

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

PRHDN-Primary Care **Notice of Intent - CFI Cyber-infrastructure Challenge 1 Competition 1**

Summary

- The PRHDN-Primary Care project will establish primary health care research data infrastructure as part of the broader Pan-Canadian Real-world Health Data Network (PRHDN) led by senior representatives from the Institute for Clinical Evaluative Sciences, the Manitoba Centre for Health Policy, Population Data BC, and the Canadian Institute for Health Information.
- PRHDN-Primary Care will establish distributed data networks that allow researchers and policy/decision makers across Canada to make full use of linked administrative (real-world) data holdings in multi-province studies while data remain within provincial boundaries.
- PRHDN-Primary Care core research data infrastructure will include algorithms, harmonized common data and common analytic protocols.
- PRHDN-Primary Care objectives are to:
 - Complement and build on existing expertise, structures and processes; providing connecting, not duplicative infrastructure.
 - Support knowledge creation that is informed by local connections between researchers and policy/decision makers; building on the strength of provincial resources.
 - Increase the efficiency of research by establishing standing research data infrastructure and processes for multi-province studies.
 - Increase the robustness of research findings, and the capacity of researchers and policy/decision makers to use real-world data, by bringing together experts from across Canada to establish and implement common approaches to multi-province analyses.

Criterion 1: Canada has significant research strengths and globally-competitive expertise in the area of the proposal

Many jurisdictions track health care metrics such as the number of people that get laboratory tests like hemoglobin A1c (HbA1c) for diabetes and the number of people who visit emergency departments. Because we have population-wide de-identified data holdings that are linked at the individual level, Canada is one of the few places in the world where we can capture that information for the same individuals over time, and study the relationship between, for example, not having regular diabetes lab tests and emergency department use. With few exceptions, Canadian linked health data have been used in single province studies, but there is growing demand from health researchers, policy makers and decision makers to advance to national-level analyses and comparative research that capitalizes on Canada's individual-level linked data.

To address this opportunity, the Institute for Clinical Evaluative Sciences (ICES), the Manitoba Centre for Health Policy (MCHP), Population Data BC (PopData) and the Canadian Institute for Health Information (CIHI) have developed the PRHDN proposal to bring together diverse data and expertise and unite them through a Health & Social Data Council and National Coordinating Centre (see the PRHDN website: www.prhdn.ca for additional details)). As a first step toward PRHDN, CFI Cyber-infrastructure project funding would establish PRHDN-Primary Care research data infrastructure.

34968: Pan-Canadian Real-world Health Data Network – Primary Care (PRHDN-Primary Care)

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

In the last 15 years there has been unprecedented investment in primary care reform across Canada, totaling several billion dollars. These reforms range from the establishment of team-based primary care, to incentives for physicians, new and varied payment models (fee-for-service, capitation and blended models) and formal linkages between primary care and other health system sectors. While there are some commonalities across provinces, most notably in the objectives of the reforms to improve care, improve health, and control costs, the specific policies differ. The learnings from these natural and ongoing policy experiments involving millions of Canadians can, and should, inform policymaking across Canada and internationally. Studying them will require that (a) we identify all the relevant individual-level linked data that are already comparable among provinces (b) we harmonize new data relevant to priority questions and (c) we develop and share methods that allow us to perform analyses in “apples to oranges” cases where data in different provinces cannot be harmonized perfectly. PRHDN-Primary Care will use an inclusive process to establish research data infrastructure on all three of these fronts.

The PRHDN-Primary Care Consortium Leads Team (Leads Team) is the decision-making body and management team that will oversee all project work. PRHDN core personnel (not less than 3 FTEs) will be funded to coordinate and provide logistics, document management, communications and secretariat support. The broader PRHDN Consortium, (89 members as of May 21, 2015), comprises researchers, policy makers (e.g., staff and management at ministries of health) and decision makers (e.g., staff and management from regional health authorities, hospitals, and research funding organizations) including senior leaders who have agreed to serve as advisors to the Leads Team. The general PRHDN-Primary Care process for the establishment of new research data infrastructure will be:

- PRHDN core personnel will perform background groundwork, gathering information from the literature, Leads Team and key advisors to PRHDN to develop short lists of specific potential PRHDN-Primary Care activities (e.g., identifying specific algorithms that could be established and/or validated).
- Background information and potential activities will be shared with users of infrastructure in the PRHDN Consortium, both researcher and policy/decision makers, and final decisions about which activities to pursue will be made by the Leads Team incorporating the priorities and feedback of PRHDN Consortium members and input from advisors.
- Working groups will be established to perform the activities (e.g., groups that will work together and collectively agree whether an existing algorithm validation is sufficient, undertaking new validations as required).
- Once infrastructure is ready for use, information about it will be communicated to the PRHDN Consortium and posted on www.prhdn.ca. The introduction of new research data infrastructure will be accompanied by full documentation, step-by-step process guides for using it in multi-province studies, and webcasts and in-person training seminars. As in-person training sessions represent opportunities to bring members of the PRHDN Consortium together, they will generally be paired and combined with policy dialogues and other knowledge mobilization activities that help build relationships and capacity across the PRHDN Consortium. Web-based training sessions will be recorded and posted to www.prhdn.ca to serve as standing resources.

This inclusive process will not only ensure that the infrastructure established is high quality and useful, it will also serve as a mechanism to build buy-in and increase awareness and use of PRHDN-Primary Care research data infrastructure.

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

Criterion 2: The research data infrastructure project is essential in maintaining Canada at the leading-edge internationally

Canada has benefited from a health data advantage based on our large single-payer health systems, but other jurisdictions are gaining ground. Recently established and notable international “competitors” (who are also potential collaborators) include the HMO research network in the US and UK’s Farr Institute, both of which bring together data from multiple organizations and/or sectors to enable leading edge multi-jurisdiction analyses. In the absence of core research data infrastructure funding, Canadian researchers undertaking multi-province studies have borne the burden of establishing comparable multi-province data one study at a time. This has deterred them from undertaking multi-province studies that would ensure Canada’s position as a global leader in real-world health data research.

Because health care delivery is a provincial responsibility, there are few data sets that are comparable across provinces. While the Canada Health Act enshrines public payment of medically necessary physician and hospital services, the manner in which those services are paid, and the data collected to document payments and services, differ by province. Provinces have naturally focused on research data infrastructure that makes full use of all of their own detailed (and often unique) real-world data holdings; however, in doing so we have collectively underexplored opportunities to establish common definitions and methods that can be applied across multiple provinces. CIHI has made significant investments in standards for data collection which have paid dividends for tracking and studying institution-based care, but challenges remain for primary health care, where community-based physician data can vary significantly between provinces. PRHDN-Primary Care will establish the three types of primary care research data infrastructure that are essential if Canada is to maintain its leading edge internationally.

1. Algorithms

PRHDN-Primary Care Commitment: At least 15 algorithms, nine of which will be available to the users within the first 6 months of CFI funding.

An algorithm is a reusable process, ideally in the form of documented code, that implements a common approach or definition. Algorithms can be used for a variety of purposes, such as to define “cases” for certain conditions or to create derived variables. There are many published algorithms, including those developed through the Canadian Chronic Disease Surveillance System (CCDSS)¹, but a much smaller subset that have been subjected to rigorous testing to ensure comparability across different provincial data holdings. The work of PRHDN-Primary Care will be in confirming and making available existing comparable algorithms, and building and testing new algorithms for key concepts in primary care where existing ones do not exist.

The first phase of algorithm work will focus on existing algorithms that could be applied without further validation to produce data that are comparable between different provinces. The specific number and nature of algorithms for this phase will depend on input and advice from the PRHDN Consortium on factors such as: feasibility (including data quality), priority (does the algorithm relate to priority questions), opportunity (based on existing inter-provincial policy differences).

¹ The Canadian Chronic Disease Surveillance System (CCDSS) uses linked administrative data sources from provinces and territories to estimate the incidence, prevalence, use of health services and health outcomes for chronic conditions. 18 CCDSS algorithms are available through <http://open.canada.ca/en/open-data>.

34968: Pan-Canadian Real-world Health Data Network – Primary Care (PRHDN-Primary Care)

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

Taking all these factors into consideration, it is expected that nine or more algorithms may be identified as being immediately ready for use in multi-province studies, such as:

- The subset of case definition algorithms published by the CCDSS that are already comparable between provinces because they are based on CIHI standardized hospital data, e.g., acute myocardial infarction – defined by CCDSS as one or more hospitalizations within a year.
- New algorithms for high-level simple derived variables, (i.e., variables that do not exist in an original data set but can be created by combining or otherwise manipulating existing information), that can be established without validation, e.g., primary care services – binary (yes/no) indicating whether an individual has one or more billed service with any primary care provider in a defined period of time.

Though the CCDSS case definition algorithms from this phase of work will be based on hospital data, they can be combined with primary care data in studies that are directly relevant to primary care policy but not easily performed in the absence of PRHDN-Primary Care infrastructure, e.g., testing the hypothesis, *Is an individual more likely to be hospitalized for hip fracture (in any province) if they haven't visited a primary care provider for more than one year?*

The second phase of PRHDN-Primary Care algorithm work will focus on CCDSS algorithms that include both hospital and physician billing data, e.g., diabetes – defined by CCDSS as one or more hospitalizations or two or more physician claims within two years. Because physician billing data are much more likely to vary between provinces, work will begin with compiling information about previous validation(s) of these algorithms (if any) to determine the extent to which the scientific community has confirmed that they produce comparable data in different provinces. The Leads Team will lead and coordinate work with advisors to determine if further validation is required. In cases where it is determined that no additional validation is required, the algorithms will be posted to www.prhdn.ca with phase 1 algorithms. Where validation is required, it will be performed as described below.

The third phase of work will focus on establishing algorithms that convert dissimilar provincial data into data that are equivalent through “retrospective harmonization” i.e., the process through which variables from different time periods, sources, or jurisdictions are made to be inferentially equivalent and thus can be considered common data. It is expected that new algorithms will make use of agreed-on fields in the domains of: hospital data, prescription drug claims data, physician services (billing) data, emergency room services data and mortality data. Potential priority algorithms for the third phase include:

- Algorithms for chronic conditions that do not currently have CCDSS algorithms (e.g., chronic kidney disease)
- Specific definitions of multi-morbidity and/or high users of health care
- Algorithms that define the model of primary care that was in effect at a point in time
- Comparable total cost of care
- Derived variables for emergent properties that cannot be measured at a service encounter level but emerge by considering individual experiences over time or by aggregating the experience of many individuals, e.g., comprehensiveness, accessibility and equity of care

Where validation is required, it will make innovative use of Electronic Medical Record (EMR) data; for example using specific test results in the EMRs to identify individuals with diabetes, and comparing that number to the number obtained when the PRHDN diabetes algorithm is applied to the administrative data records of the same individuals. Retrospective validations will also make use of work underway at CIHI to harmonize physician billing codes in four provinces. The

34968: Pan-Canadian Real-world Health Data Network – Primary Care (PRHDN-Primary Care)

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

PRHDN approach to validation will be innovative in itself, and peer-reviewed publications of validations are planned. Performing the first validations on relatively simple algorithms will allow the validation working groups to establish processes that serve as the foundation for more complex validations in later PRHDN-Primary Care work.

In addition to the utility of the three phases of algorithms for generating harmonized common data for multi-province or pan-Canadian studies (see below), it is anticipated that researchers will also use PRHDN algorithms in single province studies, the result being studies that generate results that can be compared, or even combined, with other independently-performed single province studies.

2. Harmonized Common Data

PRHDN-Primary Care Commitment: The first PRHDN harmonized common data will be in researchers and policy/decision makers' hands within 9 months of the start of CFI funding; new previously unavailable harmonized data (in cases where only dissimilar data are available currently) will be in users' hands within 2.5 years of the start of CFI funding.

Harmonized common data are collections of variables that are comparable across jurisdictions; they include data that are equivalent from the time that they are collected by the primary source (common data) and data that have been harmonized. Beginning with a foundation of common data collected using CIHI standards, the four Leads Team centres will establish standardized sets of harmonized common data and make them available to researchers and policy/decision makers. Under PRHDN-Primary Care, harmonized common data will generally stay within provincial boundaries and be accessed by researchers using existing mechanisms (the exception being data that are already provided to CIHI by provinces). This will prompt studies to be organized with one research lead in each province who works closely with their local provincial policy and practice stakeholders, consistent with the PRHDN objective of building on existing regional strengths.

In Ontario, standing linkages will be established so that researchers can access research ready datasets based on PRHDN algorithms through ICES Data and Analytic Services. The MCHP, PopData and CIHI will use their current practice of generating linked datasets in response to specific research requests. Generally, the four centres will begin offering harmonized common data to researchers and policy/decision makers within three to five months of the date that an algorithm is finalized. Importantly, algorithms will be established to enable use by researchers and policy/decision makers from across Canada, and the long-term PRHDN vision is to have additional provinces and territories provide harmonized common data as future funding allows.

3. Common Analytic Protocols

PRHDN Commitment: core Common Analytic Protocol template and repository of Common Analytic Protocols used in specific studies within 2 years of the start of CFI funding.

There are important research questions that cannot be addressed through harmonized common data because provincial data holdings are too dissimilar. Following the approach of the Canadian Network for Observational Drug Effect Studies (CNODES)², a common analytic protocol is a

² CNODES is a federally funded collaboration that uses real-world drug and other data for 40 million Canadians in drug safety and effectiveness studies. CNODES uses distributed analyses with common analytic protocol followed by meta-analysis of results to provide pan-Canadian estimates. More details at www.cnodes.ca.

34968: Pan-Canadian Real-world Health Data Network – Primary Care (PRHDN-Primary Care)

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

series of technical and process documents detailing the steps for analyses designed to yield results that can be pooled or combined even though the studies are performed on data in non-harmonized (unaltered) state. Common analytic protocols may include simple statistical methods, more complex techniques, or both. Employing common analytic protocols has the advantage of building capacity as scientists in each participating province learn and apply sophisticated techniques, such as high dimensional propensity score matching, to analyses of their own provincial data holdings.

PRHDN-Primary Care will establish a common analytic protocol template that includes a core dataset creation plan which each province participating in a multi-province study would adapt and apply to their relevant data holdings. The dataset creation plan and process guidance of the PRHDN core common analytic protocol template will be designed to make full use of common or harmonized data, acknowledging that within-province tailoring will also be required.

PRHDN-Primary Care will also establish a repository of completed common analytic protocols so that researchers undertaking new multi-province studies that rely on non-harmonized data can benefit from the experience and methods of previous studies.

Criterion 3: The consortium is composed of a critical mass of leading Canadian researchers in the domain who are fully engaged in the project and who are capable of exploiting the full potential of the research data infrastructure

The Leads Team has leading primary care researchers, first-hand knowledge of how to bring evidence into practice and policymaking, experience with establishing standardized data infrastructure and making it accessible to researchers while maintaining the highest standards for privacy and security, and extensive management skills (see Table 1). Significantly, the Leads Team has representatives from CNODES (Chateau: Manitoba CNODES site lead) and initiatives under the national Strategy for Patient-Oriented Research (SPOR) funded by the Canadian Institutes of Health Research (CIHR) (Schull: lead Ontario SPOR SUPPORT Unit Data Platform Working Group, Katz: co-lead Manitoba SPOR Primary and Integrated Health Care Innovations [PIHCI] Network).

With respect to primary-care research expertise, recent publications authored by members of the Leads Team include:

- A population-based study comparing patterns of care delivery on the quality of care for persons living with HIV in Ontario, *British Medical Journal (BMJ) Open*, (2015)
- Classifying physician practice style: a new approach using administrative data in British Columbia, *Medical Care*, 2015
- Developing a non-categorical measure of child health using administrative data. *Health Reports*, 2015
- Effect of scattered-site housing using rent supplements and intensive case management on housing stability among homeless adults with mental illness. A randomized trial. *Journal of the American Medical Association (JAMA)*, 2015
- Continuity of Care With Family Medicine Physicians: Why It Matters, *CIHI Report*, 2015
- Does access to a colorectal cancer screening website and/or a nurse-managed telephone help line provided to patients by their family physician increase fecal occult blood test uptake? A pragmatic cluster randomized controlled trial study protocol. *BiomedCentral Cancer* 2014
- All the right intentions but few of the desired results: lessons on access to primary care from Ontario's patient enrolment models, *Healthcare Quarterly* 2012

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

| | Primary Care Research | Evidence-Informed Policy | Evidence-based Practice | Infrastructure and Development | Management Standards |
|--|-----------------------|--------------------------|-------------------------|--------------------------------|----------------------|
| Michael Schull, MSc, MD, FRCPC President and CEO and Senior Scientists, ICES; Emergency Department Clinician, Sunnybrook Health Sciences Centre | | | | | |
| Richard H. Glazier, MD, MPH Senior Scientist and Program Lead, ICES; Family Physician and Research Director, St. Michael's Hospital | | | | | |
| Alison Paprica, PhD, PMP Director Strategic Partnerships, ICES; former Director of Planning, Research & Analysis Branch, Ontario Ministry of Health (2010-13) | | | | | |
| Alan Katz, MBChB, MSc, CCFP Director, MCHP, Co-Lead, Manitoba, SPOR PIHCI Network | | | | | |
| Mark Smith, MSc Associate Director, Repository, MCHP | | | | | |
| Dan Chateau, PhD Research Scientist, MCHP, University of Manitoba | | | | | |
| Nancy Meagher, MA Executive Director, PopData | | | | | |
| Kimberlyn McGrail, MPH, PhD Scientific Advisor and Chair of the Advisory Board, PopData, Faculty, Centre for Health Services and Policy Research, UBC | | | | | |
| Brent Diverty, MA Vice President, Programs, CIHI | | | | | |
| Jean Harvey, MHSc Director, Canadian Population Health Initiative, CIHI | | | | | |

Beginning on May 11, 2015, the Leads Team began inviting health researchers and policy/decision makers to join the PRHDN Consortium. As of May 21, 2015, the PRHDN Consortium includes 89 researchers and/or policy/decision makers including members from BC, Alberta, Saskatchewan, Manitoba, Ontario, Quebec, New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador, Australia and the UK.

During development of this NOI, the Leads Team was in contact and shared drafts with the leads of other submissions: *Ongoing Development and use of a National Primary Care Electronic Medical Record Network for Research* (Birtwhistle); *Data and information platform on Francophone health* (Arnaud) and *Patient-Centric Research Data Infrastructure* (Earle); and *DATA: Databases, Analytical pipelines, Tools and Approaches for “Big Data”* (Wasserman). Noting those submissions are complementary, not duplicating PRHDN-Primary Care infrastructure, we commit to working closely with those consortium leads to ensure that our distinct and separate CFI-funded cyber-infrastructure projects would enable future integration.

Note: this NOI was not invited to the full application stage of Competition 1, it may be refined and resubmitted to Competition 2

Criterion 4: The consortium has the required expertise to design and construct the research data infrastructure and develop the analytical tools

The Leads Team includes the senior leaders of long-established and internationally-recognized research centres that have dedicated expert staff and faculty who specialize in designing and constructing research data infrastructure. In total, CFI Cyber-infrastructure funding would leverage and benefit from the expertise of more than 250 scientists and analysts at ICES, MCHP, PopData and CIHI. PRHDN-Primary Care represents a rare opportunity to mobilize that much expertise with project-level funding. Moreover, the four Leads Team centres collectively have decades of experience with mandates focused on achieving excellence through data infrastructure, and those centres have many prior and existing collaborations with each other. PRHDN-Primary Care would thus build on the longevity, mandate, commitment, collaboration and excellence of four highly successful organizations.

Criterion 5: The scope of the project has been clearly defined and allows for commissioning within two to three years

The Gantt chart (Figure 1) illustrates the schedule for completion of deliverables described in this NOI. By beginning with algorithms that do not require validation, PRHDN-Primary Care will have common data in the hands of researchers and policy/decision makers within the first nine months of CFI Cyber-infrastructure funding, and all infrastructure constructed within 2.5 years.

Figure 1: Gantt Chart Schedule for PRHDN-Primary Care

| | Year 1 | | | | Year 2 | | | | Year 3 | | Post-Grant | |
|---|-------------|-------------|-------------|-------------|-------------|----|----|----|-------------|-------------|------------|--|
| | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | | |
| PRHDN personnel funded | [Green bar] | | | | | | | | | | | |
| 1st phase algorithms (no validation) | [Green bar] | | | | | | | | | | | |
| 1st phase common data available for use | | | [Green bar] | | | | | | | | | |
| 2nd phase algorithms (including validation) | | [Green bar] | | | | | | | | | | |
| 2nd phase harmonized common data available for use | | | | | [Green bar] | | | | | | | |
| 3rd phase algorithms (including validation) | | | | [Green bar] | | | | | | | | |
| 3rd phase harmonized common data available for use | | | | | | | | | | [Green bar] | | |
| Core common analytic protocol template developed | | [Green bar] | | | | | | | | | | |
| Repository with core template and common analytic protocols | | | | | | | | | [Green bar] | | | |
| Training and knowledge mobilization sessions | | [Green bar] | | | [Green bar] | | | | [Green bar] | | | |

Because studies based on administrative data do not require patient recruitment, and standard processes are in place for submissions for Research Ethics Board approval, it is expected that studies based on the first phase of PRHDN-identified common data will be completed within 2.5 years, and the researcher members of the Leads Team will make it a priority to achieve that target. Accordingly, it is a realistic goal to have outcomes in terms of evidence to inform practice and policy changes within three years of the start of PRHDN-Primary Care funding.

With respect to sustainability, once established, PRHDN-Primary Care algorithms, harmonized common data, common analytic protocols and recordings of web-based training will continue to provide value to the research and policy/decision making communities at no cost after the CFI grant ends.