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

Taylor McGuckin, Katelynn Crick, Tyler W Myroniuk, Brock Setchell, Roseanne O Yeung, Denise Campbell-Scherer

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

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

Continued
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 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.” 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,
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. 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

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

<table>
<thead>
<tr>
<th>Data source</th>
<th>Description</th>
</tr>
</thead>
</table>
| 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 |
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. 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

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.

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

<table>
<thead>
<tr>
<th>Project questions and data challenges encountered</th>
<th>Challenge 1: Are the data field(s) needed to answer the clinical question available in administrative databases?</th>
<th>Challenge 2: If the data field needed to answer the clinical question is available, is the information complete and accurate?</th>
<th>Challenge 3: Can the number of visits for a particular medical condition be accurately measured using administrative data?</th>
<th>Challenge 4: Can laboratory tests across the province be identified, harmonised, and analysed?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adult Diabetes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What proportion of patients seen in the outpatient Edmonton zone have a comorbidity?</td>
<td>✗</td>
<td>✗</td>
<td>N/A</td>
<td>✗</td>
</tr>
<tr>
<td>What is the breakdown of diabetes by diagnosis among the patients visiting the outpatient clinics in the Edmonton zone?</td>
<td>✔</td>
<td>✗</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>What are the processes of care for lab and biometric measurement at each of the diabetes clinics?</td>
<td>✗</td>
<td>✗</td>
<td>N/A</td>
<td>✗</td>
</tr>
<tr>
<td>How well are patients’ diabetes being managed including lipids, blood sugar control, blood pressure control, and renal protection?</td>
<td>✔</td>
<td>✗</td>
<td>N/A</td>
<td>✗</td>
</tr>
<tr>
<td><strong>Paediatric Diabetic Ketoacidosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What medications, fluids, and electrolytes are administered during admission for diabetic ketoacidosis?</td>
<td>✗</td>
<td>✗</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Is care for diabetic ketoacidosis concordant with national diabetes guidelines?</td>
<td>✗</td>
<td>✗</td>
<td>N/A</td>
<td>✗</td>
</tr>
<tr>
<td><strong>Adrenal Insufficiency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the 5-year period prevalence of adrenal insufficiency in Alberta, Canada?</td>
<td>✔</td>
<td>✗</td>
<td>N/A</td>
<td>✔</td>
</tr>
<tr>
<td>What is the rate of emergency room and outpatient healthcare utilisation among patients with adrenal insufficiency?</td>
<td>✔</td>
<td>✗</td>
<td>N/A</td>
<td>✔</td>
</tr>
<tr>
<td><strong>Beta-Lactam Allergy and Surgical Prophylaxis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are patients with a beta-lactam allergy receiving the correct antimicrobial prophylaxis in-hospital for their surgery according to guidelines?</td>
<td>✗</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Are patients receiving antimicrobial prophylaxis within the guideline recommended time frame?</td>
<td>✗</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Continued
visits were likely duplicates. Of the 211 visits analysed, 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 HbA1c, 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 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 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 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 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.

**Table 3** Continued

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

✔ The challenge was encountered in the project.
 ✗ The challenge was not encountered in the project.
 N/A, not applicable.
## Table 4  Strengths and limitations of the databases as elucidated by our example projects

<table>
<thead>
<tr>
<th>Database name</th>
<th>Strengths encountered</th>
<th>Limitations/challenges/cautions encountered</th>
</tr>
</thead>
</table>
| 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 identification 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
Increased coordination and leveraging the opportunity of a
issues presented. We acknowledge the importance of collab-
PLP and with relevant stakeholder groups to address the
across the country. Ongoing work is being conducted by the
and NACRS, and thus these challenges are likely to exist
able across Canada, including Discharge Abstract Database
meaningful clinical data are and how to mobilise and act on
clinicians, and administrators to fully understand what the
orating with various stakeholders including data scientists,
variation in use and therefore this data source must be used with caution.\textsuperscript{33}
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)
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 labora-
tory 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 limita-
tions of administrative and electronic medical record health
databases have been described extensively, for instance in
the work of Burles \textit{et al}, Clement \textit{et al} and Edmondson
and Reimer.\textsuperscript{18–20} The inability to analyse data in real time is
not a problem unique to the Canadian context, with chal-
enges being documented in other jurisdictions including
the USA.\textsuperscript{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.\textsuperscript{22–27} Several of the databases outlined are avail-
able 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 collab-
orating 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.

\textbf{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.\textsuperscript{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 prac-
tice quality improvement.\textsuperscript{33} The overarching purpose of
these efforts is to support the development of a learning
health system and to achieve improvements in the quin-
tuple aim of improving population health, patients’ expe-
rience of care, equity, cost-effectiveness, and sustainability
of healthcare workforce.\textsuperscript{1–3} We strongly believe that the
long-term benefits of improved data capture would signifi-
cantly 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.\textsuperscript{34}

Addressing these four identified challenges is funda-
mental 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 practice, 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.

**Contributors**

TM, TWM ad KC conceived the project idea and drafted the manuscript. BS ensured data accuracy and contributed to the project methods. ROY and DC-S provided local clinical expertise. DC-S, TM, TWM, KC and ROY edited the manuscript.

**Funding**

Supported by a financial contribution from the Government of Alberta via the Physician Learning Programme.

**Disclaimer**

The views expressed herein do not necessarily represent the official policy of the Government of Alberta (no award/grant number).

**Competing interests**

None declared.

**Patient consent for publication**

Not applicable.

**Ethics approval**

This study was approved by Each project included the appropriate ethics approval from the Research Ethics Board-Health Panel at the University of Alberta, Edmonton, Alberta, Canada. The ethics approval numbers are as follows: Pediatric Diabetic Ketoadiposis-Pro00001652 Diabetes Management-Pro00085358 Beta-lactam Allergy and Surgical Prophylaxis-Pro00089539 Adrenal Insufficiency-Pro00088478. Three of our projects included secondary retrospective analyses of routinely collected health data. The Physician Learning Programme only works with unidentified data. The third project was a paper chart audit and also was deidentified.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Supplemental material**

This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access**

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**ORCID ID**

Taylor McGuckin

http://orcid.org/0000-0002-8800-6081

**REFERENCES**


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.


