Population Health Record: An Informatics Infrastructure for Management, Integration, and Analysis of Large Scale Population Health Data

Masoumeh Izadi, Arash Shaban-Nejad,
Anya Okhmatovskaia, Luke Mondor, David Buckeridge
McGill University
1140 Pine Ave West, Montreal, QC

Abstract

Practitioners and researchers in health services and public health routinely estimate population health indicators from a range of data sources. These indicators are used in many settings to describe health status, monitor quality of care, and evaluate the effect of interventions. The data and knowledge necessary to calculate indicators, however, are scattered across different health settings, resulting in inconsistent and fragmented indicators and an inefficient use of population health information in research and practice. The Population Health Record (PopHR) described in this paper is an informatics platform for semi-automated integration of disparate data to enable measurement and monitoring of population health status and determinants. The research and development to build the PopHR uses AI methods to perform many tasks, including calculation of indicators and interaction with users.

Introduction

There is a pressing need in public health practice for access to representative and timely data about population health. Access to such data is necessary for describing population health status, identifying targets for public health intervention, and evaluating interventions. Currently available population health data tend to be limited either by their lack of timeliness (e.g., administrative data), their lack of spatial resolution (e.g., survey data), or their lack of representativeness (e.g., clinical data). The implications of these data limitations are most apparent for chronic diseases, such as obesity, diabetes, hypertension, and ischemic heart disease because monitoring these disease generally requires access to multiple linked databases over time. Also known as non-communicable diseases, these conditions account for a large proportion of global illness, disability, and death (Wang et al. 2012). 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 decrease their impact. Demographic and lifestyle changes, together with increases in the prevalence of risk factors, have contributed to the rising incidence of chronic diseases (Crews and Gerber 1994). Public health interventions can modify behaviors to prevent and control chronic diseases (Mirolla 2004).

Examples of such interventions include tobacco control through taxation and limits on sales, sodium intake control through labeling and awareness campaigns, and coordination of disease management programs (Luo et al. 2007). However, the effectiveness of these interventions is not always well documented and is rarely demonstrated across different populations due to barriers in access to current population health data (Sanson-Fisher et al. 2008)

Our approach to address these limitations is to build a population health record, (PopHR),(Buckeridge et al. 2012) which is an infrastructure that retrieves and integrates heterogeneous data from multiple sources (administrative, clinical, and survey) in almost real- time, links these records to demographic data for a representative cohort, and supports intelligent analysis, and visualization of a portrait of population health through a comprehensive set of indicators. In addition to describing the health status of a defined population, this system is designed to allow monitoring of an indicator with the application of statistical algorithms to detect changes prospectively in the indicator over time and space. The PopHR can therefore be used to evaluate the effect of a public health intervention by comparing changes in indicators over time or between regions.

In our initial work, we have focused on health indicators related to obesity. Relevant indicators describe diabetes, hypertension, coronary 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), and to preventive measures (e.g., disease-related screening). Three user groups are anticipated to interact with this system: public health professionals, clinicians, and the public. The infrastructure provides a platform for sharing population health data with clinicians and individuals to examine the role of population health data in clinical practice and disease self-management by allowing population context to assist diagnostic and therapeutic decisions, making population health information directly available at the point of care such as individual's home, a community, or medical clinic. In addition to supporting these users, the infrastructure also provides a platform for public health informatics research.

From a health informatics research perspective, this infrastructure enables evaluation of the effect of different data linkage strategies on the accuracy and usability of population health indicators for complex combinations of disease status, evaluation of the effect of using multiple clinical and administrative data sources on the accuracy of population health indicators, and evaluation of how access to health indicators influences work patterns and outcomes in different settings. In last few years, there have been major advances in artificial intelligence methods including those for integration of multiple data streams, analysis of spatial and temporal data, and mining heterogenous data sources. While the future holds great potential for utilizing such tools in medical and health informatics systems, their actual deployment in practice settings remains a challenge. Within the PopHR, there are several compelling use cases for artificial intelligence techniques, from ontologies and natural language interfaces for managing and accessing health indicators to probabilistic reasoning for analysis and monitoring of indicators.

This paper briefly describes the PopHR architecture and presents the work in progress towards implementing AI methods in three main components of PopHR: a public health indicator ontology (PHIO), a natural language interface (NLI) query system, and an analysis module for deriving inference from indicator values. Some possible samples of navigation through PopHR system in our current prototype development are presented in the result section.

Background

A population health indicator is calculated by application of an algorithm to raw data. In this context, an algorithm is usually a set of rules to identify individuals with a given health condition. The elements of such algorithms for health status indicators usually include the type of administrative data source used, the relevant diagnostic or medication codes, the required frequency of the codes, and the length of time considered when searching for codes.

There is generally little standardization of public health indicators, which can consequently be estimated through a variety of algorithms. Even for a single algorithm, different definitions may include different arrangements of codes for disease diagnosis, drug prescribing and dispensing, and medical procedures. To further complicate matters, definitions can vary in a country within regions, or over different years in the same region. Therefore, having a representation of the concepts rather than only system specific details is critical for information sharing and reusability. This is one of the motivating factors in the developmental choice of architecture for PopHR.

Overview of PopHR System

The system architecture includes four structural components: databases and interfaces to other systems; an ontology that provides a semantic framework for defining population health indicators and supporting interactive browsing of indicators by capturing the domain knowledge in the form of concepts, instances, and a rich variety of relationships and axioms: a natural language interface for user queries; and an analysis module that provides tools for analyzing population health indicators.

The data integrated within the PopHR originate from a wide variety of resources such as structured and non-structured textual resources, databases, charts, tables, images, and existing controlled vocabularies in the domain. Also the cross-disciplinary and dynamic nature of the domain gives rise to volatility, heterogeneity and ambiguity at both the conceptual level and the data level. Indexing, classifying, integrating and interpreting these data is far from trivial.

PopHR relies on ontologies, where the concepts, relations and instances are meaningful while accompanied by a descriptive data set, to provide users with flexible access to different levels of information. Ontologies support consistent data access practices, both within our PopHR application and across the public health domain, and enable semantic analysis through reasoning and inference using a logical reasoner.

In addition, ontologies can assist in reducing ambiguity, overcoming redundancies, and enabling inference through reasoning (e.g. satisfiability and consistency checking) and querying.

PopHR Cohort

In the first phase of PopHR research and development, we are using a 25% random sample of people residing in the Montreal Census Metropolitan Area (CMA). The population in the CMA was 3.8 million in the year 2011. For sampled individuals, we are developing a mechanism to obtain bi-weekly updates of data describing physician billing and drugs dispensed, and periodic updates of hospitalization records and death certificates. In the current system prototype, we use a database that follows an open cohort from 1998 to 2006. Data from the Census and other sources are used to define geographical boundaries of administrative regions and to provide demographic data describing the population. These data enable population health assessment with a flexible spatial resolution including postal and census geographies, neighborhoods, and urban/rural areas. 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 the provincial health insurance agency in Quebec, the Regie de l'Assurance Maladie du Quebec (RAMQ). 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) in 1998, an in each subsequent year we refreshed the cohort by sampling from new immigrants and births to replace those leaving through emigration or death.

Data

Administrative data for PopHR cohort include physician services claims and prescription medication claims data obtained directly RAMQ as used to process payment claims. Hospitalization discharge abstract data are also obtained from a provincial registry in Quebec, MED-ECHO, and death data are obtained directly from the provincial vital statistics agency, ISQ. None of these data sources includes patient identifiers such as name, Social Insurance Number (SIN), or exact address. The system is also designed to accept clinical data such as laboratory test results from microbiology and clinical chemistry labs, but these data are not yet available in the PopHR.

Individual-level data from RAMQ, MED-ECHO, and ISQ are linked by a unique encrypted code. Physicians' name and the name of care providing institutions are also anonymised. We have access to three digit postal code of the cohort member's place of residence. These data enable population health assessments with a flexible spatial resolution, including three digit postal codes, neighborhoods, and urban/rural areas.

PHIO Ontology

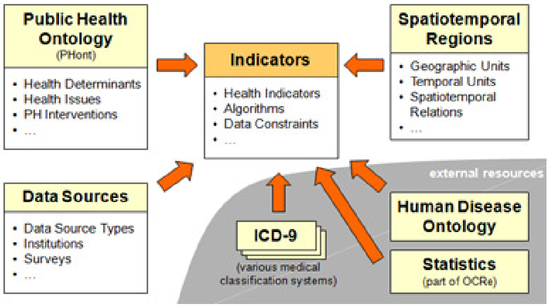


Figure 1: The major components of public health indicator ontology (PHIO). The Indicator box represents the application ontology developed in PopHR, and the other boxes demonstrate other data sources, databases and existing biomedical ontologies that have been reused or imported by PHIO.

The population health indicator ontology (PHIO) is an integrated application ontology implemented in OWL DL. Figure 1 illustrates the components imported or incorporated in PHIO. This ontology encodes knowledge about the epidemiological indicators used in the PopHR, defining categories of indicators and describing in a consistent manner the algorithms for calculating indicators. The indicators taxonomy developed by the Canadian Institute of Health Information CIHI (CIHI 2006) is incorporated into PHIO, focusing on *health status indicators* class and its subclasses (see Figure 2). The Semantic Science Integrated Ontology (SIO) (SIO 2008) is also used in PHIO as an upper ontology for consistent knowledge representation across physical, procedural and informational entities. The logical consistency and satisfiability are controlled using a logical reasoner (Pellet and Fact++). In PHIO, *health indicator* and *algorithm* are both defined concepts and a relation between their individuals is that an algorithm is used to *calculate* the value of a health indicator. A Public Health domain ontology (Jorm, Gruszin, and Churches 2009) is extended and incorporated into PHIO to describe diseases and their determinants noting, for example, *Obesity* is a *risk factor* of *Diabetes*. Indicators are linked explicitly to disease and risk factor concepts in the public health ontology through an *indicator of* relation. PHIO is used by the PopHR system to guide the selection of indicators by users, to automate the calculation of indicator values, and to support the interpretation of indicator values. The use of PHIO within the PopHR facilitates data and knowledge integration, enables knowledge discovery and exploration, and also serves as computable repository of knowledge for driving data manipulation and analysis functions. Development of the system of indicators started with a review of approaches to measuring health determinants (e.g., food supply chain, dietary intake) and health outcomes (e.g., disease burden, therapy, secondary prevention, and complications). In addition to defining a taxonomy of population health indicators, PHIO encapsulates all concepts, axioms and relationships necessary to calculate population health indicators. We also developed the GeoPopHR ontology to assist establishing the relationships between administrative data to geospatial data, with emphasize on mapping variations in disease rates. Once integrated geographically, these individual-level health data and regional data enable the flexible assessment of population health status and determinants across different geographic regions, temporal intervals, and population sub-groups.

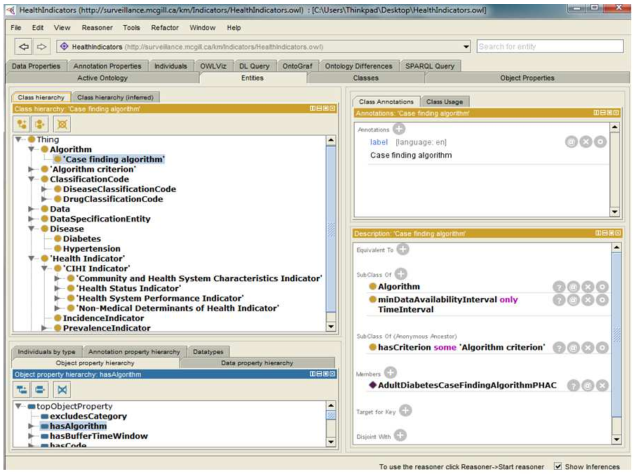
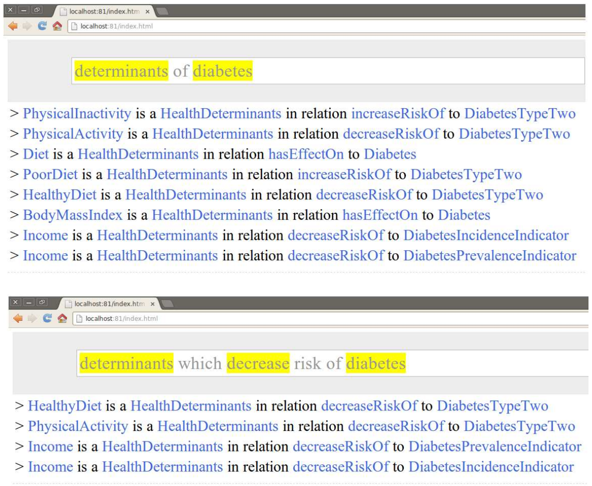

HealthIndicators (http://surveillance.mcgill.ca/km/Indicators/HealthIndicators.owl) : [C:\Users\Thinkpad\Desktop\HealthIndicators.owl]
File Edit View Reasoner Tools Refactor Window Help
HealthIndicators (http://surveillance.mcgill.ca/km/Indicators/HealthIndicators.owl)
Search for entity
Data Properties Annotation Properties Individuals OWLViz DL Query OntoGraf Ontology Differences SPARQL Query
Active Ontology Entities Classes Object Properties
Class hierarchy Class hierarchy (inferred)
Class hierarchy: 'Case finding algorithm'
Thing
Algorithm
'Case finding algorithm'
'Algorithm criterion'
ClassificationCode
DiseaseClassificationCode
DrugClassificationCode
Data
DataSpecificationEntity
Disease
Diabetes
Hypertension
'Health Indicator'
'CIHI Indicator'
'Community and Health System Characteristics Indicator'
'Health Status Indicator'
'Health System Performance Indicator'
'Non-Medical Determinants of Health Indicator'
IncidenceIndicator
PrevalenceIndicator
Individuals by type Annotation property hierarchy Datatypes
Object property hierarchy Data property hierarchy
Object property hierarchy: hasAlgorithm
topObjectProperty
excludesCategory
hasAlgorithm
hasBufferTimeWindow
hasCode
Class Annotations Class Usage
Annotations: 'Case finding algorithm'
Annotations
label [language: en]
Case finding algorithm
Description: 'Case finding algorithm'
Equivalent To
SubClass Of
Algorithm
minDataAvailabilityInterval only TimeInterval
SubClass Of (Anonymous Ancestor)
hasCriterion some 'Algorithm criterion'
Members
AdultDiabetesCaseFindingAlgorithmPHAC
Target for Key
Disjoint With
To use the reasoner click Reasoner->Start reasoner
☑ Show Inferences

Figure 2: A partial view of the taxonomy of indicators based on CIHI framework and indicator instances in Protege.

Natural Language Query Processing

Processing natural language queries about population health indicators is another unique component for PopHR. This feature allows users to access information about a particular health outcome or determinant from a massive repository of heterogeneous and distributed data without learning to use a complicated interface or a specialized query language. However, natural language queries can be ambiguous and potentially correspond to many or no indicators available in the system (Vallet, Fernndez, and Castells 2005; Guha, McCool, and Miller 2003). Therefore, the PopHR query module must check the completeness, consistency, and relevance of a user query. This module attempts to anticipate a user's information need and deliver the relevant population health information automatically. The indicator retrieval model in the PopHR query system works in two parts. In the first part, the natural language query is parsed and query components are matched to concepts in the ontology. Matching concepts are presented to the use to verify that their query was interpreted correctly. In the second part, the knowledge base is queried to return indicator values and if necessary, statistical methods are applied to the returned indicator values. Figure 3 presents this phase developed for our prototype. We used a guided and controlled Natural Language Interface (NLI), based on our defined semantic framework captured by the ontology. The PopHR visualization module is currently being designed to present the final results of the query module in a format that corresponds to the type of results. It will support a range of options including charts, graphs, maps, and tables.



determinants of diabetes

> PhysicalInactivity is a HealthDeterminants in relation increaseRiskOf to DiabetesTypeTwo
> PhysicalActivity is a HealthDeterminants in relation decreaseRiskOf to DiabetesTypeTwo
> Diet is a HealthDeterminants in relation hasEffectOn to Diabetes
> PoorDiet is a HealthDeterminants in relation increaseRiskOf to DiabetesTypeTwo
> HealthyDiet is a HealthDeterminants in relation decreaseRiskOf to DiabetesTypeTwo
> BodyMassIndex is a HealthDeterminants in relation hasEffectOn to Diabetes
> Income is a HealthDeterminants in relation decreaseRiskOf to DiabetesIncidenceIndicator
> Income is a HealthDeterminants in relation decreaseRiskOf to DiabetesPrevalenceIndicator

determinants which decrease risk of diabetes

> HealthyDiet is a HealthDeterminants in relation decreaseRiskOf to DiabetesTypeTwo
> PhysicalActivity is a HealthDeterminants in relation decreaseRiskOf to DiabetesTypeTwo
> Income is a HealthDeterminants in relation decreaseRiskOf to DiabetesPrevalenceIndicator
> Income is a HealthDeterminants in relation decreaseRiskOf to DiabetesIncidenceIndicator

Figure 3: Query examples from the natural language interface part of PopHR prototype.

Analysis Module

Methods for statistical analysis and data mining are also core components of the PopHR. These methods are used to develop a portrait of health status for a defined population, to monitor indicators of population health over time, and to evaluate changes in indicators to determine the effect of an intervention.

To perform the functions of describing, monitoring, and evaluating, the PopHR employs methods that are efficient and scalable. We are developing the PopHR to use machine learning methods that exploit the linkage of indicator values with domain knowledge to perform prediction, classification, and pattern recognition. Future research will evaluate the effectiveness of these methods for tracking disease burdens over time, identifying high risk populations, localizing diseases and therapies, and ascertaining important variations in health services utilization by geographic regions.

Within the PopHR project, we are also exploring the use of machine learning to develop algorithms for classifying individuals according to their disease status. In practical terms, an algorithm is needed to compute an indicator. To date, algorithms for computing health indicators have been defined as Boolean combinations of different codes for diseases, drugs, and other health events. However, these definitions are essentially an encoding of expert opinion regarding what pattern of healthcare utilization is likely to reflect a true case of disease. Although this simple approach to defining algorithms can work reasonably well for simple indicators such as disease prevalence, it does not scale well to more complicated indicators such as incidence, and adherence to therapy. Moreover, algorithms developed in this manner are not capable of considering the richness of high-dimensional administrative data, including time-varying patterns of health services utilization, healthcare establishments consulted, and health providers' characteristics. A wide spectrum of advanced classification methods, statistical machine learning and probabilistic reasoning techniques can be used to understand and quantify relationships in administrative data that indicate a case of a disease and to generalize this findings to develop an automated case detection method with a potentially profound increase in sensitivity and specificity over currently used algorithms.

Monitoring an indicator includes comparing changes in an indicator over time or between regions using the application of statistical algorithms to a time series or space-time series in order to detect significant changes in the indicator value. Probabilistic graphical models such as Bayesian networks variation (Lin, Chiu, and Wu 2002; Aliferis and Cooper 2013; Tawfik1 and Neufeld2 2002; Diard, Bessiere, and Mazer 2003) and Markov models (Scott 2002; Martinis and Twele 2010; Mari and Ber 2006) have demonstrated considerable success in other fields for temporal pattern discovery. We intend to employ these techniques to improve the accuracy of monitoring disease burden indicators. Similarly, the advanced methodologies in case-based reasoning (Bichindaritz and Montani 2012; Nilsson and Sollenborn 2004) can contribute to targeting potential interventions to high-risk populations identified by PopHR and to evaluating disease interventions already deployed.

Results

We are currently developing a prototype system that implements the architecture described above. An initial version of the population health indicator ontology has been created, and we have defined algorithms and calculated values of indicators related to diabetes. To demonstrate this prototype, we present a use case where a decision-maker uses the PopHR to understand neighbourhood changes in diabetes determinants and outcomes. The general interaction with the user is as follows. First, the user is presented with a search box and enters a query in normal language. PopHR interprets and executes the query and then returns the indicator values in a tabular for- mat. The user can then modify the query or perform a new query. This prototype currently allows the users to access only information related to diabetes in the greater Montreal region. The taxonomies of disease and health determinants in PHIO are used to expand queries so that related and more specific indicators are matched to queries, not only indicators with an exact textual match to the query. For example, if a user is interested in find- ing indicators in the system related to diabetes, PHIO will expand the search term Diabetes to identify all indicators describing determinants and outcomes related to that disease. Similarly, the knowledge captured by PHIO can also support interpretation of indicator values. As example queries, consider the following questions:

- Which neighbourhoods in Montreal had a high prevalence of diabetes in 2005?
  prevalence is identified as a type of *indicator* in PHIO and *diabetes* is identified as a *disease* defined in PHIO, so the PopHR infers that the request is about the indicator that measures the prevalence of diabetes; *high* is interpreted to mean values of the diabetes prevalence indicator that ares statistically significantly greater than the mean value for all regions taken together; *neighborhoods* are recognized as a type of a region with defined boundaries and *Montreal* is recognized as an individual region in PHIO. To answer this query, prevalence values for all regions must be compared to the distribution of value for all regions and an appropriate statistical test (e.g., t-test, percentile) must be applied to the values to identify the regions with a value that is statistically significantly higher or lower than the average of all regions. Based on this analysis, regions are categorized as low, average, or high diabetes prevalence and the classified results are returned to the user;
- Which indicators describe determinants of diabetes?
  This query will use relations in PHIO to identify determinants of diabetes and display the determinants to the user;
- Which neighbourhoods with a high prevalence of diabetes in 2005 also have high purchases of sugary soft drinks?
  Answering the query would require the same series of steps as for the first query, but for both the outcome and the indicator, followed by an intersection operation to identify neighbourhoods with high values in both;
- Among neighbourhoods with a high prevalence of diabetes in 2005, what determinants of diabetes are significantly elevated?
  This query combines elements of the second and the third queries.

Discussion

The PopHR infrastructure provides a data management, integration, and analytical platform for a wide range of population health data sources. We presented the requirements and architecture for the system and we described our initial work to implement the population health record, focusing on indicators related to diabetes.

The work presented in this paper is part of prototyping for PopHR. Rapid prototyping of systems is an effective development approach, and the work to date has centered on experimentation with different system components and the automation of communication and data flow between them. Scaling-up the PopHR system in the future development cycles will require further use of techniques for the management, integration, and analysis of *big data*.

Additional relevant data sources include electronic health records and non-traditional public health data sources such as weblogs and twitter feeds. Incorporation of these and other massive data sources into PopHR will require, highly efficient algorithms to estimate and analyze indicators.

Our experience with validating indicators of diabetes prevalence in comparison with population surveys suggests that choices made in the definition of indicator algorithms (i.e., decisions to rely upon different codes and different data sources) for identifying individuals with diabetes can have a considerable impact in the population estimate of disease prevalence (Buckeridge et al. 2012). For example, prescription drugs are specific measures of diabetes, while physician billing are sensitive measures. We expect that the incorporation of clinical data into PopHR and these algorithms, planned for future phases of this project, will further improve the accuracy of case detection.

PopHR is an innovative application, which demonstrates the value of applying artificial intelligence methods such as machine learning and knowledge representation to the management and analysis of heterogeneous sources of population health data.

The prototype of the analysis module is currently focused on methods for the description of population health status. In the future, we will extend this module to include method for monitoring indicators over time and evaluating the effects of interventions. AI methods in temporal reasoning can support monitoring through the detection of possible trends and the characterization of these trends. The incorporation of temporal reasoning methods into the PopHR will enable the system to answer queries more challenging than typical cross-sectional queries presented in this paper. For instance, one could ask: what determinants of diabetes have changed significantly over the last five years? We anticipate that methods such as these will further assist public health decision-making.

The ontology development of PHIO creates explicit, formal, and multipurpose catalogs of knowledge that can be reused by other intelligent systems for population health research and practice. One of the limitations of our work at this stage is performing the spatiotemporal reasoning using the existing logical reasoners. Our future work on PHIO will be focused on enriching the ontological structure, by defining meaningful logical axioms using the knowledge derived from our statistical inference module.

Conclusion

The PopHR is at the center of a multi-year research program to create an innovative informatics platform for developing and evaluating novel methods for population health surveillance. The research and development to build the PopHR will generate knowledge about how best to exploit existing data repositories in clinical, administrative, and commercial settings for assessing population health determinants and outcomes. The current prototype that we have developed is focused on the determinants and outcomes of diabetes. In future research, we plan to extend the current prototype to cover a range of non-communicable diseases and integrate additional data sources and machine learning models. We are also exploring the application of our architecture to infectious disease surveillance in resource-constrained settings. We expect that the system of indicators will support the development of timely, accurate, and sharable descriptions of population health, and facilitate monitoring changes in health determinants and outcomes over space and time. The capacity to assess population health in this manner is critical for identifying health inequalities and evaluating interventions to enhance the prevention and management of chronic and infectious diseases in vulnerable groups and the population at large. The PopHR is being developed to provide an advanced infrastructure for research on population health monitoring, policy making, and decision making.

References

Aliferis, C. F., and Cooper, G. F. 2013. A structurally and temporally extended bayesian belief network model definitions, properties, and modeling techniques. *CoRR*.

Bichindaritz, I., and Montani, S. 2012. Report on the eighteenth international conference on case-based reasoning. *AI Magazine* 33(1).

Buckeridge, D. L.; Izadi, M. T.; ShabanNejad, A.; Mondor, L.; Jauvin, C.; Dub?, L.; Jang, Y.; and Tamblyn, R. 2012. An infrastructure for real time population health assessment and monitoring. *IBM Journal of Research and Development* 56(5).

CIHI. 2006. Hierarchy of standard geographic units for dissemination. http://www.statcan.gc.ca/pub/92-195-x/2011001/other-autre/hierarch/h-eng.htm.

Crews, E., and Gerber, M. 1994. Chronic degenerative diseases and aging. *Biological Anthropology and Aging: Perspectives on Human Variation Over the Life Span* 174?–208.

Diard, J.; Bessiere, P.; and Mazer, E. 2003. A survey of probabilistic models using the bayesian programming methodology as a unifying framework. In *International Conference on Computational Intelligence, Robotics and Autonomous Systems, Singapore*.

Guha, R.; McCool, R.; and Miller, E. 2003. Semantic search. In *Proceedings of the 12th international conference on World Wide Web*, 700–709. ACM.

Jorm, L.; Gruszin, S.; and Churches, T. 2009. A multidimensional classification of public health activity in australia. *Australia and New Zealand Health Policy* 6(1):9.

Lin, F.-R.; Chiu, C.-H.; and Wu, S.-C. 2002. Using bayesian networks for discovering temporal-state transition patterns in hemodialysis. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 153.

Luo, H.; Morrison; Groh, M.; Waters, C.; DesMeules, M.; Jones-McLean, E.; Ugnat, A.; Desjardins, S.; Lim, M.; and Mao, Y. 2007. The burden of adult obesity in canada. *J. Chronic Dis.* 27(4):134?–144.

Mari, J.-F., and Ber, F. L. 2006. Temporal and spatial data mining with second-order hidden markov models. *Soft Comput.* 10(5):406–414.

Martinis, S., and Twele, A. 2010. A hierarchical spatio-temporal markov model for improved flood mapping using multi-temporal x-band sar data. *Remote Sensing* 2(9):2240–2258.

Mirolla, M. 2004. The cost of chronic diseases in canada. *The Chronic Disease Prevention Alliance of Canada*.

Nilsson, M., and Sollenborn, M. 2004. Advancements and trends in medical case-based reasoning: An overview of systems and system development. In *Proceedings of the 17th International FLAIRS Conference, Special Track on Case-Based Reasoning*, 178–183. AAAI.

Sanson-Fisher, R.; Campbell, E.; Bailey, L.; Htun, C.; and Millar, J. 2008. We are what we do. research outputs of public health. *Am. J. Preventive Med.* 35(4):380–385.

Scott, S. L. 2002. Bayesian methods for hidden markov models : Recursive computing in the 21st century. *Journal of the American Statistical Association* 97(457):337–351.

SIO. 2008. Semanticscience integrated ontology. http://semanticscience.org.

Tawfik1, A. Y., and Neufeld2, E. M. 2002. Temporal reasoning and bayesian networks. *Computational Intelligence* 16(3).

Vallet, D.; Fernndez, M.; and Castells, P. 2005. An ontology based information retrieval model. In *In ESWC*, 455–470. Springer.

Wang, H.; Dwyer-Lindgren, L.; Lofgren, K.; Rajaratnam, J.; Marcus, J. R.; Levin-Rector, A.; Levitz, C. E.; Lopez, A. D.; and Murray, C. 2012. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* 380(9859):2071–2094.