007/s00296-010-1591-2>

23. Quinn RR, Laupacis A, Hux JE, Oliver MJ, Austin PC. Predicting the risk of 1-year mortality in incident dialysis patients: Accounting for case-mix severity in studies using administrative data. *Med Care*. 2011;49(3):257–66. Available from: <http://dx.doi.org/10.1097/MLR.0b013e318202aa0b>
24. Katz A, Enns J, Wong ST, Williamson T, Singer A, McGrail K, et al. Challenges Associated with Cross-Jurisdictional Analyses using Administrative Health Data and Primary Care Electronic Medical Records in Canada. *Int J Popul Data Sci*. 2018;3(3). Available from: <https://dx.doi.org/10.23889/ijpds.v3i3.437>
25. Lugg-Widger F, Angel L, Cannings-John R, Hood K, Hughes K, Moody G, et al. Challenges in accessing routinely collected data from multiple providers in the UK for primary studies: Managing the morass. *Int J Popul Data Sci*. 2018;3(3):2–14. Available from: <https://dx.doi.org/10.23889/ijpds.v3i3.432>
26. Walker J, Lovett R, Kukutai T, Jones C, Henry D. Indigenous health data and the path to healing. *Lancet* [Internet]. 2017;390(10107):2022–3. Available from: [http://dx.doi.org/10.1016/S0140-6736\(17\)32755-1](http://dx.doi.org/10.1016/S0140-6736(17)32755-1)
27. Inuit Tapiriit Kanatami. National Inuit Strategy on Research. 2018. 48 p. Available from: <https://www.itk.ca/wp-content/uploads/2018/03/National-Inuit-Strategy-on-Research.pdf>
28. First Nations Information Governance Centre. The First Nations Principles of OCAP [Internet]. 2020 [cited 2020 Jun 8]. Available from: <https://fnigc.ca/ocap>
29. Kukutai T, Taylor J. Indigenous Data Sovereignty: Toward an Agenda. Anu Press; 2016. 585 p. Available from: <https://press-files.anu.edu.au/downloads/press/n2140/pdf/book.pdf>
30. Cumyn A, Barton A, Dault R, Cloutier AM, Jalbert R, Ethier JF. Informed consent within a learning health system: A scoping review. *Learn Heal Syst*. 2020;4(2):E10206. Available from: <https://dx.doi.org/10.1002/lrh2.10206>
31. Walker JD, Pyper E, Jones C, Khan S, Chong N, Legge D, et al. Unlocking First Nations health information through data linkage. *Int J Popul Data Sci*. 2018;3(1). Available from: <https://dx.doi.org/10.23889/ijpds.v3i1.450>
32. Lix LM, Ayles J, Bartholomew S, Cooke CA, Ellison J, Emond V, et al. The Canadian Chronic Disease Surveillance System: A model for collaborative surveillance. *Int J Popul Data Sci*. 2018;3(3):1–11. Available from: <https://dx.doi.org/10.23889/ijpds.v3i3.433>
33. Smith R, Roberts I. Time for sharing data to become routine: The seven excuses for not doing so are all invalid. *F1000Research*. 2016;5(781). Available from: <http://dx.doi.org/10.12688/f1000research.8422.1>
34. Canadian Institutes of Health Research. Strategy for patient-oriented research: Patient engagement framework. 2014;1–11. Available from: <http://www.cihr-irsc.gc.ca/e/48413.html>, Accessed on April 26, 2015.
35. Smith M, Schull M, McGrail K, Katz A, Diverty B, McDonald T, et al. Building a Pan-Canadian Real World Health Data Network. *Int J Popul Data Sci*. 2018;3(2):23889. Available from: <https://dx.doi.org/10.23889/ijpds.v3i4.984>
36. Canadian Institutes of Health Research. SPOR Support Units [Internet]. [cited 2020 Aug 4]. Available from: <https://cihr-irsc.gc.ca/e/45859.html>
37. Council of Canadian Academies. Accessing Health-Related Data in Canada. Ottawa (ON): The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation. 2015. 1–260 p. Available from: <https://cca-reports.ca/wp-content/uploads/2018/10/healthdatafullreporten.pdf>
38. SPOR Canadian Data Platform [Internet]. [cited 2020 Jun 9]. Available from: <https://cdp.hdrn.ca/en/home>
39. Data Access Support Hub (DASH) [Internet]. [cited 2020 Jun 9]. Available from: <https://dash.hdrn.ca/>
40. Canadian Partnerships for Tomorrow's Health [Internet]. [cited 2020 Jun 9]. Available from: <https://canpath.ca/>

Abbreviations

CIHR:	Canadian Institutes of Health Research
DASH:	Data Access Support Hub
HDRN:	Health Data Research Network Canada
PRHDN:	Pan-Canadian Real-World Health Data Network
SPOR-CDP:	Strategy for Patient Oriented Research Canadian Data Platform

