

Individuals, communities and population health digital research infrastructure: incorporating primary care electronic medical record data as part of a primary health care information system.

A white paper submitted to the New Digital Research Infrastructure Organization (NDRIO)

December 14, 2020

Sabrina T. Wong¹, David Barber, Michelle Greiver, Donna Manca, John Queenan, Ayat Salman

Current Issues:

Primary care. Strong community-based primary health care (CBPHC) leads to a more equitable system of care with better population health outcomes at reduced cost [1]. Primary care, an essential part of CBPHC delivers general first-contact health care that incorporates both episodic (e.g. urinary tract infection) and chronic (e.g. diabetes management) care in addition to health promotion and disease prevention through an enduring relationship between clinicians and patients and their families. High performing primary care is foundational to achieving the triple aim of health reform—better health, improved patient experience, and more affordable costs [1]. Bodenheimer and colleagues [2] suggest 10 building blocks of high-performing primary care; data-driven improvement was one of four foundational building blocks necessary before achieving success in the higher order blocks. Yet, much of what we know about high performing primary care is based on analyses using health administrative data [3–6] and chart audits [7,8], not data from patients or practices' electronic medical records (EMRs). Across Canada and internationally we have poor infrastructure to regularly collect survey or EMR data from primary care practices. The environment for collecting data from multiple sources, or collecting data once and using it many times (COUMT) across different organizations and regions remains challenging and fragmented.

There is *incredible potential of EMR data* to be used as a tool for quality improvement, communicable and non-communicable disease surveillance and health system planning in CBPHC. For example, large population-based biobanks such as the UK Biobank, BiobankJapan, FinnGen and Precision Medicine Initiative, have started linking EMRs of their biobank participants, mostly from tertiary care, with genetic data obtained by DNA genotyping or sequencing biospecimens. These rich datasets match patient phenotype and genotype, providing the possibility to evaluate genetic risks simultaneously. While it is becoming a game changer in linking phenotype to major diseases such as cancer, heart disease and psychiatric disorders, primary care is ahead of the game in recognizing the power of EMR data. Linking primary care EMR data to population-based biobanks is still in the works [9–11], though linking EMR data to health administrative data has already taken place in Canada.

Primary care EMR data are an important piece of research infrastructure supported in other countries. For example, in the United Kingdom, data collected in primary care research databases are now increasingly used for research in many areas, and for providing information on patterns of disease [12]. These data have value in that clinical and prescription data can provide information to support pharmacovigilance, including information on demographics, medical symptoms, therapy (medicines, vaccines, devices) and treatment outcomes. The UK supports three such data efforts: 'Clinical Practice Research Datalink' (CPRD), 'QResearch' and 'The Health Improvement Network' (THIN). The data made available to researchers are anonymised, and strong patient identifiers such as name, address and postcode, date of birth and NHS number are removed. Notably, The CPRD is jointly funded by the NHS

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name

National Institute for Health Research (NIHR) and the Medicine and Healthcare Products Regulatory Agency (MHRA) [13]. QResearch data currently come from over 950 general practices using the Egton Medical Information Systems (EMIS) clinical computer system that is used throughout the UK. Finally, THIN is a collaboration between two companies; In Practice Systems Ltd. (INPS), who developed Vision software used by General Practitioners in the UK to manage patient data, and Cegedim Healthcare Software [14].

In Canada, improving patient care through the regular use of EMR data has been widely recognized at the practice-patient interface, where initiatives such as the College of Family Physicians of Canada's Practice Improvement Initiative support family physicians in implementing quality improvement [15]. Beyond the patient-practice interface, considerable work has taken place in Canada to develop a pan-Canadian primary care EMR data repository that can be used for research, epidemiology and health system monitoring. In Canada, rather than having multiple extractions and different resulting primary care EMR databases, we have an opportunity to create a countrywide resource. The Canadian Primary Care Sentinel Surveillance Network (CPCSSN), established in 2008, has developed a pan-Canadian primary care EMR data repository [16]. The CPCSSN has successfully built trusting relationships between primary care clinicians and researchers over the past 12 years. The CPCSSN draws on technological expertise to securely extract EMR data from the practices and includes close to 1500 participating family physicians, nurse practitioners and other primary care clinicians. CPCSSN applies standardized ontologies and terminologies to transform data from various EMR vendors into a common data schema. The source EMR data undergoes an extract, transform and load process that results in a deidentified, cleaned and harmonized set of pan-Canadian data that can be used for research, surveillance and quality improvement at local and national levels. CPCSSN has also developed processes that allow participating clinics to securely reidentify and view their own data to enable them to prepare customized lists of patients in specific risk populations [17].

Data from primary care EMRs forms the backbone for practice based research and learning networks. One of the more recent, ongoing studies is the "Structured Process Informed by Data, Evidence and Research (SPIDER)" intervention [18]. The objective of this work is to assess whether the intervention reduces targeted potentially inappropriate prescriptions for these patients compared with usual care. It is a collaboration between quality improvement (QI) and research programs. Primary care teams form interprofessional learning collaboratives and work with QI coaches to review EMR data provided by their regional practice based research and learning networks, identify areas of improvement, and develop and implement changes. Indeed there are other examples where CPCSSN also has been used for health services planning, disease surveillance and epidemiological studies and other research (e.g., health informatics).

Some of CPCSSN's challenges include: (1) lack of infrastructure support to become a representative, population based pan-Canadian resource where EMR data could be linked to other data across multiple jurisdictions; (2) increasing monetization of primary care EMR data by private sector EMR vendors through the imposition of access barriers and toll fees, despite the willingness of clinicians and patients to have their data used for research in the public benefit (e.g., long haul COVID-19 impacts on health); (3) lack of legislation preventing private sector blocking or monetization of health data and lack of oversight of private sector activities relating to interoperability failures; and (4) legislative, regulatory and administrative barriers to data acquisition for public sector research and linkage.

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name

Future State:

Opportunity beckons for inclusion of primary care EMR data as part of a cohesive Canadian Digital Research Infrastructure (DRI) ecosystem. *Primary care EMR data should be used for public good.* Part of our vision for the Canadian DRI ecosystem would be to include primary care EMR data for widespread research, in addition to meaningful uses such as for quality improvement, case finding and disease surveillance. Primary care EMR data contains information about sociodemographic characteristics, medications, laboratory results, comorbidities and risk factors (e.g., smoking status, alcohol use, blood pressure), and information about the reason for the primary care visit. Data from outside of hospital settings will be essential for Canada to understand and contain COVID-19 (and other future communicable diseases), but national, provincial, and territorial data platforms currently only have linked EMR records for a portion of the populations they cover; some completely lack EMR data. The integration of primary care EMR data with other critical health data through partnerships will significantly advance the establishment of the integrated health data infrastructure in Canada which is required not just for COVID-19, but for multiple essential uses long into the future.

Tools, services and resources. The ability to automate EMR extraction and tools to improve processing and standardization of these data could improve research data management. Properly funding a Canada-wide primary care EMR data curator, such as CPCSSN or other interoperable components, as part of DRI's research data management system could be vastly change the Canadian health data landscape; this would facilitate the inclusion of linkable population-based EMR data across Canada. This support would recognize the fundamental importance of EMR data to multiple forms of health research (e.g., health informatics, chronic disease prevalence, health services, planning and evaluation, pragmatic randomized clinical trials, use of new analytic techniques) and the development of a learning health system in Canada. A successful learning health system seems doomed to fail without the inclusion of primary care, including robust support for EMR data.

Another resource would be to include harmonizing metadata across data repositories/datasets which could catalyze larger-scale studies and ultimately lead to discoveries. Existing health-related datasets encompass a broad range of types of data: clinical data collected during care, administrative data collected for operational and reporting purposes, cohort data collected often longitudinally from targeted population sections, and research data, often with a great depth of information, focused on a specific scientific question. Within each dimension, metadata harmonization (e.g., mapping of variables to a common vocabulary) of similar datasets from different sources (e.g., admin data from different provinces, cohort data from different cohorts) will increase the number of cases that can be studied.

With the rapid evolution of data protection laws in Canada and the increased request for biospecimens from a primary care population, streamlining and standardizing the process of consent and, further, the collection of biospecimens, will control the costs and resolve some of the constraints currently identified with the interoperability of data. In addition, having access to a more detailed catalogue documenting the specific variables (e.g., including particular question and answer categories), biospecimens collected and associated consent constraints is therefore increasingly considered a prerequisite to the establishment of cost-effective research networks and a vital instrument to leverage multi-centre collaborations.

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name

Continuing to build Canada's capacity for advanced analytics using large amounts of data is a service that could be delivered through DRI partnerships. One example is the use of CPCSSN data linked to CANUE data and the use of analytics such as inductive logic programming (ILP). Using methods from this subfield of relational machine learning could make it possible to generate clusters of environmental exposures (air pollution, walkability, food outlets, park density, greenness, area-based social and material-based deprivation measures) for those patients who have one or multiple chronic conditions. In turn, this kind of work provides the opportunity for discovery of new features and patient profiles, which could then be integrated within clinical care and represent target groups within communities for interventions. We are highly supportive of the NDRIO working to strengthen collaboration with individuals and organizations that are submitting White Papers to NDRIO that focus on the components and enablers of this integrated infrastructure: Health Data Research Network Canada (HDRN Canada, P. Alison Paprica), Canadian Research Data Centre Network (CRDCN, Martin Taylor), Diabetes Action Canada (Gary Lewis), the Canadian Patient Information Network (William Hogg), and the Data analytics platform (Laura Rosella).

The single largest existential *challenge* in Canada to using primary care EMR data is the privatization and monetization of patient health information. For many reasons, effective health information exchange and quality improvement for clinicians and use of these data for research has failed to materialize. In part, EMR vendors and health systems have engaged in information blocking in order to increase revenue and market share. This is a challenge that is already faced by our US counterparts. The Office of the National Coordinator, which leads and coordinates US health information technology efforts, reports that information blocking "occurs when persons or entities knowingly and unreasonably interfere with the exchange or use of electronic health information" [19]. One-half of US national health information exchange leaders that responded to a survey reported that EMR vendors routinely block the flow of information. The most common practices include deploying products with limited interoperability, charging disproportionately high fees for health information exchange, and making third party access to standardized data difficult [20].

These practices have implications for quality improvement, research, and patient care. Providers and scholars increasingly rely on EMR data to track diseases, study interventions, and refine the delivery of care. These activities are critical as health systems embrace learning health care principles [21]. Within primary care, organizations are aggregating demographic, social determinants of health, clinical, and billing data extracted from EMRs to coordinate care, conduct comparative effectiveness research, and accelerate learning across practices [16,22].

Recognizing the importance of these data for the future of primary care (and other health related) research, the North American Primary Care Research Group's Research Advocacy Committee drafted and passed a position statement condemning information blocking. This statement aligns with similar calls from the College of Family Physicians Canada and American Academy of Family Physicians [23].

How to Bridge the Gap:

In order to achieve the desired future state of including primary care EMR data in the Canadian DRI ecosystem, NDRIO should leverage already existing tools, services and resources that have started to form building blocks towards a primary care information ecosystem. The below are a few of the existing pieces that would form Canada's DRI:

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name

HDRN Canada: The recently formed Health Data Research Network, brings together various national, provincial, and territorial health data partners to jointly make data accessible under strong governance so that data can be used for public benefit while privacy is protected. The more sites, provinces and territories that contribute health data to a study or analysis, the more we will all have the ability to learn what is working well within and across health systems. HDRN Canada makes it easier for researchers, analysts and policy makers to use data from multiple provinces and territories for benchmarking, research, health system planning and innovation.

CPIN: The Canadian Patient Information Network (CPIN: <https://en.cpin-rcip.com/>) allows primary care practices to better communicate with their patients and improve the operational efficiencies of practice. CPIN provides a service to primary care clinicians AND can be used for automated collection of patient reported experience and outcome measures.

SPOR Primary and Integrated Health Care Innovation Network (PICHIN) and chronic disease networks such as Diabetes Action Canada. These research networks form the “front facing” infrastructure needed to carry out research. They have worked to create clinician, science and policy engagement and be inclusive of patients as partners in carrying out the research. DRI can form an integral part of the “backbone” infrastructure which is sorely needed by individual researchers, groups of researchers and stakeholders and organizations.

CPCSSN, with its functions noted above, is closely aligned with these organizations and has the ability to: link EMR data to other health data, report data back to practices for purposes of quality improvement, provide information for purposes of disease surveillance and health system planning, and be used for research.

Data analytics platform (Laura Rosella) [see white paper].

Bridging the gap means working with stakeholders to create a DRI that would serve the public good. Data would be collected once and used many times. It means building on current capacity, such as:

- a) collecting [primary care EMR data] once and using many times for public good;
- b) expanding data comprehensiveness and interoperability by working together on metadata-level information on existing data sources and standardizing these resources to make them more accessible and usable;
- c) expanding the existing data toolbox by developing automated data tools to facilitate data cleaning, integrating, harmonizing, processing and analysing using integrated visualization and computation, independent of what the data source is or where data are located; and
- d) standardising data governance structures or establishing guidelines for the development of data governance across the ecosystem will facilitate exchanges and bring solutions to common difficulties encountered with current ecosystems that provide tools, services and resources for primary care data.

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name

References

1. Berwick DM, Nolan TW, Whittington J. The triple aim: Care, health, and cost. *Health Affairs*. 2008;27(3):759-769. doi:10.1377/hlthaff.27.3.759
2. Bodenheimer T, Ghorob A, Willard-Grace R, Grumbach K. The 10 building blocks of high-performing primary care. *Annals of Family Medicine*. 2014;12(2):166-171. doi:10.1370/afm.1616
3. Hogg W, Johnston S, Russell G, Dahrouge S, Gyorfi-Dyke E, Kristjansson E. Conducting waiting room surveys in practice-based primary care research: a user's guide. *Canadian Family Physician*. 2010;56(12):1375-1376. <http://www.ncbi.nlm.nih.gov/pubmed/21156900>.
4. Dahrouge S, Hogg W, Russell G, et al. The Comparison of Models of Primary Care in Ontario (COMP-PC) study: methodology of a multifaceted cross-sectional practice-based study. *Open Medicine*. 2009;3(3):149-164. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3090123/>.
5. Jaakkimainen L, Klein-Geltink JE, Guttman A, et al. Indicators of primary care based on administrative data. In: Jaakkimainen L, Upshur REG, Klein-Geltink JE, et al., eds. *Primary Care in Ontario: ICES Atlas*. Toronto ON: Institute for Clinical Evaluative Sciences (ICES); 2006:207-250. http://www.jcb.utoronto.ca/people/publications/PC_atlas_complete.pdf.
6. Stukel T, Croxford R, Rahman F, Bierman AS, Glazier RH. *Variations in Quality Indicators across Ontario Physician Networks*.; 2016. <https://www.ices.on.ca/Publications/Atlases-and-Reports/2016/Variations-in-Quality-Indicators-Across-Ontario-Physician-Networks>.
7. Tu K, Mitiku TF, Ivers NM, et al. Evaluation of Electronic Medical Record Administrative data Linked Database (EMRALD). *American Journal of Managed Care*. 2014;20(1):e15-e21. <https://pubmed.ncbi.nlm.nih.gov/24669409/>.
8. Green ME, Hogg W, Savage C, et al. Assessing methods for measurement of clinical outcomes and quality of care in primary care practices. *BMC Health Services Research*. 2012;12:214. doi:10.1186/1472-6963-12-214
9. Li R, Chen Y, Ritchie MD, Moore JH. Electronic health records and polygenic risk scores for predicting disease risk. *Nature Reviews Genetics*. 2020;21(8):493-502. doi:10.1038/s41576-020-0224-1
10. Reus LM, Shen X, Gibson J, et al. Association of polygenic risk for major psychiatric illness with subcortical volumes and white matter integrity in UK Biobank. *Scientific Reports*. 2017;7:1-8. doi:10.1038/srep42140
11. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *American Journal of Human Genetics*. 2019;104(1):21-34. doi:10.1016/j.ajhg.2018.11.002
12. Mannan F, Chaudhry Z, Gibson-White A, et al. Outputs and growth of primary care databases in the United Kingdom: bibliometric analysis. *Journal of Innovation in Health Informatics*. 2017;24(3):284-290. doi:10.14236/jhi.v24i3.942
13. Clinical Practice Research Datalink. UK Data Driving Real World Evidence. <https://www.cprd.com/>. Accessed December 11, 2020.
14. The Health Improvement Network. No Title. <https://www.the-health-improvement->

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name

- network.com/en/. Accessed December 11, 2020.
15. Pimlott N, Katz A. Ecology of family physicians' research engagement. *Canadian Family Physician*. 2016;62(5):385-390. <https://www.cfp.ca/content/62/5/385>.
 16. Garies S, Birtwhistle R, Drummond N, Queenan J, Williamson T. Data Resource Profile: National electronic medical record data from the Canadian Primary Care Sentinel Surveillance Network (CPCSSN). *International Journal of Epidemiology*. 2017;46(4):1091-1092f. doi:10.1093/ije/dyw248
 17. Queenan JA, Birtwhistle R, Drummond N. Supporting primary care public health functions. *Canadian Family Physician*. 2016;62(7):603-603. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4955092/>.
 18. Greiver M, Dahrouge S, O'Brien P, et al. Improving care for elderly patients living with polypharmacy: Protocol for a pragmatic cluster randomized trial in community-based primary care practices in Canada. *Implementation Science*. 2019;14(1):55. doi:10.1186/s13012-019-0904-4
 19. (ONC) Office of the National Coordinator for Health Information Technology. *Report on Health Information Blocking: Report to Congress*.; 2015. https://www.healthit.gov/sites/default/files/reports/info_blocking_040915.pdf.
 20. Adler-Milstein J, Pfeifer E. Information Blocking: Is it occurring and what policy strategies can address it? *Milbank Quarterly*. 2017;95(1):117-135. doi:10.1111/1468-0009.12247
 21. Smith M, Saunders R, Stuckhardt L, McGinnis MJ. *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. Washington, DC: National Academy of Sciences; 2013. doi:10.17226/13444
 22. DeVoe JE, Gold R, Cottrell E, et al. The ADVANCE network: Accelerating data value across a national community health center network. *Journal of the American Medical Informatics Association*. 2014;21(4):591-595. doi:10.1136/amiajnl-2014-002744
 23. Liaw W, Wong S. North American primary care research group position statement on electronic health record information blocking. *Annals of Family Medicine*. 2019;17(1):91-92. doi:10.1370/afm.2345

¹ Designated contact person, Sabrina.wong@ubc.ca, other authors listed in alphabetical order according to last name