Understanding challenges of using routinely collected health data to address clinical care gaps: a case study in Alberta, Canada

Taylor McGuckin [iD],[1] Katelynn Crick,[1] Tyler W Myroniuk,[2] Brock Setchell,[1] Roseanne O Yeung,[1,3] Denise Campbell-Scherer[1,4]

[1]Faculty of Medicine & Dentistry - Lifelong Learning & Physician Learning Program, University of Alberta, Edmonton, Alberta, Canada
[2]Public Health, University of Missouri, Columbia, Missouri, USA
[3]Division of Endocrinology & Metabolism, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada
[4]Department of Family Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada

Correspondence to
Dr Denise Campbell-Scherer; denise.campbell-scherer@ualberta.ca

ABSTRACT

High-quality data are fundamental to healthcare research, future applications of artificial intelligence and advancing healthcare delivery and outcomes through a learning health system. Although routinely collected administrative health and electronic medical record data are rich sources of information, they have significant limitations. Through four example projects from the Physician Learning Program in Edmonton, Alberta, Canada, we illustrate barriers to using routinely collected health data to conduct research and engage in clinical quality improvement. These include challenges with data availability for variables of clinical interest, data completeness within a clinical visit, missing and duplicate visits, and variability of data capture systems. We make four recommendations that highlight the need for increased clinical engagement to improve the collection and coding of routinely collected data. Advancing the quality and usability of health systems data will support the continuous quality improvement needed to achieve the quintuple aim.

INTRODUCTION

A learning health system is foundational to achieving the quintuple aim of advancing patient care, population health, equity, cost-effectiveness, healthcare worker experience, and, ultimately, future goals such as precision health.[1–3] To be able to rapidly answer important clinical questions, the structure of, and data capture in, electronic medical records and health administrative databases needs to be improved. Alberta, Canada is a globally recognised jurisdiction for its health data infrastructure and capture. However, health service researchers have identified important limitations to its use.[4–8] Reasons for these limitations include the historic use of different health information systems across Alberta's regions,[9] and the creation of administrative health databases for non-clinical functions such as payment.[10]

The Physician Learning Program (PLP)[11] is a provincial programme that works to understand gaps in clinical practice, create clinically actionable information and cocreate sustainable solutions with physicians, allied health teams, patients and community, and health system partners to advance practice. Here, we share four examples of PLP projects on a range of rare to common medical conditions that highlight some of the current challenges of using routinely collected health data to inform real-world clinical problems and support quality improvement. These four projects demonstrate areas where we encountered limitations in data capture, which if rectified, would provide needed information to help advance care of Albertans. We offer guidance in improving routinely collected health data that is broadly relevant to health systems by addressing issues of data completeness, availability, missingness and duplication, and variability in capture. Improvements in these areas are necessary to increase the usability of data for healthcare, health services research, and, eventually, future applications of artificial intelligence and precision health.

METHODS

The primary objective of this work was to capture, categorise, and label overarching and recurring problematic data patterns in electronic health records and administrative databases observed through work conducted at PLP. The four projects presented were conducted to understand gaps in clinical care and develop baseline data for quality improvement initiatives. Each project is described in table 1, with notes on data sources in table 2. For each project, a series of questions were co-created with clinicians to provide information of importance for clinical quality improvement. We identified whether secondary data from electronic medical records and administrative databases was available or whether primary data collection was necessary. Routinely collected

Table 1 Description of the Physician Learning Program projects including purpose, representative questions, whether a challenge was encountered, and databases used

Project	Clinical questions	Challenge encountered in answering the question	Databases used
Adult Diabetes (n=77 782 patient-visits, 11 714 unique patients) Objective: To understand the state of electronic medical record and administrative health data as the foundation for cocreating meaningful quality improvement projects.	(1) What are the demographic characteristics of patients seen in the outpatient diabetes clinics in the Edmonton zone?	No	▶ eClinician electronic medical record ▶ Physician Claims ▶ Alberta Health Services Labs ▶ Pharmaceutical Information Network (PIN)
	(2) What proportion of patients seen in the outpatient Edmonton zone have a comorbidity?	Yes	
	(3) What is the breakdown of diabetes by diagnosis among the patients visiting the outpatient clinics in the Edmonton zone?	Yes	
	(4) What are the processes of care for lab and biometric measurement at each of the diabetes clinics?	Yes	
	(5) What medications have patients been dispensed from community pharmacies?	No	
	(6) How well is diabetes being managed including lipids, blood sugar control, blood pressure control and renal protection?	Yes	
Paediatric Diabetic Ketoacidosis (n=929 patient admissions) Objective: To evaluate the degree to which paediatric patients in Alberta are being managed according to the 2018 Diabetes Canada Clinical Practice Guidelines.[35]	(1) How many admissions at Alberta hospitals are for diabetic ketoacidosis from 1 January 2015 to 31 December 2018?	No	▶ Discharge Abstract Database ▶ National Ambulatory Care Reporting System (NACRS) ▶ PIN ▶ Sunrise Clinical Manager ▶ Alberta Health Services Labs ▶ Diagnostic Imaging
	(2) What are the demographic characteristics of patients being admitted for diabetic ketoacidosis?	No	
	(3) Where are patients admitted to hospitals in Alberta with diabetic ketoacidosis being cared for and what are the referral pathways?	No	
	(4) What medications, fluids, and electrolytes are administered during admission for diabetic ketoacidosis?	Yes	
	(5) Is care for diabetic ketoacidosis concordant with national diabetes guidelines?	Yes	
Adrenal Insufficiency (n=211 207 patient visits) Objective: To estimate the prevalence of adrenal insufficiency in Alberta and visit rates among this patient population.	(1) What is the 5 year period prevalence of adrenal insufficiency in Alberta, Canada?	Yes	▶ NACRS ▶ Physician Claims ▶ PIN
	(2) What are the rates of emergency room and outpatient healthcare utilisation among patients with adrenal insufficiency?	Yes	
	(3) What proportion of patients with adrenal insufficiency have been dispensed a glucocorticoid and/or mineralocorticoid during the study period?	No	

Continued

Table 1 Continued

Project	Clinical questions	Challenge encountered in answering the question	Databases used
Beta-Lactam Allergy and Surgical Prophylaxis (n=3 218 patient surgeries) Objective: To assess whether 2018 surgical prophylaxis recommendations, indicating use of cefazolin in patients with a beta-lactam allergy are being followed.[36]	(1) Are patients with a beta-lactam allergy receiving the correct antimicrobial prophylaxis in-hospital for their surgery according to guidelines?	Yes	N/A (Paper chart audit required)
	(2) Are patients receiving antimicrobial prophylaxis within the guideline recommended timeframe?	Yes	
	(3) Are patients receiving postoperative prophylaxis in accordance with local guidelines?	Yes	

N/A, not available.

health data from electronic medical records and other administrative databases was feasible and extracted for three projects: (1) Adult Diabetes; (2) Paediatric Diabetic Ketoacidosis, a serious complication of diabetes and (3) Adrenal Insufficiency, a rare, life-threatening hormonal disorder. For the Beta-Lactam Allergy and Surgical Prophylaxis project, the required clinical information was not routinely collected into an administrative database. Thus, primary data collection was required and included manually extracting information from paper charts.

Figure 1 represents the iterative process used to identify, collect, clean and synthesise routinely collected health information needed for clinical quality improvement. Detailed methods and results of the four projects will be published elsewhere. The data collected and analysed for this paper is not the quantitative data of the four projects, but our observations while conducting them. Briefly, for the projects that used routinely collected health data, we formulated a data query to find and pull the raw data needed to answer each project question. A trained analyst employed by Alberta Health Services extracted the data. Once extracted, the raw data were cleaned and analysed using standard statistical software (Oracle SQL Developer, Python V.3.4, SAS V.9.4 and RStudio V.1.2.5033). The clinicians working on the project reviewed the results to assess the validity and completeness of the data in comparison with their knowledge of clinical workflow and processes. The results were compiled into various formats, including presentations, reports, infographics, and clinical tools, and then disseminated to relevant stakeholder groups. Their purpose ultimately is to inform clinical quality improvement and co-creation of interventions to address clinical gaps in care.

Systematic approach used to capture and categorise main challenges and identify root causes

Over a 2-year period, recurring difficulties arose when obtaining and analysing the administrative data needed to answer clinical questions for the four projects. We undertook a systematic approach to identify and capture whenever problems arose and then categorise them into main challenges. This systematic approach included: (1) capturing whenever a data problem occurred in a project; (2) discussing the problem within our interdisciplinary team of researchers and clinical experts; (3) discussing recurring issues and patterns through team meetings and with key informant discussions; and (4) synthesising them into main categories that spanned projects, healthcare settings, and health conditions. We identified and verified the root cause whenever possible by: (1) talking to clinical, administrative, and analytical staff within Alberta Health Services and Alberta Health (two regulatory government bodies that oversee the delivery of healthcare within the province of Alberta); (2) reading publicly available database documentation[12–14] and (3) talking to front-line healthcare staff with deep knowledge of the healthcare setting and clinical systems. Our systematic approach is summarised in box 1.

Patient and public involvement

At the PLP, we have the mission to create "actionable clinical information and engage with physicians, teams and partners to cocreate sustainable solutions to advance practice."[11] Inherent in this process is the involvement of broader networks outside of the project team including community physicians, physician networks, policy-makers, patients, researchers, and other healthcare professionals. Involvement of stakeholders starts at project conception with physicians and clinical teams cocreating project ideas with the PLP based on health system gaps. Engagement continues through to the dissemination of project outcomes where we integrate with networks to engage in knowledge translation activities, codesign sustainable solutions, and implement them with health system partners.

RESULTS

Through our systematic approach of capturing and categorising recurring problems, as outlined in detail above,

Table 2 Descriptions of the data sources used to complete the projects

Data source	Description
National Ambulatory Care Reporting System (NACRS)	▶ Governed nationally by the Canadian Institute for Health Information. ▶ Contains data for hospital-based and community-based ambulatory care including day surgery, outpatient and community-based clinics, and emergency departments. ▶ Submission of emergency room visit data to NACRS is mandatory in Alberta. ▶ Emergency room data are abstracted from patient charts by trained data extractors following standards set by the Canadian Institute for Health Information. ▶ Outpatient data are sent via non-abstracted formats, and data collection methods vary by clinic. ▶ Submission requirements determined at the clinic level for outpatient settings.
Discharge Abstract Database	▶ Governed nationally by the Canadian Institute for Health Information. ▶ Captures information from admissions to acute care facilities in the province. ▶ Mandatory for all acute care facilities to submit data. ▶ Mandatory fields vary by geographic location.
Diagnostic Imaging	▶ Provincial database made up of the 3 Radiology Information Systems (RIS) in Alberta: Cerner Millennium (Calgary), Agfa RIS (Edmonton), and Meditech (Aspen, Chinook, David Thompson, East Central, Northern Lights, Palliser and Peace Country). ▶ Contains information on diagnostic imaging tests (eg, MRI and CT scans).
Alberta Health Services Labs	▶ Contains lab results from the 4 Lab Information Systems in the province: Meditech, Millennium, Sunquest, LabFusion. ▶ Data is captured in both standardised (eg, categorical) and unstandardised (eg, free text) formats.
Physician Claims	▶ Captures 'claims submitted for payment of Alberta service providers for health services delivered under the Alberta Healthcare Insurance Plan'. ▶ Data elements include patient information, provider information, and service information such as health service code, date of service, amount paid, facility, up to three diagnostic codes, and shadow billed claims (service data optionally submitted by physicians on alternative payment plans). ▶ Mandatory for fee-for-service physicians to submit visit information. ▶ Practical differences in reporting processes between fee-for-service and alternative payment plan physician results in inconsistent data capture.
Pharmaceutical Information Network	▶ Community pharmacies are mandated to report prescription medication dispenses within 24 hours of dispensing. ▶ Includes information such as drug dispense date and drug information details (eg, drug identification number).
Sunrise Clinical Manager	▶ Clinical Information System used exclusively in the Calgary Zone. ▶ Captures rich information such as demographics, allergies, orders (eg, lab, diagnostic imaging, medications), medication administrations, results and diagnoses for patients in acute care facilities, emergency departments and some outpatient clinics.
eClinician electronic medical record	▶ Used at all outpatient diabetes clinics in the Edmonton Zone. ▶ Is an 'integrated information management platform supporting the collection, access, use and sharing of information supporting the delivery of health services to persons and populations in multiple settings across the continuum of care'.[37]

we identified four broad challenges of using routinely collected health data to address real-world clinical questions. We present them here framed in four example projects. These four challenges and example project questions are summarised in table 3.

Description of challenges

Challenge 1: are the data field(s) needed to answer the clinical question available in administrative databases?

Not all information collected at a patient encounter has a corresponding data field in an administrative database; some information, although available, is not abstracted from the patient chart into a database. In the beta-lactam allergy and surgical prophylaxis project, 0 out of 3218 audited surgical cases contained allergy information in an available administrative database because there was no routinely populated data field for this information. However, for all cases, we found that allergies were recorded in paper charts. Importantly, inappropriate antibiotic prophylaxis, due to allergy status, is associated with a 50% increase odds in surgical site infections and increased costs to the system.[15] Assessing care using paper chart audits is sometimes justified but is not sustainable or scalable on a large basis because of its resource intensiveness. We are now working with health system delivery
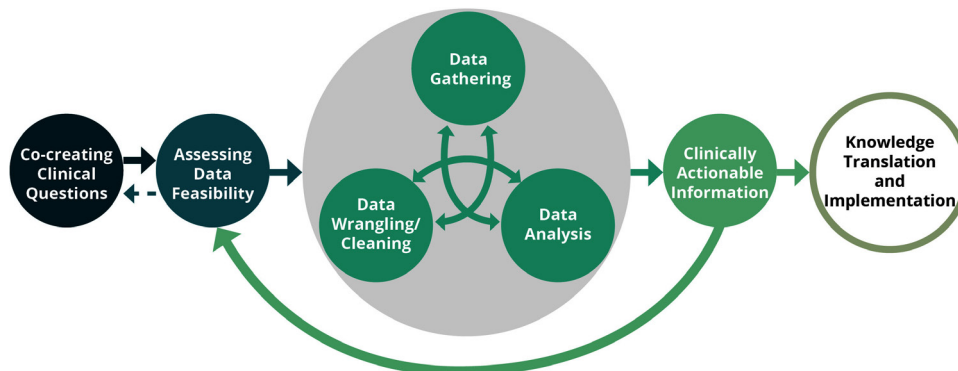
Figure 1 The Physician Learning Program's non-linear process of quality improvement using routinely collected health data. The key elements are: (1) cocreating clinical questions and identifying whether secondary data are available or if primary data collection is necessary; (2) gathering data from databases or completing primary data collection; (3) deep cleaning of the data; (4) conducting analyses and further data cleaning; and (5) effectively communicating findings that serve as the basis for quality improvement.

stakeholders to develop more sustainable solutions for this problem of antibiotic allergy and prophylaxis information not being electronically captured and available, specifically.

For the paediatric diabetic ketoacidosis project, only 28.6% of children's admissions across Alberta contained data on medication, electrolyte, and fluid administration. Guideline concordance of care for this life-threatening condition cannot be assessed without this information. This information was only available for patients whose encounter was at a site that used Sunrise Clinical Manager, a specific Clinical Information System. Only five Alberta Hospitals and Health Centres, out of over 100 included in our project, used this system inhibiting the feasibility of assessing guideline concordant care across the whole system.

When assessing patient comorbidities in the adult diabetes project, we could not determine whether patients had a history of hyperosmolar hyperglycaemic state. Despite the International Classification of Diseases-9 (ICD-9) having a corresponding code for this condition, Alberta Health's coding taxonomy, which is used

to capture visit information to pay providers across the province, does not include all ICD-9 codes.[16] Thus, this comorbidity could not be assessed for any of the patients.

Challenge 2: if the data field needed to answer the clinical question is available, is the information complete and accurate?

The completeness of extracted data was problematic in two of our projects. When assessing lab results in the paediatric diabetic ketoacidosis project, we found that 46.6%, 94.5% and 12.6% of admissions at one of the children's hospitals in the province had no results for blood pH, blood bicarbonate, and blood glucose, respectively. These laboratory results are central to guiding diabetes care and confirming a diagnosis of diabetic ketoacidosis. Through our root cause analysis, which included consulting with experts in the hospital laboratory, we uncovered that laboratory tests completed from capillary blood sources may not flow from bedside instruments to administrative databases; a historical legacy of funding restrictions when the system was developed. Additionally, we observed incomplete medication, fluid, and electrolyte administration data, which are all necessary for assessing quality of care in relation to established guidelines.

In the adult diabetes project, routinely collected health data were often missing for measures such as blood pressure, an important clinical assessment for predicting disease complications. In one clinic, 65.5% of visits did not have a blood pressure measurement recorded in a database. Through consultation, we determined that while front line staff are entering these measures into the electronic medical record, it does not flow into administrative databases.

Challenge 3: can the number of visits for a particular medical condition be accurately measured using administrative data?

We were unable to accurately estimate the number of outpatient visits for the treatment of adrenal insufficiency due to visits missing from the databases. Missing visits are a consequence of both imprecise codes used at the time of data submission (eg, visits coded as 'follow-up') and

Box 1 Methods used to identify, collect and analyse the raw data (ie, problems arising in using administrative data to answer the clinical questions)

Methods to identify the raw data
► Observe when there was a problem while conducting each of the steps in figure 1.
► Verify if there was a challenge by checking against known published problems and discussing with data analysts and clinicians to see if it matches reality.

Methods to collect the raw data
► Formally document the problem encountered and how it was verified .

Methods used to analyse the raw data
► Discuss the problems from each project and collate and summarise them into overarching themes (main challenges).

Table 3 Data challenges encountered while answering clinical questions

Project questions and data challenges encountered	Challenge 1: Are the data field(s) needed to answer the clinical question available in administrative databases?	Challenge 2: If the data field needed to answer the clinical question is available, is the information complete and accurate?	Challenge 3: Can the no of visits for a particular medical condition be accurately measured using administrative data?	Challenge 4: Can laboratory tests across the province be identified, harmonised, and analysed?
Adult Diabetes				
What proportion of patients seen in the outpatient Edmonton zone have a comorbidity?	✗	✗	N/A	✗
What is the breakdown of diabetes by diagnosis among the patients visiting the outpatient clinics in the Edmonton zone?	✔	✗	N/A	N/A
What are the processes of care for lab and biometric measurement at each of the diabetes clinics?	✗	✗	N/A	✗
How well are patients' diabetes being managed including lipids, blood sugar control, blood pressure control, and renal protection?	✔	✗	N/A	✗
Paediatric Diabetic Ketoacidosis				
What medications, fluids, and electrolytes are administered during admission for diabetic ketoacidosis?	✗	✗	N/A	N/A
Is care for diabetic ketoacidosis concordant with national diabetes guidelines?	✗	✗	N/A	✗
Adrenal Insufficiency				
What is the 5-year period prevalence of adrenal insufficiency in Alberta, Canada?	✔	✗	✗	N/A
What is the rate of emergency room and outpatient healthcare utilisation among patients with adrenal insufficiency?	✔	✗	✗	N/A
Beta-Lactam Allergy and Surgical Prophylaxis				
Are patients with a beta-lactam allergy receiving the correct antimicrobial prophylaxis in-hospital for their surgery according to guidelines?	✗	N/A	N/A	N/A
Are patients receiving antimicrobial prophylaxis within the guideline recommended time frame?	✗	N/A	N/A	N/A

Continued

Table 3 Continued

Project questions and data challenges encountered	Challenge 1: Are the data field(s) needed to answer the clinical question available in administrative databases?	Challenge 2: If the data field needed to answer the clinical question is available, is the information complete and accurate?	Challenge 3: Can the no of visits for a particular medical condition be accurately measured using administrative data?	Challenge 4: Can laboratory tests across the province be identified, harmonised, and analysed?
Are patients receiving postoperative prophylaxis in accordance with local guidelines?	✗	N/A	N/A	N/A

✔ The challenge was encountered in the project.
✗ The challenge was not encountered in the project.
N/A, not applicable.

variation in data submission requirements in which not all visits are required to be submitted and thus captured. Variation in data submission requirements are a result of various payment structures (eg, alternative payments plans) across and within regions of the province. Thus, it is uncertain of how to compare regional data.

Furthermore, we encountered difficulty reconciling duplicate entries within and between databases housing different aspects of clinical visits. In this example, both Physician Claims and the National Ambulatory Care Reporting System (NACRS) database are used to capture outpatient visit data. They capture much of the same information but use different taxonomies to capture diagnostic information: one uses ICD-9 where the other uses ICD-10. Some visits are captured only in Physician Claims or NACRS, some in neither, and some in both.[12–14 17] There is no official reconciliation for visits captured in both. We found that at least 27% of adrenal insufficiency visits were likely duplicates. Of the 211 207 visits analysed, only 5.7% had a diagnostic code for adrenal insufficiency; clinical colleagues insisted this was implausibly low. This raised concerns that an indeterminate number of visits were missing from both databases, which may be due to visits for more than one medical condition not capturing all of the relevant diagnoses. In 78% of visits, there was only one code provided for the visit. Most codes used for the analysed visits were vague such as 'general examination' and 'follow-up', making it difficult to identify visits related to the treatment of adrenal insufficiency, which likely contributed to this discrepancy.

Challenge 4: can laboratory tests across the province be identified, harmonised, and analysed?

Three laboratory information systems are used across Alberta–a historical legacy of healthcare regionalisation. Laboratory codes are not harmonised across any of the laboratory databases within the province of Alberta. Each laboratory information system uses different laboratory codes, and thus, identifying and matching relevant codes across databases is not a trivial task. For example, haemoglobin A1c, a diabetes test, was found to be coded as HbA1, ZHBA1C and HBA1X depending on where the lab test was completed. One major consequence was

that 919 laboratory codes had to be reviewed to identify and harmonise the codes used in the paediatric diabetic ketoacidosis project. This was also problematic for the adult diabetes project (online supplemental table 1).

Strengths and limitations of the databases used

Through completing these four projects, we identified both strengths of limitations of the administrative databases for informing clinical quality improvement projects. Strengths and limitations in relation to our example projects and questions specifically, and cautions for their use are summarised in table 4. This is not a comprehensive overview of the strengths and limitations of these databases, but rather a summation of our experiences.

DISCUSSION

Rapid access to clinically important information is crucial to building a powerful learning health system[3] in pursuit of the quintuple aim. Health data infrastructure that supports rapid access to clinically important information for evidence-informed care and clinical quality improvement is key to supporting practice reflection and innovations to meet patient needs. Our PLP projects illuminate four challenges of using routinely collected health data to achieve these aims. First, we found that not all information collected in a patient encounter has a corresponding data field in an administrative database; costly, time-consuming primary data collection is then needed to assess important clinical questions prohibiting the feasibility of continual monitoring. Second, when data fields are available, they may be absent or not uniformly populated. For instance, we observed this problem when clinical evaluations or readings from bedside instruments are used and the information does not flow to administrative databases. Third, establishing prevalence of medical conditions and number of visits was difficult due to missing records, complexity reconciling various databases that contain the same information, inconsistent diagnostic coding practices, and differing taxonomies used between databases. A key element of this challenge was that imprecise diagnostic codes, such as 'follow-up', did not permit clarity as to the topics addressed in the

Table 4 Strengths and limitations of the databases as elucidated by our example projects

Database name	Strengths encountered	Limitations/challenges/cautions encountered
National Ambulatory Care Reporting System (NACRS)	► Captures emergency room visit data in Alberta ► Emergency room data are abstracted in a standardised fashion by trained data extractors ► Includes up to 10 diagnostic fields ► National database allows for interprovincial comparisons ► Quality control by the Canadian Institute for Health Information	► Not mandatory for all outpatient visit data to be submitted in Alberta, therefore outpatient visits may be missed ► Unstandardised data capture and coding for outpatient visits may lead to missing data and makes analysis and interpretation difficult ► No reconciliation with Physician claims database
Discharge Abstract Database	► Captures acute care facility discharges in the province ► National database allows for interprovincial comparisons ► Quality control by the Canadian Institute for Health Information	► None identified
Diagnostic imaging	► Contains information about diagnostic imaging (eg, CT and MRI)	► None identified
Alberta Health Services Labs	► Access to lab data collected across the province is available for labs ordered and paid for by Provincial Health Authority. Labs ordered and paid for by other parties are removed	► Use of 3 different systems across the province making province-wide analysis difficult ► Labs taken using beside instruments may not flow into administrative databases ► Heavy use of free text fields making analysis difficult without proper cleaning and data analytic skills
Physician claims	► Captures data on emergency, community and in-hospital physician services provided across the province ► Captures all services provided by fee-for-service physicians and some services provided by physicians on alternative payment plans (ie, shadow billed claims)	► Does not capture all visits as shadow bill submissions by alternative payment plan physicians to Provincial Health Authority varies by clinic ► No reconciliation with the NACRS database makes the identifcation of duplicate data challenging ► Only up to three diagnostic codes are captured, with only one being mandatory for outpatient visits, therefore not all conditions treated within a visit may be captured ► Unspecific billing codes used (eg, general follow-up) ► Variation in coding practice among physicians
Pharmaceutical Information Network	► Captures prescription dispenses from community pharmacies ► Includes information such as drug dispense date and drug information details (eg, drug identification number).	► Does not capture in-hospital medication dispenses or whether medication was taken by the patient ► Cannot make conclusions about physicians prescribing patterns as unfilled prescriptions are not captured

Continued

Table 4 Continued

Database name	Strengths encountered	Limitations/challenges/cautions encountered
Sunrise Clinical Manager	► Rich source of information for clinical bedside care	► Provincial Health Authorities warn that there is variation in use and therefore this data source must be used with caution.[13] ► Trauma room may not be captured ► Not used across the province ► Heavy use of free text fields requiring advanced and resource-intensive analytical skills ► Contains both tasks that were performed and tasks that were ordered but not performed (eg, medications)
eClinician electronic medical record	► Rich source of information for clinical bedside care	► Variation in coding behaviours across clinics ► Unclear dataflow and mapping from bedside entry to extracted databases ► Incomplete data capture in some fields. May reflect variation in use across clinics ► Multiple fields capture similar information (eg, problem list vs encounter table) ► Not used across the province

visit. The fourth challenge was the multiplicity of laboratory diagnostic codes used for the same test which made it difficult to develop data queries that capture all relevant tests.

The mission of the PLP is to create actionable clinical information and engage with physicians, teams, patients, and partners to cocreate sustainable solutions to advance practice. The creation of clinically actionable information from routinely available health data is hindered when there are substantial gaps in the information, as measuring improvement requires relevant baseline data and measurement over time to assess change. The strengths and limitations of administrative and electronic medical record health databases have been described extensively, for instance in the work of Burles *et al*, Clement *et al* and Edmondson and Reimer.[18–20] The inability to analyse data in real time is not a problem unique to the Canadian context, with challenges being documented in other jurisdictions including the USA.[21] The overarching issues relating to data capture, completeness, accuracy, and harmonisation, exist across healthcare systems and settings and challenges with data capture in clinical electronic medical records have been well documented.[22–27] Several of the databases outlined are available across Canada, including Discharge Abstract Database and NACRS, and thus these challenges are likely to exist across the country. Ongoing work is being conducted by the PLP and with relevant stakeholder groups to address the issues presented. We acknowledge the importance of collaborating with various stakeholders including data scientists, clinicians, and administrators to fully understand what the meaningful clinical data are and how to mobilise and act on them so that data-driven quality improvement is supported. Increased coordination and leveraging the opportunity of a new provincial acute care electronic medical record should continue to advance this work, particularly as efforts evolve across the care continuum.

Future directions

Advancing the quality of health systems data is crucial not only for current quality improvement projects, but also in realising the utility of precision health and artificial intelligence to advance healthcare in the future.[28–33] Health system data are necessary to meet the Federation of Medical Regulatory Authorities in Canada's goal that all Canadian physicians participate in data-driven practice quality improvement.[33] The overarching purpose of these efforts is to support the development of a learning health system and to achieve improvements in the quintuple aim of improving population health, patients' experience of care, equity, cost-effectiveness, and sustainability of healthcare workforce.[1–3] We strongly believe that the long-term benefits of improved data capture would significantly offset upfront investments. Importantly, supporting these efforts requires mobilising clinical information in a way that does not overwhelm the clinical workforce and contribute to physician burnout.[34]

Addressing these four identified challenges is fundamental to creating a learning health system and to advancing healthcare delivery and health outcomes. We recommend the following:

1. To have more clinically important data available in readily extractable formats, we suggest expanding and harmonising mandatory data submission requirements with increased clinician engagement to ensure data that is captured is clinically meaningful.

2. To increase the quality and validity of the data available to assess patient care, we suggest the use of more specific codes and consistent taxonomies across the healthcare system to capture encounter diagnoses; standardisation of data entry processes with clear mechanisms of training and maintenance; and, ensuring the flow of clinically important information from bedside instruments, laboratory settings, and diagnostic imaging results to administrative databases in analysable formats.

3. To enhance efficiency and speed of data capture so that upgrading data quality, quantity, and structure is not at the cost of the clinical user, we suggest the incorporation of technologies like natural language processing, cross-platform interoperability, and application of human-centred design for workflow process improvement.

4. To promote real-time usability of data, we propose integrating technologies such as natural language processing and artificial intelligence to automate routinised functions to support appropriate real-time clinical decisions and reduce clinician burden.

Limitations

The challenges we identified in our routinely collected health data are specific to Alberta, Canada, however, they are commonly encountered in conducting QI and research work using administrative data and are generalisable internationally.[22–27] As information technology advances, integration into different health systems is variable leading to different local challenges in deriving solutions. We submit that the principles stated here may be of interest for consideration but additional factors will exist in different jurisdictions.

CONCLUSION

Through practical, real-world projects, we have identified four challenges in using administrative health and electronic medical record data to address clinical care gaps. Improving data infrastructure and quality will enable more nimble quality improvement efforts and real-world evidence studies. Improving this infrastructure, and the reliability and validity of data, is a necessary precondition for emergent technologies in precision health and artificial intelligence, and to developing a learning health system.

ORCID iD
Taylor McGuckin http://orcid.org/0000-0002-8800-6081

REFERENCES

1 Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014;12:573–6. doi:10.1370/afm.1713

2 R. Privitera M. Addressing Human Factors in Burnout and the Delivery of Healthcare: Quality & Safety Imperative of the Quadruple Aim. *Health* 2018;10:629–44. doi:10.4236/health.2018.105049

3 Friedman C, Rubin J, Brown J, *et al*. Toward a science of learning systems: a research agenda for the high-functioning learning health system. *J Am Med Inform Assoc* 2015;22:43–50. doi:10.1136/amiajnl-2014-002977

4 Doyle CM, Lix LM, Hemmelgarn BR. Data variability across Canadian administrative health databases: differences in content, coding, and completeness. *Pharmacoepidemiol Drug Saf* 2020;29 Suppl 1. doi:10.1002/pds.4889

5 Drews SJ, Simmonds K, Usman HR, *et al*. Characterization of enterovirus activity, including that of enterovirus D68, in pediatric patients in Alberta, Canada, in 2014. *J Clin Microbiol* 2015;53:1042–5. doi:10.1128/JCM.02982-14

6 Ho C, Guilcher SJT, McKenzie N, *et al*. Validation of algorithm to identify persons with non-traumatic spinal cord dysfunction in Canada using administrative health data. *Top Spinal Cord Inj Rehabil* 2017;23:333–42. doi:10.1310/sci2304-333

7 Jolley RJ, Quan H, Jetté N, *et al*. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open* 2015;5:e009487. doi:10.1136/bmjopen-2015-009487

8 Peng M, Lee S, D'Souza AG, *et al*. Development and validation of data quality rules in administrative health data using association rule mining. *BMC Med Inform Decis Mak* 2020;20:75. doi:10.1186/s12911-020-1089-0

9 Donaldson C, Fire DC. Fire, Aim… ready? Alberta's big bang approach to healthcare disintegration. *Healthc Policy* 2010;6:22–31.

10 Lucyk K, Lu M, Sajobi T, *et al*. Administrative health data in Canada: lessons from history. *BMC Med Inform Decis Mak* 2015;15:69. doi:10.1186/s12911-015-0196-9

11 Physician learning program, 2021. Available: https://www.albertaplp.ca/ [Accessed 13 Jul 2021].

12 Government of Alberta. Overview of administrative health databases, 2017. Available: https://open.alberta.ca/dataset/657ed26d-eb2c-4432-b9cb-0ca2158f165d/resource/38f47433-b33d-4d1e-b959-df312e9d9855/download/research-health-datasets.pdf

13 Alberta Health Services. Analytics (DIMR) Alberta health services data Repository for reporting, 2016. Available: https://www.ualberta.ca/medicine/media-library/research/faculty/clin-res/spor-available-datasets.pdf

14 Canadian Institute of Health Information. Data quality documentation national ambulatory care reporting system, 2020. Available: https://

www.cihi.ca/sites/default/files/document/nacrs-data-quality-current-year-information-2019-2020-en.pdf

15 Blumenthal KG, Ryan EE, Li Y, *et al*. The impact of a reported penicillin allergy on surgical site infection risk. *Clin Infect Dis* 2018;66:329–36. doi:10.1093/cid/cix794

16 Government of Alberta. Alberta health diagnostic codes: claims assessment, 2018. Available: https://open.alberta.ca/dataset/2e226bac-1b8b-4b6a-b26c-546089a39bab/resource/f06ba4a2-7ff6-41ed-a70f-35dc28b826e6/download/diagnostic-code-icd-9.pdf

17 Cunningham CT, Cai P, Topps D, *et al*. Mining rich health data from Canadian physician claims: features and face validity. *BMC Res Notes* 2014;7:682. doi:10.1186/1756-0500-7-682

18 Burles K, Innes G, Senior K, *et al*. Limitations of pulmonary embolism ICD-10 codes in emergency department administrative data: let the buyer beware. *BMC Med Res Methodol* 2017;17:89. doi:10.1186/s12874-017-0361-1

19 Clement FM, James MT, Chin R, *et al*. Validation of a case definition to define chronic dialysis using outpatient administrative data. *BMC Med Res Methodol* 2011;11:25. doi:10.1186/1471-2288-11-25

20 Edmondson ME, Reimer AP. Challenges frequently encountered in the secondary use of electronic medical record data for research. *Computers, Informatics, Nursing* 2020;38:338–48. doi:10.1097/CIN.0000000000000609

21 Hersh WR, Weiner MG, Embi PJ, *et al*. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 2013;51:S30–7. doi:10.1097/MLR.0b013e31829b1dbd

22 Madden JM, Lakoma MD, Rusinak D, *et al*. Missing clinical and behavioral health data in a large electronic health record (EHR) system. *J Am Med Inform Assoc* 2016;23:1143–9. doi:10.1093/jamia/ocw021

23 Wright A, McCoy AB, Hickman T-TT, *et al*. Problem list completeness in electronic health records: a multi-site study and assessment of success factors. *Int J Med Inform* 2015;84:784–90. doi:10.1016/j.ijmedinf.2015.06.011

24 Kern LM, Malhotra S, Barrón Y, *et al*. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. *Ann Intern Med* 2013;158:77-83. doi:10.7326/0003-4819-158-2-201301150-00001

25 Wang EC-H, Wright A. Characterizing outpatient problem list completeness and duplications in the electronic health record. *J Am Med Inform Assoc* 2020;27:1190–7. doi:10.1093/jamia/ocaa125

26 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 Serv Res* 2017;17:766.

27 Feder SL. Data quality in electronic health records research: quality domains and assessment methods. *West J Nurs Res* 2018;40:753–66.

28 Cohoon TJ, Bhavnani SP. Toward precision health: applying artificial intelligence analytics to digital health biometric datasets. *Per Med* 2020;17:307–16. doi:10.2217/pme-2019-0113

29 Hudaa S, Setiyadi DBP, Lydia SE. Natural language processing utilization in healthcare. *International Journal of Engineering and Advance Technology* 2019;8.

30 Shaban-Nejad A, Michalowksi M, eds. *Precision health and medicine: A digital revolution in healthcare*. Switzerland: Springer Nature, 2020.

31 Maddox TM, Rumsfeld JS, Payne PRO. Questions for artificial intelligence in health care. *JAMA* 2019;321:31. doi:10.1001/jama.2018.18932

32 Wang F, Casalino LP, Khullar D. Deep learning in Medicine-Promise, progress, and challenges. *JAMA Intern Med* 2019;179:293. doi:10.1001/jamainternmed.2018.7117

33 Federation of Medicine Regulatory Authorities Canada. Physician practice improvement, 2016. Available: http://fmrac.ca/wp-content/uploads/2016/04/PPI-System_ENG.pdf [Accessed 24 Feb 2021].

34 Ye J. The impact of electronic health record-integrated patient-generated health data on clinician burnout. *J Am Med Inform Assoc* 2021;28:1051–6.

35 Diabetes Canada clinical practice guidelines expert Committee. Diabetes Canada 2018 clinical practice guidelines for the prevention and management of diabetes in Canada. *Can J Diabetes* 2018;42:S1–325.

36 Alberta Health Services. AHS recommended drug regimens for surgical prophylaxis in adult patients, 2018. Available: https://www.albertahealthservices.ca/assets/info/hp/as/if-hp-as-surgical-prophylaxis.pdf [Accessed 24 Feb 2021].

37 Alberta Health Services. Connect Care - Glossary, 2021. Available: https://www.albertahealthservices.ca/info/Page15455.aspx#cis [Accessed 24 Feb 2021].