An infrastructure for real-time population health assessment and monitoring

The fragmented nature of population health information is a barrier to public health practice. Despite repeated demands by policymakers, administrators, and practitioners to develop information systems that provide a coherent view of population health status, there has been limited progress toward developing such an infrastructure. We are creating an informatics platform for describing and monitoring the health status of a defined population by integrating multiple clinical and administrative data sources. This infrastructure, which involves a population health record, is designed to enable development of detailed portraits of population health, facilitate monitoring of population health indicators, enable evaluation of interventions, and provide clinicians and patients with population context to assist diagnostic and therapeutic decision-making.

In addition to supporting public health professionals, clinicians, and the public, we are designing the infrastructure to provide a platform for public health informatics research. This early report presents the requirements and architecture for the infrastructure and describes the initial implementation of the population health record, focusing on indicators of chronic diseases related to obesity.

Introduction

In many countries, including the United States, governments are directing considerable resources toward increasing the adoption of electronic medical records in clinical practice [1]. At the same time, the increasing use of data and messaging standards within clinical and administrative information systems is improving interoperability and allowing data to be shared between providers and institutions [2]. These data are increasingly available for use in public health practice, and they offer enormous potential for describing and monitoring population health and for making population health data available to clinicians and the public. There is, however, limited infrastructure in public health settings to make effective use of these data.

One approach to managing the increasing amount of data available in public health settings is to develop a system analogous to an electronic medical record but with the data organized around defined populations as opposed to individuals. Some researchers have called such an infrastructure a population health record [3], and a conceptual framework for the development of such an infrastructure was proposed recently [4]. A fundamental element of such an infrastructure is a repository of health-related indicators for a population. An indicator in this context is an epidemiological measure of health status, healthcare utilization, or a determinant of health. For example, an indicator may estimate the burden of disease, adherence to therapy, rates of hospitalization or complication for a disease, or screening to prevent conditions in a population. Despite the appeal of the population health record concept, such systems are not available in practice. One main reason for this lack of infrastructure is the difficulties in accessing representative and timely data.

This preliminary report presents the design and initial implementation of a population health record that integrates administrative and clinical data for a representative sample of a population and calculates population health indicators. This population health record is designed so that users can query, visualize, and analyze indicators related to specific diseases or factors that influence health; identify disease outbreaks and geographic clusters of disease or influencing factors; and display temporal trends of diseases. The system
Background

Burden of chronic diseases and the need for population health systems
There is a pressing need in public health practice for access to representative and timely data about population health. Such data are required to describe population health status, to identify targets for public health interventions, and to evaluate interventions. The data currently available through existing systems tend to be limited by their lack of representativeness, their lack of geographical resolution, or their lack of representativeness. The implications of such data limitations are most apparent for chronic diseases, such as obesity, hypertension, and ischemic heart disease. Also known as noncommunicable diseases, these conditions account for a large proportion of global illness, disability, and death. Demographic and lifestyle changes, together with increases in the prevalence of risk factors, have contributed to the rising incidence of chronic diseases [5]. As the rates of illness and death related to chronic diseases continue to grow worldwide, researchers and public health practitioners are struggling to develop and evaluate interventions to limit their impact. In Canada, the prevalence of obesity has increased from 10% in 1970 to 23% in 2004, and the proportion of major chronic diseases attributable to obesity has increased over the same interval by 138% for men and by 60% for women [6]. This increase is shocking, but in theory it can be reversed, given that 40% of chronic diseases are attributable to risk factors such as smoking, obesity, physical inactivity, and poor nutrition, which are preventable [7].

Public health interventions can modify behaviors to prevent and control chronic diseases. Examples of such interventions include tobacco control through taxation and limiting sales, sodium intake control through labeling and awareness campaigns, and coordination of disease management programs. However, the effectiveness of these interventions is not always well documented and is rarely demonstrated across different populations. One recent survey noted that less than 10% of published public health research describes evaluations of interventions, and very few of these studies use current population health data to evaluate chronic disease interventions in the last three decades [8].

Existing population health indicator systems
Numerous systems currently exist for making health indicators accessible over the Internet, but none of these systems provides access to timely, representative indicators at a high level of geographical resolution. A good example of such a system is SAVI (Social Assets and Vulnerabilities Indicators), which compiles indicators for communities in central Indiana and provides tools to analyze and visualize the data collected [9]. The indicators describe social, physical, and economic conditions of these communities at various levels of geographic resolution. Although this system provides a considerable amount of data in a useful format, limitations include timeliness (most recent data are more than 3 years old) and uneven geographical coverage. In the Canadian context, a range of federal, provincial, and regional agencies have created online repositories of health indicators. For example, the Public Health Agency of Canada has created Infobase [10] to provide the public with information about indicators of chronic disease. This system makes available health-related indicators calculated mainly from surveys. Although the data are relatively current (most recent data are 1 year old), they are available for mostly self-reported outcomes and only at a low level of geographical resolution. At the provincial level, Alberta has created the Interactive Health Data Application to provide access to health indicators derived from physician claims data, hospital discharge abstracts, and other sources [11]. While these data are representative, they are not timely (most recent data are more than 3 years old) and do not provide indicators at a high geographical resolution. At a regional level, the Montreal Health Region has created the Espace Montréalais d’information sur la santé (EMIS), which is a repository of indicators of population health status and healthcare system utilization in Montreal [12]. This system uses an interactive mapping interface and provides data at the resolution of the local health areas within the region. As with the other systems considered, however, the data are not current (most recent data are more than 3 years old).

Population health record system
We propose an architecture for a population health record that will provide timely estimates of health-related indicators at high geographical resolution for a defined population. Our initial work to develop the system is focused on the population of Quebec, particularly the Montreal region, which has a population of 3.6 million people. In this section, we describe the anticipated user groups and their requirements, present an overview of the architecture, describe the data sources used in the system, explain how indicators are derived from the data, present the ontology used in the system, and discuss issues related to privacy.

User groups and their requirements
We intend to consider three main groups of users for this system: public health professionals, clinicians, and the general public. Although we recognize the importance of population health information for clinical [13, 14] and individual decision-making, we focus on public health professionals in the initial stage of this project. For this user group, we consider three main functions: describing,
monitoring, and evaluating. *Describing* refers to the development of a portrait of health status for a population. In practical terms, this might entail creating a map or table of age-specific prevalence of a disease, such as ischemic heart disease, and a similar description of risk factors, such as diabetes. *Monitoring* refers to establishing an ongoing analysis of an indicator, with the application of statistical algorithms to a time series or space-time series in order to detect changes in the indicator. *Evaluating* refers to determining the effect of an intervention or other event on an indicator, most frequently by comparing changes in an indicator over time or between regions. To perform the functions described (i.e., describing, monitoring, and evaluating), the system must have certain characteristics. These are described in the following subsections.

**Data/Indicators**

- **Timely**—When describing the health status of a population, public health personnel must work with data that are current. Knowing the state of population health 3 or 4 years ago is of limited help when monitoring health status or when planning interventions. Timeliness is also important for evaluating the effect of past interventions. If analysts must wait for years to evaluate a program, it becomes difficult to manage programs effectively.

- **Representative**—Indicators must be accurate and unbiased for assessing differences in health status across geographical regions and subgroups of a population. If the data are not representative, public health professionals may make incorrect conclusions when describing and monitoring populations and evaluating interventions.

- **High geographical resolution**—Public health services are delivered at a local level, and many interventions are also targeted at a local level, affecting neighborhoods within cities. In order to understand the need for health interventions and to evaluate the effect of these interventions, indicators must be available at the same level of geographical resolution as the intervention. Monitoring to detect spatial variations in disease activity also requires indicators at a level of geographical resolution that is consistent with the scale of clustering.

**System**

- **Easy to access**—Users of the system must be able to find indicators of interest with minimal effort. Experience in medical informatics, and other fields, has clearly demonstrated that people will not adopt a system if they cannot accomplish necessary tasks without disrupting their workflow [15]. This requirement means that there must be an intuitive approach to interacting with the system and finding indicators that relate to commonly understood health conditions.

- **Explanation of indicators**—Users must be able to understand how indicators are calculated, including the elements used by algorithms (e.g., the type of data source used, diagnostic or medication codes used, number of records with a diagnostic or medication code, and length of time). This level of exposition is necessary for users to trust the indicators in the system. In addition, users should be able to identify the evidence of the validity of a specific indicator.

**Overview of the architecture**

To address the data and system requirements, we have designed an architecture that consists of three major components. The first component includes databases and interfaces to other systems. The second component is an ontology that provides a semantic framework for defining indicators and supporting interactive browsing of indicators by users. The third component is an application suite of query and analysis tools to allow users to analyze population health indicators.

**Figure 1** illustrates the overall architecture of the population health record system. Anonymized administrative data are linked with clinical data and are stored in a patient-level database. Algorithms are applied to these individual data to classify patients according to indicators of interest. Individual data are then aggregated to estimate indicators at the population level, and these estimates are stored in the population-level database. An ontology is used to capture the domain knowledge for indicators, by logically defining the concepts, instances, and relationships. This highly structured approach to defining how indicators are constructed and what indicators mean facilitates identification and interpretation of indicators by humans and automated calculation of indicators in software. In addition, because indicators are defined using logic, the ontology supports queries to ensure the integrity of indicator definitions as the system grows, and it also enables reasoning to discover new knowledge.

**Data sources**

A main requirement of the infrastructure is that the indicators must be representative of the population. To achieve this requirement, we use as the foundation for our data the beneficiary file from RAMQ (the Régie de l’Assurance Maladie du Québec), which is the provincial health insurance agency in Quebec. This beneficiary file includes 99% of residents in the province and is therefore representative of the entire population. Using this file, we selected a 25% random sample of people living in the Montreal Census Metropolitan Area (CMA) for the years 1996 to the present. For sampled individuals, administrative data (i.e., physician billing, drugs dispensed, and hospitalization records), death certificates, and laboratory test results can be updated frequently, as suggested in Figure 1. Data describing the
population at risk are obtained from the Canadian Census, and data describing determinants of health can be obtained from various sources. Our design calls for cohort membership to be updated twice yearly, sampling new individuals to replace those lost to death and out-migration. In the remainder of this section, we give a brief description of each of the data sources.

**Administrative and vital statistics data**

Physician services claims and prescription medication claims data are obtained directly from systems used to process payment claims maintained by RAMQ. Hospitalization discharge abstract data are obtained from MED-ECHO (Maintenance et exploitation des données pour l’étude de la clientèle hospitalière), which is a provincial registry for these data, and death certificate data are obtained directly from the ISQ (Institut de la statistique de Québec), which is the provincial vital statistics agency. The physician services and prescription medication claims data files include demographic information such as age and gender, the diagnostic and procedure codes associated with claims for ambulatory and inpatient care, and the specific medications that were dispensed to the patients by outpatient pharmacies. The claims data use standardized codes including International Statistical Classification of Diseases and Related Health Problems, 9th and 10th Revisions (ICD-9 and ICD-10, respectively), Drug Identification Number (DIN), and American Hospital Formulary Service (AHFS**). None of these data sources includes patient identifiers such as name, Social Insurance Number, or exact address.

**Clinical data**

Laboratory test results from microbiology and clinical chemistry labs in Montreal hospitals can be updated routinely through interfaces to hospital information systems or via a regional repository. Clinical and administrative data must be linked at the individual level. In the administrative data, cohort members are uniquely identified by a unique encrypted code, but in the clinical data held by hospitals, individuals are identified by an unencrypted code. To address this problem, our design uses a web service operated by RAMQ that returns an encrypted identifier for an

---

**Figure 1**

Schematic of a population health record demonstrating data flow and communications. Anonymized administrative data are linked with clinical data from hospital laboratories (labeled “H” in this diagram) and stored in a patient-level database (labeled “Individual data”) in our system. Algorithms are applied to these individual data to classify individuals according to indicators of interest. Individual data are then aggregated to estimate population-level indicators, which are stored in the population-level database. An ontology is used to define indicators and support interactive browsing of indicators by users. Note that the two unlabeled boxes at the bottom, with double arrows to “Lab,” are look-up tables or a web service to identify an anonymized identifier for nominal data. Solid-line arrows correspond to a timing of every 2 weeks (or real-time), and dashed-line arrows refer to a timing of every year.
unencrypted identifier. Hospital laboratories or regional repositories can use this service to replace the encrypted identifier with the unencrypted identifier in clinical data before sending data to the population health record infrastructure.

Spatial and census data
Data from the Canadian Census and other sources are used to define geographical boundaries of administrative regions and to provide demographic data describing the populations residing within these regions. These data enable population health assessments with a flexible spatial resolution, including postal codes, census tracts, local health region boundaries, neighborhoods, and urban/rural areas.

Population health indicators
Many indicators are available to describe the health and healthcare utilization of a population. In our initial work, we have focused on health indicators related to determinants and outcomes of obesity. The indicators measure different aspects of diabetes, hypertension, ischemic heart disease, and stroke. They span a continuum from disease burden (e.g., incidence, prevalence, and mortality), to measures of therapy (e.g., prescription, persistence, and treatment risk factors), to outcomes (e.g., complications and hospitalizations), to preventive measures (e.g., disease-related screening). In our infrastructure, these indicators are defined in an ontology (see the next subsection) that enables data and knowledge integration, facilitates knowledge discovery and exploration by users, and serves as a computable repository of knowledge for driving data manipulation and analysis.

Since a public health indicator is generally expressed as a ratio, we have defined three components of an algorithm for calculating an indicator in the population health record. These include the following: 1) the algorithm component that identifies individual cases of a disease, 2) the algorithm component that calculates the numerator (case counts) of an indicator, and 3) the algorithm component that calculates the denominator (population at risk of a condition) of an indicator. Taken together, these components create an algorithm that can be applied to administrative, clinical, or linked data to identify patients diagnosed with chronic diseases such as diabetes and hypertension. Table 1 shows a typical example of an algorithm to identify eligible patients from claims data and to define the numerator of an indicator for prevalence of diabetes. Algorithms can vary with respect to the type of data sources they use, the time window they consider to determine a condition, demographic constraints to extract a subpopulation, and the required frequency of codes in a time window for each data source. For instance, by looking at the medication history and searching for a specific number of prescriptions, or by looking at the physician visit history and searching for a particular number of diagnostic codes or laboratory test results within a certain time period, an algorithm might identify a person as a diabetic patient. Changing the elements of an algorithm can change the estimate of an indicator. Figure 2 shows the results of 20 different algorithms for estimating the prevalence of diabetes. These 20 algorithms differ in terms of three parameters: data

Table 1

<table>
<thead>
<tr>
<th>ID</th>
<th>Disease</th>
<th>Denominator</th>
<th>Numerator</th>
<th>Diagnostic code</th>
<th>RX code</th>
<th>Case finding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind234</td>
<td>Diabetes</td>
<td>+ Algorithm a23</td>
<td>+ Algorithm d11</td>
<td>Inclusion</td>
<td>ATC Code Generic Drug Name</td>
<td>+ Algorithm CS4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ICD-9 = 250</td>
<td>A10A Insulin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diabetes</td>
<td>A10BA02 Metformin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exclusion</td>
<td>A10BB01 Glibenclamide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ICD-9 = 630–677</td>
<td>A10BB02 Chlorpropamide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pregnancy related</td>
<td>A10BB03 Tolbutamide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BB09 Gliclazide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BB12 Glimepiride</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BB31 Azeotbrohexamide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BD03 Metformin and rosiglitazone</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BF01 Acarbose</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BG02 Rosiglitazone</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BG02 Pioglitazone</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BX02 Repaglinide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A10BX03 Nateglinide</td>
<td></td>
</tr>
</tbody>
</table>
sources considered, time window examined, and frequency of codes required.

For any condition, there are several indicators, each of which can be estimated by many different algorithms. On average, we include in the system 10 algorithms for estimating an indicator, selected on the basis of evidence published in scientific and gray literature (e.g., reports that are not peer-reviewed and are prepared by government agencies and research groups). The system is designed to allow users to access the profile of an indicator and inspect the algorithms used to estimate an indicator, to access indicator values for the entire population or any desired subpopulation, as well as to access the hierarchical relationship of the indicator in the ontology. These relationships describe the data sources used in algorithms to define the indicator, the codes in the data sources, the references in the literature, and the accuracy of the indicator if it has been validated.

**Indicator ontology**

In general, ontologies define basic concepts in a domain and relations among their instances in a common vocabulary for researchers and others who need to share information [16, 17]. In the context of public health indicators, coding schemes used for diagnosis of a disease, for drugs prescribed and dispensed, or for medical procedures can vary demographically, geographically, or over time. Therefore, having a representation of concepts rather than system-specific details supports information sharing and reusability. We classified indicators into four major categories, as illustrated in Figure 3. These categories include indicators representing measures of disease burden (e.g., incidence, prevalence, and mortality), measures of therapy (e.g., rate at which medications appropriate for a condition are prescribed or dispensed, compliance, and treatment risk factors), outcome measures (e.g., rate of different complications, risk of hospitalization), and preventive measures (e.g., rate at which physicians perform appropriate preventive actions, such as requesting screening tests or prescribing preventive medications for patients).

In addition to creating a taxonomy of public health indicators, which facilitates finding information and realizing indicator values, the ontology encapsulates all concepts, conceptual models, instance data, axioms, and

---

**Figure 2**

Indicator values for diabetes prevalence in Montreal metropolitan area in 2006 calculated using 20 different algorithms defined by the data sources they draw upon (on the y-axis, “Rx” is dispensed medications; “MD Bill” is physician billing records; and “Admit” is discharge abstracts from hospital admissions). The algorithms are also defined by the historical duration searched in the database for each data source (duration: 1 year; duration: 3 years) and the number of times an eligible code had to appear in the related data source during the historical duration (>1 code; >2 codes). 

---

![Figure 2](image-url)
relationships used to estimate public health indicators. The ontology defines the semantic relationship between different components and regulates their interactions through a set of axioms and rules. Defining appropriate rules and axioms (e.g., cardinality restrictions, which constrain indicator values) along with different associated relationships between the individuals of those categories supports querying and inference using a set of logical reasoners (e.g., RACER [18] and Pellet [19]). The indicator ontology has been developed using OWL 2 Web Ontology Language [20], which provides rich semantics to represent the expressive structure of the ontological framework. Employing OWL 2 and Description Logics (DL) enables us to apply different reasoning and inference tasks (i.e., subsumption and concept satisfiability) under the OWL 2 Direct Semantics [21].

**Data security and privacy**

We have obtained permission from the McGill University Faculty of Medicine Institution Review Board to implement our system prototype and from the Quebec Commission d’accès à l’information (CAI) to obtain and link the data sources used in our system prototype. A proposal to extend our data holdings to implement the full system design is currently under review by the CAI.

The system architecture defines a clear separation between individual records and population-level data expressed as indicators. Users do not have access to individual records and can access only the indicator values. There still remains a risk that an individual might be identified statistically from population-level indicators, and we are examining the possibility of placing constraints on the resolution with which certain indicators can be accessed through the system.

**Current implementation and initial results**

We are developing a prototype implementation of the architecture described above. In the current prototype, administrative data are assembled for the cohort for 11 years (1996–2006), individual and aggregate data models are established, an initial ontology has been created, and we have defined algorithms and calculated estimates of diabetes prevalence among individuals 19 years of age and older living in the Montreal CMA for 2006 (see Figure 2). An example indicator definition is described in terms of the ontology in Figure 4, and the specific codes used for the indicator are shown in Table 1. The indicator algorithms are defined in terms of codes for diabetes from physician billing and hospital admissions (ICD-9) and dispensed prescription medications for treating diabetes [Anatomical...
Therapeutic Chemical (ATC) Classification System, including insulin, analogs, and other blood glucose-lowering drugs. The indicator definition also excludes gestational diabetes, which has a separate diagnostic code but may not be properly coded. These indicators of diabetes do not discriminate between type 1 and type 2 diabetes mellitus because the ICD-9 coding system does not distinguish between the two types of diabetes. Many of the indicator algorithms we use to identify cases of diabetes have been described in the literature [22, 23].

The preliminary results shown in Figure 2 indicate that the use of different data sources for identifying individuals with diabetes can have a considerable impact on the population estimate of disease prevalence. Prescription drugs appear to contribute the most to case detection for this indicator, although the inclusion of physician billing codes and hospital discharge codes increases the estimated prevalence. Looking back for the presence of drugs and diagnostic codes over 3 years as opposed to 1 year also gave higher prevalence. Similarly, requiring greater than one occurrence of a code gave a higher prevalence than requiring more than two occurrences of a code. We are currently conducting research to validate these and other indicators.

Discussion and conclusion
This early report describes the design and prototype implementation of a novel public health informatics infrastructure to integrate clinical, administrative, and other data sources to produce estimates of population health. The architecture addresses requirements of public health users for easy access to indicators that are timely, representative, and available at a high geographical resolution. The infrastructure is intended to enhance information retrieval, analysis, and decision-making related to describing and monitoring populations and evaluating interventions. More importantly, this platform is designed to play a central role in enhancing preparedness and evaluating health system effectiveness in response to noncommunicable diseases.
The architecture also includes a semantic layer, built on top of the indicator definitions and the clinical and administrative data. This semantic layer makes it possible to employ logical reasoning to reveal implicit relationships, hidden dependencies, inconsistencies, and unsatisfiability within the semantic infrastructure. This feature is also intended to facilitate the evaluation and assessment of inherently complex, dynamic, and interrelated healthcare systems. The indicator ontology should also enable us to manage scalability/complexity of our integrated system via reusable modules.

The prototype implementation of our architecture using administrative data demonstrates that for one chronic condition, the accuracy of indicator estimates differs by data sources. Our estimates of diabetes mellitus prevalence ranged from 1% to nearly 5%, depending on the algorithm used. Self-reported prevalence of diabetes in Canada in 2006 was approximately 5% but was lower in Quebec [24]. We expect the inclusion of clinical data, such as laboratory test results, to improve the accuracy of case detection. While the indicators do not require validation prior to incorporation within the infrastructure, we are working to validate indicators against external measures such as health surveys and electronic medical records.

Performance characteristics of the platform and evaluation of the effect of the infrastructure on decision-making and outcomes are areas of focus for future research. Although there is evidence in the literature regarding the acceptability of population health information systems in clinical practice, future research will also elicit the requirements for these population data in clinical decision-making and the potential effects on the overall performance of the health system.

While the infrastructure is ultimately intended to improve the measurement of population health, it will also provide a platform for public health informatics research. Initial research areas that we have identified include the following: evaluating the effect of different data linkage strategies on the accuracy and usability of population health indicators for complex combinations of disease status, evaluating the effect of using multiple clinical and administrative data sources on the accuracy of population health indicators, and evaluating the effect of access to health indicators on work patterns and productivity in public health settings. Other research avenues that we are pursuing include the use of machine learning to identify cases and the efficient dynamic identification of cases following data updates.

We anticipate a number of challenges in advancing our population health record infrastructure. There are limitations to the data sources and coding systems used in terms of accurately identifying from our cohort cases or individuals that are at risk. For example, not all individuals may have access to the same prescription medications, and this variable eligibility must be accounted for when using prescription drug data within algorithms. In addition, little evidence currently exists in scientific and gray literature facilitating the development of indicators beyond measures of disease burden (i.e., prevalence, incidence, and mortality) and across a range of chronic conditions. The development and validation of indicators relevant to rare or comorbid conditions may be limited by statistical power, despite our large overall sample size. Finally, there are tradeoffs between providing indicator estimates with a high degree of geographic and demographic resolution and respecting privacy so that individuals are not statistically identifiable from our infrastructure. Despite these challenges, our infrastructure holds considerable promise for fulfilling the needs of public health users and others while providing a platform for advancing informatics research.

**Acknowledgment**

The Canadian Foundation for Innovation funded the development of the research infrastructure.

**References**


Received September 16, 2011; accepted for publication November 9, 2011.

David L. Buckeridge Surveillance Lab, Clinical and Health Informatics Research Group, McGill University, Montreal, QC H3A 1A3 Canada (david.buckeridge@mcgill.ca). Dr. Buckeridge is an Associate Professor of Epidemiology and Biostatistics at McGill University in Montreal where he holds a Canada Research Chair in Public Health Informatics. He is also a Medical Consultant to the Institut national de santé publique du Québec and the Direction de santé publique de Montréal. His research focuses on public health informatics and particularly on the informatics of public health surveillance. Current research projects include developing and evaluating systems for automated surveillance in community and hospital settings. In addition to his research activities, Dr. Buckeridge provides advice on surveillance to groups such as the World Health Organization, the Institute of Medicine, and the U.S. and Chinese Centers for Disease Control and Prevention. He has an M.D. degree from Queen’s University in Canada, an M.Sc. degree in epidemiology from the University of Toronto, and a Ph.D. degree in biomedical informatics from Stanford University. He is also a Fellow of the Royal College of Physicians and Surgeons of Canada, with specialty training in Public Health and Preventive Medicine.

Masoumeh Izadi Surveillance Lab, Clinical and Health Informatics Research Group, McGill University, Montreal, QC H3A 1A3 Canada (mtabae@cs.mcgill.ca). Dr. Izadi is a Research Staff Member in the Surveillance Lab at the Health Informatics Research Group. She received a Ph.D. degree in artificial intelligence from McGill University, an M.Sc. degree in advanced computing from King’s College University of London, and a B.Sc. degree in applied mathematics from the University of Tehran. Her research contributions in the investigation, development, and application of computational theories of reasoning and decision-making under uncertainty and in applications of artificial intelligence in health informatics and medicine have been published in various journals and conferences. She is a member of editorial board of Journal of Bioterrorism and Biodefense.

Arash Shaban-Nejad McGill Clinical and Health Informatics Group, Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montreal, QC H3A 1A3 Canada (arash.shaban-nejad@mcgill.ca). Dr. Shaban-Nejad is a Postdoctoral Fellow at the McGill Clinical and Health Informatics Centre at McGill University, where he conducts research on the knowledge representation and modeling for different applications in healthcare. He received M.S. and Ph.D. degrees in computer science from Concordia University in 2005 and 2010, respectively. His primary research interest is knowledge representation and Semantic Web, particularly ontologies and knowledge bases, description logics, category theory, agents, reasoning, and inferencing, with special emphasis on applications from health and biomedical domains. His Ph.D. dissertation focused on change management in distributed biomedical ontologies through a multi-agent-based framework, formalized using category theory. He is author or coauthor of several scientific papers and technical manuscripts, which have appeared in various international journals, conference/workshop proceedings, and book chapters. Dr. Shaban-Nejad is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Association for Computing Machinery (ACM), the National Institutes of Health Informatics (NIHI), and the American Medical Informatics Association (AMIA).

Luke Mondor Surveillance Lab, McGill Clinical and Health Informatics Research Group, McGill University, Montreal, QC H3A 1A3 Canada (luke.mondon@mail.mcgill.ca). Mr. Mondor is a Research Assistant at the Surveillance Lab, part of McGill University’s Clinical and Health Informatics Research Group. He received a B.A. degree in anthropology from the University of Manitoba in 2008 and recently completed an M.Sc. in epidemiology from McGill University.

Christian Jauvin Surveillance Lab, McGill Clinical and Health Informatics Research Group, McGill University, Montreal, QC H3A 1A3 Canada (cjauvin@gmail.com). Dr. Jauvin received an M.Sc. degree in computer science and machine learning from Université de Montréal in 2003. Since then, he has focused mainly on scientific computing and is now a full-time independent researcher and consultant, working in different scientific environments, on many topics including geographic information systems, high-performance computing, databases, machine learning, and visualization.

Laurette Dubé McGill University, Montreal, QC H3A 1G5 Canada (laurette.dube@mcgill.ca). Dr. Dubé is Full Professor and holds the James McGill Chair of consumer and lifestyle psychology and marketing at the Desautels Faculty of Management of McGill University, which she joined in 1995. She is also the Founding Chair and Scientific Director of the McGill World Platform for Health and Economic Convergence, a unique initiative to promote convergence between academic disciplines, innovations, and initiatives by individuals, communities, enterprises, and governments, at all levels, to address the most pressing problems through systemic change. Dr. Dubé’s lifetime research interest bears on the study of what affects and behavioral economic processes underlie consumption and lifestyle behavior and how such knowledge can inspire more effective health and marketing innovations. A leader in marketing research in Canada, and among the world leaders in her areas of expertise, she has made outstanding theoretical contributions while working creatively and rigorously on issues of managerial and social significance. Beyond scientific publications in books and in leading scientific journals of her field, including Journal of Consumer Research.
and Journal of Marketing Research, her work has been presented in leading general audience and business publications such as Maclean's, The Globe and Mail, USA Today, and The Wall Street Journal. Dr. Dubé received her Ph.D. degree in marketing from Cornell University in Ithaca, New York (1990). She has an M.P.S. degree in services marketing management from the same university, an M.B.A. degree in finance from HEC Montréal, Montreal, Quebec, Canada, and a B.Sc. degree in health sciences (nutrition) from Laval University, Quebec, Canada, with 10 years of experience in hospital management. She holds a research career award jointly funded by the Canadian Institutes of Health Research and the Social Sciences and Humanities Research Council of Canada. She is a Fellow of the Royal Society of Canada.

**Yeona Jang**  
McGill University, Desautels Faculty of Management, Montréal, QC, H3A 1G5 Canada  
(yeona.jang@mcgill.ca). Dr. Jang is a professor of practice in management in the Desautels Faculty of Management at McGill University. She received M.S. and Ph.D. degrees in computer science from MIT and an M.S. degree in business management from the MIT Sloan School of Management. She has more than 15 years of experience as a senior IT (information technology) executive and decision-maker in various companies, in innovative use of IT in business. She is experienced in various aspects of IT management, including IT-enabled business innovation, large-scale systems integration, enterprise architecture, large-scale data warehouse and business intelligence, and knowledge management. She joined the Desautels Faculty of Management at McGill University in June 2008. Her current research interest in health IT includes advancing research in medical informatics and business models for electronic health records to improve patient safety and health outcomes. Dr. Jang is a member of the Canada/U.S. Research on Healthcare IT team in collaboration with McGill Medical Department, University of Ottawa Medical Department, and Brigham and Women’s Hospital/Harvard Medical School, and an international consortia of the Scleroderma Patient-centered Intervention Network.

**Robyn Tamblyn**  
Departments of Medicine and Epidemiology and Biostatistics, McGill University, Faculty of Medicine, Montreal, QC H3A 1A3 Canada (robyn.tamblyn@mcgill.ca). Dr. Tamblyn is a Professor in the Department of Medicine and the Department of Epidemiology and Biostatistics at McGill University. She is a James McGill Chair, a Medical Scientist at the McGill University Health Center Research Institute, and the Scientific Director of the Clinical and Health Informatics Research Group at McGill University. Dr. Tamblyn’s ground-breaking research on educational outcomes has elucidated important relationships between health professional training, licensure, and practice that have subsequently guided credentialing policies. Her work on prescription drug use, its determinants, and computerized interventions to improve drug safety [Medical Office of the XXIst Century (MOXXI)] have been recognized internationally. She leads a team funded by the Canadian Institutes of Health Research to investigate the use of e-health technologies to support integrated care for chronic disease, and she co-leads a Canadian Foundation for Innovation Informatics Innovation Laboratory to create advanced technologies to monitor adverse events in populations and create new tools to improve the safety and effectiveness of healthcare. Her work is published in the Journal of the American Medical Association, the Annals of Internal Medicine, the British Medical Journal, Medical Care, and Health Services Research, among others. She has received the Canadian Health Services Research Foundation Knowledge Translation award for her research in improving the use of medication, as well as the L’Association canadienne-française pour l’avancement des sciences (ACFAS) Bombardier award for innovation in the development of a computerized drug management system. As of January 2011, she became the Scientific Director of the Institute of Health Services and Policy Research at the Canadian Institutes of Health Research.