FACTORS AFFECTING ACCESS TO ADMINISTRATIVE HEALTH DATA FOR RESEARCH: A MULTIPLE CASE STUDY OF THREE CANADIAN PROVINCES

by

Cynthia Kendell

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Nova Scotia, Canada
June 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Background: Administrative health data are increasingly recognized as an invaluable resource for health research and an important source of real-world evidence to inform healthcare delivery and policy. However, a growing body of literature indicates that there are barriers to accessing these data for research, resulting in delays and interprovincial variations in access to data. This study set out to identify the factors affecting access to administrative health data for research with the dual aims of identifying the specific barriers affecting access to data and gaining insight into the factors contributing to inter-provincial variations in the timeliness of access. **Approach:** A qualitative multiple case study was undertaken. Access to administrative health data was examined in three Canadian provinces (Nova Scotia, British Columbia, and Ontario) with a focus on research where the point of access was the provincial data center (Health Data Nova Scotia, PopData, or ICES). Data were collected from case documents and semi-structured interviews with regulatory and research stakeholders (n=46). Data analysis was carried out separately for each case with findings compared across cases to identify similarities and differences. **Results:** A total of 32 inter-related factors spanning seven common categories were identified as affecting access to administrative health data for research: study-related, researcher-related, regulatory-stakeholder related, relational, organizational, regulatory, and contextual. The factors affecting access to administrative health data were largely similar across cases but varied in terms of how they affected access (e.g., as a barrier or facilitator), and in the magnitude of their impact. **Conclusion:** As the first in-depth study examining factors affecting access to administrative health data for research in Canada, this study provides evidence that may be used to inform ongoing local and national efforts to improve access to these data. The variation in barriers across provinces provides insight into reported inter-provincial variations in the timeliness of data access and highlights the need for context-specific strategies to improve data access.

LIST OF ABBREVIATIONS USED

AB	Alberta
AHRQ	Applied Health Research Question
BC	British Columbia
CAHSPR	Canadian Association of Health Services and Policy Research
CDP	Canadian Data Platform
CEO	Chief Executive Officer
CNODES	Canadian Network for Observational Drug Effect Studies
CIHR	Canadian Institutes for Health Research
DAC	Data Access Committee
DAS	Data and Analytic Services
DASH	Data Access Support Hub
DAR	Data Access Request
DAU	Data Access Unit
DCP	Dataset Creation Plan
GC	Government of Canada
HDNS	Health Data Nova Scotia
HDRN	Health Data Research Network
ICES	Institute for Clinical and Evaluative Sciences, now referred to as "ICES" ("eye-see-ee-ess")
IHSPR	Institute of Health Services and Policy Research
IKN	ICES Key Number
IPC	Information and Privacy Commissioner
MB	Manitoba
MCHP	Manitoba Centre for Health Policy
MRC	Medical Research Council
NLCHI	Newfoundland and Labrador Centre for Health Information
NB	New Brunswick
NB-IRDT	New Brunswick Institute for Research, Data, and Training

NL	Newfoundland and Labrador
NS	Nova Scotia
NSERC	National Sciences and Engineering Research Council
NU	Nunavut
ON	Ontario
PEI	Prince Edward Island
PHAC	Public Health Agency of Canada
PHI	Personal Health Information
PIA	Privacy Impact Assessment
PIPEDA	Personal Information and Protection of Electronics Document Act
PSHA	Provincial Services Health Authority
QC	Quebec
REB	Research Ethics Board
SIDR	Secure Island Data Repository
SPOR	Strategy for Patient Oriented Research
SSHRC	Social Sciences and Humanities Research Council
TCPS	Tri-Council Policy Statement
WADLS	Western Australia Data Linkage System

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Dr. Geoff Porter and Dr. Robin Urquhart, for their support and guidance. Geoff and Robin, I could not have asked for better mentors throughout this process. There are no words to express my appreciation for the opportunities you have given me over the years, or my gratitude for the time and energy you have invested in helping me succeed. Thank you for seeing potential in me early on and encouraging me to pursue this next step in my career. Most importantly, thank you both for the incredible support I received since becoming a parent. Your understanding and flexibility have meant so much.

I am also grateful for the contributions of my committee members, Professor Elaine Gibson and Dr. Adrian Levy. Elaine, thank you for taking me on all those years ago and for so generously sharing your time and legal expertise. Thank you for helping me build a strong foundation early on, for bringing a fresh perspective to every discussion, and for your kindness. It's been a pleasure to work with you and learn from you. Adrian, thank you for your guidance over the years and for somehow always having the perfect heuristic on hand. I appreciate your unique insights on the study topic and enjoyed our many conversations over the years.

Next, I would like to thank Dr. George Kephart and Dr. David Wiljer. I sincerely appreciate you both taking the time out of your busy schedules to be part of the examining committee. Thank you both for posing such thoughtful questions during my defence, and for making the overall experience an overwhelmingly positive one. Additionally, I appreciate the feedback you provided on this document, which resulted in a much-improved final version.

I would also like to acknowledge the Cancer Outcomes Research Program, which provided financial support for this research, and Margaret Jorgensen for her administrative assistance, research ethics expertise, and for lending a listening ear on more than one occasion—thank you!

Certainly, none of this would have been possible without the individuals who so graciously agreed to participate in this study. Thank you all for sharing your experiences

and perspectives with me and for being so engaged in and supportive of this research. Trying to recruit participants and collect data during the height of the pandemic were daunting tasks, but each interview I carried out reinforced the importance of this study and motivated me to keep going.

On a more personal note, I would like to thank my parents, Debbie and Sidney Kendell. Mom and Dad, I have so much appreciation for how hard you worked to raise David and I and to give us opportunities that you never had. Thank you for always emphasizing the importance of education and for teaching me to never shy away from hard work.

Finally, thank you to Lee and Ciaran. Lee, your steadfast support has been the most incredible gift. Thank you for being by my side through this, for keeping me anchored in the here and now, and for taking on so much extra so I could focus on my research. Ciaran, you arrived right in the middle of this whole thing and I am so glad you did. You've been my greatest source of joy, and my greatest motivator. Thank you for filling these last five years with such happiness and more laughter than I could ever dream of. I love you both and can't wait to see what the next chapter holds for the three of us.

CHAPTER 1: INTRODUCTION

1.1. Attribution and permissions

A protocol for the study presented in this document has been published in the International Journal of Population Data Science under a Creative Commons License (CC-BY 4.0). Permissions to incorporate content from this publication into the dissertation has been provided by the co-authors (the supervisory committee) and are provided in Appendix A. The complete article citation is as follows:

Kendell C, Levy AR, Porter G, Gibson E, Urquhart R. Factors affecting access to administrative health data for research in Canada: a study protocol. *International Journal of Population Data Science* 2021;6(1):1653.

1.2. Background/context

Administrative health data are generated through the routine delivery of healthcare programs and services [1]. These data are primarily used to facilitate the administration of health care (e.g., to regulate patient flow, determine resource-use, distribute funds to hospitals, and for physician billings [2, 3]). As a result of having publicly administered health care, a wealth of administrative health data are collected in Canada and stored in electronic databases [2]. While there are variations in the data that are captured at the provincial/territorial level, and the databases in which they are stored, common databases typically include provincial/territorial insurance registries, physician billing claims, and those capturing inpatient hospitalizations, day surgeries, vital statistics, and prescription medications [2, 3].

Increasingly, administrative health data are being recognized as an invaluable resource for health research, and as an important source of real-world evidence to inform healthcare delivery and policy [4-10]. These data are particularly useful for large-scale observational studies, offering a number of methodological and practical advantages compared to primary data collection [5, 11, 12]. In several provinces, provincial repositories containing a variety of administrative health databases have been established to facilitate the use of these data for research purposes. These include: Population Data BC (PopData) [13], the Manitoba Center for Health Policy (MCHP) [14], ICES (formerly, the Institute for Clinical and Evaluative Sciences; Ontario) [15], the New Brunswick Institute for Research, Data, and Training (NB-IRDT) [16], Health Data Nova Scotia (HDNS) [17], the Secure Island Data Repository (SIDR; Prince Edward Island) [18], and the Newfoundland and Labrador Center for Health Information (NLCHI) [19].

Despite this "information rich" environment [20], and the existence of infrastructure and resources dedicated to facilitating data access, evidence from disparate sources indicates that there are barriers to accessing these data for research [5, 21-26]. Moreover, substantial inter-provincial variations in the timeliness of access to these data have been reported [21, 25-28]— including one report that data access timelines ranged from 1 to 18 months across provinces [21]— which may reflect variations in the extent to which barriers exist across jurisdictions. Importantly, there is a lack of recent and empirically based literature examining access to administrative health data for research in Canada. As such, the issues researchers are facing when attempting to access administrative health data for research are not well understood,

nor are the reasons for the reported variations in the timeliness of data access. Using a qualitative, multiple case study approach, this study set out to identify the factors affecting access to administrative health data for research, with a view to identifying the barriers affecting access to data and gaining an improved understanding of inter-provincial variations in the timeliness of access.

1.3. Research questions and objectives

This study was guided by three research questions:

(1) What are the factors affecting access to administrative health data for research purposes in Canada?

(2) How do these vary across provinces?

(3) Why?

To address these questions, this study examined access to administrative health data for research across three provinces: Nova Scotia (NS), British Columbia (BC), and Ontario (ON). The specific objectives of this study were to:

1) Describe the policies and processes for accessing administrative health data for research purposes in each province in terms of:
    a) the key actors and approval bodies involved in governance,
    b) the number and nature of required reviews and approvals,
    c) required documentation (e.g., specific forms, applications, letters of support),
    d) sequence and duration of steps,
    e) relevant organizational policies and federal and/or provincial legislation.

2) Explore researchers' experiences with accessing administrative health data for research purposes in each province, including their ability to obtain access and the timeliness of data access.
3) Explore the perspectives of individuals involved in the regulation and oversight of access to administrative health data for research (i.e., privacy officers, data access committee members, data custodians) regarding:
    a) the use of administrative health data for research purposes,
    b) the regulatory processes and policies in place,
    c) their regulatory role (e.g., training, expertise, resources, supports),
4) Compare and contrast (1)-(3) across selected provinces.

1.4. Scope of study

This study focused on academic research involving administrative health data where the point of access was a provincial data centre. This encompassed research involving the data centre's internal data holdings, as well as research involving "external" data linkages—that is, where data held by the provincial data centre were linked to data that were held by an external data provider. Research that did not involve data held by a provincial data centre, and/or where the provincial data centre was not the point of access, was not within the scope of this study. Additionally, research involving linkages to data held in biobanks was beyond the scope of this study due to the unique ethical and privacy concerns associated with the use of biospecimens in research (e.g., biological samples can never truly be de-identified, and contain information relevant to other family members [29]). Finally, given that the primary issue being addressed in this

study was access to administrative health data for research purposes, access to these data for health system planning and management or quality assurance purposes was also beyond the scope of this study.

1.5. Document outline

This document is organized based on a traditional dissertation format and contains five chapters. Chapter 1 has introduced the topic, objectives, and scope of the study. Chapter 2 provides a review of the literature, including an overview of administrative health data, the regulatory framework that governs access to these data for research, and the challenges with access. The chapter concludes by identifying gaps in the existing literature. Chapter 3 is divided into two components. The first describes the study methodology, including the philosophical underpinnings of the study, and the selected methodological approach, while the second contains a description of methods. Chapter 4 presents the study results, and is comprised of a description of data sources, a comparison of the cases studied, and the results of the cross-case analysis. Chapter 5 provides a brief study summary followed by a discussion of the key findings, implications, study strengths and limitations, reflexivity, and ethical considerations. Following a proposed dissemination plan and recommendations for future research, the document ends with a brief conclusion, reiterating the motivation for the study, key findings, and next steps.

CHAPTER 2: REVIEW OF LITERATURE

2.1. Use of administrative health data for research

2.1.1. Benefits of using administrative health data for research

Administrative health databases have a number of characteristics that make them appealing for use in research: 1) they are often population-based (i.e., contain data for all members of a population, or the vast majority thereof); 2) they contain a wide range of variables, such as patient demographics, diagnoses, interventions and outcomes; 3) they contain data collected over long periods of time, from several years to several decades; and 4) they contain unique identifiers (i.e., all records within the database that belong to a unique individual have the same identifier). As a result, the use of administrative health data for research offers a wide range of benefits [5, 11, 12], including:

- *Reduction of bias.* Quantitative research often employs statistical generalization, which involves sample-to-population extrapolation [30]. For this to be valid, the study sample must be representative of the broader population. When primary data collection occurs, recruitment and consent processes may introduce non-response or participant bias [31] (also referred to as authorization bias [32]) into the study, reducing the representativeness of the study sample and undermining study validity [33]. This issue is minimized when the researcher uses data contained in population-based databases.
- *Improvements in statistical power*. Administrative health databases, whether or not they are population-based, often contain data for thousands or even tens of

thousands of individuals. Such sample sizes cannot be easily attained via primary data collection. A larger sample size improves the ability to detect a difference between groups (e.g., experimental and control), where a true difference exists [34]. If a study sample is too small (i.e., "lacks power"), it may incorrectly conclude that there is no difference between groups. Thus, the large sample sizes available within administrative health databases, allow differences to be detected that may have otherwise remained unknown.

- *Creation of comprehensive datasets.* Data linkage refers to "the bringing together from two or more different sources, data that relates to the same individual, family, place or event"[11]. As individuals access a variety of health services across the system, information about them may be captured in many different administrative health databases. The process of data linkage (via a common unique identifier such as health card number) allows records pertaining to an individual to be identified across multiple databases and brought together into one comprehensive dataset, providing a complete view of their encounters with the healthcare system.
- *Examination of social determinants of health.* Where individual identifiers are available, administrative health databases can be linked to other health-*related* databases (e.g., judicial, educational, social services) that can then be used to examine the social determinants of health (i.e., income, social status, social support networks, education, employment/working conditions, social and physical environments, personal health practices and coping skills, healthy child development, gender, and culture [35]) [5, 12].

- *Longitudinal analyses.* Longitudinal studies are used to examine temporal trends, however, primary data collection is time and resource intensive and such studies are prone to participant attrition [36, 37] . Where administrative data have been collected continuously over an extended period of time, they provide an effective and efficient alternative for carrying out longitudinal research [38, 39].
- *Lower costs compared to primary data collection.* The use of administrative health data for research is often less expensive in comparison to primary data collection, particularly where large samples sizes and/or longitudinal data are required [11, 12, 38]. Moreover, the costs of the infrastructure required to support research using administrative health data has been shown to be substantially less compared to other types of research (specifically, biomedical research) [11].
- *Improved privacy protection.* With regard to privacy protection, the use of administrative health data often negates the need for researchers to access direct identifiers such a name, birthdate, social insurance number, or health card number. Instead, direct identifiers can be removed, and unique encrypted identifiers used in their place (e.g., encrypted health card number or unique study identifier). These help facilitate data linkage while eliminating the need for researchers to assess direct patient identifiers [40, 41].
- *Reduction of participant burden.* Compared to studies involving primary data collection, where individuals must complete surveys or interviews, the use of administrative data reduces demands on individuals' time [5].

- *Improvements in data quality.* Using administrative health data for research provides an opportunity for researchers to closely examine the data that have been collected and to identify and correct data errors and technical glitches, improving data quality for future uses [11].

Given these attributes, the data contained within administrative health databases are particularly useful for large-scale observational studies, such as those commonly employed in health services research (i.e., study of the use, costs, quality, accessibility, delivery, organization, financing, and outcomes of health care services [42]), epidemiological research (i.e., the study of the "distribution and determinants of health-related states or events in specified populations" [43] (p.62)), and population health research (i.e., the study of "health outcomes, patterns of health determinants, and the policies and interventions that link these two" [44] (p.380)).

2.1.2. Limitations of using administrative health data for research

While the use of administrative health data offers a range of advantages, a variety of limitations have also been identified within the literature [45-49], including:

- *Data inaccuracies*. Several studies have identified issues with the accuracy or correctness of administrative health data, such that the data captured may not be an accurate representation the diagnosis that was received or procedure that was performed [50-53]. This may occur for several reasons, including human error [52, 53].

- *Data may not be "research-ready"*. Administrative health data are not collected for research purposes. Depending on the dataset, substantial processing may be required before the data can be used for research, which can require extensive time and resources to address [48].
- *Potential for misclassification bias*. When defining study cohorts, the researcher must develop variable definitions or algorithms to categorize individuals on the basis of having a specific condition, intervention, or outcome. A number of studies have reported on the limited ability to identify specific subgroups from within claims data [51, 54-58]. Without adequate validation efforts, this may result in the misclassification of individuals to an incorrect group, leading to "information bias" [59] or "misclassification bias" [60].
- *Exclusion of covariates and confounders.* Administrative health databases do not always contain all data that may be relevant in the context of a specific study. For example, data related to the social determinants of health (e.g., income, education, race) and health behaviours are not typically captured within administrative health data. As such, important covariates may be excluded from multivariate analyses, subsequently limiting the ability to identify and account for important confounders [46, 61].
- *Data linkage errors.* When attempting to perform linkage across different data sources, errors can occur, including missed linkages (false negatives) and incorrect linkages (false positives) [62-64]. These errors have been shown to introduce substantial bias into the study and impact the validity of results [65, 66].

- *Changes in how data are captured over time.* How data are collected and coded can change over time, which can create data measurement artifacts and produce misleading data [46, 67]. For example, changes in how specific diseases or events are coded in a specific database may lead to what appears as an increase or decrease in disease prevalence or event rates, when in fact such changes did not actually occur.
- *Variations across jurisdictions*. How data are collected and recorded varies across jurisdictions, including across Canadian provinces, which has implications for the comparability of research findings [68, 69]. Even when using common study protocols and analysis plans across provinces, variations in provincial data (i.e., content, coding, and completeness) have been shown to contribute to variations in case definitions, event rates, and effect size estimates [69].
- *Statistical significance*. The use of very large sample sizes increases the likelihood of finding a statistically significant relationship that is not practically or clinically significant [70, 71]. As such, caution must be exercised during interpretation.

An understanding of these limitations is required to ensure researchers select appropriate data sources, make informed methodological decisions (i.e., to minimize bias/optimize study validity), and correctly interpret the data [46, 47] .

2.1.3. Potential impacts of research involving administrative health data

Administrative health data have been used to study a wide range of healthcare topics in Canada, including disease incidence and prevalence [72-75], patterns of care [76-78], quality [79, 80] and timeliness of care [81, 82], health outcomes [83, 84], and the costs

associated with care delivery [85, 86]. While the findings of these individual studies may have led to local improvements in care, the overall impacts of research involving administrative health data on health care, health policy, and the health outcomes of Canadians are unclear. Nonetheless, reports from ICES and MCHP have clearly demonstrated the potential of administrative health data research to make substantial contributions to knowledge, and for this new knowledge to inform improvements in health care delivery and policy development [87-89]. In a 2019 profile of ICES [89], the authors reported that ICES researchers published over 2200 peer-reviewed articles and over 200 additional reports in the five years prior. This work included research using ICES data that contributed to the development of a risk projection tool for kidney failure [90] which is used to evaluate living kidney donor candidates [91] and has led to an increase in the number of eligible donors. ICES data has also informed legislation on cellphone use while driving, return to sport after concussion, and firearm control [89]. A profile of the MCHP [88] highlighted the role of MCHP data in various health and social initiatives, including the evaluation and long-term sustainment of a prenatal income supplement for low-income pregnant women (Healthy Baby Prenatal Benefit [92]), and long-term healthcare planning by the Manitoba (MB) government.

Similarly, efforts to evaluate the impacts of the Western Australia Data Linkage System (WADLS)[93] revealed substantial contributions. In the first 10 years the WADLS was in place, more than 400 new studies were carried out, resulting in over 250 journal publications [11], as well as hundreds of other academic products [38]. Numerous studies were credited with improvements in clinical care delivery and health policy. One

area where this was particularly evident was in the area of mental health. According to Brook et al [38], as a direct result of research carried out using data held by WADLS, the government committed a $173 million funding package to improve mental health services, amendments were made to mental health legislation, and recommendations for improved follow-up processes for psychiatric in-patients were implemented to reduce occurrence of suicide. Thus, where similar infrastructure exists to support research involving administrative health data, so too does the potential to use the findings of this research, (pending successful knowledge translation strategies) to trigger change within healthcare systems.

2.1.4. Privacy risks of using administrative health data for research

The risks of using administrative health data for research are primarily related to privacy and confidentiality. The following sections provide a brief overview of privacy (i.e., what it is and why it is important), followed by a discussion of the specific privacy risks associated with research involving administrative health data.

The importance of privacy

Privacy is a commonly used term but poorly understood concept. While numerous scholars from various disciplines have put forth highly nuanced legal and philosophical theories of privacy [94-106], there is no single, agreed upon definition of what privacy is or how it may be achieved. However, relevant to the flow of electronic data, privacy may be understood in terms of informational access and control [99, 103, 107]. More specifically, privacy may be viewed as a condition that is achieved when an individual

attains a desired level of informational control, and/or when others' access to information about them has been satisfactorily limited—that is, when access to the self is sufficiently regulated [97, 104, 108-110]. What is considered an acceptable level of privacy differs between individuals and changes over time and circumstances. Confidentiality is within the "umbrella" of privacy [111], referring specifically to the duty of an individual or organization to protect entrusted information from unauthorized access, use, disclosure, modification, loss or theft [112].

The protection of privacy is important for many reasons. Privacy plays an important role in individuals' development of personal identity and sense of self [103, 104, 110, 113]. The protection of privacy relevant to health information is especially important to an individual's personal identity:

> Health information contains arguably the most sensitive and intensely personal aspects of ourselves, and thus is a fundamental aspect of identity. How we choose to be known or not known, the health information we reveal or don't reveal based on how we think others will identify or label us, and the ways in which we reinvent ourselves over time are all powerful ways in which we control aspects of our identity [113](p.2).

Privacy is also key to establishing and defining interpersonal relationships [97, 101, 104, 109, 114]. This is particularly relevant in the context of health care. The confidential nature of the patient-physician relationship fosters trust and facilitates the honest and complete sharing of information that is essential to the process of care delivery [115].

When trust is broken in other social settings, it can still have implications for health care delivery. As stated by Steeves [116], "…practices that violate the social experience of privacy as it is lived in our daily lives will break down the trust that is an essential part of healthcare delivery" (p.32).

By allowing individuals to decide for themselves which information they disclose to others and under what circumstances, privacy is also a mechanism by which personal autonomy or self-determination is protected. When information about an individual that would not have otherwise been disclosed becomes known to others, that individual becomes vulnerable to manipulation by others and subject to their influence [97, 117]. This is particularly relevant in the context of personal health information (PHI), which may include information related to an individual's sexual health, drug use, mental health status, and reproductive history. Related to this, privacy also protects individuals from being stigmatized as a result of the disclosure of personal information [99].

Finally, privacy is closely tied to emotional and psychological well-being [97, 104, 118, 119]. Specifically, privacy is related to an individual's sense of competence and self-worth [104], psychological security [118], and ability to self-actualize [97, 119]. As such, privacy is closely tied to the health and well-being of individuals.

Privacy risks

The use of individual-level administrative health data for research purposes poses risks to individuals' privacy, even when there have been efforts to de-identify the data [120]. These risks may be categorized as follows [5]:

1) *Accidental release of data.* The accidental release of identifiable data to unauthorized individuals may occur when proper data handling protocols are not followed (e.g., losing a device containing sensitive data).
2) *Illicit access*. External parties (i.e., hackers) or internal parties (i.e., employees) may deliberately access data for illicit or inappropriate purposes.
3) *Inadvertent access*. In the course of doing their job, an individual may inadvertently recognize someone they know (e.g., family member, friend, or neighbor) within a dataset.
4) *Data re-identification*. If data de-identification is not done properly, enough identifiers may remain to make re-identification possible, particularly where multiple databases or data sources are involved.

These privacy risks are compounded by several factors. First of all, electronic information has been referred to as "greased", meaning it "moves quickly and is hard to hold onto" [99] (p. 27). As such, information that is improperly accessed can be quickly, easily, and broadly disseminated via the internet and used for any number of purposes [99]. Second, the nature of privacy-related violations has changed over time. In the past, these violations tended to involve easily identifiable and discrete events that were physical in nature (e.g., viewing paper files in a filing cabinet), whereas those occurring in the "information age" occur repeatedly over an extended period of time and often remain unknown [114]. Third, in contrast to physical records that can be costly to store, the cost of storing electronic data is so low that there is often very little motivation to destroy the data, increasing the risk of a privacy breach [113]. Finally, even when there

have been efforts to de-identify the dataset, it has been noted that "data contained in these research databases have been cleaned and organized in such a fashion as to maximize the efficiency of analysis. This very efficiency increases the risk of misadventure in the event of a breach"[121](p.13).

2.2. Regulation of access to administrative health data for research in Canada

In Canada, access to administrative health data for use in research is regulated by human research ethics and information legislation. For researchers to gain access to administrative health data for research purposes, they must demonstrate compliance to all relevant ethical and legal requirements. These are summarized in the following sections.

2.2.1. Human research ethics

Overview of regulatory structures

In Canada, human research ethics is primarily regulated by two key structures: (1) ethical guidelines developed by federal research funding agencies; and (2) institutional research ethics boards.

Ethical Guidelines: The Tri-Council Policy Statement

In 1994, the three federal research agencies in Canada—the Medical Research Council (MRC; now the Canadian Institutes of Health Research [CIHR]), Social Sciences and Humanities Research Council (SSHRC), and Natural Sciences and Engineering Research Council (NSERC)—set out to create a joint policy on research ethics [122]. This resulted in the publication of the *Tri-Council Policy Statement: Ethical Conduct for Research*

Involving Humans (TCPS) in 1998 [123]. The second edition of this document (TCPS 2) was published in 2010 [124] (with revisions released in December 2014 [125], December 2018 [112], and December 2022 [126]). The TCPS 2 applies to all research carried out at institutions that are eligible to hold funds from CIHR, SSHRC, or NSERC [112, 127]—that is, institutions that have entered into the "*Agreement on the Administration of Agency Grants and Awards by Research Institutions*" [128]. Thus, the TCPS 2 applies to research carried out at the vast majority of university and hospital settings across the country, and also been adapted by other private and public institutions/entities (e.g., Health Canada, the National Research Council, and the Department of National Defense)[127], making it the primary source of ethical guidance for research in Canada.

The TCPS 2 [112] is primarily concerned with ensuring respect for human dignity is upheld throughout the conduct of research involving humans participants. Respect for human dignity is expressed through three principles:

1) *Respect for persons*—respect for participants' autonomy and freedom of choice,
2) *Concern for welfare*—the protection of participants' overall well-being, and
3) *Justice*—the equitable distribution of the benefits and harms of research and minimizing power imbalances between researchers and participants.

Concern for welfare encompasses the impact of a variety of factors on an individual's quality of life, including privacy and informational control. Thus, the TCPS 2 recognizes that individuals have a privacy interest in information that is about them, while researchers have an ethical obligation to protect individuals' privacy and to treat their information in a confidential manner. Nonetheless, the guidance provided with

respect to privacy and confidentiality in the original TCPS and subsequent iterations has been rather limited. To address this gap, and to support a harmonized approach to the application of privacy laws and policies, CIHR developed a supplemental document to assist researchers with the application of the principals of the TCPS 2 entitled *CIHR Best Practices for Protecting Privacy in Health Research* [129] (hereafter "CIHR Best Practices document"). Despite being released in 2005, this document continues to be used as a source of additional guidance on privacy related matters, as evidenced by multiple citations in the most recent versions of the TCPS 2 [112, 126].

Research Ethics Boards

The TCPS 2 [112, 124-126] mandates all institutions that receive funds from the Tri-Council to have one or more research ethics boards (REBs) in place, which function as the primary mechanism for ensuring compliance with the TCPS 2. The guidelines for the establishment and operation of research ethics boards (REBs) also set out within the TCPS 2. REBs are committees comprised of a chair and various members, all of whom are volunteers. Though they are considered independent decision-making bodies, they are accountable to the highest body at the institution (e.g., president or council), which is responsible for establishing the REB, defining terms of operation, appointing members, providing resources, and ensuring members receive appropriate education and training.

The primary responsibility of REBs is to perform ethical review—that is, to assess the "ethical acceptability"—of study protocols submitted by researchers, and to determine whether the study may proceed, and if so, whether revisions are required. At

institutions that are subject to the TCPS 2, ethical review is required for all research involving human participants, where "research" is defined as "an undertaking intended to extend knowledge through a disciplined inquiry and/or systematic investigation" [112](p. 13), and "human participants" are defined as "individuals whose data, biological materials, or responses to interventions, stimuli or questions by the researcher, are relevant to answering the research question(s)" (p.14). This encompasses a broad range of activities, from studies involving physical interventions such as clinical trials, to qualitative research, to research involving the secondary use of PHI.

In certain cases, a study may be exempt from REB review; however, eligibility for exemption can only be determined by an REB. Where review is required, a study may be subjected to either full REB review or delegated review. Full REB review is required if, based on the "magnitude and probability of harms" associated with the research, it is believed that the study poses substantial risks to participants. Delegated review (by a single REB member) is required when the study is considered minimal risk, meaning "the probability and magnitude of possible harms implied by participation in the research is no greater than those encountered by participants in those aspects of their everyday life that relate to the research" (p.22). This "proportionate approach to review" aims to ensure that appropriate protections for research participants are in place, without creating unnecessary barriers to research. At the same time, it helps ensure that REB resources are directed to the most "ethically challenging" research.

Considerations Relevant to Research Involving Administrative Health Data

In addition to ensuring appropriate physical, technical, and administrative safeguards are in place to prevent data from unauthorized access, loss, or manipulation, there are two key factors that REBs must consider when reviewing a study protocol: (1) the "identifiability" of the data, and (2) the need for individual consent [125]. The REBs assessment of these factors has implications in terms of whether or not approval will be granted and will also determine whether the researcher will require consent from the individuals whom the requested information is about.

<u>Identifiability of Information</u>

REBs are instructed to consider whether the information being used in a proposed study is identifiable or non-identifiable [112, 125, 126]. Consistent with a proportionate approach to review, studies wherein researchers require access to "identifiable" information are subject to a higher level of scrutiny than studies involving "non-identifiable" information. Identifiable information is defined as information that "may reasonably be expected to identify an individual, when used alone or combined with other available information" [125](p.58), whereas non-identifiable information does not identify an individual. To assist REBs with assessing the privacy risks associated with a particular study, the TCPS 2 sets out the following five categories of information [112, 125, 126]:

1) *Directly identifying information*. Information that contains direct identifiers (e.g., name, social insurance number, or health card number) which can be used to identify a specific individual.

2) *Indirectly identifying information*. Information that contains indirect identifiers (e.g., birthdate, place of residence, and unique characteristic such as rare health conditions), the combination of which can be reasonably expected to identify an individual.
3) *Coded information*. Information from which direct identifiers have been removed and replaced with a code (e.g., a unique study identifier). The "key" (i.e., the list of codes and the individuals to which they correspond) may be kept, allowing for future re-identification of individuals.
4) *Anonymized information*. Information from which direct identifiers have been removed, the "key" has not been kept, and the remaining indirect identifiers pose minimal risk of re-identification.
5) *Anonymous information*. Information that has not had identifiers associated with it at any point and the risk of identification is minimal (e.g., anonymous surveys).

Researchers who are unclear about whether the information they plan to use in a study is identifiable are expected to consult with the REB. Determining whether information is identifiable or not it important because it has implications for the overall risk-benefits assessment that is performed by the REB during review. It is also a key factor considered by REBs when determining the consent requirements for a study. This is described in greater detail below.

Requirements for Individual Consent

In the TCPS 2, consent is the default requirement to participate in research in any capacity (i.e., whether directly or via the inclusion of information about an individual) [112, 125, 126]. Consent must be given voluntarily, be informed, and be ongoing such that participants can withdraw at any time. These requirements are absolute in research where participants experience a physical intervention; however, there are exceptions for non-interventional research exclusively involving the secondary use of information for research purposes. For studies involving non-identifiable information, including coded information where the researcher does not have access to the key, consent is not required. Where identifiable information is involved, consent may be waived if the researchers have successfully demonstrated in their REB application that: i) the research cannot be conducted without the use of identifiable information, ii) there are minimal risks to individual welfare, iii) privacy will be respected and adequate data safeguards will be in place, iv) individual preferences regarding data use will be respected when known, v) obtaining consent is either impossible or impracticable, and vi) all other permissions and approvals have been obtained.

The TCPS 2 defines impracticable as "incapable of being put into practice due to a degree of hardship or onerousness that jeopardizes the conduct of research; it does not mean mere inconvenience" [125](p.205). The CIHR Best Practices document provides specific criteria for impracticability [129]. Specifically, obtaining individual consent for the use of personal information may be considered impracticable for researchers if circumstances (i.e., the population size is large, or a large proportion have

moved or died since the personal information was originally collected, or there is no ongoing relationship between the individual and the data holder who would need to contact them to obtain consent) are such that the inability to obtain consent from segments of the population may introduce a bias into the study, or the resource implications for the research team would be so great that the research could not be done. The CIHR Best Practices document also acknowledges that the process of obtaining consent may be considered inappropriate and warrant a waiver of consent from the reviewing REB. For example, if contacting an individual poses a risk of harm, requires "re-identifying" coded data, or is prohibited under a previous data sharing agreement, policy, or law.

2.2.2. Legal regulation of personal health information

In Canada, the collection, use, and disclosure of PHI are regulated by a myriad of legislation at the federal and provincial/territorial levels (Appendix B). Administrative health data are data that are collected from individual patients during their encounters with the healthcare system, and as such, are typically regarded as PHI that is subject to information legislation. However, not all legislation that applies to PHI will necessarily apply to administrative health data. The following sections provide a summary of how federal and provincial information legislation applies to research involving administrative health data.

Federal legislation

At the federal level, there are two statutes that are involved in the regulation of PHI: the *Privacy Act* [130] and the *Personal Information Protection and Electronic Documents Act* (PIPEDA) [131]. The *Privacy Act* [130] applies to personal information held by federal institutions, including certain types of PHI (i.e., medical history and blood type). PIPEDA applies to personal information collected, used, or disclosed in the course of commercial activities [131], including PHI in some circumstances [132].

These Acts are not typically relevant with respect to research involving administrative health data. This is primarily related to the division of legislation powers, which places the operation of hospitals under provincial jurisdiction [133]. As such, health care delivery and administration, and the information collected in the course of these activities, are regulated at the provincial level. In situations where administrative data are held by the federal government, the *Privacy Act* [130] may apply to research involving administrative health data. For example, research using data from the Non-Insured Health Benefits Program [134], a federally administered program which captures information on health services utilization among First Nations and Inuit populations, must comply with the *Privacy Act* [135].

PIPEDA applies to certain types of PHI, including information collected as a result of commercial activity conducted on the premises of public institutions (e.g., by hospital-based pharmacies), well as the activities of healthcare providers in private practices (e.g., doctors, dentists and chiropractors)[132]. However, the “core activities” of public healthcare institutions (e.g., public hospitals, publicly funded long-term care

facilities) are not considered commercial in nature and are not subject to PIPEDA. Though the "core activities" criterion is not infallible [136], administrative data would not likely be considered commercial in nature. Nonetheless, several provinces (i.e., ON, New Brunswick [NB], Nova Scotia [NS], and Newfoundland and Labrador [NL]) have implemented health sector specific legislation that has been declared "substantially similar" to PIPEDA. In these provinces, organizations that are subject to provincial health information laws are typically exempt from PIPEDA [137].

Provincial/territorial legislation

At the provincial/territorial level, PHI (including administrative health data) is primarily regulated by health sector-specific legislation, except for Nunavut (Nu), BC, and Quebec (QC). In Nu, the government has stated that health sector-specific legislation will not be developed until the planned interoperable electronic health record is operational [138], which has not occurred to date. As such, PHI in Nu is regulated primarily by the *Access to Information and Protection of Privacy Act* [139]. In BC, health sector specific legislation (i.e., the *E-Health Act* [140]) has been passed, however, it applies to only designated health information banks, of which there are only three in the province (i.e., Provincial Laboratory Information Solution repository, Client Registry System, Provider Registry)[141]. The majority of PHI is therefore regulated by the *Freedom of Information and Protection of Privacy Act* [142]. Similarly, there is health sector-specific legislation in place in QC (i.e., *An Act Respecting the Sharing of Certain Health Information* [143]), however, it only applies to the establishment of a particular data platform (i.e., the Dossier Santé Quebec). The main piece of legislation pertaining to PHI in QC is *An Act*

Respecting Documents Held by Public Bodies and the Protection of Personal Information [144] .

The following sections will provide an overview of how researchers' access to administrative health data is regulated under provincial legislation. Specifically, the following sections will focus on the *Access to Information and Protection of Privacy Act* [139] in Nu, the *Freedom of Information and Protection of Privacy Act* [142] in BC, *An Act Respecting Documents Held by Public Bodies and the Protection of Personal Information* [144] in QC (herein referred to as "*An Act Respecting Documents Held by Pubic Bodies*"), and health sector-specific legislation in all other provinces. In some provinces it may be possible that other pieces of legislation come into play depending on the data being accessed and other study-related details, however, these statutes will not be addressed in the following sections.

Purpose of the legislation

The majority of provincial/territorial Acts contain explicit statements identifying the purpose(s) of the legislation [139, 142, 145-153]. Although there is variation between provinces in terms of the content and description of these purposes, in general, provincial/territorial health-sector specific legislation aims to regulate practices involving PHI (e.g., collection, use, disclosure) in such a way that recognizes the need to protect individual privacy, but also acknowledges that access to PHI by third parties may be required for legitimate uses under certain circumstances. For example, the stated purpose of NS's *Personal Health Information Act* [149] is to "govern the collection, use, disclosure, retention, disposal and destruction of PHI in a manner that recognizes both

the right of individuals to protect their PHI and the need of custodians to collect, use and disclose PHI to provide, support and manage health care" (s.2). Similar purpose statements appear in several other provincial/territorial statutes [145-148, 150, 152, 153].

Application of the legislation

The type of information that is subject to regulation under provincial/territorial legislation varies across jurisdictions. BC's *Freedom of Information and Protection of Privacy Act* [142], QC's *Act Respecting Access to Information Held by Public Bodies* [144], and Nu's *Access to Information and Protection of Privacy Act* [139] regulate all information held by public bodies, encompassing personal information and PHI. In all other provinces/territories, the legislation applies specifically to PHI [146-154], with the exception of the *Health Information Act* [145] in Alberta (AB), which applies to health information more broadly (it does not use the term "personal health information", but distinguishes between "individually identifying" and "non-identifying" health information).

Where provincial/territorial legislation regulates PHI, a variety of definitions of PHI are used. Typically, PHI is defined to include a broad range of health-related information, including information about: an individual's physical or mental health; medical history of the individuals' family; care provided to the individual; the donation by an individual of a body part or bodily substance; registration in a provincial medical insurance plan; the individuals' substitute decision-maker; the individuals' healthcare provider; any drug, device, product, or equipment prescribed to an individual [146-

154]. Importantly, PHI is typically limited to "identifiable information", though there is variation in the terms that are used in reference to the "identifiability" (e.g., "identifiable", "identifying", "non-identifying", "de-identified"), as well as how these terms are defined (Appendix C).

With regard to format, the majority of legislation applies to both recorded and oral information [147-153], while others do not specify which formats are included [139, 145, 154]. In BC's *Freedom of Information and Protection of Privacy Act* [142] and MB's *Personal Health Information Act* [146], only recorded information is considered PHI.

Disclosures for research

Although research is not explicitly mentioned in the purpose statement of any of the relevant provincial/territorial Acts, all Acts contain provisions permitting the disclosures of PHI for research purposes. Conditions for disclosure typically include a data sharing agreement between the individual researcher and data custodian or trustee [139, 142, 145-149, 152-154], and approval from a research ethics board [145, 147, 149-152, 154] or equivalent review body [146, 148, 153]. In MB, review is required by an institutional "research review committee" if the data *are not* maintained by the government or a government agency, otherwise, review by the provincial "health information privacy committee" is required [146]. Only BC's *Freedom of Information and Protection Privacy Act* [142] and QC's *Act Respecting Information Held By Public Bodies* [144] do not explicitly refer to the need for ethical approval.

Consent requirements

All provincial/territorial Acts contain provisions explicitly permitting the disclosure of PHI without individual consent for research purposes if certain conditions are met. The exceptions to this are BC's *Freedom of Information and Protection of Privacy Act* [142] and Nu' s *Access to Information and Protection of Privacy Act* [139], which do not explicitly state the consent requirements for disclosures of PHI for research purposes.

Where consent requirements for disclosures for research are addressed in the provincial/territorial legislation, decisions related to consent are typically delegated to REBs [145, 147, 149-152, 154], or to other relevant review bodies (i.e., health information privacy committee [146], institutional research review body [146, 148, 153]). Disclosures without consent are also permitted by QC's *Act Respecting Documents Held by Public Bodies* [144]. Until recently disclosures without consent required authorization from the Commission d'access a l'information du Quebec, however, this changed in Fall 2022 and organizations now have discretionary power over whether to allow disclosures without consent [155].

Where REBs or other appropriate review bodies are granted decision-making authority relevant to consent, the conditions under which disclosure without consent may occur are not well defined. Not only is there substantial variation in the terminology used to describe when consent is not required (summarized in Table 1), but the terms that are used are not defined in most of the provincial statutes. Only NS's *Personal Health Information Act* provides a definition for the term "impracticable" (i.e.,

a degree of difficulty higher than inconvenience or impracticality but lower than impossibility), while the relevant terms used in other legislation are not defined.

Table 1. Conditions for waiver of individual consent

Province/ Territory*	**Legislation**	**Consent is not required if determined to be…**
AB	Health Information Act [145]	"unreasonable", "impractical" or "not feasible"
SK	Health Information Protection Act [154]	"not reasonably practicable"
MB	Personal Health Information Act [146]	"unreasonable" or "impractical"
ON	Personal Health Information Protection Act [147]	"impractical"
NB	Personal Health Information Privacy and Access Act [148]	"unreasonable" or "impractical"
NS	Personal Health Information Act [149]	"impracticable"
PEI	Health Information Act [150]	"unreasonable", "impractical" or "not feasible"
YK	Health Information Privacy and Management Act [153]	"unreasonable" or "impractical"
NWT	Health Information Act [152]	"unreasonable", "impractical" or "not feasible"

Abbreviations: AB=Alberta, SK=Saskatchewan, MB=Manitoba, ON=Ontario, NB=New Brunswick, NS=Nova Scotia, PEI=Prince Edward Island, YK=Yukon, NWT=Northwest Territories

*Table includes only provinces wherein the relevant legislation explicitly addresses consent and/or specifies the conditions under which a waiver may be granted.

2.3. Challenges related to accessing administrative health data for research in Canada

2.3.1. Delays and variations in access to administrative health data for research

Despite the wealth of administrative health data captured across Canadian provinces and territories, the existence of provincial data centres with resources and

infrastructure to facilitate access to these data for research, and a regulatory framework that contains explicit provisions allowing researchers to access administrative health data for research purposes, there is a growing evidence to indicate that researchers are experiencing challenges when attempting to access these data for research [5, 21-28, 156, 157]. Specifically, researchers have reported delays as well as substantial variations in the timeliness of obtaining access to data when conducting multi-jurisdictional research [21, 22, 25, 27, 28].

One of the first accounts of the challenges associated with accessing administrative health data for research was published in 2002 [22]. In this report, the author described the challenges their team experienced when attempting to access linked administrative health data for a multi-province study. After two years of trying to access data, approval in one province was still pending, resulting in that province being excluded from the study. Since that time, data access challenges have persisted. Several articles have highlighted lengthy data access timelines in BC specifically. In 2016, Lavoie et al [26] reported that for one study conducted several years prior, it took 26 months to gain access to data in BC. That same year, a multi-site study involving the Canadian Network for Observational Drug Effect Studies (CNODES) reported that the BC site was not included due to “lengthy timelines” for data access [23]. More recently, researchers from the BC Children’s Hospital Research Institute, published the results of environmental scan which identified issues with timeliness as the primary challenge being faced by local researchers when attempting to access health data (including administrative health data) for research [24].

Elsewhere, researchers involved in multi-jurisdictional health services described variations in data access timelines across provinces. In a 2012 article Quan et al [25] referred to a multi-jurisdictional study wherein data access timelines varied from six months to several years. In 2018, Groome et al [27] described the challenges of conducting a multi-province study comparing breast cancer care, noting that variation in the timeliness of data access across provinces hindered efforts to perform analyses in parallel and negatively impacted project timelines. Similarly, Butler et al [28] reported that in their study the time required to access data varied across provinces (from 4-9 months), creating sequencing challenges such that some provinces were waiting for data while others were analyzing data.

Only one study has sought to systematically compare the timeliness of access to administrative health data across Canada. Meagher and McGrail [21] conducted Interviews with representatives from nine provincial agencies (e.g., provincial ministries of health and provincial data repositories) in six provinces, which revealed substantial variations in data access, ranging from 1-18 months. As acknowledged by the authors, this work had important limitations that impacted the reliability of the data. Specifically, the timeliness data that was used to examine access was self-reported by provincial agencies and not confirmed by other means. In addition, the starting point for measuring timeliness varied across agencies, likely contributing to the reported variations.

2.3.2. Factors potentially affecting access to administrative data for research

The factors affecting access to administrative health data for research in Canada have not been comprehensively assessed, however, insights about the potential factors that come into play in the Canadian context can be garnered through the examination of a disparate body of literature, comprised of reports from individual research teams, quality assurance work, legal commentary and analyses, and expert opinion. The potential factors affecting access to administrative health data identified from across these sources are summarized in the sections that follow.

Study-related factors

Meagher and McGrail [21] identified the characteristics of the study itself as having an impact on data access. Specifically, the complexity of the project was identified as a factor affecting timeliness due to the increased work involved for researchers during application preparation (i.e., to ensure clarity and consistency), and the increased time required on the part of the data provider related to dataset preparation. Project complexity also played a role in decision-making on the part of the individual or entity performing review, with more complex protocols requiring an additional "layer" of privacy risk considerations, though it was unclear what this additional "layer" involved.

Researcher-related factors

In interviews with regulatory stakeholders, researcher experience and responsiveness were identified as affecting the timeliness of data access [21]. More experienced researchers were viewed as being more familiar with the available data, and better able

to prepare data access applications, whereas less experienced researchers reportedly required more assistance from the organization to support application development. Some regulatory stakeholders also attributed at least partial responsibility for delays in access to data to a lack of responsiveness on the part of the researcher.

Data-provider-related factors

There is evidence indicating that the knowledge and expertise of individual data providers also affects access to data for research. In a 2009 study conducted in NL [158], the authors found that individuals responsible for collecting and retaining PHI who had a limited understanding of the policies processes relevant to the use of PHI for health research were reluctant to share data for research purposes (e.g., for example, 12% of health professionals surveyed indicated that they would not share de-identified data, even with individual consent). Related to individual data provider concerns, van Panhuis et al [159] undertook a systematic review exploring real and perceived barriers to "sharing" public health data. This review focused on the sharing of data that were collected by public health agencies for routine purposes (e.g., disease surveillance and program monitoring). Although it was not specific to data sharing for research purposes, many of the barriers are likely relevant to data providers across a variety of settings who are faced with sharing health information for a range of purposes. These include disagreement with the data requestor over the appropriateness or perceived risks and benefits of the intended data use, a lack of personal incentive to provide data, concerns over what the data might reveal (e.g., performance issues), and a lack of trust of the data requestor leading to fear of data misuse or misrepresentation.

Organizational factors

Data custodians have previously indicated that their ability to facilitate access to data is dependent on organizational capacity, including dedicated staff, resources, and IT infrastructure [21, 160]. In organizations where capacity is limited, data access requests are "generally met when, and if, there is time" [160](p.e259). These findings were echoed in the literature review carried out by van Panhuis et al [159], which highlighted the substantial human resources (e.g., to prepare data and communicate with recipients) and technical resources (e.g., computer equipment, software, analytic tools) required to share data, which may not be available in all organizations. Related to this, the findings also indicated that in organizations with limited resources, data collection, preservation, and sharing may not be a priority.

Factors related to the regulatory framework

Complexity of the regulatory model

The regulation of research involving human participants is notably complex, involving a variety of laws, administrative policies, and guidelines [161, 162]. Over time, ethical guidelines have become integrated into statutes and regulations in what has been referred to as "ad hoc and piece meal evolution" [162](p.18). Certainly, both of these challenges are relevant in the context of research involving administrative health data, given that the regulatory framework is comprised of two separate frameworks that have evolved separately, with distinct regulatory aims, and have been made to work together after-the-fact. Adding to the complexity of research involving administrative health data is the "patchwork" of legislation governing PHI [116](p.28)—comprised of federal,

provincial, public, private, and health-sector specific statutes—which continues to evolve as new legislation is introduced.

The complexity of the regulatory framework has several important implications relevant to data access. First, it may create uncertainty amongst researchers and oversight bodies with regard to which laws, policies, and guidelines apply in specific circumstances and undermine the effectiveness of the governance structures in place [161]. Second, the variation in legislation and policies across jurisdictions is such that inter-provincial studies may be met with resistance from REBs and data custodians who are uncertain about "legislative equivalency" [156](p.40). Finally, the complexity of the governance framework may contribute to confusion around roles and responsibilities of the many parties involved in governance [161].

Inadequate oversight and accountability

The governance of health research (not specific to research involving administrative health data) has been criticized for lacking adequate oversight and accountability [161-163]. While various parties are involved in the regulation of health research, there is no clear hierarchy of accountability [162]. Although REBs are accountable to the highest body at the institution where they are based, they are not monitored or regulated by the institution [163] or an external regulatory body [162]. Moreover, REBs are not required to evaluate their own performance, and are therefore unable to gauge the appropriateness or effectiveness of their decision-making over time [163]. Similarly, researchers are accountable to research participants, their institutions, and funders [162], yet their conduct is not closely monitored. REBs review proposals submitted by

researchers prior to the commencement of research, but typically have little involvement in the monitoring of ongoing research [163]. Although researcher compliance with ethical standards is required to retain funding, it is not clear within CIHR's enabling statute whether the institution or the funding body is responsible for monitoring compliance or investigating allegations of non-compliance [162]. Overall, there are few "checks and balances" in place to ensure that oversight bodies are exercising their power appropriately and effectively. Moreover, the involvement of so many parties with overlapping responsibilities may lead to regulatory inefficiencies. For example, the involvement of these various parties may lead to confusion around regulatory roles and responsibilities [161], cause parties to off-load their responsibilities onto others [163] (often onto REBs [161]), or result in a regulatory vacuum (i.e., everyone assumes someone else is taking responsibility which results in no one taking responsibility) [121].

Lack of clear and consistent policies and guidelines

It has been suggested that where regulations are not clear, anxiety or fear of inadvertently breaching patient confidentiality or violating data protection laws may lead to a conservative interpretation of the legislation or guidelines in question [5, 12, 156, 164, 165], which may subsequently result in "idiosyncratic" institutional policies that sometimes go beyond the requirements of the law [156] (p.40). A lack of clarity within the current regulatory framework has been noted [5, 156, 161, 165-167], particularly with respect to consent requirements [156, 166], and what constitutes identifiable information [5, 165, 168].

In a study by Willison et al [166], variation in consent requirements across REBs were attributed to lack of clear requirements in both ethical guidelines and privacy laws. He stated, "…like the TCPS, our privacy laws offer such broad concessions for non-consensual use of personal information for research that they offer little to no guidance for REBs" (p.312). Elsewhere, Willison et al [156] argued that data protection laws contained research exemptions that were so broad that REBs and data custodians did not know when it was acceptable to release data to researchers without consent, and that this created data access issues for researchers.

With regard to the identifiability of information, a recent Canadian report pointed out that researchers may access data that do not include "identifiable information", but noted that this term is not well defined and subject to differing interpretations across jurisdictions [5]. Furthermore, there are no specific criteria with regard to when data have been sufficiently de-identified so as to no longer fall under the purview of the current regulatory framework [168]. Yiannakoulias [165] posited that in the absence of clear national guidelines defining identifiability (i.e., the ability to identify a single individual from the data), it is often confused with self-identification (i.e., the ability of an individual to identify themselves from the data), which can create barriers with respect to data access. In other words, data custodians may be unduly restricting access to data about individuals based on the risk of an individual identifying themselves within the data, rather than the risk of an individual being identified by others, which is less likely.

The Role and Function of REBs

Predictive initial review

REBs are responsible for reviewing and approving research protocols prior to the commencement of research, thereby functioning as "gatekeepers" [163]. REBs focus their time and resources on initial review and approval, with little involvement in the monitoring of ongoing research [161, 163, 167, 169]. Initial review is only effective in protecting individuals from the harms of research to the extent that the REB is able to make accurate predictive judgements [163]. Without processes in place to monitor the accuracy of these judgements (i.e., ongoing monitoring by REBs), there is a lack of "virtuous learning loops" (p.6). This may result in certain types of research being unnecessarily restricted, or repeatedly permitted, as a result of erroneous risk-benefit assessments.

Bureaucratic reductionism

REBs have also been criticized for being more concerned with process and forms than with actual ethical matters of research [161, 163]. McDonald [163] refers to this phenomenon as "bureaucratic reductionism". For REBs, a focus on processing ethics applications and reviewing consent forms takes time and resources away from harms-risks assessments and consideration of major ethical concerns [170]. This raises questions about whether REBs are effectively conducting ethical review, and doing so in a way that merits public trust, or whether they have just become another layer of bureaucratic red tape [161].

Multiple and inconsistent REB reviews

Despite being required to adhere to common ethical guidelines (i.e., the TCPS 2 [125]), substantial variation in REB decision-making has been reported in Canada [166, 171, 172]. The literature suggests that ambiguities within the regulatory framework contribute to variations in the interpretation and application of regulations across jurisdictions and stakeholder groups [166]. Such inconsistencies are problematic as they may cause confusion in the research community and amongst the public, make it difficult for researchers to meet REB requirements, and lead to increased time and costs as researchers attempt to adhere to the requirements of multiple REBs [167].

Knowledge and expertise

REBs may not possess the knowledge and expertise required to perform adequate ethical review [161, 167, 169]. The issue of insufficient expertise is not unexpected—for any given study "there are few scientists who can accurately assess the scientific validity of proposed research projects and identify the harm/benefit ratio posed by the research" [167](p.168). For research involving administrative data, health information legislation has conferred both ethical and legal obligations onto REBs, so they require specific expertise to ensure compliance with relevant statutes. Moreover, the use of linked administrative health data for research is characterized by unique methodological and privacy-related concerns (e.g., de-identification, data linkage, data matching, etc.), requiring familiarity with highly technical processes in order to assess harms and benefits. Unfortunately, Canadian REBs lack specialized training in various types of research, including research with databases [161] and have identified the need for more

education on matters related to privacy, confidentiality, and security, and improved guidance for interpreting relevant laws and the TCPS 2 [166]. These gaps in knowledge and expertise raise questions about the effectiveness of REBs in ensuring compliance with ethical and legal requirements.

2.4. Gaps in the Literature

There are number of important gaps in the literature regarding access to administrative health data for research. First, despite a growing interest in the use of administrative health data for research in Canada, and evidence indicating that researchers have been experiencing barriers to access [5, 21-24, 26], *empirical* evidence is lacking. To date, there has been no systematic examination of researchers' experiences relevant to accessing administrative health data.

Second, inter-provincial variations in the timeliness of access to data have not been reliably reported and presented in the literature. While several research teams have noted variations in data access timelines when carrying out studies involving administrative health data across multiple provinces [22, 23, 27, 28], the extent of these variations and the underlying factors remain largely unknown. Specifically, it is unclear whether these variations are due to differences in data access processes, or if they are indicative of barriers to access.

Third, little is known about how relevant laws and ethical guidelines are being applied in practice and what the implications are for research involving administrative health data. Burdensome processes have been identified as a barrier to accessing administrative health data in Canada [22]; however, little is known about the specific

processes that are in place (i.e., what is happening “on the ground”), how these vary across provinces, and the impact on researchers’ ability to access administrative health data for research, and the timeliness of access.

Fourth, a broad range of factors have been identified as *potentially* affecting researchers’ access to administrative health data, however, the extent to which these factors *actually* impact data access is unknown. For example, the factors identified by van Panhuis et al [159] were specific to the sharing of public health data, so their applicability to administrative health data is not clear. Other factors were identified from within the literature related to research governance. Within this literature, various issues were highlighted that may have implications for data access; however, this literature is largely comprised of commentaries and legal analyses regarding *potential* limitations and challenges of the governance framework. Thus, empirical research is needed to identify the broad range of factors that impact access to administrative health data for research specifically.

CHAPTER 3: METHODOLOGY AND METHODS

3.1. Methodology

This study used a qualitative, multiple case study approach, underpinned by pragmatism, to explore the factors affecting access to administrative health data for research purposes in three Canadian provinces. The justification for this approach along with details related to data collection and analysis are addressed in the following sections.

3.1.1. Philosophical underpinnings: pragmatism

The methodological approach selected for this study can be best understood through an examination of the researcher's philosophical starting point. A researcher's philosophical position encompasses their beliefs and assumptions about knowledge, which underpins their approach to research—"It shapes how [they] formulate [their] problem and research questions to study, and how [they] seek information to answer the questions" [173](p.18). This study was carried out through the lens of pragmatism, which aims to address social problems using logical, common-sense approaches [174-176].

The influences of pragmatism were present early on in study conceptualization, as evidenced through the research questions that were asked. Researchers who operate from a pragmatic starting point tend to focus on "solving practical problems in the real world" [177]. The questions asked in this study, framed in terms of "factors affecting access to administrative health data for research", seeks to identify barriers to data access so that they may be mitigated, thereby improving the extent to which routinely

collected administrative health data can be leveraged to optimize the health of populations.

Decision-making regarding the selected methodological approach was also influenced by pragmatism, which emphasizes the practical aspects of research rather than ideology [177]. In fact, the approach of pragmatists has been described in two words: “whatever works” [178](p.21). Importantly, this does not mean that pragmatism is without ideology, rather that it embraces positivism and constructivism as complementary viewpoints, providing a means to explore both the objective and subjective [178]. Pragmatism posits that an external world exists separately from us and that we gain knowledge of this world through our interactions with it, though this knowledge is limited by our interpretations of these interactions [175]. As such, the “knower and the known [are] inseparable”, and the internal/subjective and external/objective are “two sides of the same coin” [179] (p.1048). Thus, researchers who view the work through the lens of pragmatism are not tied to a specific set of methods based on their ontological assumptions, and are free to use the method or methods that are best suited to answering a specific research question [178]. In this study, freedom from ontological assumptions was reflected in the use of an approach to case study methodology that drew on three established yet distinct approaches, each underpinned by a different philosophical starting point [180].

Pragmatism also informed the approach to theory development employed in this study. Because pragmatists believe that knowledge is shaped by context and experiences, what is considered the truth is what “works” in a specific context and may

be revisited (and revised) as new knowledge emerges [174, 178]. This view of the truth as changeable is congruent with scientific inquiry. Just as repeated experiments under various conditions may reveal an accepted theory to be fallible, pragmatists may accept an idea or theory to be true until new information emerges to call its truth into question. As such, a theory or idea is not simply true or false, but rather becomes "truer or falser" as new information is acquired, or as circumstances change [175]. In this study, an inductive and iterative process of theory development was employed, allowing the theory to change and evolve as new information was collected and analysed, and providing space for theory refinement in the future as the body of literature in this area expands.

3.1.2. Case study methodology

This study used case study methodology, which has been broadly defined as:

> …a qualitative approach in which the investigator explores a real-life, contemporary bounded system (a case) or multiple bounded systems (cases) over time, through detailed, in-depth data collection involving multiple sources of information (e.g., observations, interviews, audiovisual material, and documents and reports), and reports a case description and case themes. The unit of analysis in the case study might be multiple cases (a multi-site study) or a single case (a within-site study)[173](p.97).

Justification

Case study methodology was selected for several reasons. First, it was selected because it provided an appropriate framework to address the research questions. This study sought to identify the factors affecting access to administrative health data for research and to examine how these factors varied across provinces. Thus, a methodological framework was required that would facilitate the examination of data access within individual provinces, as well as comparison across provinces. Case study methodology was ideally suited for this. In case study methodology, the unit of analysis is a "case", which represents a particular instance of the social phenomenon being studied [181]. Using case study methodology (more specifically, a multiple-case study design [182-184]) facilitated the in-depth examination of three provincial "cases", as well as the comparison of findings across cases.

Second, case study methodology was selected because of its utility in situations wherein the phenomena being studied cannot be separated from its context [183]. In this study, the factors affecting access to administrative health data for research purposes (i.e., the phenomena) could only be understood in terms of the interactions of the actors involved (i.e., researchers and regulatory stakeholders), the organizations in which they were situated, and the broader regulatory landscape in which they operated.

Third, case study was selected because it is useful for gaining an in-depth understanding of a complex social phenomenon about which little is known. Case studies use "particularization" (i.e., a focus on the details and uniqueness of the case) and "thick description" [185] to provide a rich and holistic account of the phenomenon

being studied [182]. In this study, the focus on a select number of cases allowed the researcher to gain an in-depth understanding of each case, as well as insight into the interaction of significant factors characteristic of the phenomenon [182]. Moreover, through the use of multiple data collection methods, the various facets of the phenomenon were explored [178, 186]. As such, case study methodology allowed for a comprehensive examination of each of the selected cases included in this study and was conducive to the identification of the multitude of factors affecting access to administrative health data for research purposes.

Selected case study approach

There are several different approaches to case study, including those of Stake [184, 185, 187], Merriam [182, 188], and Yin [181, 183, 189, 190]. Each of these approaches is underpinned by the author's philosophical orientation, resulting in three unique approaches to case study research (Appendix D). Specifically, Stake's work is situated in constructivism, Merriam's in pragmatism, and Yin's in post-positivism [180]. These three approaches share many commonalities, including a focus on studying a case, or cases, within a real-life context using multiple data sources (e.g., interviews, observation, focus groups, documents, etc.), the selection of cases based on what can be learned, and the option of single and multiple case designs. With that said, there are differences between approaches, most notably with regard to analysis.

Of the three approaches, Yin's is the most structured, setting out a detailed, case-study specific research design. Based on a quasi-experimental study design [183], Yin's approach employs deductive analytic processes, relying on the *a priori*

identification of theoretical propositions (similar to hypotheses in quantitative research), which are subsequently confirmed, rejected, or revised. In the context of the current study, this was not appropriate for several reasons. First, the emphasis on deduction is more useful for testing or revising theory than generating new knowledge [191], which was the focus of the proposed study as little was known about the factors affecting access to administrative health data for research in Canada, or the reasons for interprovincial variations in the timeliness of access, at the time this study was undertaken. Second, the limited knowledge base was not sufficient to inform the development of theoretical propositions. Third, and perhaps most importantly, the study objectives could only be addressed by examining the first-hand experiences of individuals who both seek access to data, and the perspectives of those who regulate it, which required an inductive analytic approach.

The approaches of Stake [184, 185, 187] and Merriam [182, 188] are both fundamentally inductive [180], building conclusions from the "data-up" [192]. Stake's approach is highly interpretive and subjective [185], relying on the researcher to assign meaning to the data based on their own knowledge and experiences. Stake describes this analytic process as involving "much art and much intuitive processing" (p.72), as having a "mystical" side (p.72), and his approach as "greatly subjective" (p.77). Merriam's approach, on the other hand, sets out a stepwise process for data collection and provides practical strategies for ensuring rigor throughout the research process [182, 188]. Toward minimizing subjectivity and ensuring that study findings reflected the

experiences and perspectives of key informants, Merriam's analytic approach was determined to be most appropriate for the current study.

Consistent with the pragmatic worldview underpinning this study, a case study approach informed by the works of all three authors was used. While analysis was informed primarily by Merriam [182, 188], specific guidance from Yin and Stake was also incorporated where it was considered to strengthen the overall methodology. For example, while all three approaches included multiple case studies as potential study designs, the current study incorporated embedded units of analysis as proposed by Yin [181, 183, 189, 190] (see section 3.2.1). Similarly, case selection was based on maximizing the knowledge to be gained, as recommended across all three approaches, as well as additional practical guidance from Stake [185] (see section 3.2.2).

3.2. Methods

3.2.1. Study design

This study employed a multiple case study design (also referred to as "multi-site" [182] or "collective"[185] case study) with embedded units of analysis as set out by Yin [183] (Figure 1). A multiple case study design was used to facilitate the comparison of factors affecting variations in data access across provinces. In accordance with the study objectives, there were two embedded units of analysis, comprised of two stakeholder groups of interest: 1) researchers and research staff who have accessed or sought access to administrative health data, and 2) individuals involved in the regulation and oversight of data access. The groups differed in terms of their responsibilities, training, interests at stake, and their role within their affiliate institutions. As such, each group

was expected to provide unique insights into the barriers and facilitators to access to administrate health data. These stakeholder groups are described in more detail in section 3.2.3.

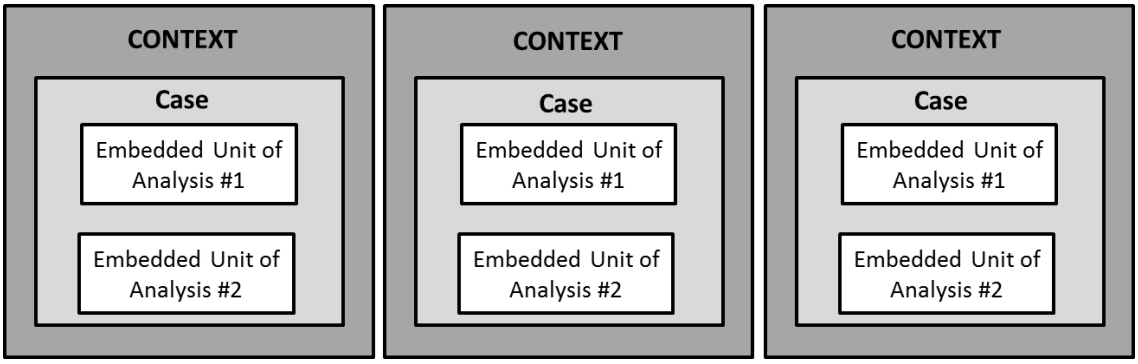


Figure 1. Multiple case study design with embedded units of analysis (modified from [183]).

3.2.2. Case selection

In case study methodology, the unit of analysis is a "case" or "cases" [182, 185]—that is, an "integrated system" with "a boundary and working parts" [185] (p.2), or a "bounded system" [182](p.40). When conducting a multiple case study, each case represents a particular instance of the social phenomenon being studied [181]. In this study, the central social phenomenon being studied was access to administrative health data for research purposes. To facilitate the examination of the impacts of various contextual factors on access to data, including differences in provincial legislation, one case was included from each of three different provinces. Each case was defined as a "research system" comprised of: (1) a data centre housing a provincial health data repository and serving as a point of access for researchers, (2) research stakeholders who sought access

to administrative health data via the data centre, and (3) relevant regulatory stakeholders (i.e., individuals involved in the regulation and oversight of access to administrative health data for research) within each province. The three provinces (and data centres) selected for inclusion were: NS (HDNS), BC (PopData), and ON (ICES).

In accordance with the methodological literature, case selection was based on maximizing the knowledge gained [182, 183, 185]. As recommended by Stake [185], this meant taking into account not only what could potentially be learned from each case, but the extent to which each case was accessible to the research team—that is, where permission could be obtained to gather data, where documentation existed and was most likely to be shared with the research team, and where people were willing to discuss the case. More specifically, case selection was based on the following considerations:

1) Existence of a data centre housing a provincial health data repository—At the time of case selection (2019), just over half of Canadian provinces/territories had a data centre/research unit that housed a provincial health data repository (i.e., BC [13], MB [14], ON [15], NB [16], NS [17], PEI [18], and NL [19]). Selecting cases based on the existence of such centres, ensured that infrastructure was in place to support research involving administrative health data as well as processes by which access to these data may be obtained.
2) Variations in case attributes and contextual factors—Cases were selected to represent a range of instances of the phenomenon of interest, consistent with a maximum variation sampling approach [193]. To facilitate the selection of cases,

information about the various data centres in place across the country was obtained from various sources, including web searches, data centre websites, and published literature. Variation in the following case attributes was sought: i) data centre structure, organization, and funding; ii) relevant legislation, and iii) reported timeliness of data access (as reported in [21]). This approach allowed the impact of a broad range of factors on data access to be examined, and helped illuminate the specific circumstances under which findings would "hold true" [183] and identify fundamental aspects of the phenomena being studied [194].

3) Ability to access the case—To help ensure sufficient access to each case, whether and the extent to which the researcher and committee members had existing professional relationships with members of the stakeholder groups of interest in each province was also considered. Relationships were important from a practical perspective, helping to ensure entry into each case, and improving the likelihood that sufficient data would be collected for theory development [185]. Relationships and access to the case were considered after the considerations described above in (1) and (2).

With no specific criteria for determining the appropriate number of cases to include in a multiple case study, the decision to include three cases was based on several attributes of case study research identified within the methodological literature. First of all, the aim of case study research is not generalizability [173, 185]. As such, the primary consideration during case selection was not how many cases to include, but rather, which cases to include and what could potentially be learned from each. Second,

given that the overall aim of case study research is to gain an in-depth understanding of a phenomenon, the inclusion of a limited number of cases was considered preferrable as it is conducive to acquiring a depth of knowledge rather than breadth of knowledge [173]. Third, one of the challenges of case study research is that it is resource intensive, generating enormous amounts of data that require analysis and management [186]. Given the complexity of the phenomenon being studied (e.g., various pieces of legislation, the involvement of numerous stakeholder groups and organizations, etc.) the inclusion of three cases was considered feasible within the study timeframe.

3.2.3. Data collection

Case study research is characterized by the use of multiple data collection sources and methods [173, 182, 183, 185] to facilitate an in-depth understanding of a phenomenon and improve the strength of evidence where findings are supported by data from multiple sources [183]. In this study, data were obtained from two sources: key informant interviews and documentary evidence. Data were collected from both sources concurrently.

Key-informant interviews

Interviews are considered an essential source of case study data, providing important historical and contextual information relevant to a case, explaining events and behaviors, and understanding participants' opinions and attitudes [182, 183]. In this study, semi-structured, in-depth interviews with key informants were conducted. Key informants are defined as individuals who are particularly knowledgeable about a case,

or aspects of a case, and are therefore valuable sources of information to a researcher [195]. In the context of this study, key informants were sought from two stakeholder groups:

1) *Individuals with experience accessing or attempting to access administrative health data for research purposes ("research stakeholders").* Key-informants in this group included academic researchers (e.g., university faculty or affiliated researchers), research trainees (e.g., students and post-doctoral fellows) and research staff (e.g., research associates, assistants, and coordinators) with experience accessing or attempting to access administrative health data held by the relevant provincial data centre in the five years prior to being interviewed.
2) *Individuals involved in the regulation and oversight of access to administrative health data for research purposes ("regulatory stakeholders").* Key informants in this group were broadly defined as individuals with a role in developing and implementing data access policies and processes in each province in the five years prior to being interviewed, for example: individuals affiliated with provincial data centres; members of relevant review bodies (e.g., data access committees, privacy review bodies, and REBs); data stewards and custodians, and privacy officers situated within relevant institutions and organizations (e.g., universities, health authorities, and provincial ministries/departments of health). Importantly, the individuals who were considered key informants varied across provinces.

Key informants from both groups were identified using two strategies. First, a general purposive sampling strategy was used to ensure that study participants included

those individuals from whom the most can be learned [194]—that is, to ensure that the individuals invited to participate were in fact key informants. A list of potential key informants was prepared by the researcher based on their knowledge of the research and regulatory landscapes in Canada as well as publicly available information obtained from online searches and relevant published documents and reports. This list was reviewed and refined by members of the supervisory committee. Secondly, as key informants were recruited and interviewed, snowball sampling [194] was used—that is, key informants were asked to identify other potential participants with relevant expertise in their province.

Individuals who were identified as potential key informants were contacted by the researcher via email and invited to participate. A study summary was included with each email (Appendix E). For individuals who responded that they would be interested in participating, an interview time was arranged, and a consent form (Appendix F) was sent via email. All participants provided written consent to participate in interviews. In addition, at the scheduled interview time, the researcher reviewed study details with the participant, addressed participant questions, and confirmed consent prior to starting the interview.

Interviews were conducted in person and via telephone by the researcher. Interview format and interview guide development were informed by the work of Patton [194] and Rubin and Rubin [196]. Interviews were semi-structured, taking on the form of a guided conversation rather than a structured interview [183]. Using this approach, a core set of interview questions was developed for each stakeholder group;

however, the order in which these questions were asked, and the specific wording used, varied from person to person. Probes, or follow-up questions, were used to gain additional information, clarification, and/or illustrative examples as needed [182]. Key-informants from both stakeholder groups were asked questions to establish the facts of the case and obtain information relevant to Objective 1 (i.e., overview of the data access process, required approvals and documentation, actors involved, and relevant policies and legislation). For research stakeholders, the remainder of the interview focused on their experiences accessing administrative health data for research purposes in each province (Objective 2), including their ability to access data and perceived factors affecting access to administrative health data for research (Interview Guide, Appendix G). For regulatory stakeholders, the remainder of the interview focused on their perspectives on the use and regulation of administrative health data for research, the factors affecting access to administrative health data for research, and their regulatory role (Objective 3; Interview Guide, Appendix H).

Interviews were audio-recorded and transcribed to retain an accurate, verbatim account of the conversation for analysis. Drawing on guidance from Stake [185], notes were taken throughout interviews to capture impressions, particularly useful quotations, key ideas or concepts to be explored in subsequent interviews, questions that need to be revised or reframed, and to capture recommendations of documentary evidence to include or other key informants to invite.

Documents

Documents were used to confirm, corroborate, and supplement information gained from interviews [183]. It was also used to obtain information that was not otherwise available, such as important historical and contextual information relevant to each case [182, 185]. Inferences arising from documents were substantiated via interviews [183].

Documents were obtained from online sources and provided by key informants. Access to a wide array of documents relevant to each case was sought, including policy documents (e.g., provincial legislation and regulations relevant to research involving administrative health data, institutional policies and guidance documents for researchers and oversight bodies); data access documents (e.g., data access forms, data sharing agreements); and evaluations of the provincial health information legislation, and provincial or institutional research reports.

3.2.4. Data analysis

Data collection and analysis occurred concurrently, as is common in qualitative research [182]. This allowed data collection to be adjusted in response to the emergent theory (e.g., to revise or add questions to gain clarity or additional information on emerging themes) and new data collection opportunities to be taken advantage of if they arose (e.g., the discovery of new documents) [197]. An analytic approach consistent with that of Merriam [182] was used. Since a multiple case study design was used, analysis was undertaken at two "levels": (1) within-case analysis, and (2) cross-case analysis, whereby findings from individual cases were compared.

Within-case analysis

During within-case analysis, each case was analyzed separately from the others and treated as a stand-alone study. This was done to gain an in-depth understanding of each individual case, which is required prior to attempting to identify commonalities or differences across cases [198]. Within-case analysis involved the analysis of data from each individual data source, the integration of data from multiple data sources, and the development of detailed case descriptions.

Development of detailed case descriptions

For each case, a detailed case description was developed to establish the facts of the case, including important contextual and historical information [173, 182]. In case study research, the phenomenon being studied cannot be separated from the context in which it takes place [182, 183], so a well-developed case description provides a useful frame of reference for interpreting findings. A detailed case description also facilitates the transferability of findings, allowing readers to have a vicarious experience of the phenomenon to determine whether and how these experiences can be applied or transferred to new settings [199]. In this study, case descriptions focused on: (1) describing the provincial data centres (i.e., legal designation, organization and structure, role, funding, and data holdings), (2) summarizing relevant provincial legislation (i.e., key provisions relevant to the secondary use of personal health information for research), and (3) describing data access processes for accessing linked administrative health data via each provincial data centre (i.e., required applications, reviews, and approvals).

Analysis of individual data sources

Interviews

Interview transcripts were analyzed using the constant comparative method developed by Glaser and Strauss [193], which involves a process of coding and categorization to develop a theory or explanation. Since it was originally developed, several specific approaches to coding and categorization have emerged. This study used an approach consistent with that of Strauss and Corbin [200], involving open, axial, and selective coding.

To guide analysis, the researcher, and a committee member (RU) with expertise in qualitative data analysis developed a codebook. Since interviews were undertaken in NS first, the codebook was developed using four interview transcripts from the NS case (2 from each stakeholder group). The four transcripts were coded independently by the researcher and committee member, who then met to review the codes, discuss their conceptual basis, and refine as needed. Once a revised list of codes had been agreed upon, the researcher resumed coding the remainder of the transcripts for the NS case. The researcher and committee member met again once all transcripts for the NS case were coded to for additional discussion and refinement of codes and concepts. This codebook was subsequently applied to the coding of transcripts for the BC and ON cases. The researcher met with RU as needed throughout the duration of the analysis process to discuss, refine, and clarify codes. Coding was facilitated by the use of qualitative data analysis software (NVivo, QSR International).

During open coding, each transcript was reviewed and codes (i.e., labels) were assigned to units or passages of text. This was followed by axial coding, during which conceptually similar codes were collapsed into categories. Axial coding was followed by selective coding, which is an interpretive process whereby core categories or themes are identified and the relationships between them are described—that is, it is the process of theorizing. As the name suggests, the constant comparative method of analysis is highly iterative. Thus, as new data were collected and analyzed, they were compared to those that has already been collected and analyzed, resulting in ongoing refinement of codes, categories, and themes. Text descriptions of each category and the relationships between them were developed and supplemented with use of illustrative quotes.

The researcher met with committee members (individually and as a group) on an as-needed basis throughout the analysis process to review and discuss the emerging theory. Data analysis (and collection) was considered complete on the researcher determined that theoretical saturation was researched—that is, when the collection of additional data did not contribute to new codes or themes [193, 201, 202].

Documents

A preliminary review of each document was performed to determine relevancy and importance. Where it was immediately clear that a document was not relevant to the case, it was excluded from the study. For each case, a database of documents was created, containing document title, date of creation, document type, author and/or institution to which it pertains, and brief summary of content (2-3 sentences). This process of reviewing and cataloging was done to facilitate the process of "triaging"

documents to identify those most relevant to each case [183]. Those identified as being most relevant were reviewed in greater frequency and detail over the course of the study. Thus, document analysis was an ongoing and iterative process rather than a one-time occurrence. Information contained within case documents were used in several ways:

1) Facts pertaining to the case (e.g., local context, organizational history, relevant policies, descriptions of data access processes) were extracted and incorporated into the development of case descriptions.
2) Names and roles of key stakeholders (individuals and organizations) contained documents were used to inform recruitment efforts.
3) Inferences and conclusions drawn on the basis of documents were explored further in interviews.
4) Information contained within documents that corroborated or contradicted information from interviews were used to facilitate the process of triangulation [183].

Integration of data from multiple sources

Interview data were collected for all three study objectives and therefore comprised the majority of data that were collected and played the largest role in theory development; however, documents were an important complement to the process of theory development, particularly given their role in the development of case descriptions (Objective 1), which provided the context for interpreting interview data, and in the triangulation of interview data (Objectives 2 and 3). The integration of evidence from

documents and interviews was “built-in” to both data collection and analysis, with each activity directly influenced the other. During data collection, the information contained in documents influenced interviews as the researcher probes and sought clarification and additional details regarding information contained, while additional study documents were identified via interviews. During analysis, interview data elucidated information obtained from documents, while documents provided specific pieces of information to confirm (i.e., triangulate), clarify, and provide a more detailed understanding of key elements of or address gaps in the emergent theory. More broadly, each individual document and stakeholder interview provided various pieces of the “puzzle”, contributing to the overall understanding of the phenomenon being studied [186].

3.2.5. Ensuring rigour and trustworthiness

A variety of strategies were used to improve study rigour (Table 3). These included case-study specific strategies to improve validity and reliability as set out by Yin [183], as well as general strategies for improving the “trustworthiness” of qualitative research [182, 185, 203, 204].

Table 2. Summary of strategies to improve rigor.

Strategy	**Description**
Development and use of a case study protocol [183]	A study protocol was developed and published, and the research was conducted in accordance with the protocol to improve transparency.

Strategy	Description
Reflexivity [182, 203, 204]	The researcher engaged in a process of reflexivity to improve credibility and transparency of the research. This involved acknowledging and explicitly stating personal biases and assumptions relevant to the research, and the potential impacts on study conduct and findings (see Section 5.5., 'Researcher reflexivity').
Triangulation of data sources [182, 183, 185, 203, 204]	Triangulation is the use of multiple evidence sources to corroborate a single phenomenon. In this study, interviews with two key stakeholder groups and documents were used to achieve convergence. Discrepancies in information from different data sources were resolved through the collection of additional data where possible.
Adequate engagement in data collection [182, 203, 204]	Data were collected until saturation was reached (i.e., no new substantive concepts or findings emerged from new data collection).
Peer review/examination [182, 203, 204]	The researcher met with committee members throughout the study to discuss study processes (i.e., practical and/or methodological issues), emergent theory, and preliminary interpretations.
Use of rich, thick descriptions [182, 203, 204]	Each case and relevant contextual information were described in as much detail as possible to facilitate the reader's ability to assess the transferability of findings to other settings.
Maximum variation sampling [182, 203, 204]	Cases were selected to represent a range of case attributes and contextual factors. This provided the opportunity to explore a broader range of circumstances under which specific study findings occur, improving the robustness of the emerging theory (see Section 3.2.2., 'Case selection').
Establish a chain of evidence or audit trail [182, 183, 203, 204]	Documents and records were retained to demonstrate a direct link between the research questions/objectives and the study findings. this includes case documents, a codebook, notes capturing reflections on data, and notes documenting justification for analytic decisions.

CHAPTER 4: RESULTS

4.1. Data sources

4.1.1. Key-informant interviews

Interviews were conducted with a total of 46 key informants across the three cases (18 in NS, 14 in BC, and 14 in ON) between February 2020 and July 2021. Data collection commenced in NS where seven interviews were carried out in-person prior to the implementation of restrictions due to the COVID-19 pandemic in mid-March 2020. The remainder of the interviews for NS, and all interviews for BC and ON, were carried out via telephone. The participation rate was highest in NS (90%), followed by BC (67%) and ON (56%).

A summary of key informants by province and stakeholder group is provided in Table 3. Across cases, the research stakeholder group was comprised of trainees, research staff, and researchers with varying levels of experience with administrative health data (ranging from less than 5 to more than 20 years of experience). All individuals were affiliated with an academic institution and had direct experience accessing administrative health data for research purposes within the study timeframe, or experience overseeing staff and/or trainees who had direct experience accessing administrative health data for research purposes. Research stakeholders' university affiliations were as follows: in NS, Dalhousie University; in BC, University of Victoria, University of British Columbia, and Simon Fraser University; in ON, University of Toronto, Queen's University, McMaster University, and University of Ottawa. Research stakeholders were engaged in health services research spanning healthcare sectors

(primary, secondary, and tertiary) and a range of topic areas including, but not limited to: oncology, nutrition, frailty, mental health and addictions, end-of-life and palliative care, occupational health and safety, geographical health inequities, and health economics.

The regulatory stakeholder group was comprised of individuals who played a role in the regulation and oversight of access to administrative health data (i.e., in the development and/or implementation of policies and processes governing access to data). Key informants in this stakeholder group varied across provinces in terms of specific role, including privacy officers/analysts, individuals responsible for ensuring compliance, legal experts, members of data access committees, and individuals in leadership roles within relevant organizations. In NS, key informants included individuals affiliated with HDNS, the NS Department of Health and Wellness, NS Health, and the IWK Health Centre. In BC, key informants included individuals affiliated with PopData, the Provincial Services Health Authority (PSHA), and the Ministry of Health. In ON, all key informants in the regulatory stakeholder group held roles at ICES.

While the majority of individuals belonged to a single stakeholder group, several individuals in each province had experience as members of both groups. For example, several members of the research stakeholder group also held current or prior regulatory roles (e.g., members of data access committees, leadership positions at provincial data centres, etc.). Where it was known to the researcher at the outset of the interview that the individual held dual roles, the individual was asked whether they would prefer to provide the researcher perspective, or the regulatory perspective. Individuals who

completed the research stakeholder interview guide were included in the research stakeholder group regardless of any other current or prior regulatory roles held, and vice versa.

Table 3. Description of stakeholder groups by case.

Case	Invited	Participated	Research stakeholders		Regulatory stakeholders	
			Number of participants	Number with regulatory experience	Number of participants	Number with research experience*
NS	20	18	9	3	9	6
BC	21	14	9	3	5	2
ON	25	14	7	1	7	7
Total	66	46	25	7	21	15

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

* Not limited to experience involving administrative health data.

4.1.2. Documents

A summary of the documents analyzed for each case is provided below (Table 4). The specific documents obtained for each case varied, reflecting differences in both the documents that existed relevant to each case and the researcher's access to documents. For example, published data centre profiles existed for PopData BC and ICES, but not for HDNS. In terms of the researcher's access to documents, the researcher had access to HDNS application documents, forms, and policies, as well as local REB applications as a result of having made multiple data access requests in recent years, whereas there was a greater reliance on key informants to gain access to these documents for PopData and ICES. In addition to these documents, the website for each data centre was a key source

of case information—and often a source of other relevant documentation—although the websites for ICES and PopData were more comprehensive than that of HDNS.

Where detailed information on specific aspects of each case could not be obtained from documents, deliberate efforts were made to address these gaps during interviews. For example, there were fewer historical documents available pertaining to the NS case compared to the BC and ON cases. This "gap" was addressed via interviews with key informants who were particularly knowledgeable about the history and evolution of HDNS.

Table 4. Summary of documents analyzed for each case.

Document type	**NS**	**BC**	**ON**
Data centre internal policies	X		X
Data centre organizational chart		X	
Feasibility and cost assessment forms	X		
Application forms (e.g., data access request forms, privacy impact assessment forms, project activation worksheets)	X	X	X
Supporting documents for applications (e.g., non-disclosure and confidentiality agreements, student application documents)	X		X
Data access guidance documents	X	X	X
Data centre annual reports			X
Research ethics board policies/standard operating procedures	X		
Research ethics board applications	X		
Statutes and regulations	X	X	X
Government documents and reports		X	X
Publications in peer-reviewed journals (e.g., data centre profiles)		X	X
Internal communications (e.g., emails, memos)			X
News articles		X	

4.2. Comparison of selected cases

Each provincial case included in this study was centred around a data centre with a provincial health data repository—HDNS in NS, PopData in BC, and ICES in ON. For publicly funded academic researchers seeking access to linked administrative health data for research, these data centres were the primary point of access during the study timeframe. Key characteristics of the data centers are summarized in Table 5 and described in greater detail in the sections that follow (4.2.1 to 4.2.6).

Table 5. Key characteristics of provincial data centres.

	Provincial Data Centre		
	HDNS [17]	**PopData [13, 205]**	**ICES [15, 89]**
Year established	1992	2009 *	1992
Legal designation	Agent	Custodian	Prescribed entity
Role	Service provider	Service provider	Research institute; Service provider
Organizational type	University-based	University-based	Independent not-for-profit
Sites	Dalhousie University	Simon Fraser University; University of Victoria; University of British Columbia	ICES Central; ICES U of T; ICES North; ICES Queens; ICES McMaster; ICES North; ICES University of Ottawa
Funding	User-fees; Other sources	Provincial government; User-fees; Other sources	Provincial government; User-fees; Other sources
Number of datasets†	9	30+	100+
Type of datasets	Administrative; Registry	Administrative; Registry; Survey	Administrative; Registry; Survey; Derived cohorts; Clinical extracts; Other

	Provincial Data Centre		
	HDNS [17]	**PopData [13, 205]**	**ICES [15, 89]**
Data sources	Nova Scotia Department of Health and Wellness	Ministry of Health; Other data providers	Ministry of Health and Long-Term Care; Other data providers
Data domains	Health; Demographic	Health; Demographic; Education; Social; Environment and resources; Work and income	Health; Demographic; Financial

Abbreviations: HDNS=Health Data Nova Scotia

* PopData was established in 2009 but has origins dating back to the late 1980s with the creation of the BC Linked Health Data set (BCLHD)[205].

†Based on recent sources: HDNS [17], PopData [13], ICES [15].

4.2.1. Organization and governance

HDNS is a university-based research unit, situated within the Department of Community of Health and Epidemiology at Dalhousie University. It is located on the Dalhousie University campus in Halifax and consists of a small team of staff, including a Database Administrator, Data Coordination Specialist (i.e., "Data Navigator"), Regulatory Compliance and Finance Officer, and several Data Analysts. These staff report to the Manager, who is accountable to the Director. This role is held by a faculty member in the Department of Community Health and Epidemiology, who reports to the Department Head.

PopData is also a university-based data centre and is physically located at the University of British Columbia in the School of Population and Public Health. However, it involves a partnership across three universities—the University of British Columbia, the University of Victoria, and Simon Fraser University [205]. There are approximately 30

staff employed in a range of roles working across a number of different “units”, including: business development, communications, data access, data partnerships, data services, education and training, operations, privacy and governance, strategic projects, and systems and security [13]. Day-to-day operations are overseen by unit Leads and the Managing Director, who reports to the Scientific Director. Two advisory groups also inform the governance and management of PopData: the Interim Advisory Board, which advises on core operational issues; and the Data Stewards Working Group, which advises on data access policies and processes.

Of the three data centres, ICES is the largest and most complex in terms of organizational structure and governance. It is an independent, not-for-profit organization that has six physical sites across Ontario. ICES Central was established in 1992 and operated as the sole site for 15 years. To expand access across the province, five new sites were added: Queen’s in 2007, uOttawa in 2010, UofT and Western in 2012, McMaster in 2016, and North in 2018 [206]. ICES employs a large and growing number of staff across sites (up from 259 in 2016 [207] to 300 hundred 2022 [208]), including data analysts, privacy and legal experts, communications staff, information security personnel, administrative and financial staff, and human resources professionals. Staff report to various Managers and Directors, who report to the Chief Executive Officer (CEO). The CEO is accountable to the Board of Directors who guide the organization’s strategic direction, oversee the CEO’s performance, and approve major financial decisions [209]. The Board of Directors is advised by a Scientific Advisory

Committee, which provides input on ICES' research agenda (e.g., scope, priorities, and direction) and reviews ICES' achievements [210].

4.2.2. Data centre role

HDNS and PopData are service providers—they do not have their own research agendas, but rather, provide a range of services to academic researchers at a cost. Services include assisting with the preparation of data access applications, providing access to their internal data holdings and linkage to external datasets, as well as providing access to a central server and secure data environment for data storage and analysis [13, 17]. ICES is unique in that it is a research institute—it has its own research agenda, and a growing membership of ICES "scientists" or affiliated researchers (up from 208 in 2016 [207] to 285 in 2022 [208]). Historically, all research carried out using ICES data was carried out internally by ICES-affiliated researchers and staff. This changed in 2014 when ICES established Data and Analytic Services (DAS), which sought to expand access to data to third-party (non-ICES-affiliated) researchers in the public and private sectors, and to knowledge users within the healthcare system [89, 211]. DAS is the service-provider division of ICES and functions similarly to HDNS and PopData, offering services to researchers at a cost. However, unlike HDNS and PopData, DAS provides data to both public and private sector researchers. Most public sector research using ICES data continues to be carried out internally by ICES-affiliated researchers (i.e., within the research institute). Notably, there is a third data access pathway at ICES, which is the "Applied Health Research Question" pathway, or AHRQ. Using this pathway, knowledge users from Ontario organizations (e.g., health system policy- and decision-makers) can

request research evidence to support health system planning and management. This pathway is not considered within the scope of the current study.

4.2.3. Legal designation

The legal designation of each data centre varies. ICES is designated as a prescribed entity under ON's *Personal Health Information Protection Act* [147] and corresponding regulations [212], which gives it the authority to collect PHI from other data custodians without individual consent and to disclose PHI for research purposes. Access to data is regulated internally, in accordance with policies and processes that have been approved by the Information and Privacy Commissioner (IPC) of Ontario [213].

HDNS is not named within the provincial health information legislation or regulations; instead, its authority comes from a data sharing agreement between the NS Department of Health and Wellness and the Department of Community Health and Epidemiology at Dalhousie University. HDNS receives all of its data from the Department of Health and Wellness, which is the legal data custodian under the *Personal Health Information Act* [149]. In the data sharing agreement, HDNS is designated as an agent of the Department of Health and Wellness, which gives it the authority to act on the Department's behalf in terms of facilitating access to data.

In BC, the disclosure of data to PopData BC is permitted under the *Freedom of Information and Protection of Privacy Act* [142]. PopData is regarded as a data custodian in that it holds or has "custody" of data from the various data providers with whom data sharing agreements are in place, though this is not codified in any particular piece of legislation. Decision-making authority regarding data use is retained by each

organization or institution that provides data (e.g., BC Cancer, Ministry of Health, WorkSafe BC, etc.), which appoints a data steward who is responsible for the review and approval of data access applications [13, 205].

4.2.4. Data holdings

Data holdings at each centre fluctuate over time as data sharing agreements change. At ICES and PopData, internal data holdings have grown over the years as a result of efforts to increase the total number of datasets, and to include data from various domains (e.g., social, educational, judicial, etc.). ICES holds over 100 datasets [208](up from 90 in 2019 [89]), including administrative, clinical, registry, and survey datasets from a variety of sources. While some of ICES' internal data holdings are considered 'general use' data, others require 'special permissions' as set out in the data sharing agreement between ICES and the original data provider. PopData's data holdings are substantially smaller, with approximately 30 datasets, including administrative, registry, and survey datasets from various data providers [13]. HDNS' data holdings consist of nine datasets, all of which are administrative health data provided by the NS Department of Health and Wellness [17].

4.2.5. Funding

Funding sources vary substantially across data centres. PopData receives operational funding from its data provider partners (e.g., Ministry of Health, Worksafe BC, BC Cancer, and others) and through cost-recovery (i.e., user fees), with additional funding from other funding partners and competitive grants (e.g., CIHR's Strategy for Patient

Oriented Research (SPOR) [214]) [205]. Similarly, ICES receives operational funding from the Ministry of Health and Long-Term Care, and through cost-recovery, and funding from various partners and grants, including CIHR SPOR [89]. HDNS is the only province that does not receive core funding from the provincial government. Since its inception, it has been funded almost entirely through cost-recovery, though in recent years it has shifted to a partial cost-recovery model as a result of additional funding received through involvement in national research initiatives (i.e., CIHR SPOR and CNODES) [215]).

4.3. Comparison of access to administrative health data across provinces

During interviews, key informants described access to data in terms of four "outcomes": (1) *data acquisition,* (2) the *timeliness* of obtaining access to the requested data, (3) the total *costs* associated with obtaining access to the requested data, and (4) the *quality* of the dataset. Each of these outcomes is described below and accompanied by illustrative quotes from both regulatory and research stakeholders.

4.3.1. Data acquisition

In each case, responses from both research and regulatory stakeholders indicated that researchers who submitted a formal data access were typically successful in obtaining the required approvals and acquiring the requested data. In each case, the processes in place at the provincial data centres to confirm feasibility (i.e., feasibility and cost assessments) meant that potential issues related to data acquisition were typically identified early on, minimizing the potential for issues to arise later. For studies

involving external linkages, research teams often connected with the data provider early on to confirm data availability, study feasibility, and establish the data provider's support. As a result of these steps, where a formal data access was made, data acquisition was likely.

Where a formal data access request was submitted, if issues were identified with aspects of the study (e.g., feasibility, methodological, ethical, or privacy concerns) or with the quality or completeness of submitted applications, researchers were given multiple opportunities to provide clarifications and/or make any necessary revisions and resubmit. In other words, it was unlikely that researchers seeking access data held by the provincial data centre would receive a hard "no". On the rare occasion that a researcher was refused access to data outright or was unable to access the data within the study timeframe, the issue was usually with an external data provider, not with the provincial data centre. If approvals were obtained, researchers typically acquired the requested data (see Table 7), but often with caveats related to timeliness, costs, and data quality.

Table 6. Illustrative quotes regarding data acquisition.

Case	Illustrative quotes
NS	I mean, every request I've put in has gotten approved. So I've never had something not approved. [P13] Like we're not in the business of just saying no. We're in the business of what can we do to make this work? [P21]
BC	Okay, from my experience, other than people who are completely not eligible to apply for data, in BC from my experience working with data requests, I do not see a lot of requests being rejected. [P20] I requested intake data from [external data provider]. And because it's not normally used for research purposes, it comes with its fair share of challenges, and it also comes with the possibility that I could have just been denied. Whereas if you apply to PopData and you answer all of the questions, and you follow the process, and you have a good rationale, and you can pay for it, you're not going to be denied because there's an infrastructure for that process. [P24]
ON	Yeah, I would say that the actual number of projects that we flat out reject are very few as an organization. [P42] So you need to have all your ducks in a row, you know, in order to gain access. But yeah, I haven't been denied any data that I can think of. [P34] I think, in general, particularly in Ontario and particularly in today's ICES environment, if someone needs to access the administrative data, they can. I think we have lots of different pathways for people now. [P46]

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

4.3.2. Timeliness of access to data

Across cases, the main challenge associated with accessing data, from the perspective of both researchers and regulatory stakeholders, was timeliness (see Table 8). Researchers who were interviewed for this study reported a wide range of experiences in terms of the time required to obtain access to a linked administrative health dataset for research. Within each case, reported data access timelines (from initial application submission to obtaining access to the linked dataset) ranged from several months to several years.

Although it did not typically take years to gain access to a linked dataset for research, multiple researchers in each case described having experiences where obtaining access took one to three years. Across cases, the unpredictable nature of data access timelines and the potential for extensive delays in access were a cause for concern among researchers. In fact, the primary concern for many researchers was not whether they would get access to data, but how long it would take.

Timeliness was a concern for several reasons, including: (1) the need to provide timely research evidence to inform health-system decision-making, (2) the impact on researchers' ability to complete studies within the funding timeframe, particularly for shorter grants (1-2 years), (3) the impact on trainees' abilities to meet deadlines for program completion, and (4) the impact on grant funds, which could be largely depleted in the pursuit of access to data, leaving inadequate resources for remaining project components.

Issues related to timeliness were also acknowledged by regulatory stakeholders affiliated with the provincial data centres (all cases), and those affiliated with external organizations (NS and BC), who identified various steps that had been taken within their respective organizations to streamline processes and reduce delays.

Table 7. Illustrative quotes regarding timeliness.

Case	Illustrative quotes
NS	We have this information. Why… I mean we're collecting it for the health of Nova Scotians. And part of that is to be able to access it quickly to get information… And I don't think it's fast enough. I don't think the access is fast enough. [P1]

Case	Illustrative quotes
	[the process of accessing data] is like a super dauntingly long process. And it makes it I think difficult when you're applying for like one-year grants. Because even if you start as soon as you get that approval, like you have a month to get it to the data access committee. And then that approval process and like all the paperwork during that time. And then you have like the few months wait for it to get assigned to an analyst. And then depending on how straightforward your project is, you have like up to weeks to a few months to actually get the data. And that's already like your year. [P9]
BC	Where we fall short is our access times are way too long, and we haven't automated enough, and it's not a flexible enough system to be nimble to bring in new data sets very quickly. [P19] We want to make data access much faster. Because sometimes for some projects, researchers, they only have one year to…like under a grant, they may only have one year to work on the data. They cannot spend six or seven months waiting for the data and going through the data access process. [P20] But generally from when I write a proposal to when I can access the data is a year. But that's not in all of PopData BC's control, right? [P23]
ON	You know, it used to be when I came here, it used to be like everybody did ICES studies because, you know, like who wants to do a trial? Right. It's complicated. And you have to deal with real human beings. And so people… you know, graduate students almost always did ICES-related studies. But now I think people just feel that it's so complicated and delays are so long that they're less likely to necessarily do an ICES study for their graduate work. [P35] I mean, truthfully, we do get complaints about our response time, but we can only go as fast as we can. [P38]

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

4.3.3. Costs of accessing data

Each provincial data centre employed a cost-recovery model to cover operational costs, although each had a unique fee schedule. Researchers did not pay for the actual data, but rather, for the staff time and resources required for project administration (e.g., application processing, contracts and agreements), dataset creation (e.g., importing, extracting, linking, cleaning, and analysing data), and access to IT infrastructure (i.e., secure data platform/environment) and support. At HDNS, the total costs related to

accessing data were reported as commonly ranging between 10 to 20 thousand dollars [17], while PopData reported between 8 and 15 thousand dollars at [13], though these amounts could be dramatically higher depending on the particulars of the study. Information provided by research stakeholders in ON also indicated high, and variable, costs associated with accessing data via ICES. Where new datasets were brought into ICES secure data environment—requiring the removal of identifiers, assignment of ICES Key Number (IKN), and required data checks—the added costs were often substantial (tens of thousands of dollars, depending on the size of the dataset). Notably, PopData, HDNS, and ICES-DAS all offered reduced rates for students. Student projects carried out via ICES' internal data access pathway were not discounted, but typically cost less because students often did their own analysis and did not need to pay for analyst time.

Importantly, in each case, the costs associated with accessing data via the provincial data centre comprised only a portion of the total costs associated with accessing data. For some studies, where researchers linked to an external dataset, the data provider also charged a fee for providing the data. Researchers also faced substantial costs as a result of employing dedicated staff to support the data access process (e.g., to prepare and submit applications, respond to feedback relevant regulatory entities, facilitate the signing of contracts and agreements, etc.). Notably, staff-related costs were frequently compounded by delays in access to the analytic dataset. Where obtaining access to data took longer than expected—whether due to extensive toing and froing with relevant regulatory entities or limited capacity on the part of an external data provider to provide data—funds that had been budgeted for

other activities, such as analysis, were allocated to staff salaries. Overall accessing administrative health data was generally considered an expensive endeavor (see Table 9), though not necessarily cost-prohibitive.

Table 8. Illustrative quotes regarding costs.

Case	Illustrative quotes
NS	So I would say the cost here in our province is fairly high. But I understand why it is. So from HDNS' perspective, I can understand why. But I mean from a system perspective, I think it's a bit too much. [P13] You need somebody to prepare the data. You need the [staff] and the managers. You know, you need the data analysts there to help you put the data set together so that it's prepared for you in a way that you can do your research work with it. And I think there's a misunderstanding that that just doesn't all happen magically behind the scenes, kind of thing, at no cost. These are highly trained professionals with good education that need decent salaries. And there seems to be a bit of a begrudging sometimes when we start talking about the finance side. [P8]
BC	I think that's good value for money. And I've never had a research review committee look at the budget and say that's an unreasonable cost. Even when it gets a bit higher and maybe it's a complicated, like $25,000 plus annual fees. I've just never had a reviewer say that's an unreasonable cost. Rather it's probably under-valued, to be frank. Yeah. [P23] I mean I find that there's general rumblings that people like to complain about PopData - about the time that it takes and the cost. But I honestly think that they provide a huge service to researchers in BC. [P27]
ON	There's no way to do this cheaply. And cost recovery, I mean you're paying for… ICES has gotten, I'll just say, you know, in my opinion is that it's gotten relatively, as they're tried to do more and be fancier, it's gotten relatively more expensive than it was earlier. I would say it's probably twice the cost now that it was 10 years ago to do research using health admin data. [P33] And while you're having to wait, you still have pay [your staff]. And obviously, you try to find productive work. But at the same time, you know, while all of these delays are going on, and more forms need to be completed and, you know, the endless bureaucratic carryon, that's all opportunity cost. Because the research assistant or the student, you know, their salaries have to be maintained even though they're waiting or they're doing another bureaucratic thing. And of course it always leads to, you know, needing a no cost extension for the grant. Now, a no cost extension sounds really like it

Case	Illustrative quotes
	doesn't cost anything. But that's completely not true. A no cost extension, you're still paying the research assistants and whatever other infrastructure you need to support from the grant. It's not free money…or free work. So I think the costs, the opportunity costs and the costs themselves are substantial. [P35]

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

4.2.6. Data access requirements

Data access requirements relevant to each case are summarized in Table 6. The details provided reflect information available on organizational websites, published in the peer-reviewed literature, and obtained from key informant interviews. Detailed depictions of data access processes at each data centre are provided in Appendix I.

Table 9. Overview of data access requirements for each case*

Case	Required reviews and approvals for research	
NS [17]	REB	▪ All studies involving HDNS data required REB approval.
	Other	▪ All studies required application review and approval from the HDNS DAC. ▪ Studies involving external linkages required additional approvals, contracts, and agreements; however, these varied by data provider.
BC [13, 205]	REB	▪ All studies involving PopData data required REB approval.
	Other	▪ All studies required review and approval from the relevant data steward for each dataset being accessed. Review was coordinated by PopData's DAU. ▪ All studies required external peer-review. ▪ Studies involving external linkages required additional approvals, contracts, and agreements; however, these varied by data provider.
ON [15, 89]	REB	▪ All research studies involving ICES data required REB approval (internal and DAS). The specific REB from which approval was required was dependent on the researcher's affiliation. ▪ Exception: For studies being carried out by ICES researchers affiliated with Sunnybrook Hospital that involved *only* ICES' internal data holdings and did *not* require "special permissions," researchers did not need to submit an REB

Case	Required reviews and approvals for research
	application. The review of the PIA form carried out by the ICES Privacy and Legal Office was considered sufficient to proceed. The Sunnybrook Hospital REB regularly reviewed a random sample of ICES studies to ensure compliance with ethical guidelines.
	Other ▪ All internal research studies required PIA review and approval from the ICES Privacy and Legal Office. ▪ For some data sources "special permissions" were required (depending on the particulars of the data sharing agreement with ICES). ▪ DAS studies require review by the Privacy and Legal Office but do not need to submit a PIA. ▪ Studies involving external linkages required additional approvals, contracts, and agreements; however, these varied depending on the datasets involved.

Abbreviations: BC=British Columbia, DAS= Data and Analytic Services, DAU=Data Access Unit, NS=Nova Scotia, ON=Ontario

*Does not reflect changes that have occurred since time of data collection (2020-21).

4.3.4. Data quality

Data quality was discussed at two levels: (1) the quality of administrative health data in general and (2) the quality of the linked dataset prepared for individual studies (see Table 10). In each province, stakeholders acknowledged that administrative health data are prone to data quality issues (e.g., subject to human error, may be incorrect, incomplete, inconsistent, and require cleaning and formatting to be used for research). At PopData and HDNS, it was noted that any issues that were identified with the data could not be addressed because the data centres did not have the authority to make changes to the data, and there were no mechanisms in place to provide feedback to the original data providers so that these issues could be addressed at the source. As a result, data issues are carried over into future studies, where they may or may not be identified

and addressed by a new research team. Across cases, concerns were also expressed about whether the linked datasets created for individual studies were prepared in accordance with the specifications set out by the research team.

Related to this, several participants described occasions where they received a linked dataset from the assigned data centre analyst and, because of their own familiarity with the data, recognized errors within the dataset. Addressing these errors had implications for project timeliness and costs, but perhaps more importantly, led some researchers to lose confidence in the data. At HDNS, steps had been taken in recent years to mitigate this issue, with the introduction of double coding (i.e., having two analysts code the data separately to ensure similar outcomes). At the time of this study, it was not apparent that this strategy, or other strategies to ensure the integrity of the data provided to the research team, had been implemented at PopData or ICES.

Table 10. Illustrative quotes regarding data quality.

Case	Illustrative quotes
NS	… at the end of the day, I'm left feeling uncertain about the quality of the dataset that's been pulled. So I really… Where I can, I avoid it. [P15] And actually, at one point I did have to go back and say, 'no, actually, that doesn't look right.' And it's only because I have datasets. Again, someone else might have just taken and ran with it. But because I have a lot of experience with what it should look like, I knew it wasn't right, and it wasn't. [P18]
BC	So you know, we had three people that, you know, had 1,000 visits in a year, right. That's clearly out of whack. But the question is, why are they out of whack? [P22] And when they do [get access to the data], it's so old, it's useless, or, you know, it's missing tons of stuff, or whatever. It's just… It's terrible. It's so much worse than people even express. [P30]

Case	Illustrative quotes
ON	I've done many, many of my own analyses. And not just analyses, but cleaning the data and linking it. And so I have a very hands-on feel of what that looks like and feels like. And it's painful. It is hard work. It is challenging. It is messy. It is prone to error. It takes many hours to hone it in. Even for someone who's expert, like who's skilled and who does it frequently, I will tell you that they make mistakes, and they have to do checks, and so on and so forth. So I mean that is the biggest issue, is you're dealing with messy data. [P44] So you know, [the analysts] would produce something, and you'd have to examine it really carefully, go back over it really carefully. And, you know, in many instances, it was the logic was flawed. [P35]

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

4.4. Factors affecting access to administrative health data for research purposes

A total of 32 inter-related factors spanning seven categories were identified as affecting access to administrative health data for research purposes which are summarized in Table 11. These include: study-related, research stakeholder-related, regulatory stakeholder-related, relational, organizational, regulatory, and contextual factors. Although categories were common across cases and factors were largely consistent, there were differences in how each factor affected access to data and the magnitude of impact.

In the following sections (4.4.1-4.4.7.), factors will be compared across provinces in terms of their impact on access to data. Illustrative quotes will be provided to facilitate comparison. The factors affecting access for each individual case are provided separately in appendices (NS, Appendix J; BC, Appendix K; ON, Appendix L).

Table 11. Factors affecting access to administrative health data for research, by category.

Category	Factors
Study-related	▪ **Requested data:** Whether the required data are available/have been captured; which specific variables are being requested and for what reason; the sensitivity (i.e., potential for harm if disclosed), level of identifiability, and quality and completeness of these variables; and whether external linkages are required. ▪ **Study design and methods:** The specific study objectives/aims, proposed methods for data collection and linkage, safeguards in place to mitigate risks, and planned analyses.
Researcher-related factors	▪ **Researcher affiliation:** Whether or not the researcher's affiliation meets eligibility criteria for accessing data. ▪ **Researcher knowledge and expertise:** The extent of the researcher's knowledge of the local context (i.e., local data holdings, data access pathways, and regulatory requirements) and their expertise relevant to working with administrative health data. ▪ **Researcher experience:** Whether and to what extent the researcher has prior experience accessing administrative health data for research. ▪ **Access to funding:** Whether or not researchers are able to access the funds required to obtain access to data.
Regulatory-stakeholder-related factors	▪ **Knowledge and expertise:** The extent to which regulatory stakeholders understand the methodological aspects of research involving administrative health data as well as the relevant regulatory requirements. ▪ **Individual perspectives on benefits:** The extent to which regulatory stakeholders view research involving administrative health data as important/beneficial (either generally or in terms of a specific study). ▪ **Individual perspectives on risks:** The extent to which regulatory stakeholders perceive risks associated with research involving administrative health data (either generally or in terms of a specific study).
Relational factors	▪ **Communication:** The extent to which relevant stakeholders (researchers and staff, regulatory stakeholders and entities, analysts) effectively communicate with each other about matters related to data access. ▪ **Relationships:** The extent to which relevant stakeholders (researchers and staff, regulatory stakeholders and entities, analysts) have pre-existing relationships, and the nature of those relationships.

Category	Factors
	▪ **Trust:** The extent to which relevant stakeholders and organizations are confident that other stakeholders and organizations engaged in research and data access governance will (1) collect, use, and disclose data in a manner that is consistent with ethical and legal requirements, (2) adhere to the terms of all relevant contracts and agreements, and (3) otherwise conduct themselves and their duties in way that is honest, transparent, and in the best interests of the individuals whom the data are about.
Organizational factors*	▪ **Organizational mandate/priorities:** Whether the provincial data centre, or other data provider, has a research mandate and the relative importance of research compared to other organizational priorities. ▪ **Data centre funding model:** Whether the provincial data centre receives operations funding from the provincial government and the extent to which cost-recovery is employed. ▪ **Organizational capacity to support research:** Whether and to what extent the provincial data centre, or other data provider from which data are being requested, has capacity in the following areas: analytic capacity (analysts to extract, prepare, and/or analyse data), regulatory capacity (established policies and processes and dedicated human resources), technical capacity (IT systems), data holdings, supports for researchers (training and resources), and supports for regulatory stakeholders (training and resources). ▪ **Organizational culture:** Organizational (data centre or other data provider) beliefs, values, and attitudes relevant to research, data sharing, and privacy protection, and the overall approach to data access governance.
Regulatory factors	▪ **Information legislation:** The content and clarity of relevant statutes with regard to the collection, use, and disclosure of health data for research. ▪ **Transparency of data access pathway:** The extent to which there is openness regarding how access to data may be obtained for research purposes, including: (1) where data are held, (2) who to contact regarding access, and (3) required steps to obtain access. ▪ **Complexity of data access pathway:** The number and sequence of steps required to obtain access to data. ▪ **Required forms and documents:** The volume and complexity of application forms and supporting documentation that researchers must prepare and submit in order to gain access to data.

Category	Factors
	▪ **Required reviews and approvals:** The number of reviews that must be performed and approvals that must be obtained for any given study. ▪ **Scope of review:** The scope of review performed (e.g., scientific, privacy, or ethical) and the extent to which there is overlap between multiple reviews. ▪ **Transparency of review:** The extent to which there is openness regarding (1) who or what entity is reviewing their data access application, (2) the purpose of the review, (3) the assessment criteria being used, (4) whether the application has met these criteria, and if not (5) what issues must be addressed in order to meet these criteria. ▪ **Application of data minimization principle:** How the data minimization principle is operationalized. ▪ **Role of data centre analysts:** The extent to which the data centre analyst is part of the research team and/or involved in the methodological aspects of the study. ▪ **Proportionality:** Whether processes vary depending on the perceived level of risk posed by the study. ▪ **Accountability:** The entities to whom the provincial data centres are accountable and the accountability mechanisms in place.
Contextual factors	▪ **Leadership:** The extent to which leaders at the provincial level have demonstrated support for research involving linked administrative health data, in terms of priority setting and strategic planning, and allocating resources and supports. ▪ **Health system organization and integration:** The health care institutions, programs, and services that exist in each province and the extent to which these function as a single system regarding data access governance. ▪ **Legislative landscape:** The various pieces of information legislation in force in each province and how these interact. ▪ **Historical events:** The impact of past events (e.g., breaches) on specific policies, processes, or the overall approach to regulating access to administrative health data for research. ▪ **Current events:** The impact of current events (e.g., COVID-19) on specific policies, processes, and/or the overall approach to regulating access to administrative health data for research.

*Organization refers to any organization that is engaged or "activated" during the data access process for a study (e.g., data provider organizations/institutions, including provincial data centres, external data providers, and academic institutions in which REBs are embedded.)

4.4.1. Study-related factors

Requested data

Across cases, the requested data impacted access in several ways. First, whether the data were captured, the timeframe for which data were available, and the quality and completeness of the data determined study feasibility. This was typically determined early on, prior to the submission of a formal data access request. Studies were sometimes abandoned at this stage—though the frequency at which this occurred was not clear— or were adapted to use alternative data sources.

Second, assuming the data were available and were considered of sufficient quality and completeness to meet the needs of the study, the specific variables that were being requested (e.g., sensitivity, identifiability), and the justification for their inclusion were considered by regulatory stakeholders in terms of ethical acceptability and privacy risks. Often, a substantial amount of toing and froing occurred between the researcher team and regulatory stakeholders in order to reach an agreement on which variables would ultimately be included in the research dataset. This contributed to longer data access timelines as well as costs given the additional staff time required to correspond with the relevant regulatory stakeholders. An exception to this was the review carried out by the ICES Privacy and Legal Office for internal studies—the specific variables being requested were not a focus of review as research teams were not limited to requesting a minimum dataset.

Third, the requested data ultimately determined the data access process. Depending on where the requested data were held (i.e., by the provincial data centre or

an external data provider), a single study could require linkages to multiple external datasets, and as a result, require review and approvals from multiple regulatory entities. Where multiple external linkages were required, data access timelines increased substantially, given the greater total number of total steps involved and potential for delays. External linkages also resulted in higher costs to the research team due to the increased research staff time required to navigate and coordinate the additional steps involved, fees charged by external data providers to access their data (occasionally), and increased data centre fees. Data centre fees increased due to the additional analyst time required to "bring in" external data (PopData, ICES, and HDNS), the development and administration of additional contracts and agreements with external data providers (PopData and ICES) and coordinating application review by external data stewards (PopData).

Study design and methodology

Across cases, study design and methodology impacted the amount of time required by the assigned data centre analyst to prepare the dataset and conduct any requested analyses. For example, more analyst time was likely to be required for studies that involved a more complex study design (e.g., quasi-experimental designs requiring the identification of case and control groups) and for those requiring more complex statistical analyses (e.g., inferential versus descriptive statistics, logistic regression versus machine learning). Analyst time was one of the main drivers of overall costs across cases, however, the extent to which the data centre analyst was involved in analysis varied. For internal ICES studies, the assigned ICES analyst was responsible for

dataset preparation and completing *all* required analyses and providing aggregate results to the research team. In contrast, HDNS analysts were responsible for dataset preparation and provided additional analytic services only upon request (at an additional cost the researcher), while researchers accessing data via PopData were expected to perform their own analyses. Thus, analyst time was more of a cost-driver at ICES compared to PopData and HDNS.

Study design and methodology were also considered during review processes, specifically, the methodological appropriateness of the study, privacy risks associated with the study, and the safeguards in place to mitigate potential risks. Depending on the specifics of the study, as well as the knowledge and perspectives of the individuals involved in reviewing the data access applications, numerous rounds of review and feedback were sometimes required for the research team to satisfactorily address the various issues identified during review, which had implications for timeliness and the costs associated with accessing data.

Table 12. Illustrative quotes regarding study-related factors.

Factor	Illustrative quotes
Requested data	The other thing is you are limited by the fact that HDNS doesn't have everything. So you can only… If you're going through HDNS, I mean it makes sense, right, if you're going shopping at a store, you can only buy what's in the store. [P11, NS] And also sensitivity of the information they request. If the researcher asks for some very sensitive information, let's say if they want the full birth date and the six character postal code for all individuals in BC, but they are not able to provide the reason why they need that for their research. And if the data steward does not see why they need it, likely their request will be rejected or will be…like the researcher may likely be asked to modify the request to ask for less. [P20, BC]

Factor	Illustrative quotes
	So there are definitely things that don't make it through because we don't have the data. [P41, ON]
Study design and methodology	So sometimes once the project charter is signed, it's really quick. Because if it's not a complicated project, it could be like a week and then they'll get their data. But sometimes a lot needs to be worked out in the charter, and then the pull is more complicated. So then it could take a few months. So it's hard to give an estimate from once it's approved to once the data is delivered because it could be a couple of months, it could be six months. [P2, NS] Typically, from when I submit a data access request, like actually get involved in the PopData BC process, to start that approval process, from request through to the data stewardship review, to accessing the data, is I think about six months. It can be shorter if it's an uncomplicated request. It gets lengthier if it's more complicated in terms of the number of databases, the number of data stewards, the number of variables, if there's an external linkage with primary data. Anything like that will complicate… Complicate might not be the right word. But it's just more review, more time involved. [P23, BC] But, yeah, it takes…it does take time. And when I meet with investigators who are coming into ICES, the first thing I tell them when they say they're going to bring in project specific data is this will take time. So if you don't have time, this isn't your path. [P43, ON]

4.4.2. Research stakeholder-related factors

Affiliation

Across cases, only individuals who met specific eligibility criteria were permitted to access data. Eligibility criteria varied substantially across cases and were primarily related to the researcher's organizational affiliations. Access to data via PopData was open to all researchers across Canada with an academic affiliation, although they required peer-reviewed funding. In contrast, access to data via HDNS was limited to researchers with a Dalhousie University affiliation. ICES' approach was unique in that researchers could access data via one of two pathways depending on their affiliation:

researchers with an ICES affiliation and ICES students could access data via ICES' internal data access pathway, while third-party researchers (i.e., academic researchers without an ICES affiliation and private sector researchers) were required to access data via DAS. Notably, compared to researchers accessing data via the internal data access pathway, third-party researchers accessing data had access to fewer datasets as well as access to less granular data. Key informants indicated that researchers who did not have an ICES affiliation sometimes sought out an ICES researcher to collaborate with as this enabled them to get a higher level of access to the data and made data access process easier to navigate.

Knowledge and expertise

Across cases, researchers' ability to gain access to data was affected by their level of knowledge and expertise relevant to administrative health data (e.g., strengths and limitations, content of specific datasets, nuances of the data), local data holdings and data access pathways, and the regulatory requirements relevant to accessing administrative health data for research. This impacted researchers' ability to determine which data to request, to put together a high-quality application, and to initiate and navigate the data access process. Overall, those with more knowledge and expertise had improved access to data. However, given limited training opportunities specific to administrative health data, particularly within academic institutions, knowledge and expertise were primarily gained through first-hand experience.

Experience

Across cases, individual's access to linked administrative health data tended to improve over time as they gained more direct experience accessing data. Rather than spending time and resources identifying potential data sources and how to access them, experienced researchers (and/or their staff) were able to initiate the required data access approval processes in a timelier manner and were better prepared to navigate these processes (e.g., they knew what to submit, to whom, and when). Through experience, researchers became skilled at determining which data were required to address their study objectives and preparing high quality applications that clearly communicated the particulars of their research and met the expectations of the relevant regulatory/oversight bodies. Moreover, experienced researchers had an improved ability to access supports because they were more aware of the supports that were available and/or had established personal connections and were able to seek assistance from others when needed.

For individuals who were new to research involving linked administrative health data and to accessing data, such as trainees and early career investigators, mentorship was critical. For trainees working with administrative health data, students typically worked under the supervision of a more senior researcher with experience specific to administrative health data. Master's and Doctoral students in ON who wanted to use ICES data for their thesis research were required to work under the supervision of an ICES-affiliated researcher. For early career researchers who sought an ICES affiliation, specific requirements had to be met, including the completion of research under the

supervision of an existing ICES-affiliated researcher. Similar mentorship programs for early career researchers did not exist at PopData or HDNS.

Access to funding

Given the substantial costs associated with obtaining access to administrative health data for research, access to funding was required to access data. For researchers seeking access to data via PopData, peer-reviewed funding was a requirement to access data, so the inability to obtain such funding would prevent the study from moving forward, unless proxy review could be arranged, which was uncommon. In contrast, researchers seeking access to data via ICES and HDNS were not limited to peer-reviewed funding and could therefore seek funding from a broader range of sources (e.g., government sponsorship, departmental funds, etc.). Regardless of the source of funding, in each case, studies could only proceed if sufficient funds were available. The reliance on funding meant certain groups of researchers were placed at a disadvantage. Across cases, key informants identified early career researchers, individuals with "one-off" research questions (specifically, clinician researchers), and trainees as groups of researchers from whom limited access to funding was more likely to be a barrier to accessing data. Early career researchers were acknowledged as having lower funding success rates compared to their more senior counterparts, while individuals with a one-off research question (as opposed to an established program of research involving administrative health data) were thought to have limited time, expertise, and/or support to pursue funding. Key-informants also recognized that even where discounts were available for eligible trainees (offered by HDNS, PopData, and ICES-DAS), only

those with access to grant funding through their supervisor were able to obtain access to data through the provincial data centres.

Table 13. Illustrative quotes regarding research stakeholder-related factors.

Factor	Illustrative quotes
Affiliation	So, if someone's not an ICES scientist, they need to work with an ICES scientist to gain access. And so, I've always tried to take the approach of, you know, I'll be the ICES scientist, and you can do your project, and that's fine. But you know, I may have minimal involvement aside from being the… You know, there's some stuff that I have to do. But I tried… I personally tried to facilitate that sort of thing. So, it does just, you know, for the unaffiliated scientist, it means reaching out to somebody and, you know, forming some sort of a loose collaboration there. [P40, ON]
Knowledge and expertise	The fact that people don't understand that the best use of these data is for health services research and health service delivery evaluation is lost on people. They want to do clinical studies or population health studies. And the data are okay for that, but they're ideally suited for health… So there's a misunderstanding there. Second, people misunderstand how administrative data are used, what's required, the complexity involved in using them correctly to get real answers we can stand behind. And so there's a whole education piece that has to happen in the researcher community. [P3, NS]
Experience	So, it does I think create a bit of an unequitable kind of environment where if you're in the know, you get access to data. But if you're a newer researcher or if you don't have the connections or the history, it will take you quite a while to figure out what's possible. [P6, NS] So, it's definitely one of the things where it's a complicated enough process that it's very difficult to go through it for the first time. And that once you've done it and it becomes familiar, it just becomes… Like there are still hoops that you jump through, but you know what all of those hoops look like, and you know which ones are on fire and which ones are not. [P26, NS] I mean there's no substitute for hands-on experience going through the DAR process and then working with the data to understand, you know, what's actually required and what the data can do, basically. I really think there's no substitute for that. [P27, BC]
Access to funding	The others it excludes are people who have fundable projects and might even have money in a bank account but without the peer review,

Factor	Illustrative quotes
	meaning unless you've received funding from CIHR, you don't have the ability to go forward. [P19, BC] Really, funding was the main thing. But I've been quite successful at grants. And that's the main barrier, right, is… And then you just have to build a good budget justification to be able to do it, and don't under-fund. And if you don't, if you're not successful in grants then there's no access. [P33, ON]

4.4.3. Regulatory stakeholder-related factors

Knowledge and expertise

Individuals involved in the regulation and oversight of access to administrative health data for research came from a wide range of backgrounds and had different areas and levels of knowledge and expertise relevant to administrative health data research and governance. Thus, individuals in regulatory roles, particularly those who were new to their role, did not always have a strong grasp on data access requirements and processes which sometimes led to incorrect information being communicated to researchers. In addition, this led to a lack of consistency in how applications were reviewed within and between regulatory entities, particularly in NS and BC.

Where applications were reviewed by an individual (e.g., a data steward in BC) rather than a group (e.g., full REB review, or review by the HDNS DAC), the results of the review were much more dependent on the individual's knowledge and expertise. For example, when review was performed by individuals who were not well versed in the methodological aspects of research using administrative health data, extensive toing and froing with researchers often occurred as a result, leading to longer data access timelines and higher overall costs. Where review and approval by multiple regulatory

entities were required, the potential impacts on timeliness and costs were even more pronounced as researchers were challenged to address a broader range of comments from multiple entities, which were sometimes contradictory.

At ICES, heterogeneity among regulatory stakeholders was less of an issue since Privacy Impact Assessment (PIA) forms were vetted by a small group of highly trained legal and privacy personnel through ICES' Privacy and Legal Office. Nonetheless, in ON, as in BC and NS, variability was still encountered given that REB members and individuals in various regulatory roles across external data provider organizations had varying levels of knowledge and expertise relevant to administrative health data research and governance.

Perspectives on the benefits and risks of administrative health data research

As with knowledge and expertise, regulatory stakeholders had varying perspectives on the risks and benefits of administrative health data research (certainly, the former influenced the latter) which impacted how they assessed relevant applications. Across cases, the regulatory stakeholders who were interviewed were highly supportive of research involving administrative health data, as long as the required safeguards were in place, and all regulatory requirements were met. Privacy risks were acknowledged—with re-identification being of particular concern—but were considered low given the policies, processes, and technical safeguards in place. Other concerns identified by regulatory stakeholders were related to the misuse and misinterpretation of data and conduct of "spurious" research.

Table 14. Illustrative quotes regarding regulatory stakeholder-related factors.

Factor	Illustrative quotes
Knowledge and expertise	You know, [the REB application] wouldn't go to the whole committee, it would go out to a reviewer and the co-chair, for example. And if you got a reviewer who wasn't terribly conversant with data access then it could really be, you know, a challenge to navigate because you would have to educate them basically on basic data access. [P15, NS] And in BC, usually for each organization, they appoint one person as the data steward. And each person may have a different background, right. Some data stewards come in from more like a legal background. Some data stewards, they have a more like research background, or they are a data expert. And some data stewards, they are having some management background or administration background, or privacy background. So, you can see that they all see… Like even their reviewing the request from the researcher or distinct project, they may look at the request from different perspectives and make different conclusions, or ask specific questions, or identify different issues. Which is very common. [P20, BC] …the sort of manager person that you're dealing with, they often didn't understand the bureaucratic steps. So, you know, you would be told that you have to do X, Y or Z in terms of the administrative steps. And then, lo and behold, you know, some months later, all of a sudden there's another administrative step that you had to go through. [P35, ON]
Perspectives on the benefits and risks of administrative health data research	But I think the bigger risk is when we start to actually think about the ethics of data use and around using data to effect public policy, to make decisions about people's healthcare, to take away programs. I think that to me is a really big risk. Because we're making assumptions. How we design our algorithms, how we interpret information, I think it's really, really important that those people that we are doing the data work about have a say in how we're using the data. Along those lines, I think that's where our biggest risks are. [P8, NS] So there are obviously privacy risks. I would underline that I do not believe that those only relate to single individuals and the risk of your personal information being exposed in one way or another. I think they really relate as well to groups, the potential for group identification, the potential for stigma, and the potential just for not

	even necessarily privacy exposure but uses of data that don't conform with your personal values and expectations. In terms of actual privacy risk, honestly I think it's very, very low in the way that we work currently. [P19, BC] I think one of the risks is that researchers assume it's easy to work with administrative data. And I think there then the risk is that people produce spurious findings… Because you can, as you know, you can generate any number of statistics. But whether or not they're rigorous and meaningful is really important. [P46, ON]

4.4.4. Organizational factors

Mandate

Provincial data centres were established specifically to facilitate access to administrative health data for research, and as such, had dedicated resources to building capacity in key areas to support this mandate. Studies involving external data linkages were often requested from organizations/entities where research was not the main priority, such as health care organizations or entities (e.g., provincial programs, clinical services, disease registries). Where organizational mandates prioritized health care delivery over research, and where fiscal resources were limited, there was often limited capacity to support research and high opportunity costs for doing so. These opportunity costs were sometimes a deterrent to supporting research. Where organizations or entities agreed to provide data, limited resources (e.g., staff time to review applications, extract and clean the data, correspond with the research team) sometimes contributed to delays in dataset acquisition, increasing the overall project timeline and costs.

Data centre funding model

The data centre funding model directly impacted the resources available within the organization, which subsequently affected the centre's capacity to support research. PopData and ICES received funding from their respective provincial governments, a variety of funding partners, and through cost-recovery (i.e., research fees). In contrast, HDNS lacked stable funding and relied entirely on cost-recovery for many years, which limited its ability to meet the demands of the local research community. At the time of this study, HDNS continued to operate without financial support from the provincial government, although additional funding had been secured through participation in national research initiatives, improving HDNS' operational capacity.

Capacity

Despite existing specifically to facilitate access to data for research, provincial data centres have faced various challenges related to capacity that have had implications for access to data. For example, analytic capacity was identified as an issue at all three data centres, which sometimes resulted in projects sitting in queue for extended periods of times until an analyst was free to start working on the request. At PopData, staff have been challenged to provide one-on-one support to a growing number of researchers seeking to access data, while simultaneously coordinating review and managing the feedback provided by various data stewards. At ICES, regulatory capacity was identified as a challenge given the increasing number of data access requests each year and the limited number of staff within the organization responsible for performing the required reviews, alongside the reporting requirements set out by various data providers and the

IPC. At HDNS, capacity was highlighted as a particularly challenging and long-standing issue, due to a reliance on cost-recovery as the primary source of operational funding for many years. Without stable funding, retaining staff was a challenge, particularly skilled analysts, which led to a number of challenges, including incorrectly prepared datasets and increased data access timelines. Importantly, these issues were largely addressed in recent years with a shift to a partial cost-recovery model as additional funding was obtained (i.e., through the CIHR SPOR initiative and CNODES). This enabled the hiring of additional analysts and allowed for the implementation of "double coding" to minimize the likelihood that errors would be made during dataset preparation.

Organizational culture

Although ICES and HDNS were both characterized as having conservative approaches to data access (i.e., described as risk-adverse, having a low-risk tolerance, cautious, etc.), all three data centres were regarded as supportive of research, which was consistent with their organizational mandates. Key informants across all three cases described variations across external data provider organizations in terms of willingness to share data and/or support for research to provide access to data for research. Of the limited number of regulatory stakeholders from external data provider organizations that participated in this study, it was clear that this "willingness" and "support" was often tied to mandate, resources, and organizational capacity (i.e., limited capacity impacted willingness).

Table 15. Illustrative quotes regarding organizational factors.

Factor	Illustrative quotes
Mandate	I can't speak to other teams but on my team like we do do our best to meet researcher requests in a timely manner. But that would be generally prioritized below our kind of day-to-day work. [P6, NS] But, you know, it's this person's…it's off the side of someone's desk, right. It's not… It's usually an executive, just like Ministry of Health, where research is not their prime focus…They want to know the protocol, you know, whether it's been funded, approved by research ethics, and there's security in place. And basically, they tend to, you know, once that's given to them, they read it all through, they may have a question, but they usually…they then sign off. But there are delays there because it's not their primary activity. [P25, BC]
Data centre funding model	First there was no source of permanent funding. So employees were like moving through a revolving door. There was no ability to train people and keep them in place. As soon as they were trained, they were leaving. [P3, NS]
Capacity	But I think it's just the lack of resource. It's not… I think we're highly under-resourced when you think about the amount of information that goes through the organization, the amount of patients that come through, and patient visits, the amount of employees, to have one director and four privacy officers have to manage all of the breaches, the questions, the PIAs, the audits, data access requests, we simply do not have enough people. We don't have enough resources to manage it. [P21, NS] But some data providers may need some help on that because they may not have a fully developed data access policy within the organization…But the Ministry of Health, of course they probably have a legal team there to help them. But for some smaller organizations, they do not have any formal process in place. So, it's a bit of a learning curve for them. Like when they see a request coming in, they may not even know how to deal with it. [P20, BC] PopData does have some really helpful resources though, and three training videos on their website. I think, you know, in comparison to other things I've seen, they've done a really good job of catering towards an audience that doesn't necessarily have experience in this area. That's not to say that it's easy, but there are things that you can access for free to teach you how to go through the DAR process and understanding what even linked administrative data means, and of those kinds of things. [P24, BC]

Factor	Illustrative quotes
	I think still the area that needs to be addressed is on, um, just like the staff side I guess. Like, making sure that we have enough people to support the work, and as work increases, that we're able to increase the number of staff required to support the project. So, in looking over the past decade, the projects at ICES, now we're close to over a thousand. We started maybe a decade ago with two to three hundred. It's increased but we've not necessarily be able to increase the staff to support the projects, and they're just getting more complex. [P38, ON]
Organizational culture	So, there's nothing in our regulatory structure that hinders what it is that we're doing. It just requires certain activities and, I don't know, assurances, if you will. That's very different from the policies and processes that actually get implemented by the health sector and others, which are sometimes fine and sometimes I would say tend toward the, "No, you can't do that unless you prove to me why you should," versus, "Yes, we want to support you to do that but let's make sure that it meets the criteria." And that may sound like a subtle distinction but it actually has a lot of bearing on how things work. [P19, BC] It's really not about the interpretation. It truly isn't. It's about using privacy, using ethics, using these things to retain power and control. And it's a really effective thing to say to people that they're being privacy non-compliant or they're being unethical. That's a very powerful thing to say to people. And you get your way. And it's bullying. [P30, BC] I guess conservative is the word that comes to mind. So you know, conservative, risk averse, etc. Which, you know, is probably appropriate to a certain extent. But it's… You know, it could be… It needs to be a little more can do in terms of how can we help you do things? [P40, ON] But an administrative audit, you know, examine all of the rules and the processes, and determine like what here really needs to be done versus this obsessive, you know, cover your butt attitude that people have. [ON, P35]

4.4.5. Relational factors

Communication

Across cases, the importance of clear and effective communication between the researcher and relevant stakeholders was considered important for several reasons. First, communication played a role in establishing credibility with data providers and relevant review bodies. In the submitted application documents, as well as initial conversations with relevant stakeholders, researchers were expected to clearly demonstrate that they had carefully thought through the data they needed and how the data would be used, that they understood the regulatory requirements and planned to adhere, and that they were capable of undertaking the planned study. Second, clear and effective communication ensured the review process proceeded smoothly. The more clearly the researcher had articulated the details of their study and particulars of their request, the less toing and froing occurred between the researcher and relevant review bodies. Simply put, higher quality applications raised fewer questions and required fewer clarifications, and typically obtained approval more quickly than those of poorer quality. Similarly, where review bodies clearly communicated with researchers regarding issues with the application, these could be more readily addressed by the research team. Third, clear and effective communication between the researcher and data centre analysts, as well as analysts at other data provider organizations, was necessary to ensure that the research team and analysts had a common understanding of the data being requested and the scope of work to be done. Finally, it was also important that organizations communicated to researchers any ongoing changes to data access policies

and procedures. Where this did not occur, researchers were unsure how to proceed, or proceeded incorrectly, which contributed to delays in accessing data.

Relationships

Researchers that were well-connected in the local research community and had established positive working relationships with other researchers, data providers, analysts, regulatory stakeholders, and review bodies had improved access to data. These relationships tended to develop over time as researchers got to know others who worked with administrative health data, and as they made repeated data access requests. Where these relationships existed, researchers were better able to access formal and informal supports throughout the data access process—they knew who to contact when issues arose, and felt comfortable asking for help, or inquiring about the status of their applications, etc. In contrast, researchers who were not as well connected had a more difficult time navigating the data access process. For this reason, mentorship for trainees and early career researchers was considered essential across all cases. Moreover, where researchers and regulatory stakeholders had established a positive working relationship, regulatory stakeholders sometimes had an improved understanding of the research, a higher level of trust for the researcher, and an increased willingness to provide data.

Relationships between other stakeholders in each province also had implications for data access. For example, in NS, there have been instances where relationships between personnel at HDNS and NS Health have facilitated joint data access review for studies involving a linkage of NS Health data to HDNS data holdings. Where relationships

have been established between PopData and external data providers, standing data agreements have been reached (i.e., rather than creating a new agreement on project-by-project basis). Finally, ICES ability to hold external datasets was, in some cases, perceived to be dependent on the relationship between ICES and the data provider.

Trust

Across cases, key informants acknowledged that the regulatory model in place in each province was dependent on trust. In NS and BC, the disclosure of data to researchers ultimately required trust that the researcher would comply with the relevant policies, conduct their research in accordance with the approved study protocol, and adhere to the terms of the various data sharing agreements and contracts that had been negotiated. Trust was typically established where researchers demonstrated that they possessed the appropriate knowledge and expertise to conduct the research and to mitigate any potential risks that might arise (privacy-related or other) and enhanced where a "track record" of regulatory compliance had been established, and positive relationships had been established with data providers and/or regulatory stakeholders. Nonetheless, several experienced and well-connected research stakeholders in NS reported that, despite their knowledge and expertise, track records, and interest in safeguarding data, they were treated by regulatory entities as untrustworthy and subjected to unreasonable levels of scrutiny when seeking access to data.

At ICES, the trustworthiness of researchers did not factor as prominently into decision-making around data access, which was likely attributable to two things. First, researchers did not typically receive access to individual, line-level data—only aggregate

data were disclosed to ICES-affiliated researchers (with rare exception), and third-party researchers received risk-reduced datasets via ICES-DAS. Second, all ICES-affiliated researchers were required to meet specific criteria, including training under an ICES mentor, likely reducing concerns around trustworthiness.

In additional to the trustworthiness of researchers, the extent to which provincial data centres were trusted also had implications for data access. Key informants noted that these entities have been entrusted by various data providers and custodians with the responsibility of facilitating access to data for research while simultaneously safeguarding these data to ensure that individual privacy is protected. Maintaining this trust was considered critical to the continued operation of the data centres through the ongoing receipt of data from various data providers (at HDNS, PopData, and ICES), as well as the continued receipt of funding from provincial governments and other funding partners (at PopData and ICES). Similarly, trust was described as essential for external data linkages. According to key informants, knowing that the data centres had rigorous processes in place governing data access, and that the data would be stored and accessed on the data centre's secure data environment, played a key role in data providers' decisions to allow access to data for research. The need to maintain this trust impacted the data centres' overall regulatory approach, with legislative compliance and adherence to the terms and conditions of the various data sharing agreements that had been negotiated, being considered paramount.

Table 16. Illustrative quotes regarding relational factors.

Factor	Illustrative quotes
Communication	Because in the many projects we've submitted to HDNS, we've found more and more that the more flushed out your ideas are, the more you've thought about it obviously ahead of time, the easier it is to articulate it and the less back and forth you're going to have to go through. [P9, NS] And I have learned how to say things. Because you may be just writing something that you're thinking about, and that can create flags. And it's amazing how sometimes new researchers are just saying, well, we may do this, we may do that. And all of a sudden it sounds like fishing. Well, you can't fish in the data…So even just the way they write something. And that's not really what they meant, but I understand where they were going from, it flagged something to the data steward, going no, they just want a free for all here. [P31, BC] I guess another minor factor is the difference in language used by three people – policy-makers or stakeholders, you can call them, scientists and analysts. Those three people, although they may all be doing the same work, like working on the same project, speak entirely different languages, right… So for a scientist, you're thinking about exposures and outcomes and confounders, right. For stakeholders, you're thinking about so what? What is the policy? What is the implication? What is the cost? What is the big picture? How am I going to implement this if you do find something? And for the analysts, it's like where do I find this data? Like where… Exactly how do you want me to code that again? Because like that's not doable, or that it only goes up to this year but you want it up to this year. Like they really want to know the nitty-gritty technical… [P44, ON]
Relationships	On the flip side, you know, I think sometimes we form really good relationships with some of these analysts. Right. And they know exactly what we want, and they know the cohort we want, they know exactly the outcomes we want. And it's the same cohort, the same outcomes, you know, slightly different exposures or slightly different outcomes, like different cohorts, you know, but there's a whole structure there. Right. And when that happens, things can get done very quickly. [P37, ON] Look, I mean it's not for the faint of heart. I have had a really good experience in that my like interpersonal relationships with PopData

Factor	Illustrative quotes
	are really strong. And I've always felt that, at least with the folks that I've engaged with, that they've been really helpful, really responsive. I've been really fortunate that I've primarily worked with kind of the lead. And I'm not sure why that happened. It could just be that that's sort of the lead of data access at PopData. And so my experience has been positive in that I've felt supported to have questions answered, that like response times have been good via email. [P24, BC]
Trust	But it just seems to be a difference in trust and understanding that, you know, we're really out to do good work, that we have no sort of personal motive to violate anybody's privacy… [P15, NS] [the REBs] attitude is that I am generally trying to do something untoward, and their job is to protect the public from me. And it's not a collaborative attitude, it's not positive in any sense of the word… But more to the point, it creates more inefficiencies. [P3, NS] when [researchers] want to go use PopData, I'm like, "Approved." I don't look at anything. I know their processes. They're pre-approved in my mind. [P30, BC] And what I want to highlight, in the data access world, working with academic researchers, there's an element of trust as well. Which is very important. Because once we send the data to the researchers, even if we do it through the [secure research environment], we have no control of what they do with the data. [P20, BC] We need these policies and procedures in place in order to make sure that we maintain a trust level with the clients, with Ontario clients, with Ontarians who have an OHIP number. That, you know, every effort is made all the time. And with our data custodians. You know, data is not just personal health information. The ICES repository is now over 100 data holdings, including personal health information, socioeconomic data, etc. And we're growing it. And that only comes with trust that the policies and procedures we have a place, we're compliant to, and that they are appropriate and necessary. [P43, ON]

4.4.6. Regulatory factors

Legislation

Provincial information legislation was primarily a facilitator to accessing administrative health data for research. In addition to enabling the use and disclosure of data for research, health information legislation in NS and ON facilitated data access in other ways. For example, the "chapter" of provisions dedicated to research in NS' *Personal Health Information Act* [149] facilitated access to data by signalling the importance of research, and by providing guidance for researchers and regulatory stakeholders in terms of the conditions for data disclosure, particularly with regard to required documentation. In ON, the designation of ICES as a prescribed entity under the *Personal Health Information Protection Act* [147, 212] allowed the organization to collect, use, and disclose personal health information for research without individual consent, and without researchers establishing impracticability. Although there was no health sector specific legislation in place in BC, the public sector information legislation (the *Freedom of Information and Protection of Privacy Act* [142]) permitted public bodies to disclose personal information to PopData for research purposes.

Despite facilitating access to administrative health data, access to data was sometimes hindered as a result of how the legislation was interpreted and applied. Vague or ambiguous legislation sometimes led to uncertainty among regulatory stakeholders, resulting in both broad and narrow interpretations and applications of the legislation. Rather than risk a breach or do something that may reflect poorly on themselves or their institutions, regulatory stakeholders often erred on the side of

caution which challenged researchers to meet stringent requirements. Moreover, ambiguity contributed to variations in the interpretation and application of legislative provisions across regulatory stakeholders and organizations. For example, in BC, the *Freedom of Information and Protection of Privacy Act* [142] addressed research quite generally and contained few research-specific provisions. As a result, the various organizations in the province developed their own policies and processes for data access based on their interpretation of the legislation. Researchers were therefore challenged to meet the requirements of multiple data stewards with unique requirements, leading to multiple rounds of review and feedback between researchers and data stewards (or other relevant stakeholders) in an attempt to address extensive and sometimes contradictory feedback.

Similar issues were identified in NS and ON. In NS, the lack of clear criteria for establishing the "impracticability" of consent in the *Personal Health Information Act* [142] resulted in disagreement between regulatory stakeholders about whether the criteria for impracticability were met. In Ontario, the lack of a clear distinction between "research" and "health system planning and management" in the *Personal Health Information and Protection of Privacy Act* [147] had led to instances where research was carried out under the banner of health system monitoring and evaluation, which has less oversight. To avoid such scenarios, at least one site (ICES Queen's) implemented a policy requiring all projects to undergo REB review, demonstrating variations in interpretation and application of the legislation within ICES itself.

Table 17. Illustrative quotes regarding regulatory factors (information legislation).

Case	Illustrative quotes
NS	So everybody has different rules. Everybody has different procedures. Some of it is legal. Some of it is that people have different ideas about it. Some of it is that they have different interpretations about how detailed they need to be with respect to what the laws are. They interpret them differently. People interpret PHIA differently. And the tendency is always to err on the side of caution. [P7] I don't think the legislation itself creates any barriers. It's how the legislation is interpreted in various different organizations. I think that that's changed over time as well. I think there were more barriers maybe when it was first implemented and organizations were trying to figure out, you know, where they were going to sit on that sort of conservative to less conservative continuum. So, yeah, so I think that it may be… I think that NS organizations probably took a fairly conservative route to its implementation. I don't think that's... I'm not going to put a judgment there per se, but I think it does… I think that conservative interpretation does make accessing data maybe a little more complicated. [P13]
BC	Our legislation is fabulous. It's very permissive, it's very smart when it comes to research. But because we have all of these health authorities, they're all distinct legal entities so they all interpret that their own way. There is no consistency at all. And there's tons and tons of research ethics board. There's no consistency there. So poor PopData BC is trying to meet the needs of this Medusa head, and none of them agree, none of them interpret things the same. [P30]
ON	So the legislation is cited, and then, you know, you have lawyers who can look at the same legislation and say it permits this, and others who say, well, you can't do that with this. So it's actually probably more about the interpretation, and which interpretation organizations tend to go with. Which is probably the more conservative interpretations of the legislation. [P40] I think the other piece is people's interpretation of legislation. So that is where a lot of the rules are in place of an interpretation of the legislation, not actually the intention of the legislation. And this is where different people will interpret it differently. And therefore, you kind…you almost need policies that will cover anybody's interpretation. Which is why it gets broader, it gets more convoluted. [P41]

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

Application of data minimization principle

The minimum dataset requirement was applied differently in each case. When accessing data via PopData and HDNS, researchers were only permitted access to the minimum dataset required to address the study objectives. In their data access applications, research teams were required to specify the exact variables requested from individual datasets and to provide a justification for their inclusion in the study. Often, what ensued was a process of extensive toing and froing with the relevant regulatory entities (e.g., HDNS DAC and external data custodians in NS, and data stewards in BC) about which variables were absolutely required, and at HDNS, whether the variables could be categorized or "rolled up" in some way to minimize the identifiability of the data provided to the research team. In the end, researchers accessing data via PopData received a dataset consisting of variables that had undergone minimal processing by the assigned analyst. In contrast, researchers accessing data via HDNS received a dataset consisting of unprocessed variables, as well as variables provided in the least identifiable format possible (e.g., categorical variables and derived variables calculated using a definition provided by the research team).

Key informants highlighted several issues with this approach. First, it required the research team to determine which variables to include in their request *a priori,* incorrectly assuming that the team has a sufficiently in-depth understanding of the data within each dataset, and the nuances of the individual variables (i.e., in terms of content, format, quality, and completeness) to determine those that should be included in the data access request. Even for very experienced researchers, the inability to see or

work with the data in advance, combined with a lack of sufficiently detailed data documentation (e.g., metadata and data dictionaries), made it difficult to ensure that the requested data were suitable for addressing the study objectives. For researchers accessing data via HDNS, an additional layer of difficulty was added—they were required to anticipate how variables should be categorized for analysis before they were able to work with the data.

Second, this approach frequently created the need for researchers to seek amendments to their data access requests. After receiving access to the research dataset and having the opportunity to work with the data, it often became evident that key variables were missing, the quality and completeness of certain variables were not as expected, and/or that variables did not capture the information that they were expected to capture. When this occurred, the research team was required to submit an amendment to request changes to the research dataset, increasing both the time and costs associated with gaining access to the final research dataset, particularly where a change to the research dataset triggered amendments to multiple regulatory entities.

Third, this approach did not lend itself to exploratory research, sometimes conflated with “fishing”, and did not consider new analytic techniques, such as machine learning, where the minimum dataset cannot be specified in advance. Moreover, without the ability to work with entire datasets to gain a more in-depth understanding of the relationships between, and relative strengths and limitations of, specific variables, important analyses were sometimes neglected.

A very different approach was employed at ICES for researchers accessing data via the internal data access pathway. At ICES, the assigned analyst was responsible for completing all required analyses (exceptions for student projects) and was therefore the only individual with access to line-level data. The research team was responsible for developing a document called a Dataset Creation Plan (DCP), which outlined the study objectives, timeframe, analytic variables, and planned analyses. Based on the DCP, the analyst prepared the dataset (e.g., cleaning, coding, and linking the data), performed the required analyses, and provided the PI and approved team members with aggregate data, usually in the form of results tables. This process was iterative in nature, with ongoing communication between the PI and assigned analyst and changes (i.e., to variable definitions, analyses, etc.) made as needed. The DCP, which was considered a "living" document, was updated continually to reflect these changes, with amendments required only in limited circumstances. This approach offered a number of benefits: researchers were not required to request a minimum dataset; with access to entire datasets, ICES analysts were able ensure the most appropriate variables were included in the dataset and in the analyses; and there was flexibility to make adjustments to the dataset and analyses in real time, allowing studies to proceed in a more streamlined and timely manner. A downfall of this approach was that by having the ICES analyst involved to such an extent, overall study costs were driven by the costs of analyst time to a greater extent compared to HDNS and PopData.

Table 18. Illustrative quotes regarding regulatory factors (minimum dataset requirement).

Case	Illustrative quotes
NS	I mean the only thing I'd like to see is just a little loosening in the legislation that allows people not to do like variable-based hypothesis where you have to tie exact variables to their hypothesis. That they don't give them enough leeway to say oh, can we just look at this one too, you know, because we probably picked the wrong one first. [P1] I understand that there's a responsibility and that the onus is on the researcher. And I take that very seriously to think things through. But you're doing that when you're blind to how the data are actually collected. Half the time, I don't know how many missing fields there are. You know, I can see that there's a variable in the documentation. I don't know the range of the values, I don't know how many are missing, I don't know… You know, just because it says it's there doesn't mean it's actually there…I can see what kind of things are meant to be captured in the documentation once that's shared with me, but I don't know what's actually there in the dataset. So, I can't tell you in a meaningful way how to collapse them because I don't know because I can't see it. [P15]
BC	We often spend a lot of time like labouring over exactly how to define that and what data you should include because you know how difficult it's going to be if you want to change it down the road. [P24] You have to sift through all of the datasets that they have, look at every individual variable and think through in your head like is this a piece of information that I think I might need? Why do I think I might need it? And can I rationalize that in the context of the application? And I think that can be quite limiting because, you know, sometimes as you're learning, especially as a student when you're relatively new to this process, as you're learning, you analyze something in your data that raises other issues or other questions, and then you realize that you don't have access to the variable that you need because you didn't ask for it in your original data set. And then there's a process that you have to go through to ask to get that piece of information later. And that's also a lengthy process involving…you know, involving waste. [P26]
ON	If you think about things like machine learning methods, um, you know principles like data minimization, which are better on principles of privacy, don't actually work well for machine learning techniques, which requires masses of data and where machine learning researchers prefer to have, you know, the entire dataset and they can work with it rather than needing to pre-specify which variables they are going to study. Um, there's a bit of a disconnect between some of the regulations and rules around data can be

Case	Illustrative quotes
	held, data can be used for research, and actual modern, some modern, health research techniques and methodologies. [P32] Yeah, I mean I think that's a bit redundant. Because if you have access to a database, you have access to that database. There's no sense in giving you part of it or only some of the fields of it. That's not going to really improve privacy, at least in my opinion. Yeah, most of the fields don't get used. But sometimes you don't know how you're going to use which field. Sometimes it's not perfect. Sometimes you think you're going to use one field, and then you look at your data and you're like, oh, that's really funny, that discharge date, you know, that there's… Like there's some discharge dates that are length of stay is zero. That doesn't make sense. Let's look at this or that factor. Or let's measure it differently. Let's look at this variable, right. And so there is a bit of like a search or a hunt for the best way to represent your data. [P44]

Abbreviations: NS=Nova Scotia, BC=British Columbia, ON=Ontario

Transparency of the data access pathway

Across cases, data access pathways were frequently described as being unclear. Researchers were responsible for determining which data were required for the study, who was responsible for these data, and how to obtain access. This information was not always readily available, which made it difficult for some researchers to know where to begin, especially those who were relatively new to research involving administrative health data.

At HDNS and PopData, the data access pathway and required processes for accessing internal data holdings were provided online, although some key informants felt that the information was not sufficiently detailed. Where information was lacking, the Data Navigator at HDNS, and DAU staff at PopData were available to assist researchers in determining how to proceed. In ON, the processes for accessing data via

the internal data access pathway were not communicated to researchers, although the DAS pathway was available online.

The greater challenge for researchers, especially those accessing data via HDNS or PopData, was determining which data were held by external data providers, and how these could be accessed as this information was not typically publicly available. Adding to the overall lack of clarity across cases, was the fact that data access policies and processes were constantly changing, but these changes were not always communicated to the research community, causing uncertainty amongst researchers of all experience levels about how to proceed.

In NS, key informants also described a lack of clarity regarding the data access pathway for existing datasets (i.e., datasets created for a previous study that were being stored on the HDNS platform). Given the time and costs associated with the creation of a new analytic dataset, researchers sometimes sought access to an existing dataset to address a new research question or objective; however, the circumstances under which access to an existing dataset required a new application versus an amendment were unclear.

Complexity of the data access pathway

Data access pathways were characterized as complex in all three cases, especially where linkages to external datasets were involved. At HDNS and PopData, accessing internal data holdings was described as more complex and taking longer to navigate than expected, while several key-informants at ICES reported that processes within the organization have become more complex and "convoluted" over time. Nevertheless,

across all cases, it was clear that the overall process of accessing data was much more complex when access to externally held data was required.

Compared to studies involving only data held by the provincial data centre, studies with one or more external linkages required a number of additional steps, including: (1) preparation and submission of additional applications and documents to relevant regulatory entities (e.g., external data stewards/custodians, additional REBs) by the research team, (2) review and approval of submitted applications by relevant regulatory entities, (3) development and administration of various contracts and agreements by relevant stakeholders, (4) preparation and provision of data files by the external data providers, and (5) data linkage by the assigned data centre analyst. Where multiple linkages were required, researchers (and/or research staff) often navigated multiple data access processes simultaneously. The time needed to complete all necessary steps was often extensive and required substantial human resources, subsequently leading to longer data access timelines (i.e., to receipt of the final linked dataset) as well as higher overall costs.

Required forms and documents

Data access applications varied across cases, with researchers required to submit a variety of forms and supporting documents. Nonetheless, across cases, the process of preparing these forms and documents was described as time consuming, tedious, and involving a heavy administrative workload due to the complexity and volume of paperwork involved. The required forms were regarded as requiring a high level of knowledge and skills to complete, making them particularly challenging for

inexperienced researchers. The time required to prepare these documents had implications for data access timelines (i.e., an extended preparatory phase), as well as costs given that research teams frequently hired skilled staff specifically to prepare application forms and documents and navigate the data access process.

At HDNS and PopData, researchers were required to identify every variable being requested as well as a detailed justification for each. At ICES (internal pathway), this was not required because ICES analysts were able to access complete datasets. At ICES and HDNS, the researchers were also required to develop detailed analysis plans to guide the work of the data centre analyst (i.e., a DCP and "Project Charter", respectively). At PopData, where researchers were provided unprocessed data, such documentation was not required. The administrative workload associated with completing forms and documents was exacerbated for ICES students who were also required to apply to become an ICES student (requiring additional documentation), and for researchers embarking on studies involving external data linkages. Researchers in NS in particular discussed the impact of external linkages on a project's overall administrative workload. This may have been related to the frequency of studies involving external linkage (HDNS data holdings contain 9 databases, so external linkages are commonplace) as well as the fact that researchers in NS must facilitate external linkage themselves. This is unlike ON where the Privacy and Legal Office develops and negotiates agreements with external data providers.

Number and scope of required reviews

Studies involving access to administrative health data via the provincial data centres were typically subject to multiple reviews, though the specific reviews varied by province and study (i.e., depending on the required data). At PopData, all studies were required to undergo peer-review (peer-review funding is a condition of access), REB review, and review by the relevant internal and external data stewards. At HDNS, studies often underwent peer-review (if peer-reviewed funding was obtained), and always required DAC review, REB review, and additional reviews where external linkages were involved. At ICES, studies often underwent peer-review (if peer-reviewed funding was obtained) and REB review (exceptions for some studies), and always required privacy review. In each case, depending on the data involved in the study (e.g., if there was a linkage to hospital data), multiple REB approvals were sometimes required.

Despite each review having a distinct purpose—scientific review (funding agencies), ethical review (REBs), and privacy review (DAC and data stewards)—in practice, there was often overlap in scope. For example, in NS, both REB review and DAC review incorporated scientific review, resulting in instances where researchers were asked to make changes to the study protocol. Similarly, in BC, data stewards often provided feedback on methods and analyses. In both provinces, research stakeholders considered this inappropriate, particularly for studies that had received peer-reviewed national funding (e.g., CIHR). At ICES, there was less overlap, largely due to the fact that the review performed at ICES (for studies involving the internal data access pathway) was typically well-defined in scope. Reviews performed by the Privacy and Legal Office

consisted primarily of a legislative analysis to confirm the legislative authority of ICES, and any other relevant data custodians, to collect and disclose personal health information.

The need for multiple reviews was characterized as problematic for two reasons. First, multiple reviews often resulted in variations in review, which created challenges for researchers who were required to address a wide range of comments and concerns that were sometimes contradictory in nature—a time consuming and resource intensive process. Related to this, variations in review sometimes contributed to delays in data access timeliness as a result of the back-and-forth required to reach a common understanding/agreement between various data providers. For example, in BC and NS, changes requested by a single data steward sometimes resulted in amendments to the REB application(s), and to applications submitted to other external data providers. Moreover, the overlapping scope of reviews was redundant, contributing to inefficiencies in the overall process and pointing to the potential for improved streamlining.

Transparency of review

Transparency of review was not identified as a factor affecting access to administrative health data for research at ICES but was important for research at PopData and HDNS. In BC, the lack of transparency related to data steward review was considered a barrier to accessing data. Staff at PopData's DAU coordinated the application review by relevant data stewards, who were responsible for performing review and making decisions about whether and under what conditions access to data was permitted. A

benefit of this approach was the researcher could correspond with the DAU who collated feedback from multiple data stewards, rather than communicating with each data steward individually. However, this approach hindered transparency by preventing communication between the data stewards and researchers. Without a direct line of communication to the data stewards, and in many cases without knowing the identity of the data steward, it was often difficult for researchers to gain a clear understanding of how their applications were being assessed, what the expectations of the data stewards were, and how these could be met. Moreover, data steward expectations were not always evident to PopData staff, who noted the lack of a harmonized approach to review across the various data stewards in BC. As a result, researchers often engaged in extensive toing and froing with data stewards, through the DAU, in an attempt to address data steward concerns. The time involved in this process was compounded by the competing demands on the time of individual DAU staff members, who were responsible for managing multiple projects at different stages of the data access process.

At HDNS, transparency of review was considered a facilitator of gaining access to data. Researchers who submitted applications to HDNS were invited to attend the DAC committee meeting to address the concerns of the committee in person. Following the meeting, the DAC provided the researcher with a list of requested revisions and clarifications, specifying changes to be made to the submitted application. The opportunity to speak directly with committee members and address their concerns was described as improving communication between the researcher and the DAC, with the

DAC gaining an improved understanding of the proposed study and why specified variables were necessary. Similarly, by speaking directly with the DAC, and through the written request for revisions and clarifications, researchers gained an improved understanding of the concerns of the DAC, and their expectations in terms of how these should be addressed. Together, this minimized the number of rounds of written correspondence between the researcher and the DAC prior to approval being granted. Over time, this also contributed to the development of a rapport between the DAC and the researcher, which enhanced communication even further.

Role of data centre analysts

The role of data centre analysts varied drastically across provinces. At HDNS and PopData, the analysts were not embedded in the researcher team, working under a service provider model. Analysts at both centres were responsible for data linkage and dataset preparation, however, there were important differences in terms of exactly what was involved in the latter. At HDNS, the analyst was involved in linkage, data cleaning, cohort identification, calculating derived variables, and "rolling up" variables to minimize the potential risk of re-identification. For example, HDNS commonly provided categories of diagnosis codes to researchers rather than individual diagnosis codes, and commonly calculated indices rather than provide the raw data to the research team so that they could run the analysis. As a result, the dataset provided to researchers was a mix of raw and processed variables, which could then be analyzed by the research team. At PopData, the analyst performed minimal processing and the researcher received a dataset comprised of nearly all unprocessed variables.

The approach at PopData was deemed beneficial for two reasons. First, it eliminated the need for the research team to develop a detailed dataset creation and analysis plan. At HDNS, the researcher and assigned analyst co-developed a document called a "Project Charter" which set out the scope of work to be undertaken by the analyst as well as the deliverables to be provided to the researcher. The development of this document was a time-consuming process, and once this document was executed, any changes to the researcher's request required an amendment. Second, by providing researchers with the raw data, researchers were able to do the work of defining variables and cohorts themselves, which often required multiple iterations to figure out. Thus, the potential for datasets to be prepared incorrectly, and for additional time to be spent identifying and addressing errors, was reduced. The approach employed at HDNS was such that researchers were unable to confirm whether the dataset had been prepared correctly, and if errors were discovered, there were sometimes substantial delays and costs associated with addressing these.

At ICES, specifically for studies conducted via the internal data access pathway, researchers worked with an assigned ICES analyst rather than hire an external analyst. The ICES analyst was considered part of the research team, and the only individual on the team to have line-level access to data. Challenges with this approach included the extensive time and skills required to develop a DCP to guide the work of the analyst, the inability of the researcher to confirm that the work was being done in accordance with their request, and inability to control costs associated with analysis. Working with an embedded ICES analyst also had advantages. Specifically, the extensive data and

methodological expertise provided by ICES analysts, combined with a level of security clearance that permitted them to access complete datasets, help ensure the work progressed without interruption (i.e., by minimizing the need for amendments) and that researchers were provided results in a timely manner.

Proportionality

All studies involving access to data via HDNS and PopData required REB review; however, those that involved only the secondary use of data were considered minimal risk and underwent expedited review rather than full REB review. As such, most studies obtained REB approval quickly. For studies involving access to ICES data, those deemed minimal risk also underwent expedited review. Uniquely, Sunnybrook Hospital waived the requirement for REB approval for studies involving only ICES' general use data, and instead opted to regularly audit a sample of ICES studies to ensure ethical compliance. For researchers at Sunnybrook Hospital, this mitigated potential delays encountered as a result of REB approval, thereby facilitating timely access to data. Notably, neither of the data centres employed a proportionate approach to privacy review at the time of this study—that is, there were not distinct review processes for studies that had been determined to pose different levels of risk.

Accountability

The provincial data centres were accountable to various entities, including advisory boards and other oversight bodies, and the data provider organizations with whom data sharing agreements were reached. Accountability was identified as an important factor

impacting the overall regulatory approach employed by ICES. Since ICES was permitted to collect data without consent, it was subject to a high level of oversight from the IPC and required adherence to the specific policies and procedures set out in the *Manual for the Review and Approval of Prescribed Organizations* [216] (recently updated to the *Manual for the Review and Approval of Prescribed Organizations* [213]). Substantial resources within the organization were dedicated to complying with these requirements, and to meeting the Office's reporting requirements (which were quite extensive as evidenced by ICES' 2020 report [217]). While this oversight was considered integral to maintaining transparency and public trust, it meant that ICES operated within a very specific regulatory framework and had limited ability to change existing policies and processes.

Table 19. Illustrative quotes regarding regulatory factors (other).

Factor	Illustrative quotes
Transparency of the data access pathway	… like the data access committee, they invite you to their meeting if they're reviewing your project. …Like if there's stuff that wasn't clear, you can explain it to them right then and there. They can explain why it wasn't clear so you can fix that, and all that stuff. [P9, NS] I have heard from many researchers saying when they first come to PopData, they do not know what the next steps are or when they put in a request, they do not know what's next, or like when they will hear back, or what to expect. [P20, BC]

Factor	Illustrative quotes
Complexity of the data access pathway	It just seems to me that if I know I want to end up in the backyard, why can't I just walk around the house and get to the backyard? Why do I have to go through the front door, walk down, walk through the living room, family room, kitchen, and take all the twists and turns to end up at the patio door and get through that door to end up in the backyard? And that's the situation right now. I know where it is and I can see a clear path to it. I also clearly see the path they want me to take but it's not efficient and it's not effective and it's costly. [P17, BC] But I have known ICES since its inception when it was kind of a rough and ready upstart, you know, a great idea, innovative organization, until now where it is absolutely mired into bureaucracy. Absolutely mired in bureaucracy. [P35, ON]
Required forms and documents	I've seen DHW forms. And NSHA was trying to reinvent their forms.... HDNS revised their forms. How big is this province? And we all live here. You know, we all work within minutes. [P5, NS] And their form is really long…I found one a little while ago. It was about 20 pages long. So you've got to check off all the data you want from whatever databases you want. You have to pick specific field from the whole thing. So you've got to know what it is you want, right. I don't know how people can do this without having some experience with it. [P22, BC] Filling out the forms to get access to the data doesn't take very long. It takes like a few hours. What takes more time is in order to actually access the data, you have to have something called the data creation plan. And that specifies a lot of exact details about what the analyst is going to do. That's basically an instruction guide to the analyst. Like the analyst rarely will ever understand what the research question is or why they need the different data or how it has to be set up. So the person that needs access has to set that up. And that's a skill, right. Like that's not easy to do. [P33, ON]
Number and scope of required reviews	I've got to go through [external data provider] as well as through HDNS. In some cases, I may have to go through both IWK ethics and ethics with either Dal or NSHA. So yeah, you can end up going through multiple data access procedures with multiple volumes of forms… So everybody has different rules. Everybody has different procedures. [P7, NS]

Factor	Illustrative quotes
	I will say though, which I think is part of the flavour of your question, is I do think that there are overlaps. If you think about peer review, ethics review and data steward review as three components, there's clearly a Ven diagram there. And some places and units are better than others about decreasing the overlap in what people are looking at. [P19, BC]
Transparency of review	I think that once you're in the process, like once you make it to the data access committee stage, I think it's relatively… Because you get to go in, you get to talk to everybody, you have that discussion and then you get your clarification, etc. I think that's pretty transparent. [P11, NS] It hasn't always been clear what people needed to do in order to get access to data. It hasn't always been clear how data providers review those things and what they're looking for and what they aren't. [P19, BC]
Role of data centre analysts	HDNS is basically a data service centre. So the people reviewing the applications, the programmers during the work are not part of the research team. They're following instructions. And so they don't always really understand the research question. But in return, the researchers don't understand the data and the programming part. And in my view, this is a fundamental problem in the set up. [P7, NS] And then, you know, we end up with a relationship with the programmer analyst that will go throughout the project, including as we're trying to publish, to, you know, to redo or do new analysis if they're needed. Our programmer analyst will also be co-author for papers and review them for clarity as far as the statistical method that the data…and also give us a heads up if we're reporting any small cells, if we've missed any small cells for privacy issues - that sort of thing. They also provide support on keeping logs like the risk re-assessment log – that sort of stuff. So they work as sort of partners with us throughout the process. [ON, P34]

4.4.7. Contextual factors

Leadership

In each of the cases included in this study, leadership from the provincial government

(or lack thereof) had direct implications relevant to administrative health data research,

and access to data. In ON, the provincial government played a pivotal role in the advancement of administrative health data research, and in establishing ICES' role in supporting data-driven decision-making within the province. In addition to supporting the establishment of ICES in 1992, the provincial government (i.e., the Ministry of Health and Long-Term Care) helped ensure the sustainability of ICES' operations through the ongoing provision of both data and core funding since its inception. When the *Personal Health Information and Protection Act* [147] was enacted in 2004 and ICES was designated as a prescribed entity, ICES was granted unprecedented access to data and its ability to carry out its mandate became protected by law. These actions on the part of the provincial government helped create the conditions for a productive research environment and resulted in substantial organizational growth.

Conversely, in NS, the provincial government did not demonstrate the same level of support or leadership in advancing administrative health data research. For example, in addition to not providing funding to HDNS, the provincial government (i.e., the Department of Health and Wellness) declined to enter into a long-term data sharing agreement with the Department of Community Health and Epidemiology at Dalhousie University, which has contributed to a climate of uncertainty over the future of HDNS. In the broader research community, grassroot efforts to develop a more coordinated approach to data access in the province stalled, despite engagement from key research stakeholders and data provider organizations, which was attributed in part to a lack of leadership from the Department of Health and Wellness. A lack of regard for research was also evident during the health system re-structuring of 2016. No clear direction was

provided from the province or health authority about how the restructuring would impact access to data held by the former provincial programs or who would be responsible for authorizing access to data under the new organizational structure. As a result, policies and processes for accessing data held by some of the former provincial programs were still not established several years later, with researchers and regulatory stakeholders uncertain about how to obtain access to certain datasets for which there had previously been a clear data access pathway.

In BC, the provincial government (i.e., the Ministry of Health) was supportive of PopData's operations, as evidenced through the provision of data and funding; However the fragmented data landscape, the complex regulatory landscape that existed as a result of this fragmentation (i.e., a multitude of organizations with unique data access policies and processes) created challenges for health researchers across the province. Strong leadership from the provincial government was identified as integral to fostering a more collaborative approach across organizations with regard to data sharing.

Health system organization and integration

In all three provinces, the provincial data centre holds only a portion of the administrative health data collection in the province. Other administrative health data, and data that are commonly linked to administrative health data, are held in different data systems in various healthcare organizations and institutions throughout each province and fall under the authority of various data custodians. This fragmentation of data was identified as particularly problematic in NS and BC. In these provinces healthcare organizations and institutions were described as operating in a very siloed

manner, resulting in each individual organization having its own unique data access policies and processes.

Recently, several organizations within the PSHA in BC were merged into a single corporate entity, which improved access to data across organizations, further highlighting the direct relationship between health system organization and integration and access to data for research. At the time of data collection, similar efforts were underway in ON, with the merger of various organizations and agencies under a single government agency (Ontario Health), however, the implications for data access were unclear at the time.

Legislative landscape

In BC and ON, the legislative landscape created challenges with regard to data sharing for research purposes, but in different ways. BC was described as having a "patchwork" of legislation that applied to health information. Without broad reaching health sector information legislation, organizations throughout the provinces were subject to a variety of Acts. Determining which Act (or Acts) applied in the context of a given study was identified as a challenge for researchers. In ON, the interaction of various pieces of legislation sometimes created challenges for ICES, particularly when working across sectors. ICES' designation as a prescribed entity was specific to health information, creating uncertainty about its' ability to bring in data from other sectors. In NS, the legislative landscape was not identified as a barrier to data access.

Historical events

In BC, past events have had lasting impacts on relationships and research culture within the province. In 2012, an anonymous complaint was made to the Auditor General, which claimed that a data breach had occurred within the Ministry. The actions subsequently taken by the Ministry to address the situation—which included data access suspensions and employment terminations— came under fire and resulted in a lengthy and in-depth public investigation by the BC Ombudsperson's office. While this matter was discussed to a limited extent by key informants, it was acknowledged as a having a detrimental impact on the research community in BC. The Ombudsperson's report [218] found that "the government acted unfairly and unreasonably" (p.367) in their handling of the incident, resulting in direct harms to those who were accused of misconduct, including "fear, anxiety, loss of income and financial uncertainty, harm to reputation and careers, harm to relationships and, in some cases, health problems" (p.xiv). The investigation also found a loss of productivity and morale among Ministry staff, and harm to relationships with the research community in more broadly. A number of recommendations were put forth as a result of the investigation, many of which have been or are being implemented by the Ministry of Health with the aim of achieving a more balanced approach to data access.

Current events

The impact of current events on access to data was exemplified by changes to data access policies and processes that occurred because of the COVID-19 pandemic. Due to the timing of data collection for this study (i.e., most interviews in NS were carried out

in early 2020), this was particularly evident in BC and ON. In ON, ICES was heavily involved in the provision of COVID-19 data for the province (e.g., testing data, cases, hospitalizations, etc.). The need to report these data in a timely manner and to ensure that decision-makers within the healthcare system had access to information to support evidence-based decision making led to several changes in policies and processes. For example, there was an increase in the frequency of the data feeds being received by ICES, projects involving COVID-19 data were prioritized and a COVID-19 committee was established to facilitate timely review, and a new user case was established (i.e., non-ICES-affiliated agent) to expand access (e.g., to data to individuals with specific modelling expertise). Although most of these projects were carried out under the umbrella of health system planning and management rather than research, these changes highlighted the potential to improve data access without compromising privacy. In addition, the pandemic led to changes in ICES' policies regarding remote access. Whereas individuals working directly with line-level data were previously required to do so on-site, COVID-19 restrictions meant a shift to working off-site, requiring more individuals to have remote access. Similar changes occurred in British Columbia, including the prioritization of COVID-19 projects, streamlining of forms, and development and implementation of expedited processes for COVID-19 projects, further highlighting the potential to improve access to data for research, while working within existing regulatory frameworks.

Table 20. Illustrative quotes regarding contextual factors.

Case	Illustrative quotes
Leadership	I'd say we do not have…we have an utter lack of actual direction on how to approach this area. So every program area, every "custodian" or agent does it their own way. And I've been involved in many conversations bringing all these groups together, and it just becomes a circular conversation and nobody… There is no… It really would have to be political direction saying everybody come together, figure this out, and clear up the pathways. [P6, NS]
Health system organization and integration	There are challenges in the integration of DHW and NSHA – who's in charge of what, who's responsibility for what is? And then moving down to the practice level, you know, that's a whole other level as well. The divides in data are similar to the divides in the healthcare system in general, I think. [P11, NS] So if you want to go from, you know, Children's Hospital to BC Cancer, that's fine. You see, you can do that. But that's…you have the legal ability to do that. But because they were all set up as different legal entities, all of their data processes were different. And it was a dog's breakfast about how you did things. [P30, BC]
Legislative landscape	If we want to collect personal information, not personal health information, but personal information from ministries, we need to go to a different Act. So that is the Freedom of Information and Protection of Privacy Act, that's our FIPPA legislation. Um, but the problem is, we don't have a designation in FIPPA--there is nothing that says that the ministries can disclose to an institution like ICES for any analytics. So we have to rely on research, but when a lot of the work we do is actually not necessarily research, it's analytics to advise the government. So then we're sitting in a situation where the pieces of legislation don't talk to each other and they're in silos, and then they don't enable the kind of data sharing that I think would be optimal and optimized for researchers… So it's not the legislation in and of itself makes it a barrier, it's the fact that they don't speak to each other. [P38, ON]
Historical events	You know, our legacy and our past. It's important to understand how it impacts our attitudes. So there's a huge change management that needs to happen at the executive, at the doer level, and with the public. And that all takes time to change. [P36, BC] And, you know, there's the big incident in BC, I don't know, it's probably around 10 years ago now, which people point to on both sides of the issue. Of, you know, it's so important, it's so important

Case	Illustrative quotes
	because look what happens, you know, if there's even a hint of impropriety with it. But the other part of that was that it was a totally bogus crisis where people lost their lives and careers over it. So anyway…Yeah, that sent a chill through everything. And, you know, as I say, it was held up as a cautionary tale when, you know, there was nothing there. [P40, ON]
Current events	And as you know, societal influences have made a huge difference. And most notably, the recent one is COVID. So I don't know if you're aware, but on the weekend, some incidence data was leaked. And so the public wasn't happy that the government had data at the community level but hadn't published it. Right. So you see that the public is demanding higher transparency from government around the data that it has so that it can make informed decisions around what's happening say for example with COVID. You also saw the public demanding more around the long-term care things that were happening and reporting for schools. So there's also a changing attitude from the public around what it expects around data as well. [P36, BC] I used to think that they were too strong on the privacy side of it. With COVID, you may know already, that they've opened up access so people can work from home. That's really changed things. And I don't think… You know, once the genie is out of the box, I don't think she's going to be put back in again. And I really think that has been a wonderful development over the last year. Because, you know, the processes for keeping the data safe are sophisticated enough, I think. Like I'm not an expert on privacy, but it's my perception that they're sophisticated enough to protect the data and still allow people to work more flexibly than we used to. [P34, ON]

CHAPTER 5: DISCUSSION

5.1. Summary

This study set out to identify the factors affecting access to administrative health data for research with the dual aims of identifying the specific barriers affecting access to data and gaining insight into the factors contributing to reported inter-provincial variations in the timeliness of access. Using a qualitative multiple case study approach, access to administrative health data was examined in three Canadian provinces—NS, BC, and ON—with a focus on research where the point of access was the provincial data centre (i.e., HDNS, PopData, and ICES). Data collection included interviews with key-informants (n=46), triangulated with case documents.

Study findings highlight key challenges related to accessing administrative health data for research, including timeliness, costs, and data quality. A total of 32 multi-level factors spanning seven common categories, were identified as affecting access to data. From these, eight key barriers to optimal access to administrative health data for research in Canada were identified. These barriers were not evenly distributed across cases, shedding light on inter-provincial variations in the timeliness of data access.

5.2. Key findings

5.2.1. Suboptimal access to data

Across cases, interviews with key informants indicated that eligible researchers who submitted a formal application for data were typically successful in obtaining access; however, several caveats were identified indicating access was suboptimal. Specifically, data access timelines were unpredictable and often considered lengthy, especially

where linkages to external datasets were required. At the same time, the costs associated with accessing these data were considered high, which limited access among certain subgroups of researchers, and were also compounded by longer data access timelines. Finally, issues with data quality were identified in each province, which related to both the quality of administrative health data in general, and to the quality of study-specific datasets that were prepared. This further impacted data access timeliness and overall costs to both the research team and regulatory stakeholders as data issues were identified and addressed. Given that these issues were identified in provinces with well-established provincial data centres and legislation permitting access to administrative health data for research, and they were identified by individuals with established track records in accessing data, it is likely that these issues pose an even greater challenge in other jurisdictions, and for researchers with less experience.

Challenges related to the timeliness of data access were consistent with the literature in Canada [21-28] and elsewhere [219-221] in which timeliness issues have been referenced extensively. This is indicative of the academic research context, where deadlines are paramount; researchers are required to complete research within the required funding timeframes, ensure trainees meet program deadlines, and demonstrate productivity within their institutions. At the same time, research must be conducted in a timely manner to be optimally beneficial to knowledge users [222, 223], which became increasingly clear during the COVID-19 pandemic [224]. In the context of data access, timeliness is particularly important given that delays exacerbate high study costs. As highlighted in this study, paying staff to navigate unexpectedly lengthy data

access timelines can have the effect of diverting funds from other planned research activities.

Data quality also has implications for timeliness and overall study costs. The limitations of administrative health data, including inconsistent data quality, have been widely documented [45-49]. Additionally, human error may be introduced during dataset preparation. While additional time and resources may be required to address these issues once identified, a greater concern is that these issues will go unnoticed, compromising study validity and ultimately wasting the time and resources spent. This conflicts with the ethical obligation of the research community to ensure the responsible stewardship of research resources [225, 226]. Thus, initiatives to improve access to administrative health data should aim to address all three issues: timeliness, costs/resource implications, and data quality.

5.2.2. Key barriers and strategies for improvement

The findings of this study revealed a number of overarching barriers to optimal access to administrative health data for research in Canada including: researcher knowledge and expertise, regulatory stakeholder knowledge and expertise, data access processes, organizational culture, organizational capacity, legislation, trust, and leadership. Each of these is discussed below, along with potential mitigation strategies.

Researcher knowledge and expertise

This study found that due to inconsistent training, many researchers and research staff had limited knowledge and expertise relevant to administrative health data (e.g.,

nuances and limitations), the methodological aspects of administrative health data research, and the relevant regulatory requirements. This hindered researchers' ability to make informed methodological decisions during protocol development, prepare high quality data access applications, and navigate the required data access processes. This was less of an issue at ICES where researchers were required to be trained prior to accessing data, however, there was substantial variation among researchers in terms of the overall training they received. Issues related to limited training were compounded by a lack of supports for researchers throughout the data access process, particularly with respect to external data sources. The importance of having adequate knowledge and expertise to formulate requests, navigate data access processes, and work with the data have been acknowledged in the Canadian literature [21, 227] and internationally [219, 228].

To improve the overall capacity within the research community to obtain access to data in a timely and efficient manner, conduct high quality administrative health data research, and appropriately safeguard data, additional training and supports for researchers should be considered. Statistics Canada has identified a number of data literacy competencies relevant to accessing and working with data that are applicable to administrative health data (e.g., data awareness, ethics, exploration, analysis, management, etc.) [229]. To develop these competencies, researchers may avail of training opportunities available to the public, such as those offered by Statistics Canada [230] or other organizations such as PopData [231]. A more pro-active approach to improving capacity would be to increase course offerings specific to administrative

health data research at academic institutions, focusing on developing an improved understanding of the strengths and limitations of administrative health data, concerns and considerations related to privacy, and data access governance (i.e., legal and ethical frameworks). To further ensure that researchers accessing these data have the required knowledge and expertise, and to ensure key competencies are addressed, data provider organizations may consider imposing mandatory training requirements. For example, recommendations from a public engagement event in BC included the development of a mandatory data security certificate program for researchers [232]. Importantly, such requirements must be implemented in such a way that do not perpetuate barriers to access (e.g., should not be time or cost-prohibitive).

Researchers would also benefit from improved "just-in-time" supports, including improved informational resources with respect to available data sources and data access policies and processes within individual jurisdictions, as well as more detailed metadata, as recommended elsewhere [228]. Another option that may be considered is the implementation of "Data Navigators" within data centres and external data provider organizations, to assist researchers in identifying the data they require access to, preparing applications, and navigating the data access process. This role currently exists at HDNS and has been credited by stakeholders with substantially improving the quality of applications submitted to the DAC and the efficiency of the data access process as HDNS. Finally, a national best practices document, akin to the CIHR Best Practices document released in 2005 [129], dedicated specifically to the secondary use of health data, including linked administrative health data, may be beneficial. Given the evolution

of research in the nearly two decades since the CIHR Best Practices document was released, as well as changes to the current regulatory context, more up-to-date guidance is necessary.

Regulatory stakeholder knowledge and expertise

Variations in regulatory stakeholder knowledge and expertise were reported to create several issues in the context of data access. Regulatory stakeholders who lacked expertise in certain areas were sometimes reluctant to provide approval contributing to longer data access timelines. At the same time, variations in knowledge and expertise contributed to variations in review within and between regulatory entities, resulting in delays in access and creating an increased workload for researchers. These findings are consistent with previous research linking stakeholders' knowledge of regulatory requirements to their willingness to disclose PHI for research [158], as well as research highlighting variations in review across REBs [166]. It has been suggested that REBs do not have the required expertise to perform ethical review [161, 167, 169], and reported that they have knowledge gaps with regard to database research and matters related to privacy, confidentiality and security [166]; However the findings of this study did not indicate that there were issues with REBs in particular, but rather than there were variations in knowledge and expertise across all regulatory stakeholders.

Notably, the issue of variations in application processing due to a lack of standardized training among regulatory stakeholders (specifically, data stewards) was discussed at the aforementioned stakeholder engagement event in BC [232]. Recommendations to address this issue included the standardization and certification of

regulatory stakeholder training and procedures, although concerns were also expressed that certification requirements would increase delays and decrease efficiency. Even without certification, standardized training and widespread adoption of standardized assessment criteria would improve the consistency of reviews across regulatory stakeholders, organizations, and jurisdictions. For example, the "5 Safes" framework [233] was created to inform decision-making around appropriate data uses and may facilitate a more standardized approach among regulatory stakeholders. This framework has been implemented in the UK Data Service and Health Data Research UK [234] and in the time since data collection for this study occurred, has also been implemented at PopData [235]. Finally, the development of a national best practices document (as recommended in the previous section) may also improve the consistency of stakeholder decision-making and support the development and implementation of standardized review procedures.

Data Access Processes

Data access processes have been cited as a barrier to accessing health data for research in Canada [5, 22, 24, 26, 227], although specific process-related issues that hinder access to data have not been examined in detail. In this study, data access processes were closely examined and their impact on access to data discussed at length. As noted elsewhere, data access processes vary substantially within and across provinces [26, 236] (e.g., differences in application components, the number and scope of required reviews, whether there is a disclosure of individual-level data). This study revealed that the processes in place at HDNS, PopData, and ICES were fundamentally different, each

with their own unique challenges. In all three cases, there was also substantial variation in the processes employed by external data provider organizations.

With that said, the primary overarching issue relevant to data access processes in each case was that of inefficiency, which was especially pronounced for studies that required linkage to an external dataset. Data access processes were characterized by a high administrative burden as a result of extensive documentation requirements, multiple reviews that frequently overlapped in scope, opaque processes, variations in review across organizations, lack of proportionate privacy review, and extensive toing and froing with data providers and/or regulatory stakeholders. Contributing to these inefficiencies was the inability of researchers to interact with administrative health data when planning their studies and preparing their request, the reliance on manual processes in the absence of application trackers and dashboards (e.g., staff responding to researchers' requests for application updates), insufficient informational supports to assist researchers in navigating data access processes, and a lack of mechanisms to document and address data quality issues.

An assessment of data access processes in the UK highlighted many of these same issues [228]. Stakeholder recommendations to address these included the development of informational resources to assist researchers in navigating the data access process, implementation of an application tracker, implementing agreed upon target timelines, creating standardized processes for data providers, employing a proportionate approach to review, developing improved and sufficiently detailed metadata, improved communication with data providers, and additional training for

researchers. Additional strategies to improve process efficiency in the Canadian context may include the implementation of centralized review and standardized inter-organizational processes at the provincial level to facilitate external linkages. There are also opportunities to leverage what is "working well" within individual provinces, such as the implementation of provincial Data Navigators and inviting researchers to attend privacy review meetings in person—both of which have been successful in streamlining processes at HDNS in recent years.

Streamlining processes is integral to ensuring not only timely access to data, but the timely and cost-efficient conduct of research to benefit the public. Notably, successful efforts to improve the timeliness of data access were undertaken during the COVID-19 pandemic, demonstrating not only that processes can be streamlined while maintaining adequate privacy protections, but that such changes are indeed possible where there is both sufficient public interest and political will [237].

Organizational capacity

Data held by provincial data centres are increasingly linked to data held by external organizations. Consistent with reports from Canadian data custodians [21, 160], this study revealed that many external organizations have limited capacity to support research. Participants commonly identified issues with human resources, specifically a lack of time and dedicated staff to support data access processes (e.g., to maintain documentation, facilitate application intake and adjudication, perform data extractions, and prepare files for linkage, assist researchers, etc). Organizational capacity was also identified as a barrier in the international literature. In a review on the barriers to

sharing of public health data, it was found that organizations often lack the human and technical resources required for data preparation and related activities [159]. Similarly, a recent rapid review focused on the UK, reported issues related to funding and capacity as barriers to access, linkage, and use of administrative health data held by government organizations [238]. Specific issues included a lack of personnel with the appropriate knowledge and skills and IT infrastructure.

This study also revealed capacity issues with respect to the provincial data centres. A common issue across data centres was that analytic capacity was limited (i.e., an insufficient number of analysts compared to the volume of requests), sometimes resulting in lengthy data access timelines. Capacity in other areas varied across provincial data centres. For example, issues related to regulatory capacity were most pronounced at ICES due to a high and consistently increasingly volume of annual requests. At the same time, there were also substantial differences in overall capacity across provincial data centres, although the difference between ICES and HDNS was most notable because of HDNS' reliance entirely on cost-recovery for many years.

To ensure appropriate resources are available to meet the growing needs of the research community, and of health system and government decision-makers, the issue of capacity must be addressed. The most obvious strategy to improve capacity is the provision of additional funding at the organizational level to support research. Given the fragmented data landscape and the sheer number of institutions across which data are distributed, addressing organizational capacity is a complex undertaking, highlighting the importance of centralized data infrastructure. Certainly, core funding for provincial

data centres is needed to ensure appropriate infrastructure is in place and to hire and retain skilled staff [239]. In addition to increasing funding, an integral aspect of mitigating capacity issues is to ensure that existing resources are being used in the most efficient manner possible by addressing the process inefficiencies identified above and that all stakeholders have an adequate level of knowledge and expertise to perform their role.

Organizational culture

In this study, organizational culture was a barrier primarily with respect to external data provider organizations. Defining and articulating organizational culture is challenging [240, 241]. Broadly, organizational culture consists of the shared values, assumptions, and beliefs of individuals within an organization [242]. The findings of this study highlighted the fact that administrative health data are frequently linked with data from a range of organizations, each with a unique organizational culture. Stakeholders did not explicitly use the term "organizational culture", rather, described the "personality" of the organization [243]. While some organizations were described as very supportive of research, others were described as hesitant to provide data and occasionally unwilling to provide data at all. These findings were consistent with those of a recent literature review, which identified variations across organizations in terms of their willingness to share data, and attitudes toward data linkage and sharing, as barriers to data access and linkage [238].

Insights into organizational culture also come from "artefacts", which are observable manifestations of the underlying organizational assumptions and beliefs

[244]. In the context of linked administrative health data, examples of artefacts include policies and processes, resources, and IT infrastructure. As such, differences in these artefacts across organizations were also indicative of variations in culture. A lack of established policies and processes, funding, and/or IT infrastructure may therefore be indicative of a culture wherein research is not valued or not prioritized alongside other organizational commitments. Thus, organizational capacity is closely intertwined with organizational culture.

Organizational culture has been described as the most challenging aspect of an organization to change [240]. Bringing about change requires strategies that are targeted to individual organizations based on the specific aspects of culture that need to be improved, and the underlying assumptions and beliefs. These may be guided by broader, overarching principles for implementing and sustaining organizational change [245]. Nonetheless, one area where there has been substantial agreement in the literature is regarding the importance of leadership in shaping organizational culture [240, 245, 246], providing a focal point for intervention.

Legislation

Across cases, legislation was primarily considered a facilitator to accessing administrative health data for research; however, a lack of clear legislative provisions was found to contribute to variations in how the legislation was interpreted and applied, which subsequently hindered access to data. The lack of clear policies and guidelines, and resulting variations in interpretation and application have been acknowledged within the Canadian research community previously, particularly with regard to consent

requirements and what constitutes de-identified data [5, 156, 165, 166]. In addition to the practical challenges associated with these variations (e.g., meeting different requirements across multiple organizations), unclear laws and guidelines have been thought to create uncertainty among regulatory stakeholders, leading to a conservative approach to data access that exceeds legal requirements [156, 247]. Davies and Collins [247] describe the interpretation of legislation as being "driven less by careful consideration of the likelihood of real harm for individuals than by the desire to minimise the risk of criticism for organisations". This was supported by the current study wherein several regulatory stakeholders reported erring on the side of caution rather than risk a breach or error.

These findings reiterate the need for a national best practices guideline to reduce variability across organizations in terms of policies, processes, and regulatory stakeholder decision-making. The implementation of standard training and procedures for regulatory stakeholders have also been suggested as a potential strategy for reducing variability in how data access applications are processed across organizations [232]. Notably, issues related to variations in interpretation and application were most pronounced in BC, where health sector-specific information legislation had not been implemented, and where the existing public sector legislation made brief mention of research. Certainly, one strategy to mitigate this variation would be to enact health-sector specific legislation, which would require action on the part of the provincial government.

Trust

Trust is integral to the sharing and use of health data for research [159, 238, 248]. In an in-depth examination of the role of trust relevant to the use of and access to health data for research, Sexton et al [248] highlighted the need for a "balance of trust" between researchers, data providers, and the public. They described trust as follows:

> Trust is … akin to a form of faith. It results from reflexive deliberation that is always contingent on context. It supports the shared goodwill between parties in awareness of reciprocal duties and obligations. A precondition of trust is the absence of complete assurance, and it therefore relates to an assessment of risk. [248](p.311)

In the current study, trust was discussed in terms of regulatory stakeholders' trust of researchers, the trust of provincial data centres by external data providers, and the public's trust of the research community and those charged with safeguarding health information. Although a lack of patient and public trust has been identified as a barrier to data access elsewhere [238], the findings of this study suggested that trust between regulatory stakeholders and researchers was of greatest concern, particularly in NS and BC. This was evident in statements from regulatory stakeholders who expressed discomfort with disclosing data to less experienced researchers, concerns about the misuse and misrepresentation of data, as well as concerns about privacy violations (both unintentional and deliberate). This is consistent with the findings of van

Panhuis [159] who reported that where trust is lacking, data providers may anticipate potential misinterpretation, misuse, or abuse of the data. Trust of researchers was likely less of a concern in ON because ICES researchers had all undergone a minimum level of training, and because researchers did not typically receive individual-level data.

Related to the notion of a "balance of trust", when data providers or provincial data centres implement regulatory measures to increase their own trustworthiness, there is a risk of damaging the relationship with researchers [248]. This was evident in NS, where several researchers expressed that they perceived a mistrust of researchers by local regulatory stakeholders based on the level of scrutiny they were subject to when seeking access to data (i.e., extensive documentation, extremely detailed applications, and intense questioning by regulatory stakeholders). Such breakdowns in trust equate to a breakdown in the "lubricant" that keeps data access processes moving smoothly [249]. Researchers who feel unfairly scrutinized or perceived as untrustworthy may be deterred from working with administrative health data again in the future.

With regard to building and maintaining trust between researchers and data providers, or the regulatory stakeholders who act on their behalf, the solution is complex. While one-on-one relationships were found to facilitate trust in the current study, Sexton et al [248] suggest that "standardized and impersonal mechanisms" for data access lead to fairer access compared to relying on personal relationships. At the same time, they acknowledge that a reliance entirely on standardized and impersonal mechanisms can limit researchers' ability to access support. Given that trust relates to risk assessment, ensuring all researchers complete standardized training and possess

the knowledge and expertise required to access, analyze, interpret, and safeguard data (e.g., data literacy competencies [229]) may be an important first step toward achieving a balance of trust.

Leadership

The findings of the current study clearly pointed to the need for greater leadership from provincial governments, particularly in NS and BC. In NS, leadership relevant to administrative health data research was especially wanting, as evidenced by an absence of dedicated funding to HDNS, reluctance to enter into a long-term data sharing agreement with HDNS (and talks of terminating the existing data sharing agreement), and lack of follow-through on addressing key issues to improve access to data as identified by local stakeholders. In BC, PopData received funding from the provincial government, but at the same time, the strained relationship between the broader academic research community and the provincial government, the practical challenges of accessing data held across a myriad of organizations and health authorities, and the absence of health sector specific legislation had not been addressed. In all three provinces, including ON, variations in policies and process, capacity, and culture across external data provider organizations, would benefit from provincial leadership.

Given the link between leadership and culture [240, 245, 246], leadership from the provincial government is integral to establishing a strong research culture within the province, where research evidence is valued and incorporated into decision-making and where resources are allocated to supporting research activity [250]. On a more practical level, provincial leadership is especially important given the fragmented nature of the

current data landscape, whereby data are held by a multitude of organizations spanning various sectors and legal entities—only the provincial government has sufficient authority to amend legislation, develop and enforce regulations, improve coordination across data provider organizations, and allocate provincial resources.

Affecting change within the upper echelons of the provincial government is not a straightforward task, but it does not represent an insurmountable obstacle. Just as leaders have an impact on culture, "cultural norms define how a given nation or organizations will define leadership – who will get promoted, who will get the attention of followers" [240](p.7). Thus, bringing about change at the level of the provincial government can occur by bringing about a cultural shift within the broader public. Evidence indicates that when the public is appropriately informed, they are supportive of the use of their health data for research [232, 251]. By continuing to inform and engage the public in conversations about the benefits of the secondary use of health data for research, safeguards in place to ensure these data are protected, and the barriers to leveraging these data for the benefit for Canadians, a shift in the broader culture may occur that is subsequently reflected in the provincial government.

5.2.3. Variations in the timeliness of data access

Inter-provincial variations

One of the overarching aims of this study was to gain insight into the underlying causes of inter-provincial variations in timely access to data. As cross-jurisdictional research has become more common in Canada, with funders increasingly requesting collaborations across jurisdictions, the practical challenges of working across borders have been

highlighted [22, 27, 28, 236]. As a result, much of the discussion around improving access to administrative health data in Canada has focused on addressing inter-provincial variations in the timeliness of access to data in order to facilitate multi-province research [21, 227, 236, 252, 253]. When conducting multi-province studies, inter-provincial variations in access to data may be expected to some extent given differences in the statutes governing access to administrative health data for research, and the ways in which these statutes may be interpreted and applied within each province. Nonetheless, there is cause for concern when variations in timeliness occur because of an uneven distribution of barriers to access, especially when this results in consistently longer data access timelines in some jurisdictions compared to others.

Overall, there were fewer barriers to accessing administrative health data in ON compared to NS and BC, particularly when comparing studies involving only internal data holdings. This was a direct result of provincial leadership that recognized the value of administrative health data research. Unlike PopData and HDNS, which were established to facilitate access to administrative health data by academic researchers, ICES was established in ON at the behest of the provincial government specifically to leverage health data to facilitate health system improvement [87]. To this end, the provincial government demonstrated its support of ICES through the provision of core funding, and the designation of ICES as a prescribed entity under the provincial health information legislation. The availability of core funding, in addition to funds from research grants, combined with the ability to collect and use PHI without individual consent created the conditions for a highly productive research environment. This

allowed the organization to continually grow its operational capacity by expanding to additional sites, growing its membership, adding more datasets, and developing an additional data access pathway (i.e., DAS) which opened up access to a broader range of users. At the same time, the policies and processes in place at ICES were indicative of a strong research culture [250]. Processes were streamlined such that researchers were not required to specify a minimum dataset, some studies were exempt from obtaining REB approval, the scope of PIA review was narrow and well defined, and amendments were not required in order to change or add a single variable within an approved database. Additionally, ICES scientists were required to have demonstrated experience working with ICES data or similar data from another jurisdictions, and to work with a mentor, helping to ensure all researchers had the required knowledge and expertise to access and work with administrative health data. Similar requirements did not exist at HDNS or PopData.

Importantly, provincial data centres do not operate in isolation but are part of a larger, interconnected, and dynamic system including academic institutions, external data providers, and provincial governments. As such, the ability of provincial data centres to provide timely access to data and meet the needs of the research community more broadly, is impacted by many factors outside of their control, especially for studies requiring linkage to external datasets. As such, when comparing studies involving external data linkages, the differences between provinces become less pronounced.

With that said, it is important to note that studies that involve an external linkage in NS or BC, may not require an external linkage in ON given ICES' extensive

internal data holdings. As an example, in a hypothetical study of breast cancer care in ON and NS requiring breast screening data, cancer registry data, and physician billing data, would require two external linkages in NS (i.e., the linkage of physician billing data held by HDNS to breast screening data held by the IWK Health Centre and to cancer registry data held by NS Health). In contrast, no external linkages would be required in ON since these data are all part of ICES internal data holdings (based on the ICES data dictionary [254]). Given the time required to prepare the additional application documents, obtain the required approvals, and perform data extractions and linkage, and the increased likelihood of unexpected delays, the overall data access timeline would likely be much longer in NS than in ON. If external linkages were required in both provinces, this may not be the case.

Intra-provincial variations

Although the emphasis has been on inter-provincial variations in data access, this study found that substantial intra-provincial variations also occur. Stakeholder interviews revealed that data access timelines varied substantially within each case. Researchers in each case reported a wide range of data access timelines with some individuals reporting that they obtained access to data in a relatively short timeframe (e.g., 2-3 months), while others reported protracted timelines (e.g., one or more years). Certainly, some of this variation was related to differences in the knowledge and expertise of the researchers themselves, which impacted the ability of individuals to prepare high quality applications and to navigate the data access process. However, individual researchers who had carried out multiple studies involving administrative

health data and had a range of experiences to draw from, reported substantial variations in timeliness between studies. These variations were reflective of the multitude of factors that affected access to data and could vary between studies—including study design and methods (e.g., study complexity), the required reviews and approvals, knowledge and expertise of individual regulatory stakeholders, capacity and culture of relevant data provider organizations, as well as events happening in the broader context over which researcher and regulatory stakeholders had little to no control (e.g., health system reorganization, a worldwide pandemic)—causing the data access process to unfold differently for each individual study for reasons that could not always be anticipated and accounted for by the research team.

As a result of this unpredictability, even experienced researchers with well-designed studies often encountered unanticipated barriers when seeking access to data, resulting in delays in access, unanticipated costs, and in some cases, the receipt of datasets that did not meet the needs of the research team. More broadly, the inability to anticipate barriers and the impact on access hinders researchers' ability to plan appropriately (e.g., to estimate overall timelines or total costs based on their past experiences or the experiences of others), further impacting their ability carry out the planned research within the required timeframe and budget. In other words, while the potential for delays is problematic from a research perspective, the unpredictability inherent in data access processes also presents a challenge.

5.3. Implications

Despite a growing interest in the use of administrative health data for research, and concern among various stakeholder groups regarding the accessibility of these data across the country, empirical evidence has been lacking. Comprised largely of commentary and anecdotal accounts from individual research teams, and expert opinion, the existing literature suggests suboptimal access to administrative health data; however, to borrow a phrase from van Panhuis [159], these experiences and expertise "have not been translated into scientific evidence". Thus, this study provides an empirical basis to inform both discourse and action on this topic.

Specifically, the findings of this study may be used to inform ongoing data initiatives, including the Strategy for Patient Oriented Research Canadian Data Platform (SPOR-CDP) [227, 252]. The SPOR-CDP was established in 2019 by the Health Data Research Network Canada (HDRN) who received $81 million dollars in funding from CIHR and partners. The aim of the initiative is to "become a distributed network that facilitates and accelerates multi-jurisdictional research in Canada" [227]. To address challenges related to data access, the Data Access Support Hub (DASH) was developed and implemented in 2020 [255]. Through the DASH, researchers embarking on multi-jurisdictional research can access a central data access coordination team to help them navigate the data access processes in each jurisdiction. The central coordination team facilitates communication between researchers and participating data centres in each jurisdiction (which include HDNS, PopData, and ICES).

The SPOR-CDP and DASH are currently in the early stages of implementation and their impact on the conduct of multi-jurisdictional research is not yet well understood. However, the findings of this study suggest that efforts to improve access to administrative health data *across* jurisdictions in Canada must occur in conjunction with efforts to identify and address the underlying barriers to accessing data *within* individual jurisdictions. Without a comprehensive strategy to address what is happening at the level of the individual provinces and territories, initiatives to improve access to data nationally are unlikely to achieve the desired aims. This study may provide the impetus for HDRN membership to undertake such efforts, as well as provide a framework for identifying factors affecting access to administrative health data in each participating province to improve harmonization across the network.

A second initiative to which the findings of this study may be relevant is the Pan-Canadian Health Data Strategy [256]. Led by the Public Health Agency of Canada (PHAC), and motivated by the COVID-19 pandemic, the aim of the Strategy is to "support the effective creation, exchange, and use of critical health data for the benefit of Canadians and the health and public health systems they rely on" [257]. In their most recent report, the Strategy's Expert Advisory Group [258] recommended the implementation of a Canada-wide learning health system, with action at the provincial level guided by a common set of principles. One of the guiding principles is the need for timely access to data by researchers and other health system stakeholders. This study highlights a range of barriers relevant to administrative health data that should be considered when developing provincial data strategies.

The findings of this study also have far-reaching implications for a range of other stakeholder groups. For researchers, the typology of factors affecting access to administrative health data for research identified in this study provides a framework that may be used to examine access to data in other Canadian or international jurisdictions, or who are interested in adapting it to other types of health data. The findings may also be of relevance to the funders of health research in Canada. For all funders, a better understanding of what's happening "on the ground" can inform funding calls by ensuring that the funding timeframes and amounts are sufficient to be inclusive of administrative health data research. Moreover, funders may also have a role in addressing the barriers identified in this study. For example, federal funding bodies wield a great deal of influence within the research community, which can be used to bring about positive change (e.g., to enforce mandatory training requirements, develop national best practices document, etc.). As the most comprehensive study of access to administrative health data in Canada to date, the findings of this study may be of particular interest to CIHR given their substantial investments in national data initiatives [227, 252]. Additionally, regulatory stakeholders and decisions-makers within organizations who are responsible for the regulation and oversight of administrative health data (e.g., within provincial data centres, or external data provider organizations) may use the findings of this study to develop, implement, and evaluate strategies to improve access to administrative health data, or health data more broadly, within their respective organizations. Finally, the findings of this study may help support efforts to improve data access in jurisdictions outside of Canada where data access challenges

exist [219, 220, 238, 259], particularly in those that employ a similar governance framework (i.e., an institution-based research ethics model intersecting with information legislation regulating the access, use, and disclosure of personal health information for research purposes).

5.4. Strengths and limitations

5.4.1. Strengths

This study had many strengths, several of which were related specifically to the methodology. First, the use of case study methodology provided a framework for interprovincial comparisons that was not available via other qualitative approaches. Multiple case study facilitated the identification of commonalities across cases and provided a means to explore differences. Moreover, the inclusion of multiple cases provided a range of contexts under which the study findings held true, contributing to the robustness of study findings [183] and facilitating the discovery of the fundamental aspects of the phenomena [194]. A second methodologic strength was the focus on a small number of purposefully selected cases [182], which allowed the researcher to develop an in-depth understanding of each case, to explore the heterogeneity that existed within each case, and to consider the impacts on the phenomenon of interest. The use of multiple data sources, including interviews with research and regulatory stakeholders with diverse experiences and perspectives, allowed the researcher to identify the multi-level factors affecting access to administrative health data for research and improved rigor by enabling the triangulation of data sources [183].

The study was also strengthened by the collective expertise of the researcher and committee, who had extensive experience working with administrative health data locally and across provinces (CK, GP, RU, AL) and legal expertise regarding the regulation of health data, both broadly and in the context of research (EG). Familiarity with provincial data centres across the country, as well as known differences in the regulatory policies and processes across provinces, facilitated case selection. In addition, an awareness of individuals across the country with experiencing working with administrative health data facilitated recruitment through the identification of key informants. The researcher (CK) was particularly familiar with issues related to data access as someone who frequently sought access to administrative health data for research, as well as a longstanding member of the HDNS DAC. This was conducive to obtaining rich interview data and facilitated data analysis and interpretation through an improved ability to generate concepts and theoretical insights from the data and to make meaning of these (i.e., provided "theoretical sensitivity")[260].

Related to data collection, this study relied entirely on qualitative data collection and analysis and did not attempt to quantify the timeliness of data access (e.g., the amount of time required to obtain data access approvals or gain access to a research dataset). While this may be viewed as a study limitation, there were several reasons for focusing on qualitative data. First, the provincial data centres included in this study did not all routinely capture timeliness data at the time this study was undertaken. Second, even where timeliness data were available from the provincial data centre, such data represented only a portion of the total data access timeline. For example, any time

spent by the research team preparing the study protocol or data access applications or meeting with data providers would not have been captured. Third, even if all data provider organizations captured complete timeliness data, it was not within the scope of this study to calculate data access timelines for individual projects across multiple data providers. Moreover, it would not be feasible to do so, given the potential number of projects that would have to be examined (i.e., approximately 1000 at ICES annually [207, 208]). Finally, this study was not focused on understanding the factors affecting timelines, but rather sought to understand access more broadly, including the context and circumstances that impact access. In other words, quantifying the timeliness of data access was not feasible, nor was it necessary to address the study objectives. In fact, the findings suggest that quantifying the timeliness of data access would have been a fruitless effort given the intra-provincial (and indeed, intra-researcher) variations in access.

5.4.2. Limitations

This study was underpinned by pragmatism; however, interpretivism may have allowed the researcher to examine the role of the socio-historical context in which each research system evolved in greater depth as well as the implications for data access. Additionally, an interpretive approach may have been more conducive to theory development by permitting the researcher to delve more deeply into the relationships between factors even where these connections were not explicitly articulated by participants. The findings of this study reflect the policies and processes that were in place during the

study timeframe and do not capture changes that may have occurred within each province since the completion of data collection (July 2021).

Additional limitations are related to recruitment—specifically, the introduction of participant bias, and limited recruitment of regulatory stakeholders. During recruitment, the researcher set out to capture a wide range of individual experiences with accessing administrative health data for research. When recruiting research stakeholders, the aim was to purposefully recruit individuals with a range of roles and different levels of experience, and to invite both those who were successful in obtaining access to data for research, and those who were not. Except for one individual, the research stakeholders who participated in the study had all been successful in carrying out studies using administrative health data during the study timeframe (2015/16-2020/21). While regulatory stakeholders in all cases were adamant that data access applications were rarely rejected, this study did not address situations wherein researchers were deterred from attempting to pursue access, where access to data was not pursued beyond the feasibility stage, or where applications were withdrawn. This may have introduced recruitment bias—the research team's ability to purposively recruit research stakeholders relied on their awareness of individuals in each province who were engaged in research using administrative health data, and on the identification of potential participants from institutional/organization websites (i.e., those listed as having administrative health data experience or expertise). As such, the recruitment process was biased towards identifying those who had an established track record of successfully obtaining access to data, which is reflected in the study findings. It

is therefore likely that researcher experiences with accessing data are not as positive as depicted in this study.

One of the main challenges of conducting this study was participant recruitment. Originally, the goal was to recruit 20 individuals for each case (10 regulatory stakeholders and 10 research stakeholders), with the final number of study participants based on reaching theoretical saturation. Ultimately, a total of 46 participants were recruited. As anticipated, recruitment in NS proceeded without issue, likely due to the researcher and members of the research team being based locally (i.e., in Halifax and at Dalhousie University). Recruitment in BC and ON was much more difficult. Although this was expected to some extent as a result of the researcher being less well-known in these provinces, COVID-19 created additional challenges. Recruitment in BC and ON coincided with surges in COVID-19 cases in each province, which likely impacted the ability of individuals to participate in research. This led to fewer participants being recruited in BC and ON (14 in each). Recruitment in BC may also have been impacted by a reluctance to discuss the events that had transpired at the BC Ministry of Health (i.e., those set out in the "Misfire" report [218]). While every attempt was made to reach theoretical saturation, it is possible that due the low number of participants, particularly from the regulatory stakeholder group in BC (n=5), important key informant perspectives may not have been captured.

Finally, the study design was such that key stakeholders were excluded from the study. Within each case, there were two embedded units of analysis representing distinct stakeholder groups (i.e., research stakeholders and regulatory stakeholders).

This approach was taken to capture the unique perspectives of each group toward identifying the full range of factors affecting access to data. Unfortunately, this approach resulted in the exclusion of an important stakeholder group—data centre analysts. The role of these individuals differs from that of both researchers and regulatory stakeholders and they have a great deal of influence with regard to how the data access process unfolds, especially from the time of approval to receipt of data. By not including this stakeholder group, important factors affecting access may have been missed. Nonetheless, the inclusion of individuals from a range of regulatory roles in each case, combined with researcher perspectives and case documents, facilitated the in-depth examination of each case, and the identification of a wide range of factors affecting access to administrative health data for research across three provinces, which may be refined through additional inquiry.

5.5. Researcher reflexivity

5.5.1. Acknowledging and examining assumptions and biases

In qualitative research, where the researcher functions as the data collection instrument, researchers are encouraged to identify and explicitly set out their assumptions and biases through a process of critical self-reflection referred to as reflexivity [173, 182, 261]. As part of this process, the researcher discusses their experiences with the phenomenon at hand, and reflects on how these experiences may impact the study [173]. This facilitates the transparency of research [173, 182] and subsequently contributes to the credibility of the researcher and trustworthiness of the research itself [194]. In the following paragraphs, I describe how my personal

experiences led to my decision to study access to administrative health data for research and the impacts on various aspects of the study.

In 2010, I accepted a position as a Project Coordinator for a grant-funded research program (the Cancer Outcomes Research Program [COR]) situated at Dalhousie University and NS Health. As part of this role, one of my responsibilities was to obtain the approvals required to access linked administrative health data for research purposes (e.g., approvals from REBs, data access committees, privacy officers, etc.). When the *Personal Health Information Act* [149] came into force in 2013, there was concern within the local research community with regard to how the legislation would impact health research in NS. Information sessions that I attended within Capital Health (now Central Zone, NS Health) about the upcoming implementation of the *Personal Health Information Act* indicated that new policies and processes were being put into place and suggested a move toward more stringent controls around the collection, use, and disclosure of PHI for research purposes. With direct implications for my role as Project Coordinator and knowing that similar legislation already existed in other provinces, I began searching for information about how the enactment of legislation was impacting research elsewhere in Canada. What I discovered was an international body of literature, in which researchers described various challenges related to accessing health data for research purposes, and numerous critiques of information legislation, which cemented my interest in learning more about the impacts of information legislation on health research in Canada.

At the outset of my studies, I was admittedly focused on information legislation and establishing that research was not being treated "fairly". Shortly thereafter, within the first year of my studies, I was invited to become a member of the Data Access Committee (DAC) at HDNS, the research unit at Dalhousie University that houses a provincial administrative health data repository in NS [17]. From the time I joined the DAC, to finalizing my research questions and objectives, my perspective changed substantially. Sitting on the other side of the table, I saw that there was huge variation in terms of researchers' understanding of administrative health data, the provincial legislation, and the safeguards that must in place when accessing these data for research. Whereas early on I was concerned that data were being over-regulated to the extent that research would be hindered, my experience on the DAC allowed me to develop a greater appreciation of the need for privacy protections for patient information and the role of oversight bodies in ensuring that privacy is protected. My research question evolved, focusing not on establishing information legislation as problematic, but on understanding the multitude of factors that impact access to administrative health data for research (though I remained particularly interested in the role of legislation).

By the time I started data collection for this study, I had been in my Project Coordinator role at COR for 10 years, and on the DAC for five. I had submitted a number of data access applications for COR studies, both NS-specific and multi-provincial, and reviewed dozens more as a DAC member. Through these experiences, I developed relationships with other researchers in NS who worked with administrative health data,

as well as various individuals involved in the regulation and oversight of access to administrative health data for research. Multi-provincial collaborations allowed me to connect with other researchers across the country, and to gain insight into data access policies and processes in different provinces. Taken together, these experiences impacted my research in several ways. I was better prepared to select cases from which I could learn the most (see section 3.2.2). My ability to recruit was also improved because of the relationships I had within the research community and my research track record, which I believe lent credibility to the study. When it came to doing the interviews, my knowledge of the subject matter, and confidence in that knowledge, was invaluable—I was able to engage with participants at their level, to probe appropriately, and elicit rich responses. I also felt that my own experiences, as both a researcher and a DAC member, gave me a certain level of credibility, and allowed me to establish a rapport with participants. Rather than being perceived as an outsider, I felt that participants were excited to share their perspectives with someone who understood the work they did and the challenges they faced, which was reflected in their candor. Finally, my knowledge and experiences facilitated data analysis and interpretation by improving my ability to generate concepts and theoretical insights from the data and to make meaning of these (i.e., provide improved "theoretical sensitivity")[260].

With such proximity to the topic being studied, especially to the NS case, one challenge of conducting this work was to be cognizant of and mitigate the impact of my own personal biases on the findings. Admittedly, I have experienced challenges when attempting to access administrative health data for research in Nova Scotia. While these

challenges are, in part, what led me to pursue this particular area of study, many of my biases were mitigated by my involvement with the HDNS DAC and gaining an understanding of and appreciation for data access governance—an appreciation that grew over the course of this study. As such, I entered into data collection and analysis with the aim of understanding, rather than "grinding an axe". Nonetheless, I have taken appropriate steps to ensure that the findings presented in this document are representative of the experiences and perspectives of study participants and to minimize any unconscious biases that may impact my representation of the data. These steps are outlined in detail in Section 3.2.5 ('Ensuring rigor and trustworthiness').

5.5.2. Reflections on the use of pre-existing theory (or lack thereof)

In reflecting on the work outlined in this document, I would also like to take this opportunity to acknowledge that this study was not informed by a pre-existing theory and provide some insight as to why. In qualitative research, pre-existing theory is commonly used as a lens through which the phenomenon of interest can be viewed and understood. From a practical perspective, the use of pre-existing theory can also help guide analysis, providing concepts and domains that may be incorporated into the coding framework, and highlighting relationships between these. For these reasons, I initially sought out a theory (or theories) that might be helpful in guiding the planned research, however, I was unsuccessful in identifying a theory that felt like a good "fit". This was likely due to several factors, including the interdisciplinary nature of the study and the difficulty of identifying prior theory spanning the relevant disciplines (i.e.,

medicine, law, and epidemiology), as well as the overall lack of empirical research on the topic of access to administrative health data.

After much consideration, and consultation with my committee, I decided to proceed without the use of a pre-existing *formal* theory to inform the study. While there were times when I questioned the decision to embark on a study of this size and scope without the benefit of a theoretical lens to give shape and structure to the data—especially given the vast (and at times, overwhelming) amount of data collected—I believe I was justified in doing so. While the use of pre-existing theory to inform research is often considered to be the more rigorous approach, it is not always appropriate, particularly when there is a lack of pre-existing theory to facilitate deduction. This has been discussed at length by Locke [262], who also argues that the emphasis on using pre-existing theory can result in the premature selection and application of theories before the researcher has sufficient understanding of the phenomenon being studied to assess their utility. The use of an inductive analytic approach was therefore appropriate and ensured that the study findings were grounded in the data and reflective of the experiences and perspectives of key informants with firsthand knowledge relevant to access to administrative health data for research in Canadian settings.

5.6. Ethical considerations

The consent process for this study was carried out in accordance with the general consent principles outlined in the TCPS 2 [112]. All individuals who participated in this study provided written, informed consent. Consistent with the ethical principle that

consent be treated as a process rather than a one-time occurrence, consent was also confirmed verbally prior to the start of key-informant interviews, and participants were informed that they could withdraw consent at any point during the interview.

During the consent process, participants were informed of the potential risks of participating, which were primarily related to participant privacy and confidentiality. In the consent form, the researcher's commitment to protecting participant privacy and confidentiality was emphasized, and the strategies used to mitigate the risk of identification were described; However, the following required disclaimer was also included: "Even though the risk of identifying you from the data is very small, it can never be completely eliminated". Early on in the study, it became evident to the researcher that this potential risk was of great concern to participants. Due to these concerns, many individuals requested assurances from the researcher that their privacy and confidentiality would be protected prior to participating. The researcher explained that as someone working in this space themselves, they understood and appreciated participant concerns, assured participants that every effort would be made to minimize the risk of identification, and reiterated the specific steps that would be taken to mitigate risk. All participants subsequently proceeded to participate in their scheduled interview.

Due to these participant concerns, particular care was taken during the preparation of this document not to provide information about participants that could compromise privacy and confidentiality. For example, the description of key informants provided in the results (section 4.1.1.) provides only a high-level summary of key

informant characteristics and does not provide a breakdown of characteristics by province or identify the role of key informants within their organizations. In addition, illustrative quotes were selected carefully, and potentially identifying information was masked or removed, to help prevent the identification of individuals from the data.

A discussion of ethical considerations in the context of this study must also address the researcher's relationship to the study topic. As part of their role as an employee with the Cancer Outcomes Research Program at NS Health, the researcher has been involved in various provincial and national studies involving administrative health data, and frequently requested access to data from HDNS and other data providers. The researcher has also served as a member of the HDNS DAC for nearly eight years. As described in Section 5.5. ('Researcher reflexivity'), these experiences strengthened the study by improving the researcher's ability to identify potential key informants during the recruitment process, conduct in-depth interviews, and analyse and interpret data.

Importantly, these combined experiences also mitigate any perceived conflicts of interest on the part of the researcher. As someone who obtains access to administrative health data for research, and who has experienced challenges when attempting to access data in the past, it could be argued that the researcher has a vested interest in producing a report that recommends minimizing barriers to accessing administrative health data for research purposes. While this may be true to a certain extent, it may also be argued (and is indeed being argued here) that the researcher's dual role, as someone who obtains access to administrative health data for research and as someone involved

in the regulation of access to these data for research, ideally positioned the researcher to conduct this study. That is, the researcher's interests in minimizing barriers to accessing data are balanced by their understanding of the importance of privacy protection in the context of health data and appreciation for appropriate oversight relevant to data access. In addition to bringing a balanced perspective to the research, appropriate strategies were used to ensure the findings accurately reflect the experiences and perspectives of key informants rather than the researcher (as described in section 3.2.5).

5.7. Dissemination plan and next steps

5.7.1. Dissemination of findings

For each of the cases included in this study, a summary of case-specific findings will be prepared and shared with the key informants who participated in the study, and other individuals in leadership and decision-making roles relevant to each case (i.e., within each provincial data center and provincial department/ministry of health). The researcher will offer to meet with these individuals to discuss findings and will present findings to relevant stakeholder groups (e.g., data center staff) if requested.

Findings will also be presented at conferences and other academic symposia. At the national level, presentations will be sought at the Canadian Association of Health Services and Policy Research (CAHSPR) Annual Conference, as well as the annual Government of Canada (GC) Data Conference. To reach an international audience, presentations will also be sought at the bi-annual International Population Data Linkage

Conference. Abstracts may be submitted elsewhere as additional conferences and meetings are announced.

In addition to the study protocol, which was published in the International Journal of Population Data Science in 2021 (see Section 1.1. for copyright acknowledgement and details), several additional manuscripts are currently planned for publication in scientific journals. A pair of companion manuscripts will be prepared, which will focus on data access challenges (based on Section 3.2), and factors affecting access to administrative health data for research (based on Section 3.3). Additional manuscripts focusing on specific categories of factors will also be submitted for publication, which will allow a more detailed discussion of the individual factors within each of these categories, and the implications for data access. At this time, one manuscript focusing solely on the regulatory factors and another focusing solely on organizational factors are planned. Potential journals for submission include the International Journal of Population Data Science, Health Research Policy and Systems, and Healthcare Policy (an initiative of CIHR and CAHSPR). Published manuscripts, along with executive summaries, will be circulated to individuals in leadership roles at CIHR (specifically at the Institute of Health Services and Policy Research (IHSPR)), each of the CIHR SPOR Support Units, HDRN Canada, and the Pan-Canadian Data Strategy.

5.7.2. Stakeholder dialogues

In addition to traditional dissemination activities, the researcher will seek additional funding to conduct a series of deliberative dialogues (three in total) with the aim of developing stakeholder-informed actionable recommendations for addressing the

identified barriers to access to administrative health data for research. Deliberative dialogues bring relevant stakeholders together to review the evidence on a specific topic or problem, identify potential solutions, and discuss implementation considerations [263, 264]. Each dialogue will address the following question: "What can we do to improve access to administrative health data for research while ensuring patient privacy is protected?". The researcher has prior experience organizing and facilitating deliberative dialogues using the stakeholder dialogue approach developed by the McMaster Health Forum [265]. Consistent with this approach, the following activities will be undertaken:

1) *Preparation and circulation of an "evidence brief"*—A document containing relevant research evidence on the topic of access to administrative health data will be prepared and circulated to dialogue invitees. This will include research specific to the Canadian context (including the findings of the current study) as well as international peer-reviewed studies. This document will be circulated a minimum of two-weeks prior to the dialogue, providing invitees with time to review the content and ensuring that all invitees have a baseline understanding of relevant topics (e.g., the strengths and limitations of administrative health data, how these data are governed and accessed, barriers to access, etc.) and are prepared to engage in discussion.
2) *In-person event*—A total of 20-22 stakeholders will be invited to attend each in-person deliberation event. Stakeholders will include members of the public, researchers and research staff who work with administrative health data, and

relevant decision-makers at the provincial and federal levels, including data custodians (e.g., health authority leaders), and government and funding agency representatives. Each event will be conducted over the course of two days and be organized into four deliberations, starting with a general discussion of the problem, and working toward the development, ranking, and refinement of actionable recommendations to address the problem. Facilitators will be present to guide each deliberation and capture key discussion points and recommendations. The four deliberations will address the following questions:

- Deliberation 1—What are the barriers to accessing administrative health data in Canada for research?
- Deliberation 2— What are some potential strategies for addressing these barriers?
- Deliberation 3— What are the implications of implementing these solutions?
- Deliberation 4— What are some actionable recommendations to address the barriers to access to administrative health data for research, without compromising patient privacy?

3) *Dialogue summary*—For each of the three stakeholder dialogues, a report will be prepared that summarizes the discussion that took place during each of the four deliberations, highlighting the recommendations with the highest ranking. These individual reports will be collated into a single document that will subsequently be distributed to invitees and other relevant stakeholders, including decision-makers with the authority to implement stakeholder recommendations.

A stakeholder dialogue will be held in each of the three provinces included in this study (NS, BC, and ON), and will involve local partnerships (e.g., with the provincial data centres and other organizations or individuals) as appropriate. While the primary aim of these dialogues is to inform action, the process will also facilitate the dissemination of findings to relevant stakeholders, improve stakeholders' understanding of administrative health data research and the challenges and barriers relevant to data access, and serve as an opportunity to foster trust between stakeholder groups.

5.7. Recommendations for future research

The findings of the current study point to the opportunity for additional research in several areas. First, given that this study is the first to develop an empirically based typology of factors affecting access to administrative health data for research, a reasonable next step would be the application of this framework to the study of access to administrative health data in other jurisdictions with the aim of developing a mid-range theory or explanation of data access. As the first in-depth study of factors affecting access to administrative health data for research in Canada, the current study is not the "final word" on the factors affecting access to administrative health data, but a basis for additional inquiry and theory development. Second, future research may focus on the development of context-specific strategies to address each of the specific barriers through the engagement of relevant stakeholders (e.g., via stakeholder dialogues as described above). Although a range of strategies were proposed in this document, additional work is required to identify the most appropriate strategies and how they may be adapted for different contexts. Third, given the challenges associated

with external linkages, additional investigation into the organizational culture of external data providers relevant to sharing PHI for research is required. Given the complexity of the concept of organizational culture, as well as the broad range of organizations that comprise the data landscape, the potential for greater learning exists. A fourth area of potential inquiry is whether and to what extent current challenges and barriers associated with accessing administrative health data may deter researchers from pursuing administrative health data research. While this question was not explored in the current study, it would provide a more complete understanding of the implications of the current approach to regulating access to administrative health data for research while providing insight into the issue of data non-use, which has been identified as a current gap in the literature [266].

5.8. Conclusion

Administrative health data can be used to support decision-making at all levels of the healthcare system (i.e., clinical, administrative, and policy), but only if they can be accessed. While there has been a growing body of literature indicating barriers to data access in Canada, this literature has lacked empirical evidence to understand current issues, including delays and variations in the timeliness of access, and to develop appropriate strategies to improve access. Through the use of a qualitative multiple-case study of three provinces, this gap was addressed.

This study found that access to administrative data was suboptimal in all three cases with respect to timeliness, costs and data quality. Eight key barriers to optimal access were identified. While some of the identified barriers may be relatively

straightforward to address (e.g., improving stakeholder knowledge and expertise), addressing leadership and organizational culture will require complex interventions to bring about substantial and sustainable long-term change. At the same time, barriers were not distributed evenly across provinces, providing insight into interprovincial variations in the timeliness of access to data and highlighting the need for context-specific strategies to improve access.

As the first in-depth study examining factors affecting access to administrative health data for research in Canada, this study provides evidence that may be used to inform efforts to optimize access to administrative health data for research within and across provinces. As such, the findings of this study may have implications for a range of stakeholders, including researchers who access and use administrative health data, those who have roles in the oversight and regulation of access to data, individuals engaged in ongoing efforts to improve access to data, and the funders of health research. Finally, the finding of this study may be relevant to jurisdictions outside of Canada that employ a similar regulatory framework for administrative health data research.

Future research may focus on further exploring the relationships between the various factors identified in this study toward the identification of a mid-range theory or explanation of data access, developing context-specific strategies for improving access to data, examining organizational culture relevant to administrative health data research, and gaining an improved understanding of the extent to which current

challenges and barriers to data access act as a deterrent from seeking access to administrative health data.

REFERENCES

1. **Term: Administrative Health Data** [http://mchp-appserv.cpe.umanitoba.ca/viewDefinition.php?definitionID=102210]

2. Lucyk K, Lu M, Sajobi T, Quan H: **Administrative Health Data in Canada: Lessons from History**. *BMC Medical Informatics and Decision Making* 2015, **15**(1):69.

3. Cadarette SM, Wong L: **An Introduction to Health Care Administrative Data**. *Canadian Journal of Hospital Pharmacy* 2015, **68**(3).

4. Canadian Institute for Health Information: **Better Information for Improved Health: A Vision for Health System Use of Data in Canada**. Ottawa: ON: Canadian Institute for Health Information; 2013.

5. The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation: **Accessing Health and Health-Related Data in Canada**. Ottawa:ON: Council of Canadian Academies; 2015.

6. OECD: **Health Data Governance: Privacy, Monitoring and Research-Policy Brief**. Paris; 2015.

7. Sloane PD, Mor V, Preisser JS: **Administrative Data for Research: An Increasingly Powerful Tool, but Still With Caveats**. *Journal of the American Medical Directors Association* 2018, **19**(2):97-99.

8. Canadian Institute of Health Services and Policy Research: **Strategic Plan 2021-2026**: Canadian Institute for Health Research; 2021.

9. Dai WF, de Oliveira C, Blommaert S, Pataky RE, Tran D, Aurangzeb Z, Kendell C, Folkins C, Somayaji C, Dowden J *et al*: **Mapping Canadian Data Assets to Generate Real-World Evidence: Lessons Learned from Canadian Real-World Evidence for Value of Cancer Drugs (CanREValue) Collaboration's RWE Data Working Group**. *Current Oncology* 2022, **29**(3):2046-2063.

10. Togo K, Yonemoto N: **Real World Data and Data Science in Medical Research: Present and Future**. *Japanese Journal of Statistics and Data Science* 2022, **5**(2):769-781.

11. Holman CD, Bass JA, Rosman DL, Smith MB, Semmens JB, Glasson EJ, Brook EL, Trutwein B, Rouse IL, Watson CR *et al*: **A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System**. *Australian Health Review* 2008, **32**(4):766-777.

12. Jutte DP, Roos LL, Brownell MD: **Administrative Record Linkage as a Tool for Public Health Research**. *Annual Review of Public Health* 2011, **32**(1):91-108.

13. **Population Data BC** [https://www.popdata.bc.ca/]

14. **Manitoba Centre for Health Policy** [http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/]

15. **Institute for Clinical Evaluative Sciences** [https://www.ices.on.ca/]

16. **New Brunswick Institute for Research, Data, and Training** [https://www.unb.ca/fredericton/arts/nbirdt/]

17. **Health Data Nova Scotia** [https://medicine.dal.ca/departments/department-sites/community-health/research/hdns.html]

18. **Secure Island Data Repository (SIDR)** [https://chcresearch.ca/research-initiatives/secure-island-data-repository-sidr/]

19. **Newfoundland and Labrador Center for Health Information** [https://www.nlchi.nl.ca/]

20. Roos LL, Gupta S, Soodeen R-A, Jebamani L: **Data Quality in an Information-Rich Environment: Canada as an Example**. *Canadian Journal on Aging* 2004, **24**(Suppl 1):153-170.

21. Meagher N, McGrail K: **Data Access Review Times Study: Report**: Population Data BC; 2013.

22. Kephart G: **Barriers to Accessing and Analyzing Health Information in Canada**: Canadian Institute for Health Information; 2002.

23. Azoulay L, Filion KB, Platt R, Dahl M, Dormuth CR, Clemens KK, Durand M, Juurlink DN, Targownik LE, Turin TC *et al*: **Incretin Based Drugs and the Risk of Pancreatic Cancer: International Multicentre Cohort Study**. *BMJ* 2016, **352**:i581.

24. Ho HKK, Görges M, Portales-Casamar E: **Data Access and Usage Practices Across a Cohort of Researchers at a Large Tertiary Pediatric Hospital: Qualitative Survey Study**. *JMIR Medical Informatics* 2018, **6**:e32.

25. 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**. *Canadian Journal of Cardiology* 2012, **28**(2):152-154.

26. Lavoie JG, Wong S, Katz A, Sinclair S: **Opportunities and Barriers to Rural, Remote and First Nation Health Services Research in Canada: Comparing Access to Administrative Claims Data in Manitoba and British Columbia**. *Healthcare Policy* 2016, **12**(1):52-58.

27. Groome PA, McBride M, Jiang L, Kendell C, Decker K, Grunfeld E, Krzyzanowska MK WM: **Lessons Learned: It Takes a Village to Understand Inter-Sectoral Care Using Administrative Data across Jurisdictions**. *International Journal of Population Data Science* 2018, **3**(3):1-13.

28. Butler A, Smith M, Jones W, Adair CE, Vigod SN, Lesage A, L Kurdyak: **Multi-Province Epidemiological Reserach Using Linked Administrative Data: A Case Study from Canada**. *International Journal of Population Data Science* 2018, **3**(3).

29. Clayton EW, Evans BJ, Hazel JW, Rothstein MA: **The Law of Genetic Privacy: Applications, Implications, and Limitations**. *Journal of Law and the Biosciences* 2019, **6**(1):1-36.

30. Firestone WA: **Arguments for Generalizing from Data as Applied to Qualitative Reserach**. *Educational Researcher* 1993, **22**(4):16-23.

31. Jones J: **The Effects of Non-Response on Statistical Inference**. *Journal of Health and Social Policy* 1996, **8**:49-62.

32. Al-Shahi R, Warlow C: **Using Patient-Identifiable Data for Observational Research and Audit**. *BMJ* 2000, **321**(7268):1031-1032.

33. Harris MA, Levy AR, Teschke KE: **Personal Privacy and Public Health: Potential Impacts of Privacy Legislation on Health Research in Canada**. *Canadian Journal of Public Health* 2008, **99**(4):293-296.

34. **Statistical Power And Sample Size**. In: *Statistics Applied to Clinical Trials.* 4th edn. Edited by Cleophas TJ ZA, Cleophas TF, Cleophas EP: Springer, Dordrecht; 2009.

35. **Social Determinants of Health** [http://cbpp-pcpe.phac-aspc.gc.ca/public-health-topics/social-determinants-of-health/]

36. Harris PM: **Attrition Revisited**. *American Journal of Evaluation* 1998, **19**(3):293-305.

37. Young AF, Powers JR, Bell SL: **Attrition in Longitudinal Studies: Who Do You Lose?** *Australia and New Zealand Journal of Public Health* 2006, **30**(4):353-361.

38. Brook EL, Rosman DL, Holman CD: **Public Good Through Data Linkage: Measuring Research Outputs From the Western Australian Data Linkage System**. *Australian and New Zealand Journal of Public Health* 2008, **32**(1):19-23.

39. Sibthorpe B, Kliewer E, Smith L: **Record Linkage in Australian Epidemiological Research: Health Benefits, Privacy Safeguards and Future Potential**. *Australian Journal of Public Health* 1995, **19**(3):250-256.

40. Trutwein B, Holman CD, Rosman DL: **Health Data Linkage Conserves Privacy in a Research-Rich Environment**. *Annals of Epidemiology* 2006, **16**(4):279-280.

41. Brown AP, Ferrante AM, Randall SM, Boyd JH, Semmens JB: **Ensuring Privacy When Integrating Patient-Based Datasets: New Methods and Developments in Record Linkage**. *Frontiers in Public Health* 2017, **5**(34).

42. Institute of Medicine: **Health Services Research: Work Force and Educational Issues**. Washington DC: National Academy Press; 1995.

43. Last JM: **A Dictionary of Epidemiology**, 4 edn. New York: Oxford University Press; 2001.

44. Kindig D, Stoddart G: **What is population health?** *American Journal of Public Health* 2003, **93**(3):380-383.

45. Ulrich EH, So G, Zappitelli M, Canchlani R: **A Review on the Application and Limitations of Administrative Health Care Data for the Study of Acute Kidney Injury Epidemiology and Outcomes in Children**. *Frontiers in Pediatrics* 2021, **9**.

46. Gavrielov-Yusim N, Friger M: **Use of Administrative Medical Databases in Population-Based Research**. *Journal of Epidemiology and Community Health* 2014, **68**(3):283.

47. Mazzali C, Duca P: **Use of Administrative Data in Healthcare Research**. *Internal and Emergency Medicine* 2015, **10**(4):517-524.

48. Johnson EK, Nelson CP: **Values and Pitfalls of the use of Administrative Databases for Outcomes Assessment**. *Journal of Urology* 2013, **190**(1):17-18.

49. McGuckin T, Crick K, Myroniuk TW, Setchell B, Yeung RO, Campbell-Scherer D: **Understanding Challenges of Using Routinely Collected Health Data to Address Clinical Care Gaps: A Case Study in Alberta, Canada**. *BMJ Open Quality* 2022, **11**(1):e001491.

50. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P: **Assessing the Accuracy of Administrative Data in Health Information Systems**. *Medical Care* 2004, **42**(11):1066-1072.

51. Gologorsky Y, Knightly JJ, Lu Y, Chi JH, Groff MW: **Improving Discharge Data Fidelity for Use in Large Administrative Databases**. *Neurosurgical Focus* 2014, **36**(6):E2.

52. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM: **Measuring Diagnoses: ICD Code Accuracy**. *Health Services Research* 2005, **40**(5 Pt 2):1620-1639.

53. Lucyk K, Tang K, Quan H: **Barriers to Data Quality Resulting from the Process of Coding Health Information to Administrative Data: A Qualitative Study**. *BMC Health Services Research* 2017, **17**(1):766.

54. van Mourik MS, van Duijn PJ, Moons KG, Bonten MJ, Lee GM: **Accuracy of Administrative Data for Surveillance of Healthcare-Associated Infections: A Systematic Review**. *BMJ Open* 2015, **5**(8):e008424.

55. Garland A, Marrie RA, Wunsch H, Yogendran M, Chateau D: **Accuracy of Administrative Hospital Data to Identify Use of Life Support Modalities. A Canadian Study**. *Annals of the American Thoracic Society* 2020, **17**(2):229-235.

56. Lopushinsky SR, Covarrubia KA, Rabeneck L, Austin PC, Urbach DR: **Accuracy of Administrative Health Data for the Diagnosis of Upper Gastrointestinal Diseases**. *Surgical Endoscopy* 2007, **21**(10):1733-1737.

57. Southern DA, Roberts B, Edwards A, Dean S, Norton P, Svenson LW, Larsen E, Sargious P, Lau DC, Ghali WA: **Validity of Administrative Data Claim-Based Methods for Identifying Individuals with Diabetes at a Population Level**. *Canadian Journal of Public Health* 2010, **101**(1):61-64.

58. Oskoui M, Ng P, Dorais M, Pigeon N, Koclas L, Lamarre C, Malouin F, Richards CL, Shevell M, Joseph L: **Accuracy of Administrative Claims Data for Cerebral Palsy Diagnosis: A Retrospective Cohort Study**. *CMAJ Open* 2017, **5**(3):E570-e575.

59. Hennekens CH, Buring JE: **Epidemiology in Medicine**. Boston/Toronto: Little, Brown and Company; 1987.

60. **A Dictionary of Epidemiology**, 6th edn. New York: Oxford University Press; 2014.

61. Nørgaard M, Ehrenstein V, Vandenbroucke JP: **Confounding in Observational Studies Based on Large Health Care Databases: Problems and Potential Solutions - A Primer for the Clinician**. *Clinical Epidemiology* 2017, **9**:185-193.

62. Bohensky M: **Bias in Data Linkage Studies**. In: *Methodological Developments in Data Linkage.* 1st edn. Edited by Harron K, Goldstein H, Dibben C; 2015: 63-82.

63. Harron K, Dibben C, Boyd J, Hjern A, Azimaee M, Barreto ML, Goldstein H: **Challenges in Administrative Data Linkage for Research**. *Big Data and Society* 2017, **4**(2):2053951717745678.

64. Doidge JC, Harron KL: **Reflections on Modern Methods: Linkage Error Bias**. *International Journal of Epidemiology* 2019, **48**(6):2050-2060.

65. Harron K, Wade A, Gilbert R, Muller-Pebody B, Goldstein H: **Evaluating Bias Due to Data Linkage Error in Electronic Healthcare Records**. *BMC Medical Research Methodology* 2014, **14**(1):36.

66. Rentsch CT, Harron K, Urassa M, Todd J, Reniers G, Zaba B: **Impact of Linkage Quality on Inferences Drawn From Analyses Using Data With High Rates of Linkage Errors in Rural Tanzania**. *BMC Medical Research Methodology* 2018, **18**(1):165.

67. Johnson AP MB, Whitehead M, Xu J, Parlow JL: **Potential Impact of Changes in Administrative Database Coding Methodology on Research and Policy Decisions: An Example From the Ontario Health Insurance Plan**. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie* 2020, **67**(4):487-488.

68. Lix LM, Walker R, Quan H, Nedsole R, Yang J, Chen G: **Features of Physician Services Databases in Canada**. *Chronic Diseases and Injuries in Canada* 2012, **2**(4):187-193.

69. Doyle CM, Lix LM, Hemmelgarn BR, Paterson JM, C R: **Data Variability Across Canadian Administrative Health Databases: Differences in Content, Coding, and Completeness**. *Pharmacoepidemiology and Drug Safety* 2020, **29**(S1):68-77.

70. Lantz B: **The Large Sample Size Fallacy**. *Scandinavian Journal of Caring Sciences* 2013, **27**(2):487-492.

71. Kaplan RM, Chambers DA, Glasgow RE: **Big Data and Large Sample Size: A Cautionary Note on the Potential for Bias**. *Clinical and Translational Science* 2014, **7**(4):342-346.

72. Woodrow JR, Wang P, Valcour J, Aubrey-Bassler K, Gao Z: **A Crippling Cost? The Burden of Osteoarthritis in Newfoundland and Labrador, Canada: Examining OA Prevalence and Trends in Incidence Using Health Care Administrative Data**. *Osteoarthritis and Cartilage* 2017, **25**:S185.

73. Benchimol EI, Guttmann A, Griffiths AM, Rabeneck L, Mack DR, Brill H, Howard J, Guan J, To T: **Increasing Incidence of Paediatric Inflammatory Bowel Disease in Ontario, Canada: Evidence from Health Administrative Data**. *Gut* 2009, **58**(11):1490.

74. Hux JE, Ivis F, Flintoft V, Bica A: **Diabetes in Ontario: Determination of Prevalence and Incidence Using a Validated Administrative Data Algorithm**. *Diabetes Care* 2002, **25**(3):512.

75. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, Saunders LD, Beck CA, Feasby TE, Ghali WA: **Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data**. *Medical Care* 2005, **43**(11):1130-1139.

76. Gruneir A, Cigsar C, Wang X, Newman A, Bronskill SE, Anderson GM, Rochon PA: **Repeat Emergency Department Visits by Nursing Home Residents: A Cohort Study Using Health Administrative Data**. *BMC Geriatrics* 2018, **18**(1):157.

77. Kendell C, Decker KM, Groome PA, McBride ML, Jiang L, Krzyzanowska MK, Porter G, Turner D, Urquhart R, Grunfeld E: **Use of Physician Services During the Survivorship Phase: A Multi-Province Study of Women Diagnosed with Breast Cancer**. *Current Oncology* 2017.

78. Jiang L, Lofters A, Moineddin R, Decker K, Groome P, Kendell C, Krzyzanowska M, Li D, McBride ML, Mittmann N *et al*: **Primary Care Physician Use Across the Breast Cancer Care Continuum: CanIMPACT Study Using Canadian Administrative Data**. *Canadian Family Physician* 2016, **62**(10):e589-e598.

79. Grunfeld E, Hodgson DC, Del Giudice E, Moineddin R: **Population-Based Longitudinal Study of Follow-Up Care for Breast Cancer Survivors**. *Journal of Oncology Practice* 2010, **6**(4):174-181.

80. Rayson D, Urquhart R, Cox M, Grunfeld E, Porter G: **Adherence to Clinical Practice Guidelines for Adjuvant Chemotherapy for Colorectal Cancer in a Canadian Province: A Population-Based Analysis**. *Journal of Oncology Practice* 2012, **8**(4):253-259.

81. McConnell YJ, Inglis K, Porter GA: **Timely Access and Quality of Care in Colorectal Cancer: Are They Related?** *International Journal for Quality in Health Care* 2010, **22**(3):219-228.

82. Porter GA, Urquhart R, Bu J, McConnell YJ, Grunfeld E: **Timely Access and Quality of Care in Colorectal Cancer: A Population-Based Analysis**. *Annals of Surgical Oncology* 2010, **17**(Suppl 1):S118.

83. Widdifield J, Bernatsky S, Paterson JM, Tomlinson G, Tu K, Kuriya B, Thorne JC, Pope JE, Hollands S, Bombardier C: **Trends in Excess Mortality Among Patients With Rheumatoid Arthritis in Ontario, Canada**. *Arthritis Care & Research* 2015, **67**(8):1047-1053.

84. Feir D, Akee R: **First Peoples Lost: Determining the State of Status First Nations Mortality in Canada Using Administrative Data**. *Canadian Journal of Economics/Revue canadienne d'économique* 2019, **52**(2):490-525.

85. Tanuseputro P, Wodchis WP, Fowler R, Walker P, Bai YQ, Bronskill SE, Manuel D: **The Health Care Cost of Dying: A Population-Based Retrospective Cohort Study of the Last Year of Life in Ontario, Canada**. *PLoS One* 2015, **10**(3):e0121759.

86. Sam D, Cheung WY: **A Population-Level Comparison of Cancer-Related and Non-Cancer-Related Health Care Costs Using Publicly Available Provincial Administrative Data**. *Current Oncology* 2019, **26**(2).

87. Dolan D, Grainger J, MacCallum N, Creatura D, Forrester J, Shiller S: **The Institute for Clinical Evaluative Sciences: 20 Years and Counting**. *Healthcare Quarterly* 2012, **15**(4):19-21.

88. Katz A, Enns J, Smith M, Burchill C, Turner K, Towns D: **Population Data Centre Profile: The Manitoba Centre for Health Policy**. *International Journal of Population Data Science* 2020, **4**(2):1131.

89. Schull MJ, Azimaee M, Marra M, Cartegena RG, Vermeulen MJ, Ho MM, Guttman A: **ICES: Data, Discovery, Better Health**. *International Journal of Population Data Science* 2019, **4**(2).

90. Grams ME, Sang Y, Levey AS, Matsushita K, Ballew S, Chang AR, Chow EKH, Kasiske BL, Kovesdy CP, Nadkarni GN *et al*: **Kidney-Failure Risk Projection for the Living Kidney-Donor Candidate**. *New England Journal of Medicine* 2015, **374**(5):411-421.

91. Lentine KL, Kasiske BL, Levey AS, Adams PL, Alberú J, Bakr MA, Gallon L, Garvey CA, Guleria S, Li PK *et al*: **KDIGO Clinical Practice Guideline on the Evaluation and Care of Living Kidney Donors**. *Transplantation* 2017, **101**(8S Suppl 1):S1-s109.

92. **Manitoba Prenatal Benefit** [https://www.gov.mb.ca/healthychild/healthybaby/mpb.html]

93. **Data Linkage Western Australia** [http://www.datalinkage-wa.org.au/]

94. Warren SD, Brandeis LD: **The Right to Privacy**. *Harvard Law Review* 1890, **4**(5):193-220.

95. Prosser WL: **Privacy**. *California Law Review* 1960, **48**(3):338-423.

96. Gavison R: **Privacy and the Limits of the Law**. *The Yale Law Journal* 1980, **80**(3):421-471.

97. Westin AF: **Privacy and Freedom**. New York: Athenum; 1967.

98. Moor JH: **The Ethics of Privacy Protection**. *Library Trends* 1990, **39**(1&2):69-82.

99. Moor JH: **Towards a Theory of Privacy in the Information Age**. *ACM SIGCAS Computers and Society* 1997.

100. Parent WA: **Privacy, Morality, and the Law**. *Philosophy & Public Affairs* 1983, **12**(4):269-288.

101. Rachels J: **Why Privacy is Important**. *Philosophy and Public Affairs* 1975, **4**:323–333.

102. Fried C: **An Anatomy of Values**. Cambridge: Harvard University Press; 1970.

103. Tavani HT: **Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy**. *Metaphilosophy* 2007, **38**(1):1-22.

104. Altman I: **Privacy Regulation: Culturally Universal or Culturally Specific**. *Journal of Social Issues* 1977, **33**(3):66-84.

105. Moore A: **Defining Privacy**. *Journal of Social Philosphy* 2008, **39**(3):411-428.

106. Bloustein EJ: **Privacy as an Aspect of Human Dignity: an Answer to Dean Prosser**. In: *Philosophical Dimensions of Privacy: An Anthology.* edn. Edited by Schoeman FD. Cambridge: Cambridge University Press; 1984: 156-202.

107. Elder T, Yarrison FW, Long BL: **An Empirical Investigation of Privacy: The Impact of Multiple Levels of Trust**. In: *American Sociological Association Annual Meeting.* Chicago, IL; 2015.

108. Margulis S: **On the Status and Contribution of Westin's and Altman's Theories of Privacy**. *Journal of Social Issues* 2003, **59**(2):411-429.

109. Margulis ST: **Privacy as a Social Issue and Behavioural Concept**. *Journal of Social Issues* 2003, **59**(2):243-261.

110. Altman I: **The Environment and Social Behaviour: Privacy, Personal Space, Territory, and Crowding**. Monterey, CA: Brooks/Cole; 1975.

111. Cook MA: **The Risks of Absolute Medical Confidentiality**. *Science and Engineering Ethics* 2013, **19**(1):107-122.

112. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council: **Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans**. Ottawa: ON: Government of Canada; 2018.

113. Gibson E: **Wired Identities: Retention and Destruction of Personal Health Information in an Electronic World**. *The Dalhousie Law Journal* 2015, **385**.

114. Kasper DVS: **Privacy as a Social Good**. *Social Thought and Research* 2007, **28**:165-189.

115. Canadian Medical Association: **Principles for the Protection of Patients' Personal Health Information**. In. Ottawa, ON: Canadian Medical Associate; 2011.

116. Steeves J: **Data Protection and the Promotion of Health Research**. *Healthcare Policy* 2007, **2**(3):26-38.

117. Margulis ST: **Conceptions of Privacy: Current Status and Next Steps**. *Journal of Social Issues* 1977, **33**(3):5-21.

118. Glancy DJ: **The Invention of the Right to Privacy**. *Arizona Law Review* 1979 **21**(1):1-39.

119. Moore B: **Privacy: Studies in Social and Cultural History.** Armonk: M.E. Sharpe, Inc.; 1984.

120. Rocher L, Hendrickx JM, de Montjoye YA: **Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models**. *Nature Communications* 2019, **10**(1):3069.

121. Willison DJ: **Use of Data from Electronic Health Record for Health Research—Current Governance Challenges and Potential Approaches**: Office of the Privacy Commissioner of Canada; 2009.

122. Onyemelukwe C, Downie J: **The Tunnel At the End of the Light: A Critical Analysis of the Development of the Tri-Council Statement**. *Canadian Journal of Law and Society* 2011, **26**(1):159-176.

123. Canadian Institutes of Health Research, Natural Sciences and Engineering Reserach Council of Canada, Social Sciences and Humanities Reserach Council: **Tri-Council Policy Statement: Ethical Conduct for Reserach Involving Humans, 1998 (with 2000, 2002, and 2005 amendments)**. Ottawa, ON: Government of Canada; 2005.

124. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council: **Tri-Council Policy Statment: Ethical Conduct for Research Involving Humans**. Ottawa, ON: Government of Canada; 2010.

125. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council: **Tri-Council Policy Statement: Ethical Conduct for Reserach Involving Humans** Ottawa, ON: Government of Canada; 2014.

126. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council: **Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans**. Ottawa, ON: Government of Canada; 2022.

127. **TCPS 2 Interpretations: Scope** [https://ethics.gc.ca/eng/policy-politique_interpretations_scope-portee.html]

128. **Agreement on the Administration of Agency Grants and Awards by Research Institutions** [https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/institutional-agreement]

129. Canadian Institutes of Health Research: **CIHR Best Practices for Protecting Privacy in Health Research**. Ottawa: Public Works and Government Services Canada; 2005.

130. **Privacy Act, RSC 1985, c. P-21.**

131. **Personal Information Protection and Electronic Documents Act (PIPEDA), S.C. 2000, c. 5.**

132. **The Application of PIPEDA to Municipalities, Universities, Schools, and Hospitals** [https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/r_o_p/02_05_d_25/]

133. **The Constitution Act, 1867, 30 & 31 Vict, c. 3.**

134. **Non-Insured Health Benefits for First Nations and Inuit** [https://www.canada.ca/en/indigenous-services-canada/services/non-insured-health-benefits-first-nations-inuit.html]

135. **Non-Insured Health Benefits Program for First Nations and Inuit: Protecting Personal Privacy** [https://www.sac-isc.gc.ca/eng/1578072742668/1578072802513]

136. Gibson E: **Health Information Confidentiality and Access**. In: *Canadian Health Law and Policy* 4th edn. Edited by Downie J, Caulfield T, Flood CM. Markham, ON: LexisNexis Canada; 2011: 253-293.

137. **Provincial laws that may apply instead of PIPEDA** [https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/r_o_p/prov-pipeda/]

138. Government of Nunavut: **Government of Nunavut Response to the Report on the Review of the Information and Privacy Commissioner's 2008-2009 Annual Report** 2010.

139. **Access to Information and Protection of Privacy Act, SNWT (Nu) 1994, c. 20.**

140. **E-Health (Personal Health Information and Protection of Privacy) Act, SBC 2008, c. 38.**

141. **Data Stewardship Committee** [https://www2.gov.bc.ca/gov/content/health/conducting-health-research-evaluation/data-access-health-data-central/data-stewardship-committee]

142. **Freedom of Information and Protection of Privacy Act, RSBC 1996, c. 165.**

143. **An Act Respecting the Sharing of Certain Health Information, CQLR c. P-9.0001.**

144. **Act Respecting Documents Held by Public Bodies and the Protection of Personal Information, CQLR, c. A-21.**

145. **Health Information Act, RSA 2000, c. H-5.**

146. **Personal Health Information Act, CCSM 1997, c. 51.**

147. **Personal Health Information Protection Act, SO 2004, c. 3.**

148. **Personal Health Information Privacy and Access Act, SNB 2009, c. P-7.05.**

149. **Personal Health Information Act, SNS 2010, c. 41.**

150. **Health Information Act, RSPEI 2014, c. 31.**

151. **Personal Health Information Act, SNL 2008, c. P-7.01**

152. **Health Information Act, SNWT 2014, c. 2.**

153. **Health Information Privacy and Management Act, SY 2013, c. 16.**

154. **Health Information Protection Act, SS 1999, c. H-0.021.**

155. **Disclosure of personal information without consent for research purposes** [https://www.cai.gouv.qc.ca/communication-renseignements-personnels-sans-consentement-fins-recherche/]

156. Willison DJ: **Data Protection and the Promotion of Health Reserach: If the Laws Are Not a Problem, Then What Is?** *Healthcare Policy* 2007, **2**(3):39-43.

157. Willison D, Gibson E, McGrail K: **A Roadmap to Research Uses of Electronic Health Information**. In: *Data Data Everywhere: Access and Accountability?* , 1st edn. Edited by Flood C. Montreal and Kingston: McGill-Queen's University Press; 2011.

158. Pullman D, Buehler SK, Felt L, Gallagher K, House J, Keough TM, McDonald L, Power A, Ryan A: **Sorry, You Can't Have That Information: Data Holder Confusion Regarding Privacy Requirements for Personal Health Information and the Potential Chilling Effect on Health Research**. *Healthcare Policy* 2009, **4**(4):61-76.

159. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS: **A Systematic Review of Barriers to Data Sharing in Public Health**. *BMC Public Health* 2014, **14**.

160. Doiron D, Raina P, Fortier I: **Linking Canadian Population Health Data: Maximizing the Potential of Cohort and Administrative Data**. *Canadian Journal of Public Health* 2013, **104**(3):e258-261.

161. Hirtle M: **The Governance of Research Involving Human Participants in Canada**. *Health Law Journal* 2003, **11**:137-152.

162. Zimmerman SV: **Translating Ethics into Law: Duties of Care in Health Research Involving Humans**. *Health Law Review* 2005, **13**(2-3).

163. McDonald M: **Canadian Governance of Health Research Involving Human Subjects: Is Anybody Minding the Store?** *Health Law Journal* 2001, **9**:1-21.

164. The Academy of Medical Sciences: **Personal Data for Public Good: Using Health Information in Medical Research**; 2006.

165. Yiannakoulias N: **Understanding Identifiability in Secondary Health Data**. *Canadian Journal of Public Health* 2011, **102**(4):291-293.

166. Willison DJ, Emerson C, Szala-Meneok KV, Gibson E, Schwartz L, Weisbaum KM, Fournier F, Brazil K, Coughlin MD: **Access to Medical Records for Research Purposes: Varying Perceptions Across Research Ethics Boards**. *Journal of Medical Ethics* 2008, **34**(4):308-314.

167. Downie J, McDonald F: **Revisioning the Oversight of Research Involving Humans in Canada**. *Health Law Journal* 2004, **12**:159-181.

168. Kosseim P, Brady M: **Policy by Procrastination: Secondary Use of Electronic Health Records for Health Research Purposes**. *McGill Journal of Law and Health* 2008, **2**.

169. Cranley Glass K: **Questions and Challenges in the Governance of Reserach Involving Humans: A Canadian Perspective**. In: *Law and Ethics in Biomedical Reserach: Regulation, Conflict of Interest, and Liability.* edn. Edited by Lemmons T, Waring DR. Toronto: ON: University of Toronto Press; 2006.

170. Fost N, Levine RJ: **The Dysregulation of Human Subjects Research**. *Journal of the American Medical Association* 2007, **298**(18):2196-2198.

171. Hadskis M: **The Regulation of Biomedical Research in Canada**. In: *Canadian Health Law and Policy* 4th edn. Edited by Downie J, Caulfield T, Flood CM. Markham, ON: LexisNexis Canada; 2011: 437-499.

172. Gibson E, Brazil K, Coughlin M, Emerson C, Fournier F, Schwartz L, Szala-Meneok KV, Weisbaum KM, Willison DJ: **Who's Minding the Shop? The Role of Canadian Research Ethics Boards in the Creation and Uses of Registries and Biobanks**. *BMC Medical Ethics* 2008, **9**:17-17.

173. Creswell JW: **Qualitative Inquiry and Research Design: Choosing Among Five Approaches**. Thousand Oaks, CA: SAGE Publications; 2013.

174. Ormerod R: **The History and Ideas of Pragmatism**. *The Journal of the Operational Research Society* 2006, **57**(8):892-909.

175. Popkin RH, Stroll A: **Contemporary Philosophy**. In: *Philosophy Made Simple.* Second edn. New York, NY: Doubleday; 1993.

176. Menard L: **The Metaphysical Club**. New York, NY: Farrar, Straus, and Giroux; 2001.

177. Kaushik V, Walsh CA: **Pragmatism as a Research Paradigm and Its Implications for Social Work Research**. *Social Sciences* 2019, **8**(9):255.

178. Tashakkori A, Teddlie C: **Mixed Methodology: Combining Qualitative and Quantitative Approaches**. Thousand Oaks: SAGE Publications; 1998.

179. Morgan DL: **Pragmatism as a Paradigm for Social Research**. *Qualitative Inquiry* 2014, **20**(8):1045-1053.

180. Harrison H BM, Franklin R, Mills J: **Case Study Research: Foundations and Methodological Orientations**. 2017, **18**(1).

181. Yin R: **Case Study Research: Design and Methods**, 2nd ed edn. Newbury Park, CA: Sage; 1994.

182. Merriam SB: **Qualitative Research: A Guide to Design and Implementation (3rd ed.)**. San Francisco, Ca: Jossey-Bass; 2009.

183. Yin RK: **Case Study Research: Design and Methods**, 4th edn. Thousand Oaks, CA: Sage; 2009.

184. Stake R: **Multiple Case Study Analysis**. New York, NY: Guilford Press; 2006.

185. Stake R: **The Art of Case Study Research.** Thousand Oaks, CA: Sage; 1995.

186. Baxter P, Jack S: **Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers**. *The Qualitative Report* 2008, **13**(4):544-559.

187. Stake R: **Qualitative Case Studies**. In: *The Sage Handbook of Qualitatative Research (3rd ed).* edn. Edited by Denzin N LY. Thousand Oaks, CA: Sage; 2005.

188. Merriam SB: **Case Study Research and Case Study Applications in Education**. San Francisco, CA: Jossey-Bass; 1998.

189. Yin RK: **Enhancing the Quality of Case Studies in Health Services Research**. *Health Services Research* 1999, **35**(5):1209-1224.

190. Yin RK: **Applications of Case Study Research**, 2nd edn. Thousand Oaks, CA: SAGE Publications; 2003.

191. Eisenhardt KM, Graebner ME: **Theory Building from Cases: Opportunities and Challenges**. *Academy of Management Joumal* 2007, **50**(1):25-32.

192. Carroll J, Miller FA: **An Introduction to Qualitative Research**. In: *Research Guide.* edn. Edited by Lang ES, Harvey BJ, Frank JR: Royal College of Physicians and Surgeons Canada. ProQuest Ebook Central; 2011.

193. Glaser B, Strauss A: **Discovery of Grounded Theory: Strategies for Qualitative Work**. Chicago, IL: Aldine Transaction; 1967.

194. Patton MQ: **Qualitative Research and Evaluation Methods**, 3rd edn. Thousand Oaks, CA: SAGE Publications; 2002.

195. Payne G, Payne J: **Key Informants**. In: *Key Concepts in Social Research.* edn. London: SAGE Publications; 2004.

196. Rubin H, Rubin I: **Qualitative Interviewing: The Art of Hearing Data**. Thousand Oaks, CA: Sage Publications; 1995.

197. Eisenhardt KM: **Building Theories from Case Study Research**. *The Academy of Management Review* 1989, **14**(4):532-550.

198. Kendall J: **Axial Coding and the Grounded Theory Controversy**. *Western Journal of Nursing Research* 1999, **21**(6):743-757.

199. Stake R: **Case Studies**. In: *Handbook of Qualitative Research.* edn. Edited by Denzin NK, Lincoln YS. Thousand Oaks, CA: Sage Publications; 1994.

200. Strauss A, Corbin J: **Basics of Qualitative Research: Grounded Theory Procedures and Techniques**. Newbury Park, CA: Sage Publications; 1990.

201. Urquhart C: **Grounded Theory for Qualitative Research: A Practical Guide**. Thousand Oaks: Sage; 2013.

202. Given LM: **100 Questions (and Answers) About Qualitative Research**. Thousand Oaks: Sage; 2016.

203. Guba EG: **Criteria for Assessing Trustworthiness of Naturalistic Inquiries**. *Educational Communication and Technology* 1981, **29**(2):75-91.

204. Lincoln YS, Guba EG: **Naturalistic Inquiry**. Thousand Oaks, CA: Sage; 1985.

205. Ark TK, Kesselring S, Hills B, McGrail K: **Population Data BC: Supporting Population Data Science in British Columbia**. *International Journal of Populaion Data Science* 2020, **4**(2).

206. **ICES Contacts & Sites** [https://www.ices.on.ca/About-ICES/ICES-Contacts-and-Sites]

207. ICES: **2015/16 Annual Report**; 2016.

208. ICES: **2021/22 Annual Report**; 2022.

209. **Board of Directors** [https://www.ices.on.ca/About-ICES/Board-of-Directors]

210. **Scientific Advisory Committee** [https://www.ices.on.ca/About-ICES/Board-of-Directors/Scientific-Advisory-Committee]

211. Ishiguro L, Saskin R, Vermeulen MJ, Yates E, Gunraj N, Victor JC: **Increasing Access to Health Administrative Data with ICES Data & Analytic Services**. *Healthcare Quarterly* 2016, **19**(1):7-9.

212. **Ontario Regulation 329/04.**

213. Information and Privacy Commissioner of Ontario: **Manual for the Review and Approval of Prescribed Organizations**; 2021.

214. **Strategy for Patient-Oriented Research** [https://cihr-irsc.gc.ca/e/41204.html]

215. **Canadian Network for Observational Drug Effect Studies** [https://www.cnodes.ca/]

216. Information and Privacy Commissioner of Ontario: **Manual For The Review and Approval Of Prescribed Persons And Prescribed Entities**; 2010.

217. ICES: **Report to the Information and Privacy Commissioner of Ontario: Three-Year Review as a Prescribed Entity under PHIPA**; 2020.

218. The Office of the Ombudsperson: **Misfire: The 2012 Ministry of Health Employment Terminations and Related Matters**; 2017.

219. Lugg-Widger FV, Angel L, Cannings-John R, Hood K, Hughes K, Moody G, Robling M: **Challenges in Accessing Routinely Collected Data from Multiple Providers in the UK for Primary Studies: Managing the Morass**. *International Journal of Populaion Data Science* 2018, **3**(3):432.

220. Moore H, Blyth C: **Optimising the Use of Linked Administrative Data for Infectious Diseases Research in Australia**. *Public Health Research and Practice* 2018.

221. Higgins C, Matthews K: **Electronic Linkage and Interrogation of Administrative Health, Social Care, and Criminal Justice Datasets: Feasibility Concerning Process and Content**. *Informatics for Health and Social Care* 2020, **45**(4):444-460.

222. Bowen S, Botting I, Graham ID, MacLeod M, Moissac D, Harlos K, Leduc B, Ulrich C, Knox J: **Experience of Health Leadership in Partnering With University-Based Researchers in Canada - A Call to "Re-imagine" Research**. *International Journal of Health and Policy Management* 2019, **8**(12):684-699.

223. de Jong J, Hansen J, Groenewegen P: **Why Do We Need for Timeliness of Research in Decision-Making?** *European Journal of Public Health* 2019, **29**(Supplement_4).

224. McRae AD, Archambault P, Fok P, Wiemer H, Morrison LJ, Herder M: **Reducing Barriers to Accessing Administrative Data on SARS-CoV-2 Vaccination for Research**. *Canadian Medical Association Journal* 2022, **194**(27):E943.

225. Resnik DB: **Stewardship of Research Resources**. *Accountability in Research* 2019, **26**(3):246-251.

226. Shamoo AE, Resnick DB: **Responsible Conduct of Research**. New York, NY: Oxford University Press 2015.

227. Dahl LT, Katz A, McGrail K, Diverty B, Ethier J-F, Gavin F, McDonald JT, Paprica PA, Schull M, Walker JD *et al*: **The SPOR-Canadian Data Platform: A National Initiative to Facilitate Data Rich Multi-Jurisdictional Researcher**. *International Journal of Population Data Science* 2020, **5**(1).

228. Jones K, Heys S, Tingay KS, Jackson P, Dibben C: **The Good, the Bad, the Clunky: Addressing Challenges in Using Administrative Data for Research**. *International Journal of Populaion Data Science* 2019, **4**(1).

229. **Data literacy competencies** [https://www.statcan.gc.ca/en/wtc/data-literacy/compentencies]

230. **Data literacy training** [https://www.statcan.gc.ca/en/wtc/data-literacy]

231. **Education and training** [https://www.popdata.bc.ca/etu]

232. Teng J, Bentley C, Burgess MM, O'Doherty KC, McGrail K: **Sharing Linked Data Sets for Research: Results from a Deliberative Public Engagement Event in British Columbia, Canada**. *International Journal of Population Data Science* 2019, **4**(1):13.

233. Desai T, Ritchie F, Welpton R: **Five Safes: Designing Data Access for Research**: University of the West of England; 2016.

234. **What is the Five Safes framework?** [https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/]

235. **Eligibility and the Five SAFEs model** [https://www.popdata.bc.ca/data_access/requirements/eligibility]

236. Katz A, Enns J, Wong ST, Williamson T, Singer A, McGrail K, Bakal JA, Taylor C, Peterson S: **Challenges Associated with Cross-Jurisdictional Analyses Using Administrative Health Data and Primary Care Electronic Medical Records in Canada**. *International Journal of Population Data Science* 2018, **3**(3):437.

237. Cavallaro F, Lugg-Widger F, Cannings-John R, Harron K: **Reducing Barriers to Data Access for Research in the Public Interest—Lessons from Covid-19**. In: *thebmjopinion.* vol. 2023; 2020.

238. Moorthie S, Hayat S, Zhang Y, Parkin K, Philips V, Bale A, Duschinsky R, Ford T, Moore A: **Rapid Systematic Review to Identify Key Barriers to Access, Linkage, and Use of Local Authority Administrative Data for Population Health Research, Practice, and Policy in the United Kingdom**. *BMC Public Health* 2022, **22**(1):1263.

239. Marchessault G: **The Manitoba Centre for Health Policy: A Case Study**. *Healthcare Policy* 2011, **6**(Spec Issue):29-43.

240. Azzolini E, Ricciardi W, Gray M: **Healthcare Organizational Performance: Why Changing the Culture Really Matters. Commentary**. *Annali dell'Istituto Superiore de Sanita* 2018, **54**(1):6-8.

241. Scott T, Mannion R, Davies HTO, Marshall MN: **Implementing Culture Change in Health Care: Theory and Practice**. *International Journal for Quality in Health Care* 2003, **15**(2):111-118.

242. Ostroff C, Kinicki AJ, Tamkins MM: **Organizational Culture and Climate**. In: *Handbook of Psychology: Volume 12, Industrial and Organizational Psychology. Volume 12*, edn. Edited by Borman WC, Ilgen DR, Klimoski RJ. New York: Wiley; 2003: 565–587.

243. Flamholtz EF, Randle Y: **Corporate Culture, Business Models, Competitive Advantage, Strategic Assets and the Bottom Line**. *Journal of Human Resource Costing and Accounting* 2012, **16**:76-94.

244. Schein E: **The Corporate Culture Survival Guide: Sense and Nonsense about Culture Change**. San Francisco: John Wiley & Sons Inc; 1999.

245. Willis CD, Saul J, Bevan H, Scheirer MA, Best A, Greenhalgh T, Mannion R, Cornelissen E, Howland D, Jenkins E *et al*: **Sustaining Organizational Culture Change in Health Systems**. *Journal of Health Organization and Management* 2016, **30**(1):2-30.

246. Sürücü L, Yesilada T: **The Impact of Leadership Styles on Organizational Culture**. *International Journal of Business and Management Invention* 2017, **6**:2319-8028.

247. Davies C, Collins R: **Balancing Potential Risks and Benefits of Using Confidential Data**. *BMJ* 2006, **333**(7563):349-351.

248. Sexton A, Shepherd E, Duke-Williams O EA: **A Balance of Trust in the Use of Government Administrative Data**. *Archival Science* 2017, **17**(4):305-330.

249. Six B, Van Zimmeren E, Popa F, Frison C: **Trust and Social Capital in the Design and Evolution of Institutions for Collective Action**. *International Journal of the Commons* 2015, **9**(1).

250. Harding K, Lynch L, Porter J, Taylor NF: **Organisational Benefits of a Strong Research Culture in a Health Service: A Systematic Review**. *Australian Health Review* 2017, **41**(1):45-53.

251. Parkin L, Paul C: **Public Good, Personal Privacy: A Citizens' Deliberation About Using Medical Information for Pharmacoepidemiological Research**. *Journal of Epidemiology and Community Health* 2011, **65**(2):150-156.

252. Guttman A: **The SPOR Canadian Data Platform: Opportunity for Multi-Provincial Research**. *Canadian Medical Association Journal* 2019, **191**(40):E1091.

253. Canadian Institutes of Health Research: **International Review Panel Report 2005-2010**. Ottawa, ON: Canadian Institutes of Health Research; 2011.

254. **Data Dictionary** [https://datadictionary.ices.on.ca/Applications/DataDictionary/Default.aspx]

255. **Data Access Support Hub (DASH)** [https://www.hdrn.ca/en/dash]

256. **Moving Forward on a Pan-Canadian Health Data Strategy** [https://www.canada.ca/en/public-health/programs/pan-canadian-health-data-strategy.html]

257. **The Pan-Canadian Health Data Strategy: Expert Advisory Group Overview** [https://www.canada.ca/en/public-health/corporate/mandate/about-agency/external-advisory-bodies/list/pan-canadian-health-data-strategy-overview.html]

258. Expert Advisory Group: **Pan-Canadian Health Data Strategy: Toward a World-Class Health Data System**: Public Health Agency of Canada; 2022.

259. Milea D, Azmi S, Reginald P, Verpillat P, Francois C: **A Review of Accessibility of Administrative Healthcare Databases in the Asia-Pacific Region**. *Journal of Market Access and Health Policy* 2015, **3**.

260. Glaser BG, Holton J: **Remodeling Grounded Theory**. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research* 2004, **5**(2).

261. Kuper A, Lingard L, Levinson W: **Critically Appraising Qualitative Research**. *BMJ* 2008, **337**.

262. Locke EA: **The Case for Inductive Theory Building**. *Journal of Management* 2007, **33**(6):867-890.

263. Lavis JN, Boyko JA, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for Evidence-Informed Health Policymaking (STP) 14: Organising and Using Policy Dialogues to Support Evidence-Informed Policymaking**. *Health Research Policy and Systems* 2009, **7**(1):S14.

264. Boyko JA, Lavis JN, Abelson J, Dobbins M, Carter N: **Deliberative Dialogues as a Mechanism for Knowledge Translation and Exchange in Health Systems Decision-Making**. *Social Science and Medicine* 2012, **75**(11):1938-1945.

265. **McMaster Health Forum: Stakeholder Dialogues** [www.mcmasterforum.org/spark-action/stakeholder-dialogues]

266. Jones KH, Laurie G, Stevens L, Dobbs C, Ford DV, Lea N: **The Other Side of the Coin: Harm Due to the Non-Use of Health-Related Data**. *International Journal of Medical Informatics* 2017, **97**:43-51.

267. **Personal Information Protection Act, SBC 2003, c. 63.**

268. **Freedom of Information and Protection of Privacy Act, RSA 2000, c. F-25.**

269. **Freedom of Information and Protection of Privacy Act, SS 1990-91, c. F-22.01.**

270. **Freedom of Information and Protection of Privacy Act, CCSM, c. F175.**

271. **Personal Information Protection and Identity Theft Prevention Act , SM 2013, c. 17.**

272. **Freedom of Information and Protection of Privacy Act, RSO 1990, c. F. 31.**

273. **Act Respecting the Protection of Personal Information in the Private Sector, RSQ, c. P-39.1.**

274. **Act Respecting the Sharing of Certain Health Information, CQLR, c. P-9.001.**

275. **Right to Information and Protection of Privacy Act, SNB 2009, c. R-10.6.**

276. **Freedom of Information and Protection of Privacy Act, SNS 1993, c. 5.**

277. **Personal Information International Disclosure Protection Act, SNS 2006, c. 3.**

278. **Freedom of Information and Protection of Privacy Act, RSPEI 198, c. F-15.01.**

279. **Access to Information and Protection of Privacy Act, SNL 2015, c. A-1.2.**

280. **Access to Information and Protection of Privacy Act, RSY 2002, c. 1.**

281. **Access to Information and Protection of Privacy Act, SNWT 1994, c 20.**

282. Government of Canada: **Canada Gazette: Part 1**. In., vol. 136. Ottawa, ON: Canadian Government Publishing; 2002.

283. **Data & Analytic Services** [https://www.ices.on.ca/DAS]

APPENDIX A: Copyright permissions

May 4, 2023

Geoff Porter
Department of Surgery
Nova Scotia Health/Dalhousie University
Halifax, Nova Scotia
B3H 2Y9

Dear Dr. Porter,

I am preparing my doctoral thesis for submission to the Faculty of Graduate Studies at Dalhousie University, Halifax, Nova Scotia, Canada. I am seeking your permission to incorporate content from a manuscript version of the following paper into the thesis document:

Kendell C, Levy AR, Porter G, Gibson E, Urquhart R. Factors affecting access to administrative health data for research in Canada: a study protocol. *International Journal of Population Data Science* 2021; 6(1).

Dalhousie graduate theses are collected and stored online by Dalhousie University and Library and Archives of Canada. I am seeking your permission for the material described above to be stored online in Dalhousie University's institutional repository and in Library and Archives of Canada (LAC)'s Theses Canada Collection.

Full publication details and a copy of this permission letter will be included in the thesis.

Yours sincerely,

Cynthia Kendell

Permission is granted for:

a) the inclusion of the material described above in your thesis.

b) for the material described above to be included in the copy of your thesis that is sent to the Library and Archives of Canada inclusion in Theses Canada.

c) For the material described above to be included in the copy of your thesis that is sent to Dalhousie University's institutional repository.

Name: Geoff Porter Title: May 5, 2023

Signature: Professor of Surgery Date:

May 4, 2023

Dr. Robin Urquhart
Department of Community Health and Epidemiology
Dalhousie University
Halifax, Nova Scotia
B3H 1V7

Dear Dr. Urquhart,

I am preparing my doctoral thesis for submission to the Faculty of Graduate Studies at Dalhousie University, Halifax, Nova Scotia, Canada. I am seeking your permission to incorporate content from a manuscript version of the following paper into the thesis document:

Kendell C, Levy AR, Porter G, Gibson E, Urquhart R. Factors affecting access to administrative health data for research in Canada: a study protocol. *International Journal of Population Data Science* 2021; 6(1).

Dalhousie graduate theses are collected and stored online by Dalhousie University and Library and Archives of Canada. I am seeking your permission for the material described above to be stored online in Dalhousie University's institutional repository and in Library and Archives of Canada (LAC)'s Theses Canada Collection.

Full publication details and a copy of this permission letter will be included in the thesis.

Yours sincerely,

Cynthia Kendell

Permission is granted for:

a) the inclusion of the material described above in your thesis.

b) **for the material described above to be included in the copy of your thesis that is sent to the Library and Archives of Canada inclusion in Theses Canada.**

c) **For the material described above to be included in the copy of your thesis that is sent to Dalhousie University's institutional repository.**

Name: Robin Urquhart Title: Associate Professor

Signature: Date: May 5 2023

May 4, 2023

Dr. Adrian Levy
Department of Community Health and Epidemiology
Dalhousie University
Halifax, Nova Scotia
B3H 1V7

Dear Dr. Levy,

I am preparing my doctoral thesis for submission to the Faculty of Graduate Studies at Dalhousie University, Halifax, Nova Scotia, Canada. I am seeking your permission to incorporate content from a manuscript version of the following paper into the thesis document:

Kendell C, Levy AR, Porter G, Gibson E, Urquhart R. Factors affecting access to administrative health data for research in Canada: a study protocol. *International Journal of Population Data Science* 2021; 6(1).

Dalhousie graduate theses are collected and stored online by Dalhousie University and Library and Archives of Canada. I am seeking your permission for the material described above to be stored online in Dalhousie University's institutional repository and in Library and Archives of Canada (LAC)'s Theses Canada Collection.

Full publication details and a copy of this permission letter will be included in the thesis.

Yours sincerely,

Cynthia Kendell

Permission is granted for:

a) the inclusion of the material described above in your thesis.

b) **for the material described above to be included in the copy of your thesis that is sent to the Library and Archives of Canada inclusion in Theses Canada.**

c) **For the material described above to be included in the copy of your thesis that is sent to Dalhousie University's institutional repository.**

Name: Adrian Levy Title: Professor

Signature: Date: 05-May-2023

May 4, 2023

Elaine Gibson
Schulich School of Law
Health Law Institute
Dalhousie University
Halifax, Nova Scotia
B3H 4R2

Dear Prof. Gibson,

I am preparing my doctoral thesis for submission to the Faculty of Graduate Studies at Dalhousie University, Halifax, Nova Scotia, Canada. I am seeking your permission to incorporate content from a manuscript version of the following paper into the thesis document:

Kendell C, Levy AR, Porter G, Gibson E, Urquhart R. Factors affecting access to administrative health data for research in Canada: a study protocol. *International Journal of Population Data Science* 2021; 6(1).

Dalhousie graduate theses are collected and stored online by Dalhousie University and Library and Archives of Canada. I am seeking your permission for the material described above to be stored online in Dalhousie University's institutional repository and in Library and Archives of Canada (LAC)'s Theses Canada Collection.

Full publication details and a copy of this permission letter will be included in the thesis.

Yours sincerely,

Cynthia Kendell

Permission is granted for:

a) the inclusion of the material described above in your thesis.

b) for the material described above to be included in the copy of your thesis that is sent to the Library and Archives of Canada inclusion in Theses Canada.

c) For the material described above to be included in the copy of your thesis that is sent to Dalhousie University's institutional repository.

Name: Elaine Gibson Title: Professor of Law

Signature: Date: May 5, 2023

APPENDIX B: Legislation involved in the regulation of personal health information in Canada

Jurisdiction	Information Legislation	Sector
Canada	Privacy Act [130]	Public
	Personal Information Protection and Electronic Documents Act [131]	Private
British Columbia	Freedom of Information and Protection of Privacy Act [142]	Public
	Personal Information Protection Act [267]	Private*
	E-Health (Personal Health Information and Protection of Privacy) Act [140]	Health
Alberta	Freedom of Information and Protection of Privacy Act [268]	Public
	Health Information Act [145]	Health
Saskatchewan	Freedom of Information and Protection of Privacy Act [269]	Public
	Health Information Protection Act [154]	Health
Manitoba	Freedom of Information and Protection of Privacy Act [270]	Public
	Personal Information Protection and Identity Theft Prevention Act [271]	Private
	Personal Health Information Act [146]	Health
Ontario	Freedom of Information and Protection of Privacy Act [272]	Public
	Personal Health Information Protection Act [147]	Health*
Quebec	Act Respecting Documents Held by Public Bodies and the Protection of Personal Information [144]	Public
	Act Respecting the Protection of Personal Information in the Private Sector [273]	Private*
	Act Respecting the Sharing of Certain Health Information [274]	Health
New Brunswick	Right to Information and Protection of Privacy Act [275]	Public
	Personal Health Information Privacy and Access Act [148]	Health*
Nova Scotia	Freedom of Information and Protection of Privacy Act [276]	Public
	Personal Information International Disclosure Protection Act [277]	Public
	Personal Health Information Act [149]	Health*
Prince Edward Island	Freedom of Information and Protection of Privacy Act [278]	Public
	Health Information Act [150]	Health
Newfoundland and Labrador	Access to Information and Protection of Privacy Act [279]	Public
	Personal Health Information Act [151]	Health*
Yukon	Access to Information and Protection of Privacy Act [280]	Public
	Health Information Privacy and Management Act [153]	Health
Northwest Territories	Access to Information and Protection of Privacy [281]	Public
	Health Information Act [152]	Health
Nunavut	Access to Information and Protection of Privacy Act [139]	Public

*Legislation has been declared substantially similar to PIPEDA [282].

APPENDIX C: Comparison of how "identifiable" information is addressed in provincial/territorial legislation

Province/ Territory	Legislation	Terms used	Definition
AB	Health Information Act [145]	individually identifying	identity of the individual can be readily ascertained from the information
		non-identifying	identity of the individual cannot be readily ascertained from the information
MB	Personal Health Information Act [146]	identifiable	not defined
SK	Health Information Protection Act [154]	de-identified	any information that may reasonably be expected to identify an individual has been removed
ON	Personal Health Information and Protection of Privacy Act [147]	identifying	information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify an individual
		de-identified	information that identifies the individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify the individual has been removed
NB	Personal Health Information Privacy and Access Act [148]	identifying	information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify an individual
		de-identified	information from which all identifying information has been removed
NS	Personal Health Information Act [149]	identifying	information that identifies an individual or, where it is reasonably foreseeable in the circumstances, could be utilized, either alone or with other information, to identify an individual
		de-identified	information that has had all identifiers removed that identify the individual, or where it is reasonably foreseeable in the

Province/ Territory	Legislation	Terms used	Definition
			circumstances, could be utilized, either alone or with other information, to identify the individual
PEI	Health Information Act [150]	identifying	information that identifies an individual or which it is reasonably foreseeable in the circumstances could be utilized, either alone or with other information, to identify an individual
		de-identified	information that has been stripped, encoded or otherwise transformed so as to ensure that the identity of the individual who was the subject of the personal health information cannot be readily ascertained from the de-identified information
NL	Personal Health Information Act [151]	identifying	information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or together with other information, to identify an individual
		non-identifying	not defined
YK	Health Information Privacy and Management Act [153]	identifying	information that identifies the individual or that it is reasonable to believe could be used, either alone or with other information, to identify the individual
		non-identifying	information that is not identifying
NWT	Health Information Act [152]	identifies an individual	not defined
		non-identifying	not defined

Abbreviations: AB=Alberta; MB=Manitoba; SK=Saskatchewan; ON=Ontario; NB= New Brunswick; NS=Nova Scotia; PEI=Prince Edward Island; NL=Newfoundland; YK=Yukon; NWT=Northwest Territories.

APPENDIX D: Comparison of case study approaches

	Stake [184, 185, 187]	**Merriam** [182, 188]	**Yin** [181, 183, 189, 190]
When use of case study is appropriate	▪ the unit of study is case	▪ the unit of study is case	▪ the research question is "how" or "why" ▪ the researcher has no control over behavioral events ▪ the focus is on contemporary events rather than historical
Definition of case study	▪ the study of the particularity and complexity of a single case, coming to understand its activity within important circumstances	▪ an in-depth description and analysis of a bounded system	▪ empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident
Definition of case	▪ a complex, function thing or integrated system ▪ e.g. individuals, programs, or organizations	▪ bounded system ▪ e.g. individuals, institutions, and policies	▪ ranges from concrete to abstract ▪ e.g., people, programs, communities, processes, relationships
Questions addressed	▪ "issue questions"	▪ not limited to a particular type	▪ "how" and "why" questions
Types of case studies	▪ intrinsic ▪ instrumental ▪ collective	▪ intrinsic ▪ instrumental ▪ historical ▪ observational ▪ life histories ▪ multi-site	▪ single holistic ▪ singe embedded ▪ multiple holistic ▪ multiple embedded
Case selection	▪ based on maximizing the knowledge gained	▪ based on maximizing the knowledge gained	▪ individual cases chosen based on what will yield the best data

	Stake [184, 185, 187]	**Merriam** [182, 188]	**Yin** [181, 183, 189, 190]
			▪ in multiple case study, cases chosen based on replication logic
Data collection methods	▪ multiple methods ▪ prefers interviews, observation, and document review ▪ other methods may be used	▪ any and all methods can be used ▪ identifies interviews, observation, and document review as particularly useful for qualitative research	▪ identifies six methods that are commonly employed in case study: direct observation, participant observation, interviews, documents, archival records, and physical artifacts
Analytic approach	▪ interpretive approach ▪ relies on direct interpretation and categorical aggregation ▪ may involving coding and correspondence tables ▪ views analysis as an art rather than science	▪ interpretive approach ▪ views analysis as inductive and comparative ▪ recommends a systematic approach involving coding and categorization ▪ acknowledges other approaches may be used	▪ structured, detailed, and informed by experimental design ▪ identifies case-study specific analytic strategies and techniques ▪ analysis focuses on testing and revising theoretical prepositions (akin to hypotheses) ▪ focus is on examining "variables" ▪ emphasizes replication logic in multiple case studies

APPENDIX E: Study summary

Study Summary

Factors Affecting Access to Administrative Health Data for Research: A Multiple Case Study of Three Canadian Provinces

Cynthia Kendell[1,2], Robin Urquhart[1,3], Geoff Porter[1,3], Adrian Levy[4], Elaine Gibson[5]

1. Department of Surgery, Nova Scotia Health Authority; 2. Interdisciplinary PhD Program, Dalhousie University; 3. Department of Surgery, Dalhousie University; 4. Department of Community Health and Epidemiology, Dalhousie University; 5. Schulich School of Law, Dalhousie University

Background

In recent years, administrative health data have become widely recognized as an invaluable source of information and increasingly used for secondary purposes, including use in health research. Although the secondary use of administrative health data for research is permitted under the current regulatory framework (i.e., research ethics and health information legislation), researchers across Canada are reportedly experiencing challenges when attempting to gain access to administrative health data for research, including delayed data access. In addition, substantial inter-provincial variations in the timeliness of data access across provinces have been reported.

About the study

This study is guided by two main questions:
1) What are the factors affecting access to administrative health data for research purposes in Canada?
2) How do these vary across provinces? Why?

To answer these questions, the research team is embarking on an in-depth examination of access to administrative health data for research purposes, focusing on Nova Scotia (NS), Ontario (ON), and British Columbia (BC). In each province, we will be conducting one-on-one interviews with researchers (including research staff) and individuals who are involved in the regulation and oversight of access to administrative health data for research purposes.

How will the findings be used?

The findings of this study will be used to promote timely and equitable access to administrative health data for research purposes in Canada, while taking into consideration the need for adequate informational privacy protections for patients. For example, the findings of this study may be used to inform the development of evidence-based data sharing principles or best practices specific to the use of administrative health data for research, harmonization of existing data access policies, streamlining of data access processes, and/or development of supports for researchers and regulatory/oversight bodies to facilitate access to administrative health data for research.

APPENDIX F: Consent form

Informed Consent Form Non-Interventional Study

STUDY TITLE:	Factors Affecting Access to Administrative Health Data Research: A Multiple Case Study of Three Canadian Provinces
PRINCIPAL INVESTIGATOR:	Cynthia Kendell Department of Surgery, QEII Health Sciences Centre, VG Site, Room 825 Victoria Building, Halifax, Nova Scotia, B3H 2Y9 (902) 473-7501 cynthia.kendell@ccns.nshealth.ca

1. Introduction

You have been invited to take part in a research study. A research study is a way of gathering information on a treatment, procedure or medical device or to answer a question about something that is not well understood. Taking part in this study is voluntary. It is up to you to decide whether to be in the study or not. Before you decide, you need to understand what the study is for, what risks you might take and what benefits you might receive. This consent form explains the study.

Please ask the research team to clarify anything you do not understand or would like to know more about. Make sure all your questions are answered to your satisfaction before deciding whether to participate in this research study.

The researchers will:

- Discuss the study with you
- Answer your questions
- Be available during the study to deal with problems and answer questions

This study will examine the factors impacting access to administrative health data for research purposes in three Canadian provinces. You are being asked to consider participating in this study because you have been identified as belonging to one of the following groups:

1) Academic researchers and research staff who have accessed or attempted to access administrative health data held by a provincial health data repository for research purposes in the last 5 years.
2) Individuals involved in the regulation and oversight of access to administrative health data for research purposes (e.g., individuals who are members of relevant data access committees, data custodians, privacy review bodies, and research ethics boards).

If you decide not to take part or if you leave the study early, there will be no consequence to you.

2. Why is there a need for this study?

A wealth of administrative health data are collected in Canada and stored in electronic databases. In recent years, these data have become widely recognized as an invaluable source of information and increasingly used for secondary purposes, including use in health research. Although the secondary use of administrative health data for research is permitted under the current regulatory framework (i.e., research ethics and health information legislation), a number of researchers and academics have claimed that researchers are not able to readily access administrative health data for research purposes.

While this has not been examined in depth, there is some evidence that researchers in some jurisdictions are experiencing delays when attempting to access administrative health data for research purposes, and that the time it takes for researchers to access administrative health data for research purposes varies widely across Canadian provinces/territories (i.e., from 1 to 18 months). This suggests that some researchers in Canada may be experiencing barriers to accessing administrative health data for research. At this time, the reasons for delays and inter-provincial variations in access to administrative health data for research purposes are not well understood.

This study will examine access to administrative health data in three Canadian provinces: Nova Scotia, Ontario, and British Columbia. The overall aims of this study are to identify the factors affecting access to administrative health data for research, how these vary across provinces, and why.

The findings of this study will be used to promote timely and equitable access to administrative health data for research purposes in Canada, while taking into consideration the need for adequate informational privacy protections for patients. For example, the findings of this study may be used to inform the development of data sharing principles or privacy best practices specific to the use of administrative health data for research, the evaluation of existing data access policies and processes, and/or the development of interventions to improve awareness of privacy issues related use of administrative health data for research among researchers and relevant oversight bodies.

3. How Long Will I Be In The Study?

The length of this study for participants is approximately one (1) hour. The entire study is expected to take approximately 1.5 years to complete and the results should be known in approximately 2 years.

4. How Many People Will Take Part In This Study?

It is anticipated that about 60 people will participate in this study throughout the three provinces included in this study (Nova Scotia, Ontario, and British Columbia). About 20 people will participate in this study in Nova Scotia.

5. How Is The Study Being Done?

There are two research questions guiding this study:

1) What are the factors affecting access to administrative health data for research purposes in Canada?
2) How do these vary across provinces? Why?

To help address these questions, we will be carrying out a qualitative study examining access to administrative health data in three Canadian provinces: Nova Scotia, Ontario, and British Columbia. Specifically, we will be carrying out interviews with researchers (including research staff) and individuals who are involved regulating access to administrative health data for research purposes.

Participation in this study will involve a one-time interview with the principal investigator. Interviews will be carried out in-person whenever possible. When in-person interviews cannot be arranged, they may be carried out over the phone. Each interview is expected to take approximately one (1) hour.

6. What Will Happen If I Take Part In This Study?

If you decide to participate in this study, the principal investigator will work with you to arrange a time and place to do the interview that is convenient for you. If an in-person interview is not possible, a telephone interview may be arranged.

At the scheduled interview time, the principal investigator will either meet you in person at the agreed upon location, or call you at the telephone number you have provided. Prior to commencing the interview, the researcher will discuss the study with you, respond to any questions or concerns you may have, and confirm that you are still interested in participating. Interviews will be audio-recorded.

If you are a researcher or research staff, you will be asked to describe the processes and policies required to access administrative health data for research purposes in your province. You will also be asked questions about your experiences accessing these data (e.g., ease of obtaining access, timeliness of access, barriers and facilitators, etc.).

If you are and individual who is involved regulating access to administrative health data for research purposes, you will be asked to describe the processes and policies required to access

administrative health data for research purposes in your province. You will also be asked questions about your perspectives on the use of administrative health data for research, and on the data access policies and processes in place.

During the interview you may decline to respond to any questions that you do not want to answer. You are free to withdraw from this study at any point. If you wish to withdraw from the study, you need only inform the principal investigator that you no longer wish to participate. If you decide to withdraw from the study during the interview, any responses provided up to that point will be included in the study.

In reports, presentations, or published manuscripts resulting from this work, direct quotes from individuals may be used to illustrate certain issues, or concepts. Importantly, direct quotes will not be attributed to individual participants and every effort will be made to de-identify the quote in such a way that it cannot be linked to an individual participant.

7. Are There Risks To The Study?

While it is not possible to anticipate all harms that may occur as a result of participating in this study, there may be risks associated with participation.

As with all research, there is a chance that privacy and confidentiality could be compromised; however, we are taking precautions to minimize this risk. These precautions are described in detail in Section 15.

Given the research topic, it is unlikely that during the interview you will be asked questions that are upsetting or distressing, however, you may not like all of the questions that you will be asked. You may decline to answer any questions that you would prefer not to answer.

8. Are There Benefits Of Participating In This Study?

There are no direct benefits to participating in this study.

9. What Happens at the End of the Study?

It is anticipated that the results of this study will be published and or presented in a variety of forums. In any publication and/or presentation, information will be provided in such a way that you cannot be identified, except with your express permission. Publications and reports resulting from this study will be made publicly available.

10. What Are My Responsibilities?

As a study participant you will be expected to:

- Follow the directions of the principal investigator;
- Communicate with the principal investigator if there are changes in your schedule that require you to reschedule your interview;
- Inform the individual conducting the interview if, during the interview, you decide that you wish to withdraw from the study;
- Report any problems that you experience that you think might be related to participating in the study.

11. Can My Participation in this Study End Early?

Yes. If you chose to participate and later change your mind, you can say no and stop the research at any time. If you wish to withdraw your consent please inform the research team. If you choose to withdraw from this study, your decision will have no ramifications. If you decide to withdraw from the study during the interview, any responses provided up to that point will be included in the study.

Also, the Nova Scotia Health Authority Research Ethics Board and the principal investigator have the right to stop patient recruitment or cancel the study at any time.

Lastly, the principal investigator may decide to remove you from this study without your consent for any of the following reasons:

- You do not follow the directions of the research team;
- It has been determined that you do not meet study criteria.

If you are withdrawn from this study, the principal investigator will inform you and discuss the reasons with you.

If you decide that you do not want to participate in the study, please contact the principal investigator directly, via telephone at (902) 473-7501, or via email at cynthia.kendell@ccns.nshealth.ca. If you decide during the interview that you no longer wish to continue, please inform the individual conducting the interview.

12. What Will Happen To My Information After The Study Is Over?

Once the study is over, the information you provided will be securely stored by the principal investigator. All study materials will be kept in a locked cabinet in the principal investigators' office at the QEII Health Sciences Center, VG site. Files will be retained for seven years after study completion, at which point all paper and electronic documents will be securely shredded and information stored on CDs or DVDs will be manually destroyed.

13. What About New Information?

You will be told about any other new information that might affect your willingness to stay in the study and will be asked whether you wish to continue taking part in the study or not.

14. Will It Cost Me Anything?

This study will not result in any out-of-pocket expenses to you. You will not be compensated for your participation in this study.

Research Related Injury

If you become ill or injured as a direct result of participating in this study, necessary medical treatment will be available at no additional cost to you. Your signature on this form only indicates that you have understood to your satisfaction the information regarding your participation in the study and agree to participate as a subject. In no way does this waive your legal rights nor release the principal investigator, the research staff, the study sponsor or involved institutions from their legal and professional responsibilities.

15. What About My Privacy and Confidentiality?

Protecting your privacy is an important part of this study. Every effort to protect your privacy will be made. If the results of this study are presented to the public, nobody will be able to tell that you were in the study. However, complete privacy cannot be guaranteed. For example, the principal investigator may be required by law to allow access to research records.

If you decide to participate in this study, the research team will look at your personal information and collect only the information they need for this study.

- Name,
- Email address,
- Telephone number,
- Position/title and institutional affiliations,
- Information from the study interviews and questionnaires.

Storage of Study Information

- All interviews will be audio-recorded and transcribed. Transcripts will provide the research team with a verbatim account of the interview, which will assist with data analysis. Transcription will be carried out by an experienced and authorized transcriptionist who compliant with Nova Scotia Health Authority and Dalhousie privacy and confidentiality policies and protocols.
- Your name will not appear on any study documents or files. Instead, it will be replaced with a unique alphanumeric identifier (e.g., NS Researcher #1, BC REB Member #3).

- Audio recordings and transcripts will be kept in a password protected folder on a password protected computer. Any hard copies of study materials will be securely stored in a locked cabinet in the PI's office at the QEII Health Sciences Center, VG site.

Access to Records

A transcriptionist will be granted temporary access to a copy of the audio-files of interviews for the sole purpose of transcription.

Other people may need to look at your personal information to check that the information collected for the study is correct and to make sure the study followed the required laws and guidelines. These people might include:

- The Nova Scotia Health Authority Research Ethics Board (NSHA REB) and people working for or with the NSHA REB because they oversee the ethical conduct of research studies within the Nova Scotia Health Authority.

Other than these two exceptions, only the research team will have access to your personal information and interview responses.

Use of Your Study Information

The information you provide during your participation in this study will be combined with the information provided by other participants for analysis. Aggregate results will be included in publications, presentations, and reports resulting from this study. In certain instances, direct quotes may be used to help illustrate study findings, however, if this occurs, quotes will not be attributed to an identifiable using direct identifiers. Quotes will be de-identified to the maximum extent possible to protect the identity of participants, and will be attributed to individuals using an assigned alphanumeric identifier (e.g., NS Researcher #1, BC REB Member #3).

The research team will keep the information they see or receive about you confidential, to the extent permitted by applicable laws. Even though the risk of identifying you from the study data is very small, it can never be completely eliminated.

The research team will keep any personal information about you in a secure and confidential location for and then destroy it according to NSHA policy. Your personal information will not be shared with others without your permission.

You have the right to be informed of the results of this study once the entire study is complete.

The REB and people working for or with the REB may also contact you personally for quality assurance purposes.

Your Access to Records

You have the right to access, review, and request changes to your study data.

16. Declaration of Financial Interest

This study is unfunded. The PI has no vested financial interest in conducting this study.

17. What About Questions or Problems?

For further information about the study you may contact the principal investigator, who is the person in charge of this study.

The principal investigator is Cynthia Kendell.
Telephone: (902) 473-7501
Email: cynthia.kendell@ccns.nshealth.ca

18. What Are My Rights?

You have the right to all information that could help you make a decision about participating in this study. You also have the right to ask questions about this study and your rights as a research participant, and to have them answered to your satisfaction before you make any decision. You also have the right to ask questions and to receive answers throughout this study. You have the right to withdraw your consent at any time.

If you have questions about your rights as a research participant, and/or concerns or complaints about this research study, you can contact the Nova Scotia Health Authority Research Ethics Board manager at 902-473-8426 or Patient Relations at (902) 473-2133 or 1-855-799-0990 or healthcareexperience@nshealth.ca.

In the next part you will be asked if you agree (consent) to join this study. If the answer is "yes", please sign the form.

19. Consent Form Signature Page

I have reviewed all of the information in this consent form related to the study called:

"Factors Affecting Access to Administrative Health Data for Research: A Multiple Case Study of Three Canadian Provinces"

I have been given the opportunity to discuss this study. All of my questions have been answered to my satisfaction.

This signature on this consent form means that I agree to take part in this study. I understand that I am free to withdraw at any time without experiencing negative consequences as a result.

E-messaging (email and texting) can be used by a member or members of the research team to communicate with you while you are in this study. All communication done with you will be done only through a NSHA Webmail account, or text by a phone issued to a research member through NSHA. All efforts are made to keep information sent or received private, but it is possible other people may be able to see, read, and change messages sent to or from NSHA.

☐ I give my permission to be contacted by a member or members of the research team from an NSHA Webmail account or an NSHA cell phone by research staff to communicate during this study. _____ (initials and date).

Email	yes ☐	no ☐
Text message	yes ☐	no ☐

☐ I do not wish to be contacted by email or text message, unless I otherwise give permission at another time during this study _____ (initial and date).

☐ Not applicable.

_____	_____	____ / _____ / ___
Signature of Participant	Name (Printed)	Year Month Day*
_____	_____	____ / _____ / ___
Signature of Person Conducting Consent Discussion	Name (Printed)	Year Month Day*
_____	_____	____ / _____ / ___
Signature of Principal Investigator	Name (Printed)	Year Month Day*

Note: Please fill in the dates personally

I will be given a signed copy of this consent form.

APPENDIX G: Interview guide for researchers and research staff

Participant Background

1. Describe your role (e.g., researcher, research staff, graduate student/trainee, other).
2. How long have you been in this role?
3. Briefly describe the research involving administrative heath data that you have been directly involved with (e.g., content area, databases used, single/multi-jurisdictional, etc.)
4. What is your involvement in the data access process?

Facts Related to the Case

5. What is involved in accessing administrative health data for research via [repository X]?
   -Required reviews and approvals
   -Application preparation and submission
   -Relevant policies and legislation
   -Data-related costs
6. Do these things change if you are linking to external datasets? Doing a multi-province study? If so, how?

Experiences Accessing Data

7. How would you describe your experiences accessing administrative health data for research purposes via [repository x]?
   -Challenges
   -Barriers and enablers
8. How long does it take to get access to data to your study data?
   -Is that satisfactory?
   -Why do you think it takes that amount of time?

Specific Factors Affecting Data Access/Timely Data Access

9. In your opinion, what are the main factors affecting researchers' ability to access administrative health data for research purpose in [province x]?
   -Resources
   -Supports available to assist with accessing data
   -Provincial information legislation
   -Your personal knowledge and experience
10. Regarding the overall process of accessing administrative health data for research in [province X]:
    -Is there a clear pathway?
    -Is the process transparent?
    -Are the oversight/approval bodies involved responsive?
    -Are there "bottlenecks" in the process? If so, where?
11. Does the feedback provided to you during the data access process reflect an understanding of the research and related risks? If not, please explain.
12. Are there certain databases or types of data that are harder to get access to than others? If so, please explain.

Closing

13. Do you have any final comments about accessing administrative health data for research purposes in [province x] that you would like share?

APPENDIX H: Interview guide for regulatory stakeholders

Participant Background

1. Relevant to data access, how would you describe your current role (e.g., data custodian, REB member, privacy committee member, other)?
   -What are your main responsibilities?
   -What is the main aim of objective of your role?
2. How long have you been in this role?
3. Please describe your background and how you came be in this role.

Note: All subsequent questions will be adapted based on the individual's role.

Facts related to the case

4. For researchers who require access to administrative health data for research purposes
   -What are the required reviews and approvals relevant to [regulatory body x]?
   -What happens once the researcher submits an application to [regulatory body x]?
   -What happens once to [regulatory body x] decides to approve or reject an application?

Perspectives on the use and regulation of administrative health data for research

5. What are your thoughts on the use of administrative health data for research?
6. What are your thoughts on the processes involved in gaining access to administrative health data for research?
   - What is the workload involved for researchers? Those in regulatory/oversight roles?
   - Can current processes be streamlined?
7. What are your thoughts on the policies that are in place?
   -Are they documented and accessible to you?
   -Are the requirements clear?

Factors affecting researchers' access to administrative health data for research

8. What things impact your decision/the decision of [regulatory body x] to approve an application?
9. Once approval has been granted, what are some things that impact the time it takes for the researcher get access to the requested dataset?
10. What do you think are the main factors that impact researchers' ability to access administrative health data for research?
11. What do you think are the main factors impacting the timeliness with which researchers are able to access administrative health data?

Regulatory Role

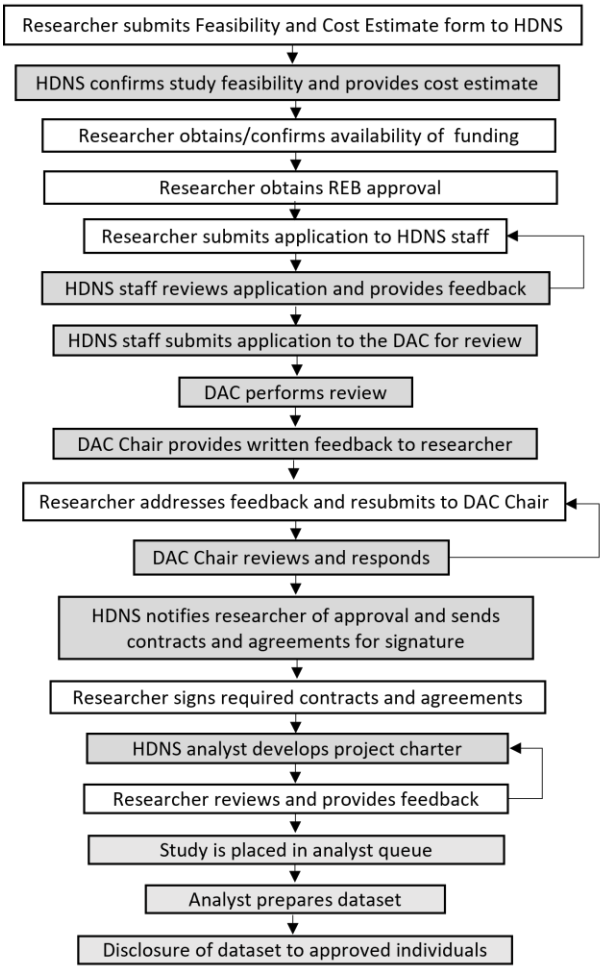
12. How would you describe your level of knowledge/expertise with regard to:
    -The methodological aspects of research involving administrative health data?
    -Ethical/legislative requirements related to disclosures of health information for research?
    -Issues related to privacy and confidentiality?
13. What organization/institutional supports are currently available to you to support you in your role?
14. What additional supports would assist you in your role?

Closing

15. Do you have any final comments about accessing administrative health data for research purposes in [province x] that you would like share?

APPENDIX I: Data access processes

Process at Health Data Nova Scotia:



Key:

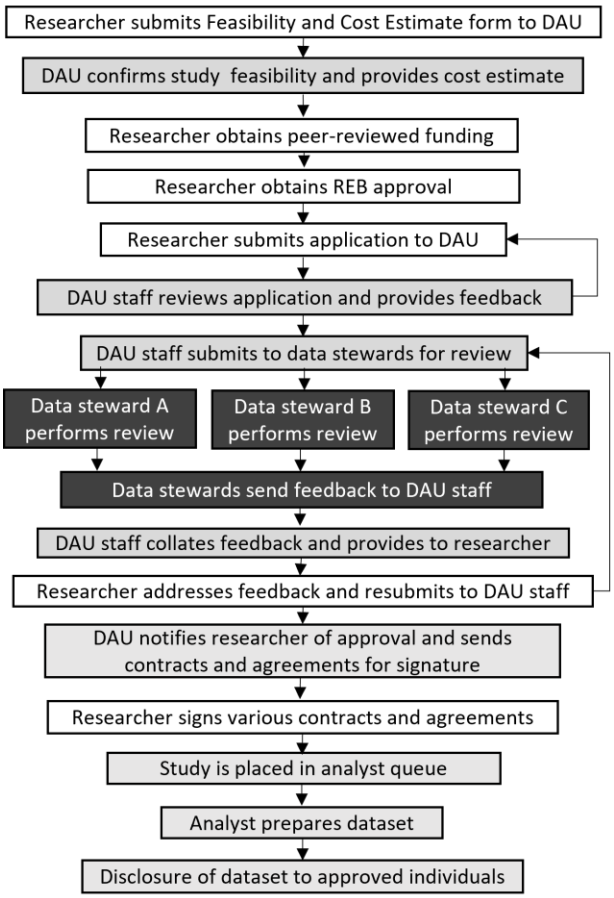
Steps carried out by the researcher (or team)

Steps carried out internally at HDNS

Abbreviations:
DAC- Data Access Committee
HDNS- Health Data Nova Scotia
REB- Research Ethics Board

Notes:
The HDNS application can be submitted once REB approval is obtained, or while REB approval is in process. REB approval is required before any data are accessed.

Process at PopData:



Key:

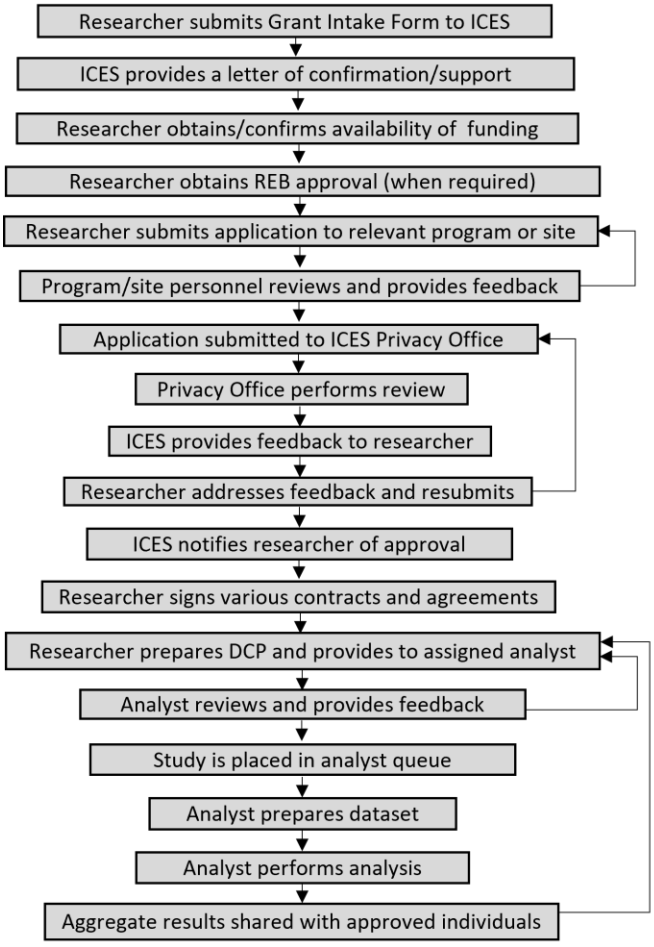
- Steps carried out by the researcher (or team)
- Steps carried out internally at PopData
- Steps carried out by data stewards (internal or external)

Abbreviations:
DAU- Data Access Unit
REB- Research Ethics Board

Notes:
Each dataset held by PopData has been assigned a data steward who is responsible for decision-making regarding access to the data. For studies involving multiple internal datasets, review by multiple data stewards may be required.

The application can be submitted to the DAU once REB approval is obtained, or while REB approval is in process. REB approval must be obtained before the application is sent to the data steward(s).

Process at ICES (internal pathway):

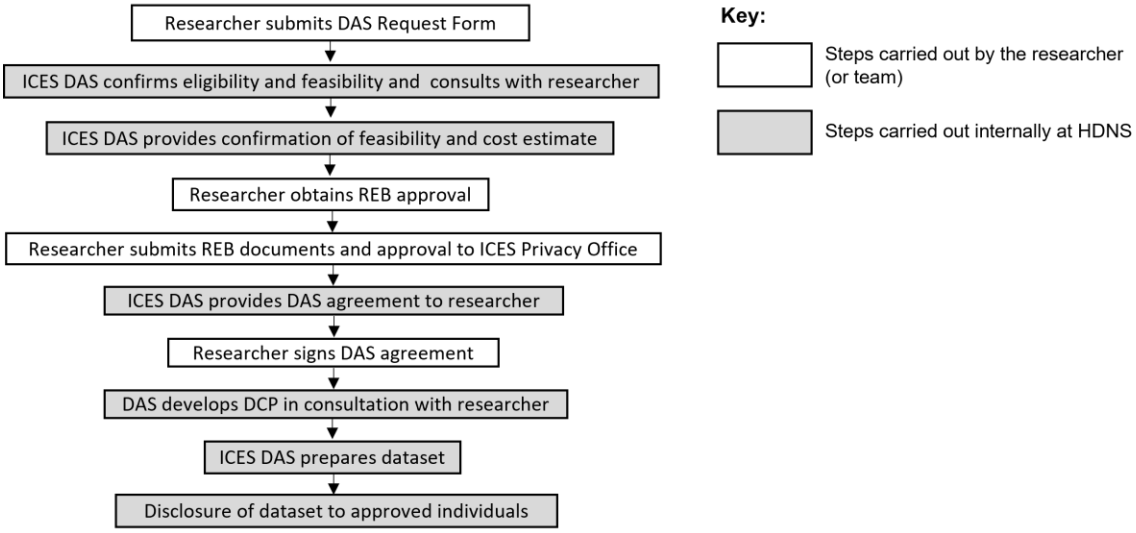


Key:

Steps carried out internally at ICES

Abbreviations:
DCP- Dataset Creation Plan
REB- Research Ethics Board

Process at ICES-DAS (third-party pathway; adapted from [283]).



APPENDIX J: Nova Scotia—Factors affecting access to administrative health data for research

Factors	Description
Study-related	
Requested data	• Data access processes were determined by where the required data were held, including which reviews and approvals were required for the study, and whether the study required one or more external linkages—both of which had implications for overall time and costs. • HDNS was not involved in facilitating external linkages. Agreements were negotiated and arrangements were made by the research team. • Researchers engaged in a process of negotiation with HDNS regarding what would be included in the research dataset. For example, HDNS would not release complete postal code if requested, but would provide either the second digit of postal code, or the first three digits. Similarly, rather than release individual diagnosis codes to the research team, HDNS calculated comorbidity indices for researchers or provided categories of diagnosis codes whenever possible.
Study design and methods	• The complexity of the request (e.g., number of linkages, amount of data to be extracted, number and complexity of derived variables, requested analyses, etc.) impacted the time required for the HDNS analyst to prepare the dataset, and given HDNS' cost recovery model, subsequently impacted costs. Increased complexity led to longer timeframes and higher costs. • Where study cohorts were pre-defined (i.e., provided by an external data provider), the timeliness of dataset preparation was sometimes improved, compared to having the HDNS analysts identify the cohort from within its data holdings. • The study design and methods were considered, although to varying extents, by reviewers (REB, DAC, or other) when assessing the scientific merit, ethical acceptability, and privacy risks posed by the study. The time required for the researcher to satisfactorily address questions and concerns raised during the review process contributed to longer data access timelines and higher overall costs.
Researcher-related factors	

Factors	Description
Researcher affiliation	• All studies involving access to data via HDNS were required to be led by a researcher with a Dalhousie University affiliation. • For studies led by researchers from other academic institutions, data could be accessed via HDNS if a local researcher was identified to take on the role of local Principal Investigator.
Researcher knowledge and expertise	• Administrative health data were described as complex and highly nuanced. A strong understanding of administrative health data, and the methodological aspects of research involving administrative health data, improved researchers' ability to determine which data were required for the study and to prepare a high-quality application. • Researchers who were knowledgeable about the local context (i.e., data holdings, regulatory requirements, available supports, key contacts) had an improved ability to initiate and navigate data access process(es) efficiently and effectively.
Researcher experience	• Researchers with prior experience accessing data were usually better prepared to navigate the data access process and had an improved understanding of what was expected of them from a regulatory perspective compared to those without. • Researchers with more experience accessing data often developed relationships with individuals (e.g., staff at HDNS, REB coordinators, external data providers, etc.) that improved access to information and supports (e.g., researchers knew who to call with questions and felt comfortable doing so). • Over time, researchers who accessed data multiple times (from HDNS or external data providers) established a "track record" which impacted their ability to access data by helping to establish trust and credibility.
Access to funding	• Given the costs associated with accessing administrative health data, researchers required access to adequate funding. • Researchers with established programs of research were recognized as typically having better access to funding compared to clinicians, early career investigators, and trainees. • Trainees' ability to access data was dependent on whether they had access to funding through their supervisor.

Factors	Description
Regulatory-stakeholder-related factors	
Knowledge and expertise	• Regulatory stakeholders in NS come from a wide range of backgrounds and had different areas and levels of expertise relevant to administrative health data, resulting in variations in review within and across regulatory entities. • Most studies involving administrative health data underwent expedited REB review. Since it was carried out by a single individual, the results of the review varied depending on the knowledge and expertise of the person conducting the review, creating uncertainty about how to approach each application to meet REB requirements.
Individual perspectives on benefits and risks of administrative health data research	• Regulatory stakeholders were largely supportive of the use of linked administrative health data for research purposes as long as all required data safeguards were in place. • Notably, many regulatory stakeholders in were trained researchers, which may have impacted their overall attitudes toward and perspectives on research involving administrative health data. • Regulatory stakeholders considered the privacy risks associated with the use of linked administrative health data to be minimal given the current policies and processes that are place to safeguard the data; however, several concerns were expressed, including the misinterpretation or misrepresentation of data by researchers and the potential for re-identification given increasingly sophisticated technologies.
Relational factors	
Communication	• The researcher's ability to communicate the details of their study and justification for the requested data impacted the quality of data access and REB applications. Higher quality applications raised fewer questions and required fewer clarifications, and typically received approval more quickly than those of poorer quality. • Researchers were expected to establish, through their applications and conversations with relevant stakeholders, that they had carefully thought through the data they need, how it will be used, and how it will be safeguarded. Failure to communicate these things was sometimes interpreted as a poor understanding of research involving administrative

Factors	Description
	health data, and/or a lack of respect for the data, leading to reluctance on the part of some data holders to provide data. • Clear and effective communication between the researcher and analysts at HDNS (and those working on behalf of external data providers/custodians) was necessary to ensure that the research team and analysts had a common understanding of the data being requested and the scope of work to be done. • Miscommunication sometimes occurred because the various stakeholders (e.g., researchers, DAC members, analysts, data custodians, etc.) did not always speak each other's "language". For example, researchers were not always adept at speaking about data at the level of detail required by an analyst. • Changes in data access policies and processes were not always communicated to researchers, leading to confusion and uncertainty about how to proceed, and sometimes leading to researchers doing the wrong "thing" (e.g., submitting the wrong version of form, missing a step in the process), ultimately leading to increased time and costs.
Relationships	• Researchers with relationships within the research community, with other researchers and with regulatory stakeholders, had improved access to information and supports relevant to the data access process. • Where researchers and regulatory stakeholders had established relationships over time, there was sometimes a greater willingness on the part of the regulatory stakeholder to share data. • Relationships between regulatory stakeholders also impacted data access in NS. Where good working relationships existed, regulatory stakeholders had, on occasion, made efforts to come together for joint review to streamline processes. • Participants noted that individuals working in the administrative health data space often knew each other. The community was small and closely connected with individuals working within proximity to one another and encountering each other in social situations. In some ways this facilitated access to data.
Trust	• Participants acknowledged that to do research involving administrative health data, trust is required between the various parties involved.

Factors	Description
	• For regulatory stakeholders, the extent to which researchers were trusted was related to several factors, including the researcher's level of experience, credibility, willingness to engage with data holders/custodians throughout the research process, and track record of adhering to regulatory requirements. • Several researchers in NS described not feeling trusted by regulatory stakeholders and entities, especially with regard to trying to re-identify individuals. However, most regulatory stakeholders believed that the risk of anyone deliberately re-identifying participants was highly unlikely. Trust is not the issue, but rather, regulatory stakeholders feel the need to do their due diligence, leading researchers to feel that they are not being trusted. • Perceived lack of trust lead to feelings of frustration among some researchers and acted as a deterrent to pursuing access to data in the future.
Organizational factors	
Organizational mandate/priorities	• HDNS' mandate was to support research, and its main priority was facilitating access to administrative health data for research purposes. • Data were often requested from organizations outside of HDNS which did not share the same organizational mandate or priorities. This limited the resources that were available to support research and how they prioritized research amidst competing priorities. • For some organizations (e.g., healthcare organizations), dedicating resources to supporting research meant diverting resources away from other areas (e.g., clinical care, health care administration), resulting in opportunity costs.
Data centre funding model	• For many years HDNS did not receive any external funding (e.g., from provincial government or university) and relied on a cost-recovery model, resulting in limited operational capacity. • In recent years, funding from national research grants allowed additional staff to be hired, including several analysts, increasing HDNS' overall operational capacity.
Organizational capacity to support research	

Factors	Description
Analytic capacity	• In the past, funding instability created challenges with staff retention, particularly with regard to analysts. Experienced analysts left for work elsewhere, while newer analysts did not always have the skillset to complete some of the more complex data pulls. At times, this resulted in delays and/or datasets that were prepared incorrectly. • More analysts were hired in recent years, though depending on the volume of requests in the queue at any given time, and whether staff were working on priority projects for the provincial government, delays in dataset preparation sometimes still occurred. • The extent to which data providers/custodians outside of HDNS had time and resources available for preparing and providing research datasets varied.
Regulatory capacity	• HDNS had an established data access committee that met regularly to review data access applications. In addition, HDNS had staff dedicated to supporting application intake, correspondence with researchers, contracts and agreements, regulatory compliance, and vetting results. • The extent to which data providers/custodians outside of HDNS had time and resources dedicated to accepting, reviewing, and adjudicating applications varied substantially. Often, this work was being carried out "off the side of one's desk". • Fluctuations in staffing levels at NS Health Research Services, particularly with regard to ethics coordinators, impacted REB review timelines. • In the past, substantial work was done with local REBs to improve their understanding of research involving administrative health data. Turnover in personnel since that time resulted in a loss of knowledge.
Technical capacity	• HDNS' Secure Data Platform provided researchers with secure data storage and access, as well as analytic tools for analyzing linked administrative health datasets. • Prior to the implementation of remote access, researchers were required to access their data on one of few secure workstations located on the Dalhousie University campus. The ability to access the HDNS Secure Data Platform remotely improved researchers' ability to access data (e.g., more flexible work hours, no space limitations).

Factors	Description
	• Where REB approval was required from NS Health, applications were completed, submitted, and managed via an online portal.
Data holdings	• HDNS' data holdings consisted of nine databases made available by the NS Department of Health and Wellness. Where the required data were not available via HDNS, an external linkage was required. • The frequency of data updates from the Department of Health and Wellness varied, limiting the timeframes that could be examined. • Since data provided by external data providers were often collected for purposes other than research, they were not always "research ready", and required substantial cleaning and processing to be useful for research.
Support for researchers	• HDNS' "Data Navigator" was considered an important source of support for researchers. The Data Navigator provided information on data, support during application preparation, and guidance during the data access process. • HDNS staff were regarded as approachable and willing to help researchers with various issues that arose. • Data dictionaries were available for most data holdings at HDNS but lacked the level of detail required to assist in decision-making regarding variable selection. • External data providers/custodians did not typically have data dictionaries or metadata available to share with researchers. • External data providers/custodians varied in terms of the information they could provide about their datasets. • Key informants were not aware of any courses focused specifically on administrative health data research that had been offered at Dalhousie University.
Mechanisms for feedback and improvement	• There was no mechanism in place to correct errors in the data held by HDNS. If errors were found, HDNS did not have the authority to correct the data themselves, and there was no way to feed the information back to the original data provider so that it could be corrected. • There were no processes in place at HDNS to evaluate and improve the accuracy of cost estimates. • Data access timelines were not actively measured at HDNS.
Organizational culture	• Consistent with its service provider role, HDNS was willing to provide data to researchers, assuming all requirements were met, and was characterized as working closely with

Factors	Description
	researchers to assist them in obtaining access. However, the DAC was described using terms such as "cautious" and "conservative" in their approach to data access. • Occasionally situations arose where external data providers/custodians declined or were reluctant to provide data for research. A variety of reasons for this were identified: they felt ownership over the data, they were concerned about how the data would be used and how it would reflect on the organization/service/program, they did not have an existing relationship with the individual(s) seeking access to data, they did not have the capacity/resources, there were other competing demands requiring attention, etc. These factors varied between data providers/custodians.
Regulatory factors	
Information legislation	• The Personal Health Information Act (PHIA) clearly recognized research as a legitimate use of personal health information. • PHIA formalized best practice and provided guidance for researchers and regulatory stakeholders, particularly around the required documentation. • PHIA was ambiguous in several respects which created the potential for variations in interpretation across stakeholders. Specifically, PHIA did not specify data retention timeframes, criteria for what constitutes sufficiently de-identified data, or clear criteria for establishing "impracticability" of consent. • One change brought about by PHIA was a renewed emphasis on the "minimum dataset". • The threat of legal action for non-compliance under PHIA was thought to encourage a conservative interpretation of the Act.
Transparency of data access pathway	• Researchers were often unsure of how to get started with regard to accessing administrative health data for research. This was especially true for studies involving access to external databases. • Information about how to access data via HDNS was documented and available on the HDNS website. The Data Navigator served as a point of contact for researchers and advised on data holdings and data access processes. • Information about external databases, including who to contact to make inquiries regarding access, was not typically publicly available.

Factors	Description
	• Data access policies, processes, and forms were constantly changing. Some of these changes were related to ongoing efforts by HDNS and other data custodians to streamline processes. These changes sometimes created delays for researchers as they learned about these changes and determined how to proceed. • The processes and requirements for accessing to existing datasets (i.e., datasets created for previous studies that were being stored on the HDNS server) were not clear.
Complexity of data access pathway	• The high-level overview of the data access process available on the HDNS website did not reflect the complexity of the process, which consisted of additional steps, particularly when external linkages were involved. • For studies involving one or more external data linkages, the total number of steps involved in gaining access to data increased with the addition of each additional data provider.
Required forms and documents	• Researchers in NS were required to submit, at minimum, an application to the HDNS DAC and an REB application. For studies involving linkage to external databases, additional applications were typically required. • Forms and documents were considered beneficial in terms of guiding the process and improving transparency; however, the volume of required paperwork was considered problematic. Applications to individual regulatory entities (e.g., the DAC, or REB, NS Health Privacy Office) consisted of various forms and extensive supporting documentation. For studies involving external linkage, the amount of paperwork to be prepared and submitted increased as the number of external data providers/custodians increased. • The application forms used by different regulatory entities overlapped substantially, highlighting the potential for documentation to be streamlined. • Applications were described as tedious to complete, with researchers required to identify each individual variable to be included in the research dataset, as to provide a detailed justification for each. • Amendments were commonly required, resulting in additional paperwork to be submitted. The amount of paperwork involved varied depending on the nature of the amendment.
Required reviews and approvals	• Researchers always required approval from the HDNS DAC and REB. When linking to external data sources, additional reviews

Factors	Description
	were often required. The need for multiple reviews meant that researchers spent extensive time and resources navigating the various review processes and responding to reviewer feedback, especially if reviewers provided conflicting feedback. • Studies involving linked administrative health data were typically considered low risk by the REB and underwent expedited review, which was timelier compared to full review. • Research approved by the NS Health REB did not also require approval from the Dalhousie University REB, which eliminated the need to submit REB applications to both institutions for many studies; However, in some instances, REB approval was required from other institutions (e.g., the IWK Health Centre).
Scope of review	• In addition to review by the HDNS DAC, REB, and external data custodians, studies that were grant funded typically underwent peer review. These reviews were broad and overlapping in scope (i.e., all involved scientific and privacy review), resulting in duplication of effort. • The various regulatory bodies/entities frequently provided feedback on or requested changes to the study methods, which researchers did not always consider appropriate, especially for peer-reviewed grant-funded studies.
Transparency of review	• HDNS invited researchers to attend the DAC meeting at which their application was being reviewed. This provided researchers with a clear understanding of the concerns of the committee and how to address them, thereby reducing subsequent rounds of review and feedback. • There was a lack of transparency with the REB process at NS Health in particular. Feedback was not always clear or consistent, leaving researchers questioning exactly what the REB is looking for.
Application of data minimization principle	• In the HDNS data access application, researchers were required to specify each variable required for the study (from HDNS and external data sources), the level of identifiability needed, and how each variable would be used in their analysis. This level of detail required extensive expertise and made applications time-consuming and challenging to complete. • Researchers and data providers/custodians often "negotiated" an agreed upon dataset. The negotiations were characterized by extensive toing and froing about which variables were

Factors	Description
	necessary to include in the final dataset, and the required level of identifiability of each variable. • Emphasis on providing researchers with the minimum dataset sometimes resulted in researchers not being provided access to certain variables at the level of identifiability that would be best for analysis (e.g., partial vs. full postal codes). • Emphasis on providing researchers with the minimum dataset created challenges for certain types of research, including exploratory research and studies using machine learning. These studies were sometimes mistakenly characterized as "data fishing." • Once researchers were granted access to the dataset, they often realized that additional variables were needed, or that one or more variables needed to be categorized differently for the analysis, which triggered one or more amendments. • Data were packaged by HDNS (and sometimes by external data providers) to minimize the level of identifiability. Without access to the raw data, researchers were unable confirm the reliability and validity of the variables within the final dataset.
Role of data centre analysts	• HDNS analysts were not embedded in the research team, so they were sometimes missing important content knowledge and study context. This sometimes hindered their ability to prepare the dataset in a timely manner and was thought to contribute to errors during dataset preparation. This sometimes resulted in extensive back and forth with the research team to get clarification regarding dataset preparation and/or address any errors in the linked dataset.
Proportionality	• All studies involving access to data via HDNS were subject to the same level of scrutiny. There was no expedited pathway for studies that were perceived to propose a lower level of risk. • In accordance with the TCPS 2, local REBs employed a proportionate approach to review with low-risk studies requiring expedited or delegated review, rather than full review by an entire REB.
Accountability	• As an agent of the NS Department of Health and Wellness, HDNS was required to adhere to the terms and conditions set out in the data sharing agreement between the two entities and to comply with the Personal Health Information Act.

Factors	Description
	• Occasionally, instances arose where HDNS required clarification around specific policies or processes from NS DHW; However, this did not typically affect access to data as responses were obtained in timely manner.
Contextual factors	
Leadership	• The provincial government did not provide core funding to HDNS. • The NS Department of Health and Wellness has declined to enter into a long-term data sharing agreement with the Department of Community Health and Epidemiology at Dalhousie University. In 2020, the Department of Health and Wellness also announced that they were considering not renewing the existing data sharing agreement moving forward. This was perceived by some as evidence of a lack of support of HDNS. (Note: The data sharing agreement has continued to be renewed annually to date.) • Grassroots efforts by local stakeholders to develop a more coordinated approach to data access in NS lacked the provincial leadership to implement system-wide change. • Changes to data custodianship as a result of health system restructuring in 2016 led provincial programs to lose longstanding and well-established data access policies and processes. No clear direction was provided from the province or health authority about how the restructuring would impact data access and who would be responsible for authorizing access to data under the new organizational structure. At the time of data collection (2020), policies and processes for accessing data held by some of the former provincial programs were still not well-established.
Health system organization and integration	• Administrative health data, and data that were commonly linked to administrative health data, were held by many different organizations and institutions across the province, representing numerous data custodians (e.g., NS DHW, the IWK, NS Health, etc.). Each of these had their own unique approach for governing access to data, resulting in incongruent policies and inefficient processes when accessing data in NS. • Within NS Health there was also a lack of coordination across the organization with regard to data access polices and processes.

Factors	Description
Legislative landscape	• Not identified as a factor.
Historical events	• Not identified as a factor.
Current events	• Not identified as a factor.

APPENDIX K: British Columbia—Factors affecting access to administrative health data for research

Factors	Description
Study-related	
Requested data	• Data access processes were determined by where the required data were held, including which reviews and approvals were required for the study, and whether the study required one or more external linkages—both of which had implications for overall time and costs. • Studies involving external data linkages took longer and cost more because there was more work to do on the part of PopData (e.g., developing contracts and agreements). • If data were requested from an external data provider that PopData frequently worked with, the process of obtaining data sometimes moved along more quickly and easily. • Certain variables were more likely to get "flagged" during data steward review, particularly those that may be perceived as sensitive in nature (e.g., abortion data, drug use data) or as posing an increased risk of re-identification (e.g., full postal code, date of birth). These were not necessarily rejected outright, but often required more toing and froing with the data stewards to establish justification.
Study design and methods	• The number of datasets that were linked, and the amount of data cleaning, coding, and analysis that PopData was asked to provide, increased the required amount of analyst time, which led to longer data access timelines and higher overall costs. • Where study cohorts were pre-defined (i.e., provided by an external data provider), the timeliness of dataset preparation was sometimes improved, compared to having the PopData analysts identify the cohort from within its data holdings. • The study design and methods were considered, to varying extents, by reviewers (i.e., REB, data stewards) when assessing the scientific merit, ethical acceptability, and privacy risks posed by the study. The time required for the researcher to satisfactorily address questions and concerns raised during the review process led to longer data access timelines and higher overall costs.
Researcher-related factors	

Factors	Description
Researcher affiliation	• Researchers required an academic appointment at a Canadian institution to access data at PopData. • Affiliations were beneficial for accessing external datasets. For example, individuals at BC Cancer had relatively easy access to in-house data (with the process being relatively simple, quick, and free). Similarly, once the organizations within the Provincial Services Health Authority (PSHA) became a single legal entity, individuals with a PSHA affiliation had improved access to these data (e.g., Children's hospital, BC Cancer, etc.).
Researcher knowledge and expertise	• Administrative health data were described as complex and highly nuanced. A strong understanding of administrative health data, and the methodological aspects of research involving administrative health data, was necessary to determine which data were required for the study and to prepare a high-quality application. • Researchers who were knowledgeable about the local context (i.e., data holdings, regulatory requirements, available supports, key contacts) had an improved ability to initiate and navigate data access process(es) efficiently and effectively. For example, they knew which applications were required, when and where to submit them, and how to prepare applications in way that met reviewer expectations. • For researchers with limited knowledge of administrative health data, the methodological aspects of research involving administrative health data, or the local context, mentorship was an important facilitator for accessing data. • Some training opportunities were available (e.g., a class on administrative health data was offered at UBC).
Researcher experience	• Researchers who had prior experience accessing data were better prepared to navigate the data access process and had an improved understanding of what was expected of them from a regulatory perspective compared to those who do not have prior experience. • Researchers with more experience accessing data sometimes established relationships with individuals (e.g., staff at PopData or external data providers) that improved access to information and supports (e.g., researchers knew who to call with questions and felt comfortable doing so).

Factors	Description
	• Researchers who had more experience were sometimes viewed more favorably by regulatory stakeholders (i.e., as more "trustworthy").
Access to funding	• Studies involving access to data via PopData required peer-reviewed funding. Researchers who scored well in peer-review but were not funded were not able to apply for access to data. • Researchers who had access to funding but did not go through peer-review were not eligible to apply for access to data. Although PopData had developed a proxy peer-review process for such cases, it was challenging to arrange and did not frequently occur. • For researchers with access to sufficient, peer-reviewed funding, the costs of accessing data were not a barrier, although researchers were less likely to have access to sufficient funding (e.g., clinician researchers, students/trainees). • For clinician researchers with smaller pots of funding to address a single question, the costs associated with accessing data were acknowledged as being cost-prohibitive. • Although students were eligible for a discount when accessing administrative health data, the remaining costs were still substantial, requiring funding support.
Regulatory-stakeholder-related factors	
Knowledge and expertise	• Regulatory stakeholders in BC were from a wide range of backgrounds and had different areas and levels of expertise relevant to administrative health data which led to variations in review within and across regulatory entities. • When accessing data via PopData, review was required from individual data stewards, so the results of the review varied depending on the knowledge and expertise of the individual performing the review. This contributed to uncertainty about the data steward expectations, and how to prepare applications so that these were met. • Individuals who were new to regulatory roles sometimes experienced a steep learning curve with respect to their knowledge of the relevant regulatory requirements, administrative health data, and methodological approaches.

Factors	Description
	Over time, and with more experience, the quality and consistency of their review improved.
Individual perspectives on benefits and risks of administrative health data research	• Participants characterized regulatory stakeholders in BC as having varying perspectives about the use of administrative health data for research. Although many were supportive of the use of linked administrative health data for research purposes as long as all required data safeguards were in place, some were described as being resistant to permitting access to data, while others were described as fearful. • Regulatory stakeholders interviewed in this study considered the privacy risks associated with the use of linked administrative health data to be minimal given the current policies and processes that are place to safeguard the data; However, several concerns were expressed, including the misinterpretation or misrepresentation of data by researchers, the potential for re-identification given increasingly sophisticated technologies, and the increased privacy risks associated with linking to more datasets and to datasets from other sectors.
Relational factors	
Communication	• The researcher's ability to communicate the details of their study and justification for the requested data impacted the quality of data access and REB applications. Higher quality applications raised fewer questions and required fewer clarifications, and typically obtained approval more quickly than those of poorer quality. • Clear and effective communication between the researcher, PopData staff, and external data providers (and those working on behalf of external data providers/custodians) was necessary to ensure that the stakeholders involved had a common understanding of the work to be done so that it progressed in a smooth and timely manner. • One of the roles of PopData was to facilitate communication between the researchers and relevant data stewards, and to minimize the amount of toing and froing occurring between the researcher and individual data stewards. A challenge with this approach was that researchers were unable to communicate directly with the data stewards, limiting their ability to gain an understanding of data stewards' expectations and how to address them.

Factors	Description
	• PopData was challenged with how to communicate with researchers to provide more timely updates and improve transparency. The primary means of communicating with researchers who sought access to data was through a generic email address, which was regarded as both inefficient and impersonal. Plans to improve communication included a ticketing system and application tracking system; However, these were not implemented at the time of the study.
Relationships	• Where researchers had established relationships with external data providers over time, there was sometimes a greater willingness on the part of the data provider to provide access to data. • Researchers who had relationships with regulatory stakeholders (including PopData staff) had an improved ability to access information and supports relevant to the data access process—that is, researchers knew who to contact for help and felt comfortable doing so, while regulatory stakeholders were more likely to respond and willing to help. Over time, regulatory stakeholders also developed an improved understanding of the research, which improved their ability to provide support. • Relationships between PopData and external data providers facilitated data access in two ways. First, where PopData and external data providers worked together on previous projects, there was a common understanding of the work to be done that helped things move forward in a smooth and timely manner. In addition, relationships between PopData and external providers sometimes resulted in a transition from project-specific data sharing agreements to a standing data sharing agreement.
Trust	• REB review was facilitated by the trust that had been built between various university based REBs and PopData. • Some external data providers were more inclined to permit access to data when they knew it was being stored on PopData's secure data platform. PopData was known to have appropriate data safeguards in place, so data providers trusted that the data would be secure. • Participants acknowledged that research involving administrative health data required trust between the various parties involved. For example, trust was required when researchers were granted access to data via PopData's secure

Factors	Description
	research environment. Once access was granted, there was no way to control what researchers did with the data, but there was trust that researchers would comply with regulatory requirements rather than jeopardize their career.
Organizational factors	
Organizational mandate/priorities	• PopData's mandate was to support research, but not to do research. • The data, infrastructure, and support available via PopData facilitated access to data and enabled the conduct of research that could not be done otherwise. • Research was not a priority for many data provider organizations. As a result there were limited resources allocated to supporting research. • For some organizations (e.g., healthcare organizations), dedicating resources to supporting research meant diverting resources away from other areas (e.g., clinical care, health care administrative), resulting in opportunity costs.
Data centre funding model	• Not identified as a factor.
Organizational capacity to support research	
Analytic capacity	• At PopData, projects were placed in queue for an analyst. The amount of time projects spent in the queue depended on the volume of projects that were in queue at that time, which varied. The amount of time that projects spent waiting in queue for an analyst was not identified as being particularly problematic. Several participants noted that once reviews were complete, PopData was usually able to prepare the dataset relatively quickly. • The extent to which data providers/custodians outside of PopData have time and resources available for preparing and providing research datasets varies.
Regulatory capacity	• PopData's capacity to review applications and provide feedback was limited. To minimize the amount of back and forth between the research team and data stewards, staff at the DAU took on more of this work themselves. This created a situation where researchers who were submitting their application for the next round of feedback had to wait in

Factors	Description
	queue for a DAU staff member to be available to meet with them. • PopData ensures that researchers are compliant with legislative requirements. • Data provider organizations varied in terms of their regulatory capacity. While some organizations had well-established policies and processes, access to legal services to support their development (e.g., the Ministry of Health), and dedicated human resources to perform review, this was not the case for all organizations. For some data stewards, their regulatory role was carried out "off the side of their desk" in addition to their main job or role.
Technical capacity	• PopData provided a secure research environment that researchers were able to access, as well as the statistical tools and software. Researchers who were not using PopData datasets were permitted to use the secure research environment to securely store their datasets. • Remote access was available to researchers, which allowed them to access data from anywhere in the country. • There were efforts underway to leverage IT systems to improve communication and transparency throughout the data access process (e.g., online submission system, ticketing system, application tracker); However, these were not fully operational at the time of interviews and many processes were occurring manually.
Data holdings	• At the time of this study, PopData's data holdings consisted of approximately 30 databases available from a variety of data providers. Where the required data were not available through PopData, an external linkage was required.
Support for researchers	• At PopData, support for researchers was primarily provided by staff at the DAU. The DAU was the primary point of contact for researchers seeking access to data via PopData. The DAU provided guidance to researchers regarding the data access process and assisted researchers with application preparation by reviewing and providing feedback on drafts prior to submission. Once submitted, DAU staff coordinated the review of applications by data stewards and facilitated subsequent rounds of communication between the researcher and data steward.

Factors	Description
	• PopData staff were described as customer-service focused, friendly, and willing to help researchers with various issues that may arise. • PopData's website is very detailed, providing a variety of information and resources for researchers. • Metadata was available for PopData's data internal data holdings, although these sometimes lacked the level of detail required to assist researchers in decision-making regarding variable selection. • PopData provided education and training to the research community through its Education and Training Unit, which offered webinars, training on specific analytic approaches, and a population health data analysis certificate program. • External data providers/custodians varied in terms of available dictionaries or metadata and their ability to provide information about data under their purview.
Mechanisms for feedback and improvement	• There were no mechanisms in place to correct errors in the data held by PopData. PopData was not able to correct the data themselves, and there was no process to feed the information back to the original data provider for correction. • Data access timelines were monitored by PopData. • PopData hired a consultant during the study timeframe who was evaluating internal processes. Their recommendations had not been implemented at the time of data collection.
Organizational culture	• Consistent with its service provider role, PopData was characterized as working with closely with researchers to assist them in obtaining access. • Participants described data stewards/providers as varying in their willingness to provide data for researchers. Some actively promoted the use of their data for research, while others were reluctant to provide data or even refused. • Where external data stewards/providers were reluctant to provide data for research, or refused, a variety of potential reasons were identified by participants: lack of support for research, limited capacity, a sense of ownership over the data, a desire to maintain power/exercise control, concerns over data quality and /or use, lack of familiarity with research, fear of a breach, and resistance to change. These factors were thought to vary between data providers/custodians.
Regulatory factors	

Factors	Description
Information legislation	• Health sector specific legislation was not in place in BC. Most research involving administrative health data was subject to public sector information legislation (Freedom of Information and Protection of Privacy Act), which permitted disclosure of public data for research purposes. • The legislative provision permitting the disclosure of data for research was very broad, leading to variations in interpretation by data stewards across the province, and subsequently, to variations in data access policies and processes across organizations. • The limited extent to which research was addressed in the legislation provided insufficient guidance for researchers and regulatory stakeholders. • Data stewards were not required to disclose data for research under the legislation and were able to exercise discretion in terms of whether to provide data for a particular study. • Jurisdictional issues were not addressed in the legislation, contributing to ambiguity regarding linkage of federal and provincial data (i.e., where it can reside) and whether data could leave the province. • Other statutes sometimes applied (e.g., the E-Health Act, or the Pharmaceutical Act) but it was not always easy for researchers to determine when this was the case.
Transparency of data access pathway	• Information about how to access data via PopData was documented and available on the PopData website. • Additional information regarding the data access pathway and processes could be obtained from the DAU. Each project was assigned a DAU staff member, who was able to help the researcher navigate the data access pathway. • Information about how to gain access to external data sets was not typically available.
Complexity of data access pathway	• The overview of the data access pathway on the PopData website did not reflect the complexity of the process, which consisted of additional steps, particularly when external linkages were involved. • For studies involving one or more external data linkages, the total number of steps involved in gaining access to data increased with the addition of each additional data provider.
Required forms and documents	• Researchers were required to submit research ethics applications to the relevant institutions (sometimes multiple

Factors	Description
	applications were required) and to submit a data access application to PopData. • PopData data access applications were lengthy and tedious to complete, with researchers required to identify each individual variable to be included in the research dataset, and to provide a detailed justification for each. • A high level of knowledge and expertise of administrative health data and the methodological aspects of administrative health data research was required to complete the required forms and documents. • Where changes to the requested dataset were necessary, an amendment was required, which sometimes resulted in a "cascade" of amendments (i.e., the need to submit amendments to all the data stewards and REBs involved in the study). • The amount of paperwork involved in accessing data (various applications, supporting documents, data sharing agreements, etc.) created a high administrative workload for both the research team and PopData.
Required reviews and approvals	• Researchers required approval from all relevant data stewards. When linking to external data sources, additional reviews were also required. REB review was required, sometimes from multiple institutions. • The need for multiple review (i.e., multiple data stewards, multiple REBs) led to researchers spending extensive time and resources navigating the various review processes and responding to feedback. • Multiple reviews also often resulted in variations in review, due to differences in the focus of the review that was carried out (e.g., REBs have different concerns that data stewards), differences in knowledge and expertise, and differences in the interpretation and application of the legislation. Without a harmonized approach, researchers were challenged to meet the requirements of each regulatory stakeholder and to navigate innumerable rounds of review and feedback to address a myriad of questions and concerns. • For studies involving external data providers, the sequence of approval sometimes led to amendments. For example, changes requested by a data steward sometimes resulted in an amendment to the researcher's REB application.

Factors	Description
Scope of review	• In addition to review by the relevant data stewards, one or more REBs, and external data custodians, studies were required to undergo peer review (i.e., by a funding agency). These reviews were broad and overlapping in scope (i.e., all involve scientific and privacy review), which resulting in duplication of effort. • Data stewards often provided feedback around methods and analysis, which was beyond the required scope of review (and often beyond the expertise of the data steward). At the same time, these comments (e.g., around the statistical analyses being done) were not relevant from a privacy perspective and had already been through both peer review and REB review.
Transparency of review	• Data steward review was coordinated by staff at PopData's DAU. As a result, researchers did not have a direct line of communication with the data steward(s), which hindered their ability to gain an understanding of the expectation of data stewards and how to address them. In addition, researchers were often unaware of who was performing the review (or when it would occur), further contributing to the lack of transparency of review.
Application of data minimization principle	• In the PopData data access application, researchers were required to request only the variables that were necessary and to provide a detailed justification for the inclusion of each variable. • The emphasis on providing researchers with the minimum dataset was considered to limit certain kinds of research, including exploratory research and studies using machine learning. • Once researchers were granted access to the dataset, they often realized that additional variables were needed, which triggered one or more amendments. • In some instances, depending on the requirements of the data steward, PopData would "roll up" certain variables that were considered particularly sensitive or that were potentially identifiable. Otherwise, the researcher received a unprocessed dataset.
Role of data centre analysts	• PopData analysts were not embedded in the research team and typically played a limited role in methodological aspects of the study (e.g., defining cohorts, deriving variables, performing analyses). The research team was typically

Factors	Description
	provided a dataset containing unprocessed data and was responsible for analysis. • Occasionally, PopData analysts were involved in cohort creation, calculation of standard variables, and "rolling up" sensitive data, despite not always have the required content knowledge or study context. This was acknowledged as hindering the ability of the analyst to prepare the dataset in a timely manner and contributing to errors during dataset preparation, which subsequently required time and resources to address.
Proportionality	• During the study timeframe, all studies involving access to data via PopData were subject to the same level of scrutiny by data stewards, independent of the perceived level of risk posed by the study. There was no expedited pathway for studies that were perceived to propose a lower level of risk. • At the time of data collection, efforts were underway to implement a proportionate approach to the data access process at PopData and BC Ministry of Health based on the "5 Safes" approach. • In accordance with the TCPS 2, REBs employed a proportionate approach to review with low-risk studies requiring expedited or delegated review, rather than full review by an entire REB.
Accountability	• Not identified as a factor.
Contextual factors	
Leadership	• There is a sense among members of the research community that the Ministry of Health is not collaborative and has not demonstrated that they value administrative health data research. • Strong leadership was identified as necessary to address the fragmented data landscape is in BC and to foster a more collaborative approach across organizations with respect to data sharing.
Health system organization and integration	• Administrative health data, and data commonly linked to administrative health data, were held by many different organizations, institutions, and health authorities within the province. These were described as working in silos and as being resistant to working collaboratively (even in the context of COVID-19). PopData sometimes got caught in the middle of what is happening between these other organizations.

Factors	Description
	• The agencies within the PSHA all became a single legal entity, allowing researchers with a PSHA affiliation to access data across agencies (e.g., Children's hospital, BC Cancer, etc.).
Legislative landscape	• The legislative context is complex. There is a patchwork of legislation that applies to health information. The Freedom of Information and Protection of Privacy Act (FIPPA) is the main piece of legislation, but other Acts may apply depending on the data sources being accessed. For example, the eHealth Act and the Pharmaceutical Act. Different organizations may be subject to different legislation, and determining which ones apply is not always straightforward.
Historical events	• In 2013, a privacy breach was reported by the Ministry of Health to the Office of the Information and Privacy Commissioner. The response to this incident, which was the subject of an extensive Ombudsperson's report, was found to have caused irreparable harm to the accused individuals and their families. • While the response to the reported breach created fear within the BC research community, it also highlighted where processes needed to be improved and promoted a shift toward a more balanced approach to privacy protection in the context of health research.
Current events	• During the pandemic, PopData made efforts to ensure researchers conducting COVID-19 research had timely access to data. These efforts included prioritizing COVID-19 studies, streamlining processes (i.e., establishing a "rapid access" process), streamlining forms, and receiving more frequent data updates. • In some cases, the prioritization of COVID-19 research at PopData led to delays in the review and approval of non-COVID-19 studies. • The pandemic highlighted the importance of timely access to public health data and changed public expectations with regard to how data are used. For example, during the pandemic people wanted more transparency around what was happening in schools and in long-term care homes. • The pandemic also shone a light on challenges with data access that were occurring throughout the province (not at PopData), including a reluctance by some data providers to share data with the Provincial Health Officer early on in the pandemic, and an inability to conduct COVID-19 research in a

Factors	Description
	timely manner due to the inability to reach an agreement across health authorities.

APPENDIX L: Ontario—Factors affecting access to administrative health data for research

Factors	Impacts on data access
Study-related	
Requested data	• Studies involving external linkages typically took longer and were more costly compared to studies involving only data held by ICES. External linkages required that data sharing agreements be established with the external data provider, which took time to develop and execute. When bringing in external datasets, additional work was also required to remove individual identifiers and assign a unique ICES identifier (i.e., an "IKN"). • For studies using certain types of data (i.e., project-specific or controlled use data), additional steps were required to obtain access. • Depending on the terms of the data sharing agreement with the original data provider, some databases could not be accessed via DAS, while others required special permissions.
Study design and methods	• When accessing data via the internal data access pathway, the amount of time required for the ICES analyst to perform the analysis and provide results was greater where the analysis was more complex. • Similarly, for researchers accessing data via DAS, the amount of time to receive the requested data was dependent on the complexity of the request, the amount of data processing required (i.e., how "packaged" the data needed to be), and whether the researcher had requested the analysis be performed by DAS as well.
Researcher-related factors	
Researcher affiliation	• Affiliation determined the data access pathway. Researchers with an ICES affiliation were able to access data via ICES' internal data access pathway. For researchers without an ICES affiliation, there were two options: (1) collaborate with an ICES researcher to gain access via the internal pathway, or (2) access data via DAS. • Affiliation impacted which datasets could be accessed, and the required permissions. Non-ICES researchers were required to obtain special permissions to access a greater number of datasets compared to ICES-affiliated researchers. There were

Factors	Impacts on data access
	also some datasets that non-ICES researchers were not permitted to access at all. • Affiliation impacted the granularity of the data that could be accessed. When data were accessed via DAS, the researcher was provided a data cut with a risk-reduced dataset, containing limited variables and lacking granularity. When data were accessed via the internal data access pathway, the ICES analyst was given line-level access to all variables within the datasets required for the study.
Researcher knowledge and expertise	• Researchers who were knowledgeable about data availability planned studies accordingly—that is, they based their research around the data that they knew they could get access to (or get access to more easily). • Not all researchers who accessed data via DAS had experience working with administrative health data. Those who did not sometimes required the DAS analysts to process the data to a greater extent compared to individuals who are more comfortable working with the data. • Within ICES was a wealth of knowledge and expertise that can be accessed by all researchers and staff. For example, researchers had access to macros and algorithms developed by ICES analysts. • Researchers were grouped by program, which facilitated researchers' access to others working in similar areas who can share their knowledge and expertise (e.g., how to define certain variables, etc.).
Researcher experience	• Researchers who had prior experience accessing data were better prepared to navigate the data access process and had an improved understanding of what was expected of them from a regulatory perspective compared to those who did not have prior experience. • Individuals who had more experience accessing data were more likely to have developed relationships with ICES staff and analysts, which facilitated data access and data analysis. • A part of the process of becoming an ICES-affiliated researcher was to carry out research using ICES data while working under the mentorship of an ICES-affiliated researcher. As such, all ICES-affiliated researchers had at least a baseline level of understanding of ICES data and data access processes.
Access to funding	• Given the costs associated with accessing data via ICES (which are driven largely by analyst time), researchers typically

Factors	Impacts on data access
	required access to grant funds. Where funding was not obtained, studies were not always feasible. • In some cases, researchers who were particularly well-funded were able to pay the salary of an individual ICES analyst. In such cases, the researcher did not have to wait in queue—they had their own queue for data and could ask the analyst to prioritize specific projects as needed. This improved the researcher's ability to carry out projects within the funding timeframe. • More experienced researchers with well-established programs of research typically had improved access to funding compared to students, clinician researchers, and new investigators, and were therefore more likely to be able to gain access to data.
Regulatory-stakeholder-related factors	
Knowledge and expertise	• Staff within the Privacy and Legal Office, who performed privacy impact assessment (PIA) reviews, were highly trained individuals with extensive expertise. • Elsewhere in the organization, there was a strong sense that anyone who held a regulatory role had the required skills and knowledge to be in that role.
Individual perspectives on benefits and risks of administrative health data research	• Regulatory stakeholders who were interviewed were largely supportive of the use of linked administrative health data for research purposes as long as all required data safeguards were in place. • Given ICES' low risk threshold and the policies and processes in place to ensure data were safeguarded, privacy risks were considered low; However, it was acknowledged that there were increasing threats to institutions such as ICES (i.e., cyberattacks) that must be mitigated. • In addition to privacy risks (e.g., re-identification), the risk of "spurious research" was identified.
Relational factors	
Communication	• Clear communication between the various roles (e.g., privacy officers, researchers, analysts, knowledge-users) was important for moving research forward in a timely manner but could be challenging as these groups often communicated "at cross purposes" (i.e., with different goals in mind).

Factors	Impacts on data access
	• The PIA contained the particulars of the data request. Where this document was unclear, one or more rounds of revisions were required, increasing the overall timeline and costs. • ICES sometimes prioritized projects based on study deadlines/funding timeframes; However, it was up to the researcher to communicate these to the relevant individuals (e.g., the Privacy and Legal Office, lead analyst), which did not always occur. • Researchers were required to develop a dataset creation plan (DCP), which served as a key communication tool between the researcher and analyst, identifying study variables and specifying the analysis to be undertaken; However, researchers varied in their ability to develop this document. If the DCP was not clear, multiple rounds of "back and forth" between the analyst and researcher occurred, increasing the overall study timeline and costs. • A formal kick-off meeting was held between the analyst and research team at the study outset, and additional meetings were held at various points throughout the study to discuss progress and address any issues that arose. These meetings helped promote a common understanding of the work to be done and ensured that it progressed smoothly. • The Privacy and Legal Office was in frequent communication with the IPC regarding IPC requirements; However, the challenge was how to communicate changes in policies and processes to ICES researchers and trainees, particularly with those who only did research at ICES occasionally.
Relationships	• ICES' ability to hold certain datasets was contingent on maintaining a good relationship with the data provider, which meant following their processes for access and meeting their reporting requirements. • Where researchers had good relationships with ICES staff, they were more comfortable asking for information and support, which facilitated access to data. • Researchers' relationships with external data providers were sometimes beneficial for facilitating linkage to external datasets. • Researchers who were less experienced sometimes sought support from their more experienced colleagues when attempting to access data via ICES (e.g., help filling out forms, information about data and processes, etc.).

Factors	Impacts on data access
	• Non-ICES researchers sometimes leveraged relationships with colleagues who had an ICES affiliation to gain access to ICES data for research (i.e., to access via the internal pathway rather than go through DAS). • Where researchers established a relationship with an analyst over multiple projects, the analyst often had an improved understanding of what the researcher required for the study and could complete the analysis in a timelier manner.
Trust	• The policies and processes in place at ICES were underpinned by the need to maintain the trust of data providers, and the trust of the public. ICES' ability to operate was dependent on their trustworthiness in terms of their ability to safeguard the information they hold. • While navigating the policies and processes in place can impact project timeliness and costs, these policies and processes were critical to ICES' status as a trusted entity, which facilitated access to data for research.
Organizational factors	
Organizational mandate/priorities	• ICES was established specifically to facilitate access to data for health system planning and management, and research. • For other public institutions that held data, supporting research, and facilitating access to data was not necessarily a priority.
Data centre funding model	• ICES received core funding from the Ontario Ministry of Health and additional support from grants secured by ICES affiliated researchers. ICES used a cost-recovery model of sorts, but the costs were based on the overall operational costs of ICES (average costs) versus the cost of an individual study (incremental or marginal costs). • As ICES grew and became a bigger organization with higher operational costs, the costs to researchers increased.
Organizational capacity to support research	
Analytic capacity	• ICES analysts had extensive expertise in working with administrative health data and were available to work with all researchers accessing via DAS or the internal pathway; However, there were only a limited number of analysts and a high volume of requests, so there were sometimes long waits for analyst time.

Factors	Impacts on data access
Regulatory capacity	• ICES received a high volume of data access requests annually with a limited number of staff to support the work (e.g., individuals to perform PIAs, update policies and processes, ensure compliance, meet reporting requirements, etc.). • The number and complexity of projects submitted to ICES increased over the years, but this was not met with a corresponding increase in staff to support the work.
Technical capacity	• The required data infrastructure was in place to support research (i.e., the Research Analytic Environment), but improved systems for communicating with researchers were required to improve transparency and efficiency. On both the DAS side and internal ICES side, a lot of personnel time was spent responding directly to researcher inquiries about the status of their application. The use of an application dashboard or tracker was suggested as one way to provide researchers will information and updates about the status of their request in a timely and efficient manner.
Data holdings	• ICES had a large and growing number of internal data holdings which enabled them to support research on a variety of topics. Data holdings had grown with efforts to bring in more datasets that could be used to assess the social determinants of health. • Despite having a large number of data holdings, there were instances where ICES did not have the data required to examine certain research questions leading researchers to seek linkages to external data sources. • The timeliness of data updates was sometimes a challenge (i.e., the "lag" was too long for some studies).
Support for researchers	• Researchers were required to go through a specific process to obtain an ICES affiliation. This helped ensure they were prepared to work with the data and understood the regulatory requirements. • ICES personnel (staff, coordinators, site leads) were generally helpful. They viewed helping researchers as part of their role, and researchers acknowledged the support they provided. However, this support was limited—with the volume of projects being submitted, it was not feasible for ICES personnel to provide one-on-one support to every researcher and trainee that sought access to data. • Data dictionaries were available for ICES data holdings but had limited detail.

Factors	Impacts on data access
	• Documentation varied for external data holdings. • DAS provided extensive support to researchers, including research coordinator support to complete the required documentation and ICES analysts and scientists who work with the research team.
Mechanisms for feedback and improvement	• ICES had processes in place for checking the quality of data that come into the data environment (all general use data). There were multiple points at which the data were vetted before being made available on the analytic environment.
Organizational culture	• As an organization, ICES was described by both regulatory stakeholders and researchers as having a low risk tolerance and being conservative in their approach to providing data. At the same time, the introduction of DAS (and its predecessor CD-Link) was demonstrative of a willingness to extend access to a broader group of stakeholders. • For external data provider organizations, participants indicated that the willingness to provide data varied across organizations.
Regulatory factors	
Information legislation	• Under the Personal Health Information Protection Act (PHIPA), ICES was designated as a prescribed entity, which allowed it to collect and use data for research without individual consent. • As a prescribed entity, ICES' work was protected by the legislation. • The legislation itself did not impede research. Where issues sometimes arose was with the interpretation of the legislation. Where there was room for interpretation, individuals often interpreted the legislation conservatively. • PHIPA was unclear regarding the difference between research and health system planning and management. The two activities were treated differently within the legislation but were not clearly defined. In some cases, researchers were known to pursue access through the non-research pathway because it was perceived as easier. • Sometimes the legislation did not align with the realities of administrative health data research. For example, the PHIPA was amended with a section (3.1) stating that data had to be de-identified then linked. If data were de-identified, they could not be linked. This reflected a limited understanding of

Factors	Impacts on data access
	ICES' operations and made it impossible to comply with the legislation. • The legislation limits the extent to which processes can be streamlined. Ultimately, there are processes that are set out in the legislation, and processes that relate to ICES' designation as a prescribed entity, that must be followed and that ICES does not have the authority to change.
Transparency of data access pathway	• There were established processes for accessing data via the internal data access pathway, but these were not clearly communicated to researchers, which sometimes created frustration and made it difficult for trainees and new investigators to determine how to access data. • It was not clear to researchers what happened to applications once they were submitted (i.e., what steps they went through, and where they were in the overall process). • The pathway for accessing data via ICES-DAS was provided on the ICES website and researchers were guided through the process by DAS staff.
Complexity of data access pathway	• Key informants acknowledged that since ICES was established, new processes have been continued to be added but none have been taken away, contributing to increasing complexity.
Required forms and documents	• Overall, the required documentation (PIAs, project activation worksheets, data sharing agreements, DCPs,and additional documents for special permissions, etc.) created a major administrative workload for both researchers and the organization. • Study related forms and documents were often difficult to complete. New investigators and trainees often needed support early on to understand the forms and learn how to complete them. Incorrectly completed documents resulted in additional back and forth between the researcher and ICES Privacy and Legal Office. • Preparing the DCP was particularly challenging. This document was very detailed and contained information about the dataset and the analysis plan, including variable definitions and specific codes. Learning how to create these documents took time.
Required reviews and approvals	• The specific reviews that were required varied by study, depending on the data access pathway (i.e., internal or DAS), the researcher's institution, the researcher's program of

Factors	Impacts on data access

	research, the type of data being accessed (general access, study-specific, or controlled use), and whether the study involved an external linkage.
	• Studies involving the internal data access pathway
	  ○ Sometimes required review at the program level (e.g., cancer program), in part to reduce duplication of effort.
	  ○ Always required PIA review by the Privacy and Legal Office.
	  ○ Often required additional review and approval if there was an external linkage.
	  ○ Sometimes required "special permissions" depending on the dataset being accessed and the agreement between ICES and the data provider.
	  ○ Did not always require REB approval. Certain institutions (e.g., Sunnybrook) had agreements in place with ICES that permitted researchers accessing only data held internally by ICES to proceed without submitting a separate research ethics application, as long as the PIA was approved.
	• All studies involving access to data via DAS underwent review by DAS for feasibility and eligibility, REB review, and review by the Privacy and Legal Office, although a PIA was not required.

Scope of review	• The scope of review of the PIA that was performed by the Privacy and Legal Office was very well-defined and focused on legislative authority. Specifically, it focused on whether ICES had the legislative authority to collect and use the data proposed for a particular project, and in the case of an external linkage, whether the data provider had authority to send the data to ICES. • Review at the program level was focused on feasibility and avoiding duplication of work rather than scientific or privacy review. • DAS review primarily focused on eligibility and feasibility, though alignment with organizational values was sometimes considered (this primarily applied to private sector research).
Transparency of review	• Researchers did not have access to information about why delays occurred or what stage of review/approval their project was in. • Researchers were not always sure what criteria were used to assess their applications, or how certain decisions were made.

Factors	Impacts on data access
Application of data minimization principle	• A key feature of ICES' internal data access pathway was that there was no "disclosure" of line-level data. An ICES analyst was assigned to the study and was the only member of the research team that was permitted to access to line-level data. Rather than being permitted to access only select variables within each dataset, the analyst was permitted to access all variables within each dataset required for the study. A benefit of this approach was that it provided analysts with the flexibility to look at different variables to determine which ones to use/include in analysis, rather than being limited to a minimum dataset identified *a priori*. A challenge with this approach was that the ICES analyst was involved in making methodological decisions while the researcher was unable to confirm whether dataset creation and analysis had been carried out correctly. • Researchers who accessed data via DAS were provided a risk-reduced dataset (e.g., no dates, no birthdates, etc.). • The data minimization principle created a challenge for new and emerging analytic techniques (e.g., machine learning), highlighting the need for policies to evolve.
Role of data centre analysts	• ICES analysts had a higher level of security clearance that allowed them to view individual level data directly. Researchers did not typically have this level of access. • Analysts had access to the complete datasets that were specified in the researcher's application documents, not just a portion of the dataset. • When data were accessed by ICES researchers (not trainees) via the internal data access pathway, the analyst was considered part of the research team. As the only individuals with access to identifiable data, analysts were responsible for cleaning and linking the data, preparing the dataset, and providing results to the research team (aggregate results). They were considered partners in the project, and co-authors on resulting publications. • When data were accessed via DAS, the ICES analyst prepared a risk-reduced dataset for the research team who performed their own analyses, although in some cases DAS provided analytic services as well. • Students were typically provided with a data "cut" that they could work with directly.

Factors	Impacts on data access
Proportionality	• Studies carried out by internal ICES scientists underwent a different level of review compared to those accessed via DAS. For the DAS pathway, a PIA review was not required because researchers received a risk-reduced dataset. • For studies carried out by researchers with a Sunnybrook Hospital affiliation and involving only ICES' general use data holdings (i.e., special permissions are not required), REB review was not required. Instead, the PIA review carried out by the ICES Privacy and Legal Office was considered sufficient. The Sunnybrook Hospital REB regularly reviewed a random sample of ICES studies to ensure compliance with ethical guidelines.
Accountability	• ICES was subject to oversight from IPC of Ontario. Since ICES was able to collect data without consent, it was required to adhere to stringent requirements set out by the IPC. A lot of time and resources within the organization were dedicated to ensuring compliance with the legislation and the requirements of the IPC, as well as meeting the IPC's reporting requirements. • Accountability to the IPC limited ICES' ability to make changes to processes.
Contextual factors	
Leadership	• ICES was established at the behest of the provincial government with the specific aim of leveraging data for health system improvement. • ICES received core funding from the Ministry of Health and Long-Term Care. • Under PHIPA, ICES was designated as a prescribed entity, which gave ICES the authority to collect, use, and disclose personal health information without consent for research.
Health system organization and integration	• In 2021, the provincial government had merged a variety of healthcare organizations into a single entity referred to as "Ontario Health". The merger included Cancer Care Ontario, which had previously been designated as a prescribed entity. These changes were expected to have implications for data access (e.g., changes in data custodianship) but these were not well understood at the time of data collection.
Legislative landscape	• When external data were brought into ICES, different pieces of legislation sometimes applied depending on the data source (e.g., Pharmacy Act, Laboratory Specimens Act, Retirement Homes Regulatory Authorities Act). These

Factors	Impacts on data access
	different pieces of legislation did not always "speak to each other". For example, ICES was not a prescribed entity in these Acts, so it was sometimes difficult to determine whether they had the legislative authority to collect the data.
Historical events	• The alleged privacy breach that occurred at the Ministry of Health in British Columbia was felt in Ontario (described as a "chill").
Current events	• Throughout the COVID-19 pandemic, ICES was heavily involved in producing provincial COVID data. To facilitate reporting, new datasets were brought in, the frequency of data feeds was increased, and COVID research was prioritized (e.g., analyst time was dedicated to COVID research, etc.). These changes demonstrated what was possible in terms of improving the timeliness of data access. • Prior to COVID, individuals working with line-level data were required to work on-site. This was no longer possible when COVID-19 restrictions were implemented, so remote access was permitted. This improved access for students as it gave them more flexibility with their schedule, particularly for clinical residents. At the same time, it demonstrated that the work could be done securely even when done remotely.