Building a Pan-Canadian Primary Care Sentinel Surveillance Network: Initial Development and Moving Forward

Richard Birtwhistle, MD, MSc, Karim Keshavjee, MD, MBA, Anita Lambert-Lanning, MLS, Marshall Godwin, MD, MSc, Michelle Greiver, MD, Donna Manca, MD, and Claudia Lagacé, MSc

The development of a pan-Canadian network of primary care research networks for studying issues in primary care has been the vision of Canadian primary care researchers for many years. With the opportunity for funding from the Public Health Agency of Canada and the support of the College of Family Physicians of Canada, we have planned and developed a project to assess the feasibility of a network of networks of family medicine practices that exclusively use electronic medical records. The Canadian Primary Care Sentinel Surveillance Network will collect longitudinal data from practices across Canada to assess the primary care epidemiology and management of 5 chronic diseases: hypertension, diabetes, depression, chronic obstructive lung disease, and osteoarthritis. This article reports on the 7-month first phase of the feasibility project of 7 regional networks in Canada to develop a business plan, including governance, mission, and vision; develop memorandum of agreements with the regional networks and their respective universities; develop and obtain approval of research ethics board applications; develop methods for data extraction, a Canadian Primary Care Sentinel Surveillance Network database, and initial assessment of the types of data that can be extracted; and recruitment of 10 practices at each network that use electronic medical records. The project will continue in phase 2 of the feasibility testing until April 2010. (J Am Board Fam Med 2009;22:412–22.)

The development of a pan-Canadian network of primary care research networks (PBRNs) for studying issues in primary care has been the vision of Canadian primary care researchers for many years.[1] We are not alone in this research goal. The Euro-

pean General Practice Research Workshop started a European General Practice Research Agenda in 2002.[2] There are a number of research networks within the European Union, such as the General Practice Research Database in the United Kingdom[3] and the Netherlands Information Network of General Practice.[4] In addition, the BEACH project in Australia[5] has been particularly successful conducting surveillance-based projects. The Distributed Network for Ambulatory Research in Therapeutics is a recent initiative in the United States to bring together practices with electronic medical records in 8 different organizations.[6] Van Weel and Rosser[7] have argued for such primary care networks to be supported on a global scale. The College of Family Physicians of Canada (CFPC) has successfully run a National Research System since 1976 and has conducted many funded projects on a variety of primary care topics.[8] Some Canadian physicians were part of the US Ambulatory Sentinel Practice Network until its demise but there has

not been a coordinated national initiative to create a central data source for primary care in Canada.

A major barrier to establishing and sustaining such a network in Canada and elsewhere is the need to build infrastructure when most funding is project based. Green et al[9] have described the necessary infrastructure requirements for an individual PBRN. In 2006, the Canadian Institutes for Health Research funded a workshop at Queen's University, bringing together primary care researchers from across the country interested in building a national network. Representatives of the Public Health Agency of Canada (PHAC) were also in attendance and were looking for opportunities to establish primary care data sources for chronic disease surveillance. In 2008, PHAC issued a request for proposal for a primary care sentinel surveillance system for chronic disease. The chronic diseases of interest were cardiovascular disease, chronic respiratory disease, mental health, arthritis, and diabetes. The CFPC's application was successful and the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) was born. The project is funded through a contribution agreement between the PHAC and the CFPC, with in-kind support from the CFPC and 7 regional primary-care/PBRNs during a 24-month period that began on April 1, 2008. The network initially involved 7 academic primary care research networks in 4 provinces (Newfoundland, Quebec, Ontario, and Alberta). Directors of these networks, CFPC representatives, PHAC representatives, and 2 expert consultants (one an expert on electronic medical records [EMRs] and the other on business planning) met to discuss the approach to developing infrastructure. An initial feasibility project was developed that involved developing a business plan and governance structure, approaches to data collection from EMRs in family doctors' offices, and appropriate management of the privacy and security of patient health information. This article is a description of the initial feasibility study we have completed in network development. We are currently into phase 2 of the pilot project.

The Need for Primary Care Data Sources in Public Health Surveillance

Although historically used for detecting infectious disease outbreaks[10] and occupational health problems,[11] the concept of sentinel surveillance is increasingly applied to the field of chronic disease surveillance.[12] It is based on the concept that one or more sites are chosen to collect clinically verified information (eg, risk factors, diagnosis) about relatively few individuals, representative of a larger population, to identify public health events of interest.

The medical care of people with chronic diseases is, by and large, managed by primary care physicians. Information about these people and their care is mostly locked away in medical records and not a readily available data source for research or surveillance. In addition, patients in primary care often have multiple chronic health conditions and family physicians are their major source of medical care. For example, in the 2007 survey of Canadians' Experiences with Chronic Illness Care, 63% of patients with 1 or more chronic condition indicated that their primary health provider explained examination results to them and 68% with 1 or more chronic conditions indicated that their primary health provider explained specific test results to them.[13] Currently, information about the health of Canadians is derived from national databases that provide details about, for example, mortality statistics, hospital discharge data, disease specific registries, or national population health surveys. However, some health conditions do not always lead to death or generate hospital admissions. Although self-reported survey data can provide useful information about chronic diseases and their risk factors, they are limited because people may not know that they have a specific condition or may not be willing to report it. Further, in Canada there is a single-payer system, and billing data from the provincial health insurance plans are unreliable sources of data. These data sources often do not tell the story of the care of people with chronic disease for many reasons, including the inability to list multiple reasons for the clinical encounter/visit. To produce a more complete picture of the health of the population, clinically verified information collected directly by health care providers is necessary. The data needed, including practice-level data about disease management, both pharmaceutical and nonpharmaceutical, are most readily available where the population receives ongoing health care for chronic diseases. It is clear that primary care practices may provide a rich source of data for chronic disease surveillance if only we can get at it.[14]

Chronic Disease Surveillance Objectives

A 2006 study of 980 patients in a family medicine practice in Sherbrooke, Quebec, determined that 89% to 100% had multiple chronic conditions. The mean number of chronic conditions of patients eligible for each applicable condition ranged from 5.5 ± 3.3 to 11.7 ± 5.3.[15] Issues around the morbidity of multiple chronic diseases underscore the need for surveillance of chronic disease based on the following objectives for the CPCSSN feasibility project:

1. Develop an infrastructure for CPCSSN that will underpin the operations of a robust, longitudinal data collection and maintenance of a primary care data repository on chronic disease.

2. Demonstrate the ability to extract relevant data from multiple EMRs in multiple primary care practice sites.

3. Create a usable CPCSSN database that will be a searchable data repository for primary care researchers and will be the basis for reports for government and others about chronic disease in Canada.

Initial Development

The founding networks for CPCSSN were the Atlantic Practice Based Research Network (Memorial University of Newfoundland); Q-NET (Centre santé et des services sociaus de Laval); the Centre for Studies in Primary Care network (Queen's University); the North Toronto Research Network (University of Toronto); the DELPHI network (Thames Valley Research Unit, University of Western Ontario); the Southern Alberta Primary Care Research network (University of Calgary); and the Alberta Family Practice Research Network (University of Alberta). All of these academic primary care research networks had associated family medicine practices that used EMRs. A recent example of an exploration of the process of change to implement clinical guidelines for primary and secondary prevention of cardiovascular disease in primary care practices used a common EMR.[16]

To limit the amount of data we would collect, the group decided to focus on 5 chronic diseases: hypertension, diabetes, chronic obstructive lung disease, depression, and osteoarthritis. We also hired consultants to support business and strategic planning and the development of the information technology and data management infrastructure.

The initial deliverables for each regional network included the development of a memorandum of agreement (MOA) between the CFPC and the local university or hospital and an ethics application for research ethics board (REB) approval for each site. Once the MOA was signed by all parties, funds flowed to the regional networks as restricted/conditional sub-grants from the CFPC. Funds continue to flow based on deliverables met, in accordance with the main contribution agreement conditions required by PHAC. With the MOAs and REB approvals in place, we then recruited practices that used EMRs. Because of the work involved with assessing data extraction capabilities from individual EMRs, each network was restricted to recruiting up to 10 practices using the same EMR.

Business Planning

Although the initial feasibility project only lasted 7 months, the CPCSSN board of directors decided it was necessary to take a long-term view of the development of the network; therefore a 5-year business plan was developed. As part of this plan, a governance structure, terms of reference for the board and each of the subcommittees, as well as a mission and vision were established. Early in the project we tried to determine who the stakeholders of this network would be and who would be potential consumers of the data repository being developed by CPCSSN. The network is organized as a sub-entity of the CFPC with its own board and a number of sub-committees, as outlined in Figure 1. The board consists of directors of each participating network and representative members of the CFPC. The CPCSSN board worked diligently in phase 1, meeting bimonthly via teleconference and having 2 face to face meetings; the data managers group met every week for 2-hour sessions via teleconference using GoTo Meeting visual technologies (Citrix Online, LLC, Goleta, CA). Funding for CPCSSN is initially through PHAC but future financial sustainability may require funding over and above that expected to come from PHAC. We continue to develop additional funding sources to maintain the infrastructure of the network.

Human Resources

To develop and maintain a robust longitudinal database, careful consideration was given to the nec-

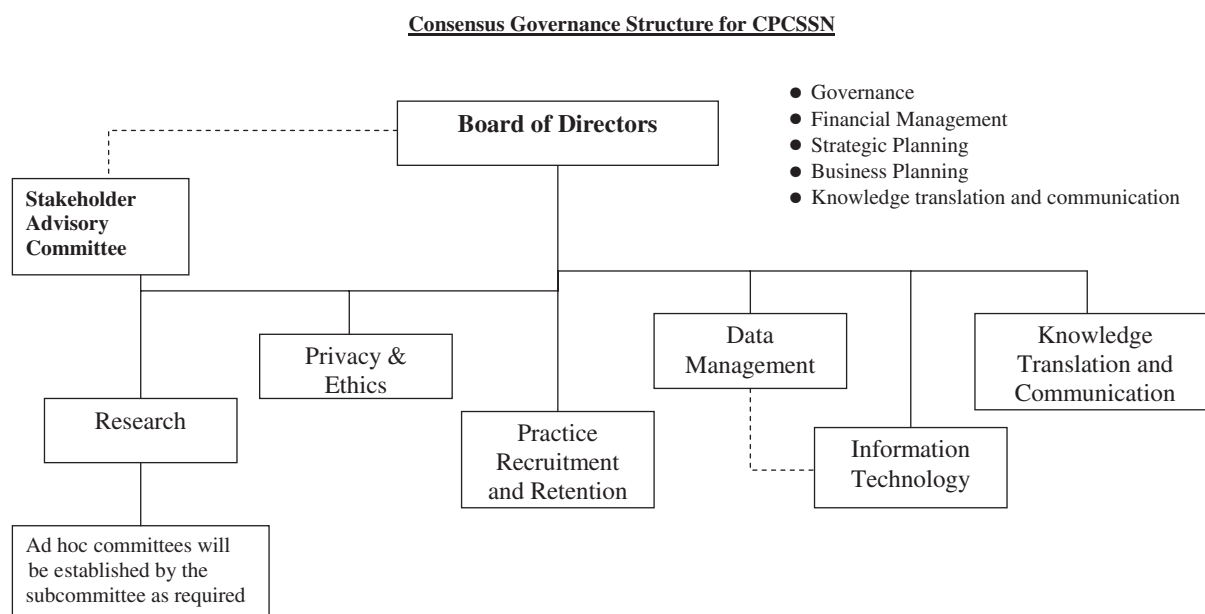**Consensus Governance Structure for CPCSSN**

Figure 1. Consensus governance structure for the Canadian Primary Care Sentinel Surveillance Network (CPCSSN).

essary human resources. We had network directors at each site but also decided that a key position was a full-time data manager at each network site. The EMR data are all sent to a regional repository where they are cleaned and stored in preparation for transfer and merging at the central repository. Because each network is working with different EMRs, having someone like a regional data manager, who could become very familiar with that system and who would be able to develop a personal relationship with the sentinel practices, is crucial. Further, having a regional data manager would help to build capacity for additional primary care research at each network. We have also employed a part-time research assistant at each site for regional research development and data reporting. In phase 2 we have established a central repository and hired a senior data manager who will look after the repository and provide assistance to regional data managers. We have also established a central office at the CFPC for the project manager/director and support staff. The chair of the board of CPCSSN is currently at Queen's University, where the central repository is housed. In the future, however, the chair may be at another network so the funding for the office of the chair will float.

Privacy and Security

In the first phase of the project, 6 out of 7 networks obtained REB approval for the project. The ethical issues focused on by the REBs were the security of patient health data, the need for patient consent, and a de-identified database. All of the approvals were based on not requiring individual patient consent for extraction of records. Signs were posted in the waiting rooms of each practice notifying patients that their personal health information is de-identified; if they do not want their health information used, they can request that they be excluded. Although this has not been a problem to date, the process for removing them from the database would happen during the data extraction and cleaning process at the regional network. These patients will have an assigned CPCSSN number that would be flagged for exclusion on data pulls. Only the practice will know the identity of these patients.

In Quebec, REB approval was not granted during the first phase of the project. The committee's major concern was related to the lack of explicit patient consent and, to obtain approval, individual patient consent would be required. This has delayed practice recruitment in Quebec but will allow for measurement of nonparticipation by patients and comparisons of participation rates as well as other differences between Quebec and other provinces.

All identifying patient information will be stripped from the database before it leaves the practice. A CPCSSN number, which consists of a network and site identifier and the unique EMR num-

ber, is assigned to each patient. Before data are transferred to the central data repository, each regional network data manager will apply risk identification software to the database to ensure that it is adequately de-identified.

EMRs

During phase 1, 6 proprietary EMRs were used: DaVinc (Montreal, Quebec); Healthscreen (Toronto, Ontario); MedAccess (Kelowna, British Columbia); Nightingale (Nightingale VantageMed Corp, Rancho Cordova, CA); P & P (P & P Data Systems, Inc, Ontario); and Wolf (Wolf Medical Systems, Surrey, British Columbia). All EMRs have different coding structures and each network decided on the best way to access the data for extraction (direct versus frontend). Initial test draws of the patient databases identified what information can be extracted. We are learning how to extract data from these EMR programs but it is clear that, as EMRs increasingly becoming a necessary tool in health care, careful selection of a vendor must also consider needs for data extraction for research.[17]

Disease Definitions

Disease definitions have been developed for each chronic disease based on International Statistical Classification of Diseases and Related Health Problems (version 9) codes for diagnosis in combination with certain drugs (from preconstructed drop down lists) and/or positive test results related to the disease. Although these codes lack the specificity preferred for primary care, this is the current coding standard for many EMRs in Canada and will be a potential limitation for the initial data extraction process. The development of metadata and data processes is explained more fully by van Vlyman and de Lusignan[18] in their discussion about a defined number of named elements that convey meaning, given that medical data are complex to process. Access to diagnostic procedures (eg, spirometry) or referrals (eg, referral to psychiatrist) will not be available in all EMRs and will be detailed, as applicable, for each practice site. The "gold standard" of disease definition is the clinical diagnosis of the primary physician. We will provide each physician with a list of their patients who have been identified from EMR data as an index case with chronic disease(s) and ask them to verify this finding. This will require some initial work by the physician but, with subsequent data draws, the number of patient verifications will be much fewer as will the consistency in EMR data input. CPCSSN disease definitions are included as Appendix 1.

CPCSSN Database Development

An entity relationship diagram has been developed that creates a structure for the types of data that is being collected for the database. Each patient has a unique CPCSSN identifier. General categories of data extraction are network and provider identifiers, patient demographics (de-identified), encounter date and encounter type, health condition, physical examination, risk factors, referrals, laboratory investigations, procedures, and medications.

Data Flow

The CPCSSN data repository is a de-identified database of patients with any of the 5 chronic diseases of interest. All patient identifiers are removed from the database before data leaves the practice and only the de-identified data will be kept on a secure regional network server. Once the data has been cleaned, it will be transferred to the central repository server (Figure 2). Data will be extracted from the EMRs every 3 months during phase 2 operations, for a total of 3 extractions. The first iteration was a test extraction of all available data in the EMRs. The second extraction will be full extraction of all available data in preparation for the third and final extraction, which will be designed to extract new information (different from that extracted in the previous extraction) to form the final data warehouse.

Assessment of Data Quality

Health surveillance, as illustrated by medication surveillance, can be achieved using EMRs.[19,20] During phase 1 of the project, we extracted data from the 6 EMRs and assessed the data for data quality issues. The data quality issues uncovered during phase 1 are listed in Table 1. Given the complexity of the data issues and the short timelines, it was not possible to quantify the number of fields in which a particular data quality issue was extant. Discussing solutions to the issue of data quality is beyond the scope of this article.
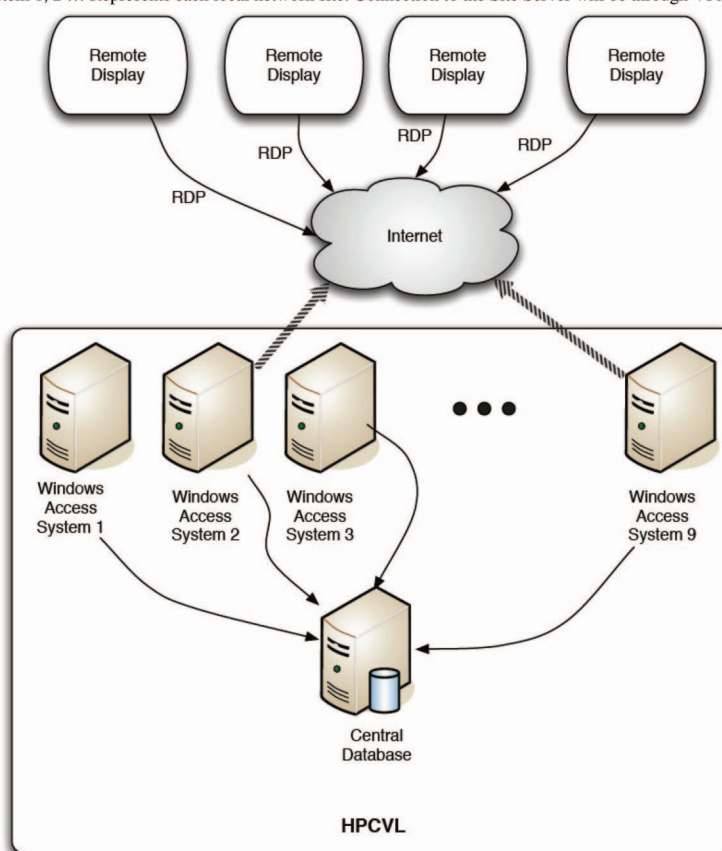
Figure 2. Data repository for regional and central sites at the Queen's University High Performance Virtual Computing Laboratory. System 1, 2, etc represents each regional network site. Connection to the site server will be through a virtual private network.

Considerations for a Data Repository

Information technology architecture

Primary care environments don't routinely have information technology staff who are able to provide high quality information technology services. Given the privacy, confidentiality, and security requirements of the surveillance network, we opted to house all servers for the network in a central server farm housed at the High Performance Virtual Computing Laboratory at Queen's University (Figure 2). Centralizing the servers allows the project to control the security practices much more tightly and to monitor any issues that may arise from a central location. To meet the requirements for accountability and responsibility for data at the regional sites, each network is assigned its own server which is under its complete contractual control.

Privacy and Security Practices

Depending on the regional network, some data managers extract data on-site in the physician of-

fice. Others get a back-up copy of the EMR database, which they use to extract data. To improve scalability over time, data managers will be encouraged to extract data using a secure, remote access utility. Data will be extracted from the physicians EMRs and uploaded directly, via secure connection, to the network's secure server at the central repository. Data managers are discouraged from saving patient data to their laptops or to removable media. All work on extracted data are to be done on the network's secure server.

Data Transformation and Loading

Once data are extracted from the physician's office and transferred to the network's secure server, it goes through 3 processes. First, the data are transformed and put into the CPCSSN database. Second, data in key fields are mapped to standard terms. Some data fields are "cleaned" using a variety of algorithms for data cleaning. Third, the data

Table 1. Canadian Primary Care Sentinel Surveillance Network Data Quality Issues Identified During Phase 1 and Our Initial Remediation Strategy

Data Quality Issue	Description	Remediation Strategy
"Dirty data"	Misspelled words, extra words in field, inconsistent strings ("ex smoker," "ex-smoker"), multiple diagnoses in a single field	Cleaned by data managers using synonym dictionaries and cleaning algorithms
Identifiable Data	Names, phone numbers, and other identifying information in diagnosis or reason for visit fields	Clean using a de-identification engine
Missing data	Dosages, dates of onset, occupation, ethnicity	Ask and train physicians and/or staff to enter appropriate data
Inconsistent data	Diagnoses stored in several different places—notes, PMH, problem list, Inconsistent Risk Factors coexisting—smoker, ex-smoker	Use physician as "gold standard" for confirming diagnoses; use dates to determine latest status of risk factor
Lacking Metadata	Referral to "Dr. Jones," but Dr. Jones' speciality is not listed	Work with EMR vendors to include specialty in address database; encourage staff to enter specialty into address databases
Inappropriate Metadata	Diagnosis not in problem list, medication in encounter notes	Ask physicians to enter "gold standard" diagnosis into Problem List for all patients with an index disease
Insufficient Meta Data	In 2 EMRs, Problem List, Risk Factors and Procedures appear in the same table with no metadata to distinguish the 3 types of data	Work with EMR vendors to separate the 3 different types of data
Lacking standardization	Multiple, changing, inconsistent names or results for lab tests, eg, HbA1C, glycosylated hemoglobin, hemoglobin A1C; 7% vs 0.07 for test results	Work with National standards bodies to encourage uptake of standards
Lacking data feeds	Lab results not coming in electronically	Encourage local labs to provide laboratory results electronically

PMH, past medical history; EMR, electronic medical record.

are processed through a de-identification engine to remove identifying information in text fields and a variety of other methods to decrease reidentification risk.[21]

Central Data Repository

The central data repository (CDR) is housed in the same facility as the network secure servers. The CDR will have a "landing zone" for data uploads from the individual networks. This area is outside the firewall to allow data to be transferred from the networks. As soon as the file arrives in the landing zone, the central server will transfer it to the "staging area," where the data will be processed before it is added to the data warehouse. The CDR will be housed in a SQL server database and will be analyzed using SAS (SAS Institute, Inc., Cary, NC) or SPSS (SPSS Inc., Chicago, IL).

Recruitment and Retention of Practices

Fundamental to the success of this project is the ability to attract and maintain primary care practices in the networks. Anecdotally, many family physicians with EMRs see—once they get over start-up frustration—the benefits from contributing patient health data for understanding chronic disease in a wider context. All physicians, however, are extremely sensitive to the privacy of their patients and want strong assurances that any health data taken from their electronic record is secure and de-identified. Some remain unconvinced despite these assurances. Other issues are well known.[22] We have dealt with the barriers in a variety of ways. We have tried to limit the requirement for physician involvement so that participation will be a minimal burden on the physician or the employees of the practice. Although we do have some practice compensation, this is not a driver for practices to participate. We are developing a process for participating physicians to get CFPC continuing professional development credits and we plan to develop regular reports to provide feedback to sentinel physicians about their practice as well as provincial and national comparators. We want participating practices to be sentinels for many

years but at the moment we do not have any estimate about "sentinel fatigue."

Stakeholder Development

The current major stakeholders working with the network are PHAC, CFPC, and the Canadian Institute for Health Information (CIHI). CIHI has been involved since the early days of the project and, because of their expertise with handling health data, they have been instrumental in the development of data element definitions and data capture processes.

Our relationship with CIHI will eventually allow for linkage studies with other health administrative databases. Developing stakeholder relationships with provincial and national health organizations and health professional data holders is also pivotal to CPCSSN′s success. There may be other groups for whom the network could provide valuable health data, such as for adverse event monitoring, patient safety, and cancer care, to name a few.

Challenges for Future Development

There are many challenges as we go forward, including developing ways to collect data that may be important such as risk factors in chronic disease that may not be normally recorded in EMRs. Examples of this are ethnicity, occupation, education, and income. Recording these risk factors in the EMR where they are easily extractable (rather than as part of the encounter text) is a future goal. A solution to this may be to develop templates that would be acceptable to the physician and be sufficient as an encounter note but would organize data for more easy retrieval from the EMR. Templates are on our list for future development.

As a surveillance system for chronic disease in primary care across Canada, the data quality is a prime consideration. We are developing processes for assessing data quality and will need constant vigilance. Other ongoing issues include refining our approach to estimating practice denominators; the establishment of representativeness within the patient, practice, and regional network populations with whom we collaborate; and to consider broadening our vision of primary care to other health professionals in independent practice.

The growth of EMR use in family medicine in Canada (22% of practices in 2004[23]) provides an important opportunity to collect more accurate, complete, and timely data than traditional billing-based surveillance systems without significantly increasing sentinel physician workloads. Given the resources and will of governments, the number of physicians using EMRs will increase during the next few years. We are also considering how many sentinel practices we will need to be able to provide reliable national and provincial estimates of chronic disease. For this we need to have networks in every province or region and sentinels in the 3 northern territories. As well rural and urban sentinels, representing provincial demographics for both physicians and patients will be needed to provide generalizable estimates. We do not expect this to be attainable in the short term, and statistical standardization techniques will be required. Finally, as we develop a larger database and gain more experience with the data, collaborations with networks in other countries will bring potential for international comparisons.

References

1. G Russell, R Geneau, S Johnston, C Liddy, W Hogg, K Hogan. Mapping the future of primary healthcare research in Canada: a report to the Canadian Health Services Research Foundation, September 2007. Available at: http://www.chsrf.ca/research_themes/pdf/mapping_future_report_2007_e.pdf. Accessed April 6, 2009.
2. Lionis C, Stoffers HE, Hummers-Pradier E, Griffiths F, Rotar-Pavlic D, Rethans JJ. Setting priorities and identifying barriers for general practice research in Europe. Results from an EGPRW meeting. Fam Pract 2004;21:587–93
3. NIHR Clinical Research Network Coordinating Centre. Primary care research network. Available at: http://www.ukcrn.org.uk/index/networks/primarycare.html. Accessed April 3, 2009.
4. Netherlands Institute for Health Services Research. Databases and information systems: Netherlands Information Network of General Practice. Available at: http://www.nivel.nl/oc2/page.asp?PageID=8599&path=/Startpunt/NIVEL%20international/Research/Data%20bases%20and%20information%20systems/National%20Information%20Network%20of%20GPs%20(LINH). Accessed April 3, 2009.
5. Australian General Practice Statistics and Classifica-

tion Centre. The BEACH project: Bettering the Evaluation and Care of Health. Available at: http://www.fmrc.org.au/beach.htm. Accessed April 3, 2009.

6. Agency for Health Care Research and Quality. New research: Distributed Network for Ambulatory Research in Therapeutics (DARTNet). Available at: http://www.ahrq.gov/about/nac/aafp.htm. Accessed April 6, 2009.

7. van Weel C, Rosser WW. Improving health care globally: a critical review of the necessity of family medicine research and recommendations to build research capacity. Ann Fam Med 2004;2(Suppl 2): S5–16.

8. The College of Family Physicians of Canada. National Research System (NaReS). Available at: http://www.cfpc.ca/nares. Accessed April 3, 2009.

9. Green LE, White LL, Barry HC, Neace DE, Hudson BL. Infrastructure requirements for practice-based research networks. Ann Fam Med 2005; 3(Suppl 1):S5–S11.

10. Public Health Agency of Canada. FluWatch. Available at: http://www.phac-aspc.gc.ca/fluwatch/index-eng.php. Accessed April 3, 2009.

11. The Canadian Centre for Occupational Health and Safety. Fatality reports. Available at: http://www.ccohs.ca/products/databases/fatalityreports.html. Accessed April 3, 2009.

12. Public Health Agency of Canada. Chronic diseases. Available at: http://www.phac-aspc.gc.ca/cd-mc/index-eng.php. Accessed April 3, 2009.

13. Health Council of Canada. Canadians' experiences with chronic illness care in 2007: Figure 10. Communicating with your primary care provider. Available at: http://www.healthcouncilcanada.ca/docs/rpts/2007/outcomes2/ExpFig10.jpg. Accessed on April 3, 2009.

14. Keshavjee K, Kerby JP, Holbrook AM. Comparison of diagnostic codes in a clinical-research database and an administrative database. Available at: http://www.compete-study.com/documents/Comparison_of_Diagnostic_Codes_in_a_Clinical-Research_Database_and_an_Administrative_Database.pdf. Accessed April 3, 2009.

15. Fortin M, Dionne J, Pinbo G, Gignac J, Almirall J, Lapointe L. Randomized controlled trials: do they have external validity for patients with multiple comorbities? Ann Fam Med 2006;4:104–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16569712. Accessed June 16, 2009.

16. Nemeth LS, Feifer C, Stuart GW, Ornstein SM. Implementing change in primary care practices using electronic medical records: a conceptual framework. Implement Sci 2008;3:3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=18199330. Accessed June 16, 2009.

17. Holbrook A, Keshavjee K, Langton K, et al. A critical pathway for electronic medical record selection. Proc AMIA Symp 2001:264–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=11825192. Accessed June 16, 2009.

18. van Vlymen J, de Lusignan S. A system of metadata to control the process of query, aggregating, cleaning andanalysing large datasets of primary care data. Inform Prim Care 2005;13:281–91. Available at: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&Cmd=DetailsSearch&Term=16510025. Accessed June 16, 2009.

19. Holbrook AM, Keshavjee K, Kerby JP, Goldsmith CH, and COMPETE Investigators. Evaluation of data quality for medication monitoring and intervention. Abstracts of the 17th International Conference on Pharmacoepidemiology. Toronto, Canada. 23-26 August 2001. Pharmacoepidemiol Drug Saf 2001; 10(Suppl 1):S130–1.

20. Troyan S, Keshavjee K, Holbrook AM. Data quality management in EMR entry. Available at: http://www.compete-study.com/documents/Data_Quality_Management_in_EMR_Entry.pdf. Accessed May 25, 2009.

21. El Emam K, Brown A, AbdeMalik Philip. Model formulation: evaluating predictors of geographic area population size cut-offs to manage re-identification risk. J Am Med Inform Assoc 2009;16:256–66. Available at: http://www.jamia.org/cgi/content/full/16/2/256. Accessed June 16, 2009.

22. Salmon P, Peters S, Rogers A, et al. Peering through the barriers in GPs' explanations for declining to participate in research: the role of professional autonomy and the economy of time. Fam Pract 2007; 24:269–75. Available at: http://fampra.oxfordjournals.org/cgi/reprint/24/3/269. Accessed June 16, 2009.

23. College of Family Physicians of Canada, Canadian Medical Association, and the Royal College of Physicians and Surgeons of Canada. National Physician Survey 2004: results for family physicians. Available at: http://www.nationalphysiciansurvey.ca/nps/results/PDF-e/SP/Specialty/Family_Med/CCFP_J.pdf. Accessed May 31, 2007.

Appendix 1. Canadian Primary Care Sentinel Surveillance Network (CPCSSN) Disease Definition Table, © CPCSSN 2008

Disease	Diabetes	Depression	Osteoarthritis	Hypertension	COPD
Case Finding					
	1. Billing: ICD9-250.x; Service code: in Ontario K030A or Q040A, in other provinces consult physician and document the code.	1. Billing: ICD9 309.0, 309.1 or 309.81 (Physicians should be asked which code they use before doing this search. Potential codes include: 296, 300, 309, 311, 648)	1. Billing: ICD9 –715.x	1. Billing: ICD9 –401–405	1. Billing: ICD9 –490–492, 496
	EMR	EMR	EMR	EMR	EMR
	2. In Problem List	2. In Problem List	2. In Problem List	2. In Problem List	2. In Problem List
	Diabetes/NIDDM/DM	Depression, post-traumatic stress disorder, 311	Osteoarthritis	Hypertension, HTN,	Bronchitis
	3. In the medication list:	3. In Medication list:	3. In Medication List	3. In Medication List	Emphysema
	Insulin	SSRI	None identified	ACEi/ARB	COPD/COLD
	Glyburide (Sulfonylureas)	MAOI		Diuretics	3. In Medication List
	Metformin (Biguanides)	Tricyclics (>75 mg/day)		Beta Blockers	Beta agonists
	4. In the lab result:			Calcium Channel Blockers Alpha blockers	Anticholinergics Xanthines
	HbA1C > 7%				Inhalant corticosteroids
	Fasting BS > 7				4. Risk Factors
					Smoker
Data Type					
Data to be extracted					
Billing Data	None identified	None identified	None identified	None identified	None identified
EMR –Problem List	All Diagnoses	All Diagnoses	All Diagnoses	All Diagnoses	All Diagnoses
EMR –Physiological*	BP	Weight	Weight	BP	Weight
	Weight,		Joint pain		PEFR
	Height				
	foot ulcer,				
	Waist:hip ratio				
	Waist circumference				
EMR -Adverse Events	Allergies to medication Adverse Events				
	Non-efficacious medication				
EMR –Medications List[†]	All medications	All medications	All medications	All medications	All medications

Appendix 1. Continued

Disease	Diabetes	Depression	Osteoarthritis	Hypertension	COPD
EMR –Risks	Smoker Obesity Exercise Diet	Smoker Alcohol	Obesity	Smoker Exercise Alcohol	Smoker
EMR –Labs	F-Glucose, HbA1c, TG LDL, HDL, TC, Alb:Creat in urine	None Identified	None Identified	None Identified	None Identified
EMR –Referral	Ophthalmology, nephrology, neurology, endocrinology Diabetes Education Dietician	Psychiatrist Psychologist	Orthopedic surgeon Physiotherapy Occupational therapy Exercise program	Cardiology Nephrology	Respirology Thoracic surgery
EMR –Diagnostic Imaging	Not being collected	Not being collected	Not being collected	Not being collected	Not being collected
EMR –Procedures	Revascularization procedure Coronary artery disease investigation	ECT	Knee-replacement Hip-Replacement	Angioplasty Coronary artery stent	Spirometry
EMR –Other	N/A	ER Crisis visits Hospitalization	ADL/IADL Education		Hospitalization

*BMI would not be collected as that should be calculated.

†We will use look-up tables to allow conversion of names of meds to particular classes of drugs. This will be done at the central repository level.

EMR, electronic medical record; NIDDM, non-insulin dependent diabetes mellitus; DM, diabetes mellitus; HTN, hypertension; SSRI, selective serotonin reuptake inhibitor; ACEi, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; COPD, chronic obstructive pulmonary disease; COLD, chronic obstructive lung disease; MAOI, monoamine oxidase inhibitor (MAOI) antidepressants; HbA1c, hemoglobin A1c levels in diabetic patients; BS, blood sugar; BP, blood pressure; PEFR, peak expiratory flow rate; TG, triglycerides; LDL, low density lipoproteins cholesterol; HDL, high density lipoproteins; TC, total blood cholesterol; Alb:Creat, albumin/creatinine ratio (ACR) urine; ECT, electroconvulsive therapy; ER, emergency room; ADL, activities of daily living; IADL, instrumental activities of daily living.