CIHI's Population Grouping Methodology 1.0

Canadian Institute
for Health Information

Institut canadien
d'information sur la santé

Table of contents

List of tables and figures

Acknowledgements

Acronyms

ADL: Activities of Daily Living

CAN-Marg: Canadian Marginalization Index

CCRS: Continuing Care Reporting System

CHESS: Changes in Health, End-Stage Disease and Signs and Symptoms

CPCD: Canadian Patient Cost Database

DAD: Discharge Abstract Database

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition

FIM^{TM}: Functional Independence Measure

HCN: Health Card Number

HCRS: Home Care Reporting System

ICD-9: International Statistical Classification of Diseases and Related Health Problems, Ninth Revision

ICD-10-CA: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada

INSPQ: Institut National de Santé Publique du Québec

RAI-MH©: Resident Assessment Instrument–Mental Health[i]

RAI-HC©: Resident Assessment Instrument–Home Care

RAI-MDS 2.0©: Resident Assessment Instrument–Minimum Dataset 2.0

NACRS: National Ambulatory Care Reporting System

NRS: National Rehabilitation Reporting System

OMHRS: Ontario Mental Health Reporting System

PCCF: Postal Code Conversion File

PLPB: Patient-Level Physician Billing

QAIPPE: Quintile of Annual Income Per Person Equivalent

RPD: Registered Persons Data

SES: Socioeconomic Status

---

[i] The RAI-MH© is a copyright of the Government of Ontario, Ontario Hospital Association and interRAI Corporation. The FIMTM instrument is the property of Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc. The RAI-MDS 2.0© and RAI-HC© are a copyright of interRAI Corporation. interRAI CA© is abbreviation for "interRAI Contact Assessment (CA)© Canadian Version Screening Level Assessment for Emergency Department and Intake from Community/Hospital"

Executive Summary

On December 15, 2016, the Canadian Institute for Health Information (CIHI) released version 1.0 of its population grouping methodology. This report highlights the methodology's clinical classification, predictive indicators, and the data used in the development and the application of the methodology.

Background

A population grouping methodology contains a case mix classification which profiles each person in the population, using person-level clinical information. This case-mix classification is then used to assign predictive indicators to each person. All persons in the population over a given time period, including healthy persons and persons who have not used the health system, are represented in the population grouping methodology.

For many years, clients encouraged CIHI to consider offering a population grouping methodology. Based on that encouragement, CIHI began the process of developing the population grouping methodology in January 2013. The development of the methodology and the acquisition of data needed for research and development represented a significant undertaking. Progress on the project was released in phases to an expert group as components become available. This phased approach provided stakeholders with the opportunity to provide feedback to CIHI on the methodology design prior to December 15, 2016 release.

The current release, version 1.0, represents our first release of the population grouping methodology to a wider audience, namely jurisdictional and regional ministries of health, health quality councils and research organizations, as well as other organizations upon request. Going forward, CIHI will continue to work closely with stakeholders to increase the breadth and depth of data used as the basis for the methodology, and to evolve, refine, and expand the case mix classification, predictive indicators, and software in future versions of the population grouping methodology.

Foundation data used as basis for research and development

Canadian clinical administrative and financial data was used as the foundation for the development of the population grouping methodology. This brings together, at the person level, information available in the Registered Persons Data (RPD), Discharge Abstract Database (DAD), Ontario Mental Health Reporting System (OMHRS)[ii], National Ambulatory Care Reporting System (NACRS), Continuing Care Reporting System (CCRS) and Patient-Level Physician Billing Data (PLPB).

The target population for the research was residents with a valid health card number (HCN) who were eligible to receive services in Ontario, Alberta, or British Columbia anytime during fiscal years 2010–2011 and 2011–2012. This 2-year interval is termed "the concurrent period" in the methodology. To build the target population, HCNs from the RPD were extracted when they were valid for use during periods that overlapped—in whole or in part—with the concurrent period. Some minor additions and removals were made to address data quality issues.

Clinical information for the target population was then extracted from the DAD, NACRS (inpatient and day surgeries only), OMHRS, CCRS and PLPB. The province where health care is provided needed to be the same as the province of residence for persons in order for the clinical data to be included in the foundation data. Additionally, the clinical data had to correspond to events that took place during the concurrent period. For records that met the above criteria, clinical diagnoses and their metadata—such as the database that reported the diagnoses—were extracted.

The population grouping methodology also uses patient-level costs that represent the health expenditures for a person over a defined time period and across defined health sectors. Patient-level costs were derived using actual patient-level cost data when available and estimation methods when actual costs were not available.

[ii] In British Columbia and Alberta inpatient hospitalizations for mental health conditions in are captured within the DAD. In Ontario, a small portion of inpatient hospitalizations for mental health conditions are captured in DAD and the remainder are captured in OMHRS.

Health condition classification

The health condition classification uses diagnosis codes reported in historical clinical administrative data. The health condition classification maps diagnosis codes to health conditions and then applies a methodology to assign health conditions to person-level clinical profiles.

The health conditions are the building blocks of the clinical classification. They set the boundaries for the grouping of clinically similar diseases. There are 226 health conditions:

- For 225 of these health conditions, diagnosis mappings are used to translate ICD-9 and ICD-10-CA codes (from PLPB, DAD, NACRS and OMHRS), diagnosis elements from the RAI-MDS 2.0 instrument (from CCRS) and diagnosis elements from the RAI-MH instrument (from OMHRS) to these 225 conditions. These mappings are many to one; an ICD code or RAI data element can map to only one of the 225 conditions. Altogether, the methodology includes some 27,992 diagnosis mappings. For diagnosis information captured through PLPB, tagging rules are also applied, that specify the criteria needed to confirm the presence of a given health condition for a person.

- For one health condition – 'Healthy Newborn', age is used to identify the presence of a health condition. All persons under 2 years of age at the end of the concurrent period are assigned to the 'Healthy Newborn' health condition. Use of age to identify Health Newbornensures that associated health care interventions and costs for newborns are captured regardless of place of delivery and/or any data quality issues that might prevent the population grouping methodology from capturing all newborns (e.g. sharing of health card numbers with mother).

The methodology also incorporates clinical overrides to remove redundancies that may exist on a person's clinical profile after the assignment of health conditions. Clinical overrides represent a transition point in the classification where the perspective shifts from looking at health conditions individually to looking at the clinical profile of the person holistically.

The determination of the 226 health conditions, their diagnosis mappings, tagging rules, and the clinical overrides are the components of the health condition classification included in the population grouping methodology.

Functional status classification

Functional status is a measure of one's ability to perform normal daily activities required to meet basic needs, fulfill usual roles, and maintain health and well-being. It is useful information to incorporate in the classification, as it would allow the methodology to differentiate health care need beyond what is achievable by looking at health conditions alone. Capture of functional

status measures for the entire population is currently not available for use in the population grouping methodology, however it is included for CCRS data where there is almost complete reporting of functional status information.

Socioeconomic status classification

Socioeconomic status is a combined measure of a person's economic and social position in relation to others. Income, education and living arrangements are examples of information used to determine socioeconomic status. The population grouping methodology incorporates three area-based measures of socioeconomic status originating from survey data that are attributed to individuals based on the information affiliated to the region they reside in. The socioeconomic status measures included in the grouping methodology include the Deprivation index (INSPQ), the Canadian Marginalization Index (CAN-Marg) and the Quintile of Annual Income per Person Equivalent (QAIPPE).

Health profile group (HPG) classification

The health profile group classification leverages the 226 health conditions to synthesize an individual's clinical profile down to the most complex and clinically relevant health condition present.

The health profile groups consist of a mapping of the 226 health conditions to 164 branches. Branches group together health conditions that are clinically similar and require similar cost resources and differentiate between severe and moderate cases of the same conditions. There is also a branch for health system users with no health conditions and a branch for non-users.

The 164 branches are ranked from the most clinically complex and resource intensive to the least resource intensive. For a person with more than one health condition, this hierarchy is used to identify the highest ranking branch and this person is then assigned to that one branch. This is the branch used to form the basis of their health profile group.

The majority of branches are then further segmented into two – those with and those without major/moderate comorbidities, reslting in a total of 239 health profile groups.

Each of the 164 branches is also mapped to one of 16 categories. The categories provide broad information about the type (e.g. acute, chronic, cancer, mental health, newborn, obstetrics, palliative) and severity (e.g. minor, moderate, major) of the health profile group.

All components of the health profile grouping methodology are inter-connected. The 164 branches can be rolled up into one of 16 categories, or further segmented into one of the 239 health profile groups. This allow users to analyze the data at the level most meaningful to them.

Predictive indicators

Population grouping methodologies commonly include predictive indicators that provide insight into the use of the health care system and the health of the population. Perhaps the most common, or core, indicators included cost weights - a relative measure of health care costs. CIHI's population grouping methodology assigns cost weights for each of the concurrent[iii] and the prospective periods[iv]. There are two sets of cost weights, one set based upon the 226 health conditions identified for the person, and one set based upon the HPG assignment of the person. These cost weights encompasses the hospital inpatient, day surgery and emergency room costs that are covered by hospital budgets, as well as physician costs.

This methodology also includes predictive indicators, estimating the:

- number of visits to an emergency department during the concurrent period;

- number of visits to an emergency department during the prospective period

- number of visits to a primary health care provider in the prospective period; and

- probability of entering long-term care for those age 65 or older during the prospective period

Deliverables for the release

The release of the population grouping methodology version 1.0 on December 15, 2016 consists of a methodology report, computer software that applies the methodology, a user guide for running the computer software. Jurisdictional and regional ministries of health, health quality councils and research organizations, as well as other organizations, may obtain the population grouping methodology upon request and subject to acceptance of a license agreement outlining how the software can be used.

[iii] Individuals who did not interact with the system during the concurrent period are assigned a cost weight of zero.
iv Persons who are not eligible for health care on the last day of the concurrent period; they are assigned a prospective cost weight of zero

1.Introduction

For many years, clients encouraged CIHI to consider offering a population grouping methodology. As a result of this encouragement, CIHI began the process of developing the population grouping methodology in January 2013. This development process was an iterative and consultative process which involved an expert working group of client stakeholders with expertise in the application of population grouping methodologies, clinical panels, and an alpha and beta version of the product that were tested internally and by members of the expert group.

After three years in development, CIHI launched version 1.0 of the population grouping methodology on December 15, 2016. This population grouping methodology contains a case mix classification which profiles each person in the population, using person-level clinical information. This case-mix classification is then used to assign predictive indicators to each person. All persons in the population over a given time period, including healthy persons and persons who have not used the health system, are represented in the population grouping methodology.

Version 1.0 consists of a methodology report, computer software that applies the methodology, a user guide for running the computer software and a license agreement outlining how the software can be used. This product is available to jurisdictional and regional ministries of health, health quality councils and research organizations, as well as other organizations upon request.

This grouping methodology provides Canadian policy makers, health system managers and funders with a Canadian methodology that reflects Canada's population and health care systems, meets the needs of Canadian clinical administrators and health care researchers, and takes full advantage of the wealth of health care data available at CIHI. The richness of CIHI's clinical data holdings provides opportunities for population grouping that are unavailable to many jurisdictions internationally.

Population grouping methodology software

Users of the population grouping methodology are able to enter person-level clinical data for:

- inpatient stays (including inpatient stays for mental health illnesses),

- day surgery,

- hospital clinic visits,

- emergency department visits,

- physician care (based on physician billing data from all physician types, so broader than primary care visits), and

- long-term care and complex-continuing care stays.

The population grouping methodology will then output health condition assignments based on diagnosis information from clinical data as well as several predictive indicators including:

- cost weights, based on costs for inpatient stays (excluding costs for mental health inpatient stays in Ontario), day surgery, hospital clinic visits (excluding British Columbia), emergency department visits (excluding British Columbia) and physician care;

- health profile group and associated cost weight;

- predicted number of visits to a family physician in the prospective year;

- predicted number of visits to an emergency department in the concurrent 2-year period and prospective year; and the

- probability of residents ages 65 or older entering a long-term care facility in the prospective year.

The goal of the population grouping methodology is to assist CIHI and its clients to monitor population health, predict health care utilization patterns, explain variations in health care resource use, and provide a foundation for funding models. The methodology will also provide a comprehensive basis for standardization when conducting comparisons across jurisdictions, such as among provinces and territories. Given the many potential applications for the methodology, the logic for assigning clinical profiles and resource indicators to a person is transparent, allowing clients to customize the methodology to meet their specific needs.

2.Foundation data

For research and development of the population grouping methodology (referred to as the foundation data) person-level linkable data was needed that is representative and comprehensive of an entire population. Data was required from multiple provinces in order to enable the development of a national methodology.   Data from key health sectors was critical to allow for the identification of both high-cost patients and health conditions that are costly at the system level.

The foundation data is composed of 3 consecutive years of Ontario, Alberta and British Columbia data for fiscal years 2010–2011 and 2011–2012 and 2012–2013; and brings together person-level information from the following sources:

- Registered Persons Data (RPD);
- "Clincal Adminstrative Data"
  - Inpatient stays from the Discharge Abstract Database (DAD institution type code not = A).
  - Mental health inpatient stays from the Ontario Mental Health Reporting System (OMHRS).
  - Day surgeries from DAD (DAD institution type code = A) and the National Ambulatory Care Reporting System (NACRS ambulatory care type code = 20, 21, 22 or 31)
  - CCRS assessments
  - Patient-Level Physician Billing Data;

The following sections of this chapter provide more information about the way the foundation data was constructed.

2. Foundation data

2.1 Data Years

The three years of foundation data is divided into two time periods – the concurrent and prospective periods. The concurrent period refers to the interval of time that historical clinical data is compiled to build clinical profiles. The methodology uses a 2-year concurrent period. Predictive indicators for the concurrent period are also provided by the methodology. The prospective period is the time period in the future for which health system activity and utilization are being predicted. The methodology defines the prospective period as the 1-year window of time following the concurrent period. In order to be included in the prospective period, individuals had to have a valid health card on the last day of the concurrent period – March 31, 2012 (reffered to as the reference date).

One of the first considerations when defining the scope of the foundation data was to assess the years of data available. CIHI took the most recent data year where comprehensive data was available and treated this as the prospective period. The prospective period was April 1, 2012 – March 31, 2013 (2012-2013 fiscal year) due to this being the most recent year of patient-level physician billing data available at CIHI. That also resulted in the concurrent period comprising the 2010–2011 and 2011–2012 fiscal years.

Figure 1   Fiscal years used in the development of the foundation data

Source
Canadian Institute for Health Information, population grouping methodology, 2016

2.2 Types of care included in foundation data

The development of the methodology and acquisition of the foundation data needed for research and development is a significant undertaking that is still underway at CIHI. Theoretically, using data that represents all types of care would lead to more complete clinical profiles and to a better assessment of homogeneity of resource needs within groups, which should result in more useful predictive indicators. However, these benefits were often offset by challenges with the availability of data.

The types of care included in the foundation data were chosen based on the data available to CIHI, their quality, coverage, and the marginal benefit of its inclusion for building clinical profiles and assigning predictive indicators to persons. High-cost sectors would ideally be represented in predictive indicators and hence the foundation data.

Version 1.0 of the population grouping methodology is created using foundation data from British Columbia, Alberta and Ontario.

While CIHI has a RPD for each of the three jurisdictions, we do not have complete clinical and cost data for all health sectors to-date. Table 1.0 provides an overview of the clinical and cost data included in the foundation data used to create the version 1.0 methodology.

2. Foundation data

Table 1 Foundation data inputs included and excluded in the development of the population grouping methodology for version 1.0

Source of clinical data	CIHI clinical database	Clinical Classification	Clinical data			Cost data		
			BC	AB	ON	BC	AB	ON
Physician care	Patient-level Physician Billing Data (PLPB)	ICD-9	✔	✔	✔	✔	✔	✔
Inpatient stays	Discharge Abstract Database (DAD)[†]	ICD-10-CA	✔	✔	✔	✔	✔	✔
Hospital in-patient mental health *	Ontario Mental Health Reporting System (OMHRS)	RAI-MH, DSM-IV	✔	✔	✔	✔	✔	✘
Day surgeries	Discharge Abstract Database (DAD) / National Ambulatory Care Reporting System (NACRS)[†]	ICD-10-CA	✔	✔	✔	✔	✔	✔
Hospital clinic visits	National Ambulatory Care Reporting System (NACRS)[†]	ICD-10-CA	✘ Insufficient data			✘	✔	✔
Emergency department visits	National Ambulatory Care Reporting System (NACRS)[†]	ICD-10-CA	✘	✘	✘	✘	✔	✔
Long-term care facilities	Continuing Care Reporting System (CCRS)	RAI-MDS 2.0	✔	✔	✔	✘	✘	✘
Hospital rehabilitation wards	National Rehabilitation Reporting System (NRS)	FIM, ICD-10-CA	✘ Insufficient data					
Home care services	Home Care Reporting System (HCRS)	RAI-HC	✘ Insufficient data					
Public drug programs	National Prescription Drug Utilization Information System (NPDUIS)	DIN, PDIN, ATC[‡]	✘ Insufficient data					

Notes
✔ indicates data is included
✘ indicates data is excluded
* Inpatient stays for mental health illnesses in BC and AB are included in DAD.
[†] The patient-level cost data for this sector is stored in a separate CIHI database: the Canadian Patient Costing Database (CPCD).
[‡] DIN: drug identification number; PDIN: pseudo-DIN, ATC: anatomical therapeutic chemical classification.
Source
Canadian Institute for Health Information, population grouping methodology, 2016

Over time, in future releases of the population grouping methodology, CIHI will work to add in missing cost data to accompany existing clinical data: (ie: Ontario mental health inpatient stays, British Columbia emergency department and hospital clinic visits and LTC and CCC stays for all three jurisdictions).

Additionally, over time and in future versions of the population grouping methodology, CIHI will work towards including:

- other sources of clinical data, such as home care, rehabilitation and prescription drug data; as well as

- data from other jurisidicitons.

2.2.1   Patient-level physician billing[v]

Patient-level physician billing data (PLPB) represents patient-level information about the services provided by a physician to individuals that are paid for by the province. It covers activity that takes place in physician offices, hospitals, and other health care settings. Physician billings to the provinces represent 22% of public-sector health expenditures.[1] In a given year, the majority of the population (88%) has at least 1 physician visit that is represented in the PLPB.

PLPB data is similar across provinces but it is not standardized; for example, billing claims and payment systems are customized for each province.

Diagnosis information describing the health problems being treated is submitted with billing claims. Many claims use ICD-9 codes to capture diagnoses but in some jurisdictions it is common to observe province-specific diagnosis codes to describe patient's health problems. Diagnosis coding in physician claims data does not use the same data standard as other data sources, such as hospital data.

PLPB data was considered critical for this project. Many people's only point of contact with the health system is with physicians through clinics or office visits; hence, the ability to identify the health conditions that exist for these people is only possible through PLPB data. Despite CIHI's limited access to PLPB data, all provinces collect it; therefore, its inclusion in the methodology would benefit CIHI's key stakeholders (i.e. the provincial and territorial ministries of health).

---

[v] Patient-level physician billing data (PLPB) is not exclusive to physician claims. PLPB includes claims from midwives, dentists, and other health care professionals. The non-physician health professionals represented in the PLPB differs by province. For this development project, claims from all health care professionals were used when integrating PLPB data. However, the vast majority of claims represent physician services. For simplicity, this report refers to PLPB data as representing physician claims only.

2.2.2   Hospital inpatient stays

Each year, around 3 million inpatient hospitalizations occur in Canada.[2] Inpatient abstracts contain detailed diagnosis information about patients. Diagnoses are represented with ICD-10-CA codes. There are more than 10 years of complete activity data available across provinces (note: Quebec reports to a different database). An exception to this is discussed in the section 2.2.3. Patient-level cost data is available in the Canadian Patient Cost Database (CPCD) from a sample of facilities in Alberta, British Columbia, and Ontario; coverage varies over time and by province. Patient-level costs are covered in section 2.5.

Inpatient data was considered another prerequisite for the methodology. High-cost patients can be identified in this database making it key for developing risk indicators for the methodology. Hospital inpatient clinical data is available at CIHI for every discharge from a hospital in Canada and its quality is high.

2.2.3   Hospital mental health

Patients in designated adult mental health hospital beds in Ontario have data about their stay reported to the Ontario Mental Health Reporting System (OMHRS). OMHRS data includes information about a patient's mental and physical health, social support, and service use. For the most part, similar patients from other provinces have data about their hospital stay reported to DAD. In OMHRS, mental health disorders and other conditions are captured using the RAI-MH instrument.

The foundation data (and the clinical classification) incorporates mental health diagnosis information from OMHRS in the population grouping methodology. The inclusion of OMHRS data in the foundation data will allow for comparable information about mental health to be available across provinces. Its inclusion in the methodology will be critical for Ontario's ability to use the methodology to create clinical profiles and risk indicators that account for mental health issues in its population. Functional status information from the OMHRS data will be incorporated into the classification in a future release (version 2.0 or beyond).

2.2.4   Hospital day surgeries

About 80% of hospital surgeries are day surgeries,[3] making them an important hospital activity to include in the methodology.

Complete day surgery data is available in all provinces for multiple (10+) years. Some provinces report data to the Discharge Abstract Database (DAD) and others to the National Ambulatory Care Reporting System (NACRS); both databases capture diagnoses using ICD-10-CA. CPCD data for day surgeries is available from a sample of facilities in Alberta, British Columbia, and Ontario. Data coverage of CPCD data varies over time and by province.

Hospital day surgery data was included in the foundation data since its clinical data is available for all provinces and is high quality.

2.2.5   Hospital emergency department visits

Hospital emergency department (ED) visits represent about 5% of hospital expenditure.[3] Around 16 million visits to the ED occur yearly in Canada.[4]

The NACRS database houses data for all ED activity taking place in Ontario and Alberta. It has partial coverage in 6 other jurisdictions. At a national level, this database collects data for about half of Canada's ED activity. Also, for provinces with full or adequate coverage, similar diagnoses are frequently captured in the physician data. That is, most attending ED physicians bill the province for treating the patients who present to the ED and the diagnoses from those billings are recorded in PLPB and already contribute to the classification. As NACRS coverage of ED grows, we will revisit the option to incorporate clinical information from ED visits in the foundation data used to build the population grouping methodology.

The main value of including ED data in the foundation data is for developing predictive indicators, such as expected ED use and associated costs. Additionally, CPCD data for ED visits is available from a sample of facilities in Alberta, British Columbia, and Ontario. Data coverage varies over time and by province. To this end, ED costs for Alberta and Ontario is included in the foundation data, allowing the scope of costs represented in the predictive indicators to include ED costs for 2 out of the 3 jurisdictions.

For fiscal years 2010-2011 to 2012-2013, there is no NACRS data for British Columbia, making it difficult to link CPCD data from British Columbia with services provided and conditions treated. CIHI is working on a methodology to identify ED visits in British Columbia using the PLPB which would allow us to link clinical data with CPCD cost data. Once this methodology is complete, cost data for ED visits will be included in future versions.

2.2.6   Continuing care services in hospitals and long-term care facilities

Some people require continuing care services. These services are offered in hospitals with complex continuing care beds and in residential care facilities (i.e. nursing homes). In Ontario, 142 thousand residents (1% of its population) require continuing care services annually;[5] although small in number, they consume about 7% of the public funds for health services.[1,6] Ontario's ministry of health pays about $153 per day for each resident in a long-term care facility;[7] this is equivalent to about $55,800 per resident-year. Complex continuing care provided in the hospital setting is more expensive.[8]

Data representing residents who are receiving continuing care services is collected by nursing staff and other health providers. Assessments document the clinical and functional characteristics of residents; these are done every 3 months for the duration of the patient's episode of care. Data is captured using the RAI-MDS 2.0 instrument and is submitted to CIHI.

Uptake of data collection using the Continuing Care Reporting System (CCRS) in Canada has improved in recent years. To date, British Columbia, Alberta, Saskatchewan, Ontario, Newfoundland and Labrador and Yukon have achieved full coverage; and Manitoba, Nova Scotia and New Brunswick have achieved partial coverage. The lack of historical pan-Canadian data (at present) for this sector was a consideration for whether to include it in the foundation data and, if it is included, the way it would be integrated such that the final grouping methodology would remain relevant for all provinces.

Despite these challenges, it remained important for this methodology to describe this high-need population. Some 7% of Canadians aged 65 and older reside in long-term care facilities.[8] For this group, the methodology would ideally address their different care needs. The inclusion of this data makes it possible to develop predictive indicators such as the likelihood of a person entering long-term care in the prospective period. Inclusion of RAI data provides the ability to further differentiate the care needs of the high-cost long-term care population by adding measures of functional status from the RAI-MDS 2.0 instrument to a person's clinical profile. For these reasons, CCRS is included in the foundation data.

2.2.7   CIHI databases excluded from foundation data

Four databases are not included in the foundation data. This section describes the rationale in postponing their inclusion in the grouping methodology.

Data on home care services was excluded due to data coverage. There is full coverage of people receiving home care in Newfoundland and Labrador, Ontario, Alberta and Yukon and sufficient coverage in British Columbia. With data starting to come in from other provinces, for

example Saskatchewan, home care data will be a candidate for inclusion in a future release (version 2.0 or beyond).

Data on prescription drugs was excluded for multiple reasons. In most provinces, persons represented in CIHI's prescription drug data are those covered by publicly financed drug benefit programs. Public drug plan design varies by province, with the most consistent plan coverage for seniors (65+ years of age) and for people covered on social assistance programs. In British Columbia, Manitoba and Saskatchewan, the majority of the population is covered by the public drug program, which makes their data more comprehensive. Additionally, British Columbia data also includes claims not financed by the public drug program. It is also noted that prescription drug data for Saskatchewan and Ontario, while already received at CIHI, is not linkable to other datasets at the patient level. These all present challenges when incorporating drug data and ultimately resulted in its exclusion from development of version 1.0 of the population grouping methodology. Drug data will be a candidate for inclusion in a future release (version 2.0 or beyond).

Hospital clinic activity was excluded since its data is available from only 2 provinces: Alberta and Ontario. The coverage of data available varies between these 2 provinces. Alberta collects all publicly funded hospital clinic activity; whereas, Ontario mandates the collection of clinic data for 3 high volume and high cost functional centres (cardiac catheterization, oncology, and renal dialysis). There are no plans to incorporate hospital clinic activity into the methodology at this time.

And lastly, hospital rehabilitation data that is reported to the National Rehabilitation Reporting System (NRS) was not included. Of the provinces where CIHI has rehabilitation data, full coverage is available in Newfoundland and Labrador, Prince Edward Island, and Ontario, and moderate coverage in the western provinces. NRS data will be a candidate for inclusion in future releases.

2. Foundation data

2.3 Registered persons data[vi]

One of the goals of the population grouping methodology was that it would be applicable to all persons who were eligible to receive publicly funded health care in a province regardless of whether they used the health system. Consequently, the foundation data needed to compile data for those persons in each province who were eligible to receive health services and track the time period over which they were eligible.

CIHI accessed this target population by leveraging databases maintained by the provinces. CIHI refers to these databases as Registered Persons Data (RPD), named so since they list persons who are "registered" for a health card number over time.

Most provinces update their RPD regularly using vital statistics and data from other information systems (to address transient populations). CIHI acquired these databases from Ontario, Alberta and British Columbia for this project. At this time, Saskatchewan's RPDB data is not available to CIHI which prohibited the inclusion of this province in the foundation data despite having access to the other databases required. CIHI will continue to work with jurisdictions to establish more data sharing agreements for RPDB and physician billing data, which will allow for more jurisdictions to be represented in the foundation data.

2.4 Defining the target population

The target population for methodology development is residents with a valid health card number (HCN) who were eligible to receive publicly funded services in Ontario, Alberta or British Columbia anytime during 2010−2011 and 2011−2012 (i.e. the concurrent period).[vii]

To build this population, HCNs that were valid for use during periods that overlapped—in whole or in part—with the concurrent period were extracted from the Registered Persons Data (RPD). HCNs for persons who died before or were born after the concurrent period were excluded from the target population.

Some additions and removals to the target population were made to address observed data quality issues with the RPDB:

- For the Alberta RPDB, newborns that were observed in the clinical data but were not on the RPDB as having a HCN eligible during the concurrent period were added.

[vi] This section refers to the Registered Persons Data which represents the jurisdictional registers of persons with valid health cards. British Columbia calls this the "B.C. Health Ideas Client Roster" and Alberta calls this the "Alberta Health Registration File."

[vii] Members of the Canadian Armed Forces, inmates of federal penitentiaries, and members of the Royal Canadian Mounted Police are covered by the federal government. These groups are not entitled to coverage with the provincial health insurance programs and are not in the RPDB.[9]

- HCNs with implausible high activity or with significant inconsistencies in demographic information between the RPDB and clinical data were excluded.

- HCNs where there was no date of birth were removed from the target population.

- HCNs where the sex was neither male nor female were removed from the target population.

The RPDB was used to build people's demographic profiles (date of birth, date of death and sex) and to assign area-based socioeconomic status (derived from postal code).

After these steps, 23,000,543 persons remained in the target population. Prospective indicators were based on 22,359,018 of these persons who had HCNs that were still eligible on March 31, 2012[viii]. Table 2 shows statistics about the target population represented in the RPD compared with population estimates from Statistics Canada.

[viii] Prospective indicators on admisisons to LTC were based on a sub-population of individuals in the prospective period – specifically, those ages 65 or older.

2. Foundation data

Table 2 Clinical information available for target population over the concurrent period (2010-2011 and 2011-2012), by province

	British Columbia	Alberta	Ontario	All 3 provinces
Foundation data based on RPDB				
A. Target population for concurrent indicators Eligible for health care services for at least one day between April 1, 2010 and March 31, 2012	**4,802,690**	**3,966,936**	**14,230,917**	**23,000,543**
B. Target population for prospective model Eligible for health care services on March 31, 2012	**4,660,893**	**3,805,574**	**13,892,551**	**22,359,018**
C. Population removed from prospective model (A − B)	**141,797**	**161,362**	**338,366**	**641,525**
Statistics Canada population estimates				
A. Target population for concurrent models (B + C)	**4,730,032**	**4,046,989**	**13,779,559**	**22,556,580**
B. Target population for prospective models Statistics Canada census population estimates for July 1, 2012[10]	**4,546,290**	**3,880,755**	**13,413,702**	**21,840,747**
C. Population to add for concurrent models	**183,742**	**166,234**	**365,857**	**715,833**
Deaths during 2010-2011 and 2011-2012[11]	63,953	42,837	179,323	286,113
Persons who migrated from province (interprovincial) during 2010-2011 and 2011-2012[12]	95,737	108,717	133,394	337,848
Persons who migrated out of Canada during during 2010-2011 and 2011-2012[11]	24,052	14,680	53,140	91,872

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
Alberta Health Registration Files, Alberta Health.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

2.5 Patient-level costs

The development of the classification and the predictive indicators relies on complete cost information at the person level. Costs need to be representative of health expenditures directed towards a person over a defined time period and across defined health sectors. To develop these cost profiles, patient-level cost data was used when available and data gaps were filled using estimation methods.

2.5.1   Physician costs

For physician billing data, patient-level costs are supplied by fee-for-service physicians and these were used without adjustment when building cost profiles for persons.

Patient-level costs are not supplied by physicians on alternative payment plans (e.g. physicians who are paid on capitation or salary). For these physicians, the missing patient-level costs were estimated by leveraging the patient-level shadow billing data and physician specialty. Shadow billing is an administrative process whereby physicians submit service provision information using provincial/territorial fee codes; however, payment is not directly linked to the services reported. Ontario supplied CIHI with an estimated cost for shadow billing so these imputed costs were adopted without modification in the foundation data. For Alberta and B.C., CIHI cross referenced the cost average for the same billing code submitted by a fee-for-service physician of the same specialty in the same province to determine the estimate of patient-level costs for a shadow billing claim. If this level of detail was not available for cost estimation CIHI used average fee-for-service payment for the fee code for the broad physician speciality group (family medicine or other specialist). This estimation approach assumes that the costs for patient visits are similar between physicians on fee-for-service payments and physicians on alternative payment plans.

Claims that are not associated with a specific HCN are excluded (i.e. not allocated to persons with HCNs). Also, shadow billing has limitations. Not all services funded via alternative payment programs are reported through shadow billing claims and the quantity is unknown. Hence, this may lead to an underestimate of physician costs in some cases. When unreported, there is no information available to impute costs so these unreported costs are also excluded from the foundation data.

2.5.2 Hospital costs

Patient-level cost data is detailed data on the resources utilized and the value of those resources at the patient level. Table 3 describes the hospital activity data from the CPCD that was used for building cost profiles in the population grouping methodology.

Table 3 Provinces where some hospitals collect and report patient-level costing data to the Canadian Patient Costing Database

Province	2010-2011	2011-2012	2012-2013
British Columbia	✖	✔	✔
Alberta	✔	✖	✔
Ontario	✔	✔	✔

Notes
✔ indicates data is available
✖ indicates data is not available
Source
Canadian Institute for Health Information, population grouping methodology, 2016

Case-mix methods were used to estimate patient-level costs for hospital activity that did not have CPCD cost data available. Resource intensity weights (RIWs) from CIHI's case mix inpatient (Case Mix Group Plus, CMG+) and outpatient (Comprehensive Ambulatory Care System, CACS) grouping methodologies were used to estimate cost.

RIWs provide an estimate of a person's health care costs, relative to a reference, or anchor, value. The next step for cost estimation was to find the best anchor value to use to convert the relative cost weights to a dollar scale.

The RIWs from CMG+ and CACS were used without modification with one exception. For typical inpatient cases in CMG+, the following formula was used to further refine the cost weights for hospital activity with the sole purpose to refine cost estimates:

$$Refined\ RIW = \frac{Base\_RIW}{Base\_ELOS} \times \frac{IP\_Perdiem}{Base\_Perdiem} \times Total\_LOS\_Days\ \text{(Equation 1)[ix]}$$

[ix] Base_RIW = Resource Intensity Weight of the CMG/age category when there is no other factor present.
Base _ELOS = Expected Length of Stay of the CMG/age category when there is no other factor present.
IP_Perdiem = The CMG- and age category–specific estimates for daily resources use, adjusted for factors present.
Base_Perdiem = The CMG- and age category–specific estimates for daily resources use, when there is no other factor present.
Total_LOS_Days = the Total Length of Stay in hospital, including Alternative-Level of Care days.

The dollar scale used to convert the relative cost weights to costs was the inpatient Cost per Standard Hospital Stay (CSHS), which were calculated by DAD-facility for each of 2010-2011, 2011−2012 and 2012−2013 fiscal years. CSHS values were not available for 2010−2011 data so 2011−2012 CSHS values were used for imputation of costs for that year as well. In cases where a facility-level CSHS value was not available for a facility, the provincial-average inpatient CSHS was used.

Ideally, CACS RIWs would be multiplied by a hospital-based ambulatory CSHS value to derive an estimated dollar cost. However CIHI does not currently calculate hospital-based ambulatory CSHS values. Therefore, provincial-average inpatient CSHS values were utilized to estimate patient-level costs in Ontario and Alberta's day surgery and emergency department activity (reported to NACRS).

Note that British Columbia residents do not have hospital emergency department costs represented in their cost profiles in the foundation data. British Columbia did not report to NACRS over the years of study for this project, which inhibited CIHI's ability to impute hospital emergency department costs for this province. Additionally, mental health costs in Ontario's designated adult inpatient mental health beds are not as a methodology to incorporate them is still under development.

Further enhancements to improve cost imputation are planned for future versions to address the missing year of CSHS values, to use facility-level CSHS values for day surgery and emergency department costs, to develop cost imputation methods for British Columbia's hospital emergency department costs, to estimate long-term care costs, to include estimates of inpatient mental health costs in Ontario, and to apply edits to address high-cost outliers.

2.5.3   Long-term care and complex continuing-care costs

Patient-level cost estimation methods for the long-term care sector and Ontario's complex continuing care beds were not developed for version 1.0. There are plans to develop a cost imputation method for the continuing care costs so these costs are represented in development of the classification and the predictive indicators for future versions.

2.6 Clinical foundation data

Clinical information for the target population was extracted from the DAD and NACRS (inpatient and day surgeries only), OMHRS, CCRS and PLPB. Foundation data for research is specific to health services provided within province to residents and had to correspond to events that took place during the concurrent period. For example, the hospital inpatient abstracts were required to have discharge dates that fell within the concurrent period and be reported from a facility that resided in the same province as the person. The same requirements applied for day surgeries (using "disposition date"), physician data (using "service date"), and OMHRS and CCRS data (using "assessment date").

For the abstracts that met the above criteria, clinical diagnoses and their relevant information—such as the time of their observation and the setting that they represent—were extracted. In the case of physician billing data, only the first diagnosis code was extracted since the other fields available for capturing additional diagnoses are rarely used.

The diagnoses extracted from the physician data had ICD-9 validation edits applied.[x] Physician data with invalid ICD-9 codes had diagnosis codes overwritten with blank values. Similarly, data with valid ICD-9 code that, when it was cross-referenced to the age and/or sex of the patient, failed the validation edit had diagnosis codes overwritten with blank values. This processing step was done specifically to the physician billing data since validation edits are not designed into its data collection systems. In contrast, CIHI's systems that collect hospital, mental health and continuing care data already have validation edits applied.

Table 4 summarizes the clinical data available for the target population, by province.

---

[x] For PLPB data, diagnoses information was often a combination of ICD-9 codes and jurisdiction-specific codes. For province-specific codes, CIHI worked with the juridictions to develop a mapping table that converts these codes to ICD-9 codes when running the software.

Table 4 Clinical information available for the creation of the foundation data for the concurrent period (2010-2011 and 2011-2012), by province

	British Columbia	Alberta	Ontario	All 3 provinces
Persons in target population	**4,802,690**	**3,966,936**	**14,230,917**	**23,000,543**
Total diagnoses in inpatient data	2,558,024	2,965,930	7,886,961	13,410,915
Total diagnoses in OMHRS data	0	0	3,014,988	3,014,988
Total diagnoses in day surgery data	1,379,307	1,096,902	5,007,191	7,483,400
Total diagnoses in physician data (valid ICD-9 codes only)[*]	91,084,177	50,603,537	211,204,006	352,891,720
Total Diagnoses in CCRS Data	1,721,801	1,377,475	8,493,858	11,593,134
Median number of diagnoses per person[†]	11	10	9	9
Interquartile range of number of diagnoses per person[†]	11 (4-25)	10 (3-22)	9 (3-19)	9 (3-21)

Notes

[*] Invalid ICD-10-CA codes or ICD-9 codes that conflicted with age/sex edits were excluded from these counts. Counts include diagnosis codes that are captured multiple times over the concurrent period. Additionally, duplicate PLPB claims were removed.

[†] Statistics for mean and interquartile range include health system non-users.

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Ontario Mental Health Reporting System/Continuing Care Reporting System, Canadian Institute for Health Information.

Patient-Level Physician Billing Data, Alberta Health Services.

Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.

Alberta Health Registration Files, Alberta Health.

B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

The work that CIHI put towards building foundation data for this project has led to a series of data preparation steps for provinces prior to inputting their data into the CIHI Population Grouping Methodology software. Many data preparation steps are now part of the software in order to improve its ease of use for stakeholders. Complete details are documented in *CIHI's Population Grouping Methodology 1.0: Software User Guide.*

2.7 Costing information in foundation data

The development of the predictive indicators uses patient-level costs that are representative of health expenditures directed towards a person over a defined time period and across defined health sectors. Cost data for version 1.0 focused on inpatient, day surgery, emergency department, and physician services. As described earlier, patient-level costs were used when available but were otherwise estimated. The costs for each record, whether real or estimated, were then aggregated for each person to represent hospital and physician costs: one aggregated cost to cover the concurrent period and another to cover the prospective period. That is, two costs were assigned to each person.

Table 5 summarizes the extent of cost estimation that was needed to build patient-level cost profiles for the prospective period.

Table 5 Real and estimated patient-level cost data used for cost profiles for the prospective period (2012-2013), by province

	British Columbia*	Alberta	Ontario[†]	All 3 provinces
Target population for prospective period	**4,660,893**	**3,805,574**	**13,892,551**	**22,359,018**
Persons with no cost estimation needed	**4,029,877**	**2,753,104**	**7,630,031**	**14,413,012**
No contact with inpatient, day surgery, emergency department or physician during prospective period	843,489	785,087	2,607,556	4,236,132
Visited fee-for-service physician(s) only and/or case costing hospitals only during prospective period	3,186,388	1,968,017	5,022,475	10,176,880
Persons with cost estimation needed	**631,016**	**1,052,470**	**6,262,520**	**7,946,006**
Physician costs estimated only	145,216	178,934	3,518,310	3,842,460
Hospital costs estimated only	428,925	742,052	553,805	1,724,782
Physician and hospital costs estimated	56,875	131,484	2,190,405	2,378,764

Notes
* BC costs exclude costs for emergency department visits.
[†] ON costs exclude costs for mental health inpatient stays that are captured in OMHRS.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
Alberta Health Registration Files, Alberta Health.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

2. Foundation data

Tables 6 and 7 show the costs in the foundation data for the concurrent and prospective periods respectively. In the prospective period, for example, hospital and physician costs in the foundation data, per-capita, are $1,424 in British Columbia, $1,949 in Alberta and $1,370 in Ontario.

Table 6 and 7 also show some comparative cost information from CIHI's National Health Expenditures Database (NHEX) and the National Physicians Database (NPDB) for the two broad categories – hospital and physicians costs.

Hospital costs in the foundation data represent publicly funded hospital services for inpatient stays, day surgeries and emergency department visits for the target population. This means that not all hospital costs are represented in the foundation data. Hospital expenditures towards other hospital functions (e.g. clinics, hospital-based long-term care, community services, research and so on) are not represented. In addition, compensation costs paid by the hospital to physicians via salary or contract and some depreciation costs are not represented. Hospital expenditures towards persons with federally issued HCNs and out-of-province patients (and for and any new residents after the reference date) are not represented since these persons are not in the target population. When taking all of this into account, the hospital costs represented in the foundation data represent about half of total hospital expenditures that are reported elsewhere by CIHI.[1]

Comparison of aggregated physician costs in the foundation data with that from other sources shows that they are relatively similar, given the different inclusion/exclusion methodologies that are applicable to each source. For example, physician costs in the foundation data represent costs associated to fee-for-service claims and shadow-billing claims. Excluded from the costs in the foundation data, but included in both NHEX and NPDB cost estimates are additional costs that can not be attributed directly to a patient health card (such as payments to physicians to be available on-call or to participate in benefit programs). Additionally NHEX cost estimates include other sources of professional income (i.e.: salaries, sessional, capitation) that would be excluded from both the foundation cost estimates and NPDB cost estimates.

2. Foundation data

Table 6 Patient-level costs in foundation data and comparisons with other cost estimates, concurrent period (2010-2011 and 2011-2012), by province

	British Columbia*	Alberta	Ontario[†]	All 3 provinces
Total costs for target population, in millions	**$13,069**	**$14,339**	**$38,040**	**$65,448**
Physician costs, in millions	**$5,676**	**$5,070**	**$17,107**	**$27,853**
Hospital costs, in millions	**$7,393**	**$9,269**	**$20,933**	**$37,594**
Inpatient	$6,443	$7,263	$15,593	$29,299
Day surgery	$950	$727	$2,483	$4,161
Emergency department	--	$1,279	$2,856	$4,135
Hospital costs reported by NHEX, in millions[1]	$14,504	$16,671	$42,919	$74,094
Physician costs reported by NHEX, in millions[1]	$7,468	$7,061	$24,034	$38,563
Physician costs reported by NPDB, in millions[13]	$5,335	$4,999	$18,180	$28,514

Notes
* BC costs exclude costs for emergency department visits.
[†] ON costs exclude costs for mental health inpatient stays that are captured in OMHRS.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.
National Health Expenditure Database, CIHI; Statistics Canada.
National Physician Database, 2014, Canadian Institute for Health Information.

Table 7   Patient-level costs in foundation data and comparisons with other cost estimates, prospective period (2012-2013), by province

	British Columbia*	Alberta	Ontario[†]	All 3 provinces
Total costs for target population, in millions	**$6,636**	**$7,418**	**$19,031**	**$33,085**
Physician costs, in millions	**$2,894**	**$2,599**	**$8,644**	**$14,136**
Hospital costs, in millions	**$3,742**	**$4,819**	**$10,387**	**$18,949**
Inpatient	$3,247	$3,732	$7,607	$14,586
Day surgery	$495	$375	$1,268	$2,139
Emergency department	--	$712	$1,512	$2,224
Hospital costs reported by NHEX, in millions[1]	$8,173	$8,975	$22,336	$39,484
Physician costs reported by NHEX, in millions[1]	$3,887	$3,823	$12,761	$20,471
Physician costs reported by NPDB, in millions[13]	$2,762	$2,699	$9,666	$15,127

Notes
* BC costs exclude costs for emergency department visits.
[†] ON costs exclude costs for mental health inpatient stays that are captured in OMHRS.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.
National Health Expenditure Database, CIHI; Statistics Canada.
National Physician Database, 2014, Canadian Institute for Health Information.

3.Health condition classification

The foundation of the population grouping methodology is the health condition classification, which consists of a set of 226 health conditions and identifies, for each person, which of the 226 conditions the person has. The health condition classification uses diagnosis information from historical clinical administrative data from a person's encounters with the health system.

The health condition classification includes both the mapping of diagnosis information to health conditions and the logic to build person-level clinical profiles. 225 of the 226 health conditions are identified via 27,478 ICD-9 and ICD-10-CA diagnosis, 118 clinical data elements from the RAI-MDS 2.0 instrument and 121 clinical data elements from the RAI-MH instrument. One health condition – healthy newborn – is identified by age. All those born within the 2 year concurrent period are assigned the healthy newborn condition (prior to overrides).

The health condition classification is used by predictive models that are predictive of current and future service needs. However, other factors shape the classification, such as considerations towards its potential use for disease surveillance and certain administrative and health system monitoring purposes. Some health conditions are part of the classification to address these other needs and do not necessarily benefit the predictive models.

For most clinical data sources (DAD, NACRS, OMHRS, CCRS), presence of a diagnosis at a single point in time is enough information to identify a health condition in the health condition classification. For PLPB data, an additional step is required to assign health conditions – namely tagging rules. The classification maps PLPB diagnoses to health conditions and then builds person-based clinical profiles by assigning health conditions to persons through the use of "tagging rules." Tagging rules specify the criteria for when a health condition is to be assigned to a person's clinical profile.

As a final step, groups of health conditions that are clinically related to one another are addressed through "clinical overrides".

The goal of the clinical classification is to create person-based clinical profiles for a population that are clinically and statistically relevant and robust to the variations in diagnosis coding. For example, physicians and nurse leaders should be able to intuitively understand the groupings of health conditions and the methods used to assign these to persons (i.e. the groupings should be clinically relevant). The health condition categories need to have sufficient sample size and be predictive of cost (i.e. statistically relevant). Furthermore, the classification should minimize the effects of coding variation by providers when building person-based clinical profiles (i.e. robust to variations in coding).

3.1 Health conditions

The health conditions are the building blocks of the clinical classification. They set the boundaries for the grouping of clinically similar diseases. Decisions about the groupings of diseases were the result of an iterative process of:

- Querying CIHI data to understand the health conditions that are frequently treated in different health sectors and/or are costly at the person or population level;

- Reviewing literature to identify health conditions that have been useful to researchers who study the burden of disease on populations;[14-22]

- Adjusting health condition groupings to address differences in diagnosis classifications across provinces; and

- Incorporating physician feedback.

The health condition classification includes a diverse set of conditions and aims to satisfy the many intended uses discussed earlier. The health conditions used in the classification are listed in Appendix A. They represent:

- **Chronic diseases:** diseases of long duration and generally slow progression.[23] Chronic diseases tend to be long-lasting conditions that can be controlled but not cured.[24]

- **Signs and symptoms:** a sign is objective evidence of an underlying disease that is apparent to the patient, physician, and others (e.g. a rash) whereas a symptom is any subjective evidence of disease that only the patient can perceive (e.g. pain).

- **Disabilities:** personal attributes that can provide insight into one's ability to live independently and make decisions regarding his or her health care needs.

- **Medical emergencies:** acute injuries or illnesses that pose an immediate risk to a person's life or long-term health.[25]

- **Short-term / minor acute illnesses:** acute injuries or illnesses that do not pose an immediate risk to a person's life or long-term health.

- **Other health states:** remaining health information that is predictive of health care need but that does not belong to any of the above categories. For example pregnancy influences a person's health care needs but is not a disease.

3.2 Diagnosis code mappings

Diagnosis data available in administrative data is captured using various clinical classifications. Diagnosis codes for ICD-10-CA and ICD-9, as well as mappings for diagnoses collected on the RAI-MDS 2.0 and RAI-MH instruments, are mapped to the health conditions. The methodology utilizes all diagnoses from the hospital inpatient and day surgery abstracts, and the first recorded diagnosis on each physician claim. Users can choose whether or not to include the CCRS diagnoses, if available. The classification uses clinical data over a 2-year window leading up to the reference date. Table 1 in section 2.2 lists the classifications used across the various databases included in the methodology.

Diagnosis mappings assign each diagnosis code to a single health condition. Some health conditions have hundreds of diagnosis codes mapped to them while others have a dozen or so. The simplest mapping is for *S01 Palliative State*, which has one ICD-10-CA code and one RAI-MDS 2.0 data element mapped to it. This mapping approach works for most of the health conditions albeit with some limitations:

- Some ICD codes represent a combination of health problems. For example, ICD-10-CA codes E50.2, E50.4 and E50.5 represent eye disorders that result from vitamin deficiency. In the beta methodology, the classification maps these to B07 Other Eye Diseases & Disorders and does not map to J08 Malnutrition & Vitamin Deficiency.

- Several health condition associated with newborns are mapped to N43 Newborn/Neonate with Other Complicating Diagnosis where the equivalent health condition for the rest of the population maps elsewhere. To illustrate, ICD codes for newborns with rubella are mapped to N43 and do not map to P43 Other Viral Infection.

- Some classifications used in the source data are captured at a level of detail that is less granular than the population grouping methodology's health conditions: the most notable of these is the CCRS diagnosis of I1rr Cancer. Without knowing the type of cancer the classification maps this to R17 Other/Unspecified Cancer. Other examples can be found in the physician billing provincial administrative codes in certain provinces.

- Some newborns are not captured through diagnosis codes, possibly due to diagnosis codes for birth and newborn care being attributed to the mother instead of the child.

3.2.1   ICD-9 and ICD-10-CA diagnosis mapping tables

Diagnosis codes in the 2012 and 2015 versions of ICD-10-CA are represented in the mapping tables as well as historical codes that have since been discontinued. Historical ICD-10-CA codes are included to allow users of the methodology to study the effects of expanding the concurrent period from the recommended 2 years or to study population changes over time by shifting the reference date. The diagnosis mapping tables enable jurisdictions to group data as far back to when version 2001 of ICD-10-CA classification was used.

Discontinued codes are not an issue with the ICD-9 classification since this classification has not recently undergone a revision. The mapping tables include an exhaustive list of ICD-9 codes and have been expanded to include 3-digit ICD-9 codes that could be missing mandatory fourth digits in the source data and 4-digit ICD-9 codes that could be missing mandatory fifth digits in the source data. ICD-9 codes with optional fifth digits have been added where identified.

The mapping tables list all ICD-9 and ICD-10-CA codes including those that are not relevant for any of the health conditions. ICD diagnosis codes that map to "ZZZZ" are considered out-of-scope for the classification as they do not represent any of the health conditions that make up the health condition classification. Examples of out-of-scope diagnosis codes are external cause codes. Also, as discussed at the beginning of chapter 3, healthy newborns are identified by age, thus all codes that represent healthy newborns are mapped to "ZZZZ."

3.3 Tagging rules

PLPB data is not inherently meant to be used as a clinical administrative data source, but as a mechanism to pay physicians. As such, it is considered to have a greater risk of false-positive disease tagging than sources of clinical administrative data for other health sectors since it does not have coding standards. Diagnosis codes reported could refer to query diagnoses for a patient when lab tests are being ordered. They could be preliminary or suspected diagnoses from a general practitioner who is referring a patient to a specialist. They could be diagnoses reported by a physician who was temporarily in contact with a patient in settings such as the emergency department. These and other factors result in an increased risk of false-positive disease tagging when using diagnoses from the PLPB data.

3. Health condition classification

To mitigate against false-positives, tagging rules are applied to PLPB data. Using both diagnosis data and the date of the encounter (service date), tagging rules are applied after the diagnosis codes are mapped to one of the 225 diagnosis-driven health conditions[xi].

There are 3 types of tagging rules:

- Tagging rule 1 assigns a health condition to a person if it is observed 1 or more times over the concurrent period. 84 of the 225 diagnosis-driven health conditions use this tagging rule.

- Tagging rule 2 requires the health condition to be observed for at least 2 physician visits, on two separate dates, prior to assigning the health condition. 136 of the diagnosis-driven health conditions use this tagging rule. These conditions are described as chronic conditions or health problems with long-lasting effects and need ongoing monitoring by a physician.

- Tagging rule 0 affects 5 of the diagnosis-driven health conditions. Four of these conditions are major acute episodes that cannot reasonably be managed through physician office visits alone. To this end, these health conditions are not assigned using PLPB data. The fifth condition is S01 Palliative State; it has been assigned this tagging rule because there are no ICD-9 codes available in the physician claims data to denote palliative care.

The definition of a physician visit used by tagging rule 2

Tagging rule 2 has been applied to differentiate whether or not a person has visited a physician two or more times over the concurrent period for a particular health condition.

When more than one claim is submitted by a physician on the same day for treating one health problem (i.e. ICD code) for a patient, the population grouping methodology software will count that set of claims as one physician visit. This ICD code is mapped to one of the health conditions in the classification. If this is a health condition uses tagging rule 2, the methodology will search elsewhere in the data for another physician visit involving that same patient and the same health problem (could be a different ICD code that maps to same health condition).

Tagging rule 2 would be satisfied if there was one or more claim for the same patient, health condition and physician but on a different day, or if there was one or more claim for the same patient, health condition, and a different physician (in which case the visit could occur on the same day).

The software automatically addresses the issues of "duplicate claims" in the PLPB data.

These tagging rules incorporate physician data more cautiously in response to the unknown quality of its diagnosis data. Other researchers who use health administrative data for disease

---

[xi] One health condition – healthy newborn – is identified by age. All those born within the 2 year concurrent period are assigned to the healthy newborn condition (prior to overrides).

surveillance also report treating physician data with greater caution due to risk of falsely attributing diseases to persons.[26-31]

Appendix A lists the tagging rules used for all health conditions.

3.4 Clinical overrides

Clinical overrides represent a transition point in the classification where the perspective shifts from looking at health conditions individually to looking at the clinical profile of the person holistically. Overrides remove redundancies that may exist on a person's clinical profile after the assignment of health conditions.

An example of a clinical override would be Seizure and Epilepsy. Should a person have both of these conditions assigned after the tagging rules are applied, Epilepsy would "override" Seizure as seizures are a symptom of epilepsy.

Clinical overrides address a variety of situations. Some are designed such that more severe manifestations of a condition override less serious ones (Cirrhosis vs. Chronic Liver Disease). Some constrain any redundancy of disease tagging to persons that are a consequence of variation in diagnosis coding practices amongst physicians and/or hospital coders (Influenza vs. Other & Unspecified Infection). Other overrides consider the health conditions that may be attributed to a person while treating a patient's symptoms during a period of recovery or, while treating a patient's underlying symptoms until such time a diagnosis is made (Abdominal Pain vs. Acute Gastrointestinal Hemorrhage).

Appendix A lists the clinical overrides that are used in the methodology. Figure 2 demonstrates how the overrides change a person's clinical profile.

3. Health condition classification

Figure 2   Clinical profile before and after applying clinical overrides, illustration only with foundation data
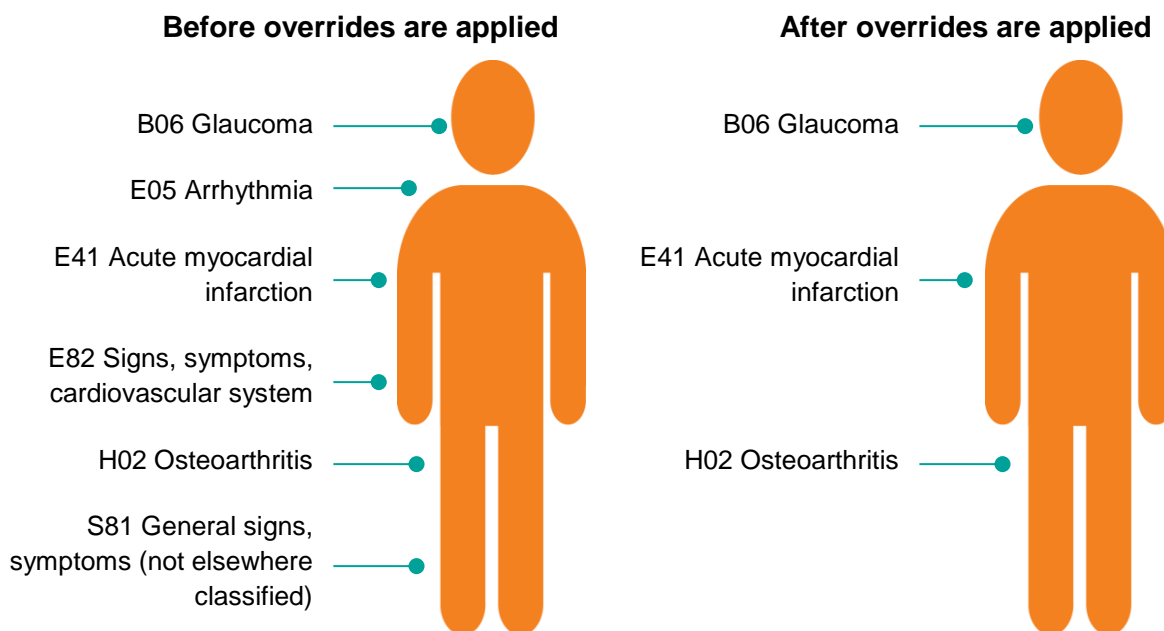
Before overrides are applied

- B06 Glaucoma
- E05 Arrhythmia
- E41 Acute myocardial infarction
- E82 Signs, symptoms, cardiovascular system
- H02 Osteoarthritis
- S81 General signs, symptoms (not elsewhere classified)

After overrides are applied

- B06 Glaucoma
- E41 Acute myocardial infarction
- H02 Osteoarthritis

Source
Canadian Institute for Health Information, population grouping methodology, 2016

The clinical overrides are similar in concept to those used by the Centers for Medicare & Medicaid Services[21] and the German Federal Insurance Office.[15] The overrides were designed such that clinical associations between conditions could be established. The project's physician panel was involved in the review of the overrides.

3.5 Health condition classification applied to foundation data

This section highlights the outputs of the health condition classification when it was applied to the foundation data. It looks at the health conditions that remain assigned to a person's clinical profile after the clinical overrides are applied. The population included in this analysis is 23,000,543 persons: those with HCNs eligible for use in Ontario, Alberta or British Columbia for all or part of the concurrent period (Table 8).

Of these persons, 11.5% are considered health system non-users—that is, during the concurrent period, they had no encounter with a physician, were never assessed in a long-term care facility or hospital complex continuing care bed, nor were hospitalized as an inpatient or day surgery patient. 4.7% persons are considered health system users (with one or more of these encounters) but had none of the 226 health conditions identified. There are a variety of reasons that can result in a person being classified as a health system user with 0 health conditions. One example is when a person's only contact with the health system during the concurrent period was with a physician where the diagnosis reported on the physician claim is considered out-of-scope for the classification. Finally, the majority of the population (83.8%) is health system users with at least 1 of the 226 health conditions.

Table 8 Distribution of non-users, users without health conditions and users with health conditions in the foundation data, by province

	Number of persons			
	British Columbia	**Alberta**	**Ontario**	**All 3 provinces**
Total population during the concurrent period (2010-2011 and 2011-2012)	**4,802,690**	**3,966,936**	**14,230,917**	**23,000,543**
Health system non-users	499,575	479,443	1,676,656	2,655,674
Health system users: 0 health conditions	181,268	176,445	718,575	1,076,288
Health system users: 1 or more health conditions	4,121,847	3,311,048	11,835,686	19,268,581

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

3. Health condition classification

While the clinical overrides do not affect the classification of persons in Table 8, they can change the number of health condition for persons who have 2 or more health conditions. Table 9 shows how the overrides change clinical profiles in terms of health condition count. It illustrates that clinical profiles for persons tend to be most affected by the overrides when more health conditions are present. For example, almost all persons with 9 health conditions identified had a change in their clinical profile—i.e. a reduction in the number of health conditions, whereas only 26.8% of the persons with 2 health conditions identified had a reduction to one health condition.

Table 9   Change in health condition count after clinical overrides applied, all 3 provinces

Health condition count before overrides	Health condition count after overrides, %									
	Number of persons	1	2	3	4	5	6	7	8	9	10+
1	2,945,520	100.0%									
2	2,994,487	26.8%	73.2%								
3	2,737,854	3.9%	42.6%	53.5%							
4	2,330,460	0.6%	12.0%	49.9%	37.5%						
5	1,888,153	0.2%	2.6%	22.2%	49.9%	25.1%					
6	1,481,465	0.2%	0.7%	7.0%	31.4%	44.5%	16.2%				
7	1,141,696	0.4%	0.2%	2.1%	13.5%	37.3%	36.5%	10.0%			
8	865,152	0.5%	0.1%	0.8%	5.0%	20.7%	38.9%	28.0%	6.0%		
9	650,034	0.8%	0.0%	0.3%	1.9%	9.4%	26.9%	36.7%	20.4%	3.6%	
10+	2,233,760	4.2%	0.0%	0.0%	0.3%	1.4%	5.2%	12.7%	18.3%	16.9%	40.9%

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

The classification aims to stratify the population into groups with varying levels of need. Figure 3 examines the concurrent needs (i.e. costs) for populations when segmented based on the number of health conditions identified. Two bar graphs are presented. In blue is the analysis based on the health conditions assigned before overrides were applied to the data. In green is the analysis based on the health conditions that remain assigned after the overrides are applied. The average per-person cost increases as the number of health conditions increases, with the trend more predominant after overrides are applied.

The exception to the trend is for persons with 1 health condition after the overrides are applied. This anomaly is driven by the 114,769 persons who have S01 Palliative State. The clinical override methodology for this health condition is unique. When S01 is present, the override methodology "untags" all other health conditions that may have been present during the concurrent period. If persons with palliative state are removed from the analysis, the green bar for the 1 health condition subset drops to $467, aligning with the pattern observed for the other health condition groupings.

The dichotomous population in the 1-health-condition category—those who have S01 Palliative State versus those who do not—is illustrated further in Table 10.

3. Health condition classification

Figure 3 Average costs for person during the concurrent period (2010-2011 and 2011-2012), by number of health conditions, all 3 provinces
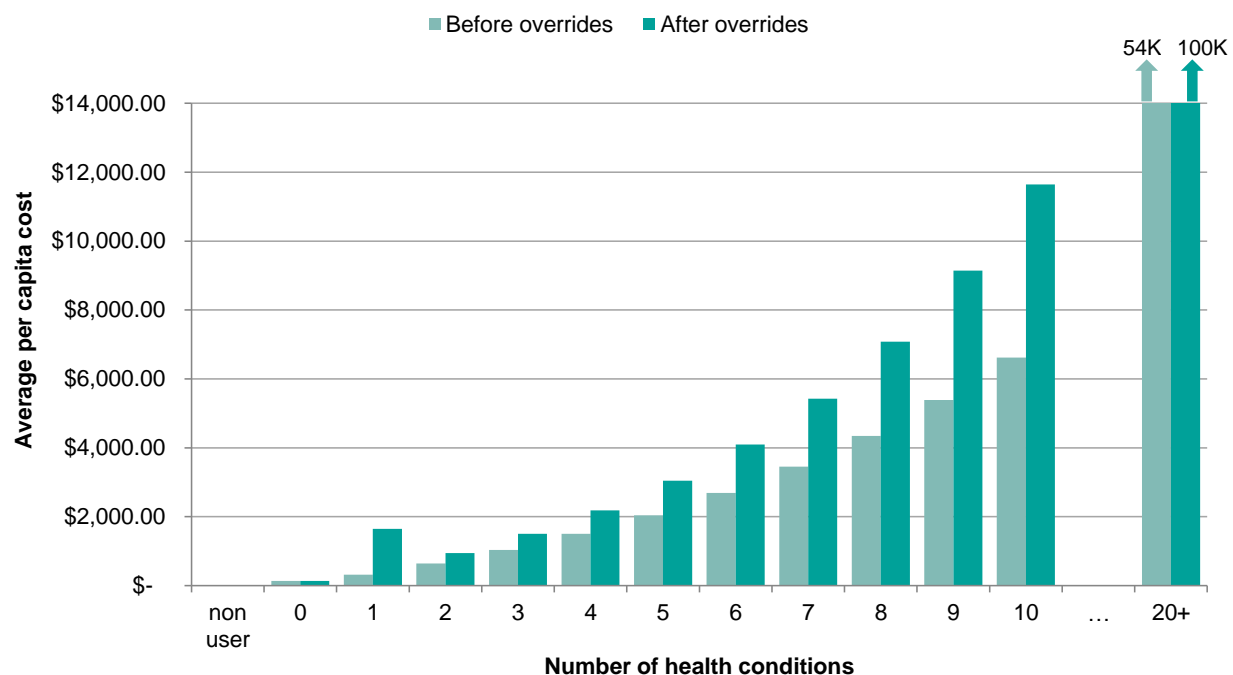
Source
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

Table 10   Persons with one health condition after clinical overrides: cost during concurrent period (2010-2011 and 2011-2012), all 3 provinces

	S01 Palliative state	Not S01 Palliative state
Number of persons with 1 health condition	**114,769**	**3,868,626**
Person-level cost at different quintiles		
20th percentile	$14,195	$76
40th percentile	$26,786	$158
60th percentile	$46,178	$275
80th percentile	$89,858	$519

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

The classification is designed to represent a sizable group of persons for each health condition. Table 11 lists the health conditions most and least commonly observed in the population after the clinical overrides were applied. Appendix B provides details for all 226 health conditions.

3. Health condition classification

Table 11 The 5 highest and 5 lowest volume health conditions after clinical overrides, during the concurrent period (2010-2011 and 2011-2012), all 3 provinces

Health conditions with:	Number of persons
Highest volume	
C42 Acute ENT, upper respiratory condition (incl. benign neoplasm, croup)	5,885,750
H46 Joint/tendon disorder and injury (incl. pain, sprain, strain)	4,734,645
F81 Signs, symptoms digestive & hepatobiliary system	3,060,448
E10 Hypertension	2,278,188
I42 Skin infection (incl. cellulitis)	2,117,408
Lowest volume	
D02 Congenital disorder of the respiratory system	2,931
R13L Ovarian cancer	2,130
J01 Cystic fibrosis	1,808
O04 Spleen disorder	1,645
I01 Autoimmune skin disorder	1,341

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

The classification's performance has face validity when calculating the average cost of persons identified with a given health condition. Table 12 lists the health conditions associated with persons most costly to treat, in the concurrent period. These health conditions tend to be treated in hospital, resulting in higher costs . More details are available in Appendix B.

Table 12 Health conditions with highest per-person average cost after clinical overrides, during the concurrent period (2010-2011 and 2011-2012), all 3 provinces

Health condition	Number of persons	Average cost
N41 – Extremely low birth weight or immaturity	3,371	$113,437
S43 – Ostomy complication	7,953	$92,783
D41 – Respiratory failure	30,916	$89,697
S41 – Transplant complication	3,086	$88,442
S47 – Shock	6,682	$80,672
P41 – Sepsis	46,148	$70,336
O04 – Spleen disorder	1,645	$61,224
S04 – Presence of ostomy	18,640	$55,078
D04 – Pulmonary hypertension	13,742	$54,325
Q81 – Delirium	29,142	$53,405

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

To understand the cost pattern for populations that predominantly rely on out-of-hospital physician care, person-level costs were divided into 3 categories: hospital costs, physician costs associated with hospital-based care, and physician costs associated with non-hospital based care. Figure 4 shows this cost breakdown and divides the population based on the number of health conditions that remain after the overrides are applied. As the number of health conditions increases, the cost share shifts from physician care for non-hospital care (in teal) to hospital care (in orange and pale teal). Note that persons with *S01 Palliative State* have been analyzed separately and have a cost pattern similar to the 20+ health condition grouping.

Figure 4 Physician and hospital costs during the concurrent period (2010-2011 and 2011-2012), by number of health conditions after clinical overrides, all 3 provinces
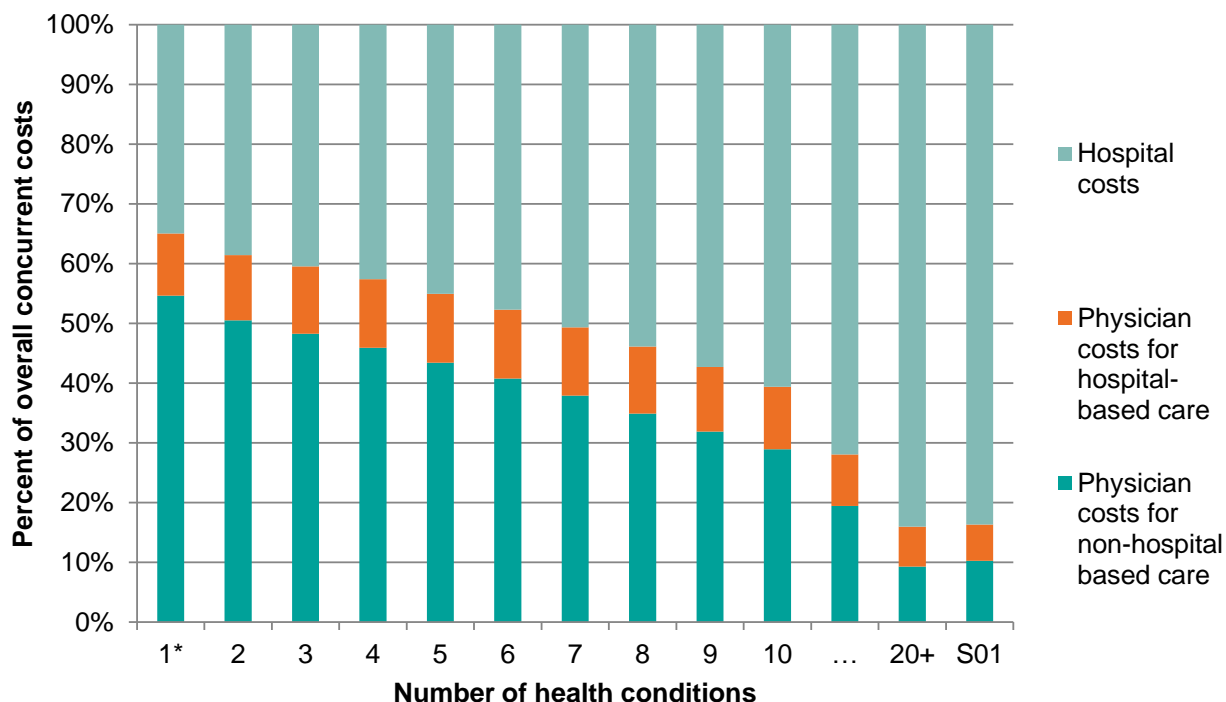
Notes
* Persons with 1 health condition, excluding those with S01 Palliative state. Persons with S01 Palliative state are presented at the end of the figure.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

Further analysis looked at the number of health, by age group. 5-year age groups were used, with the exception of . "0–5 years", which was split into "0-1 years" and "2-4 years". Figure 5 shows, for each age group, a stacked bar graph illustrating the proportion of persons with 0 health conditions (in red), then with 1, 2, 3- 4 and 5+ conditions (in green, purple, light blue and orange respectively). The bottom segment (in dark blue) represents non-users.

Figure 5 shows that, across all age groups, a notable proportion of the population has 1 or more health conditions. These health conditions encompass the full spectrum of the 226 health conditions in the classification including minor acute conditions. As age increases, the proportion of non-users (starting from age 30) and users with no conditions (starting from age 50) both decrease. The older population has more health conditions present.

The increase in the proportion of non-users for older populations, in particular those aged 85 and older, is due to data quality issues with reporting death information in the Registered Persons Data. This issue is most notable with Ontario's RPD data which makes up more than half of the foundation data. Table 13 further describes the non-users for the three provinces in the foundation data.

Figure 5 Number of conditions present during the concurrent period (2010-2011 and 2011-2012), by age group, all 3 provinces
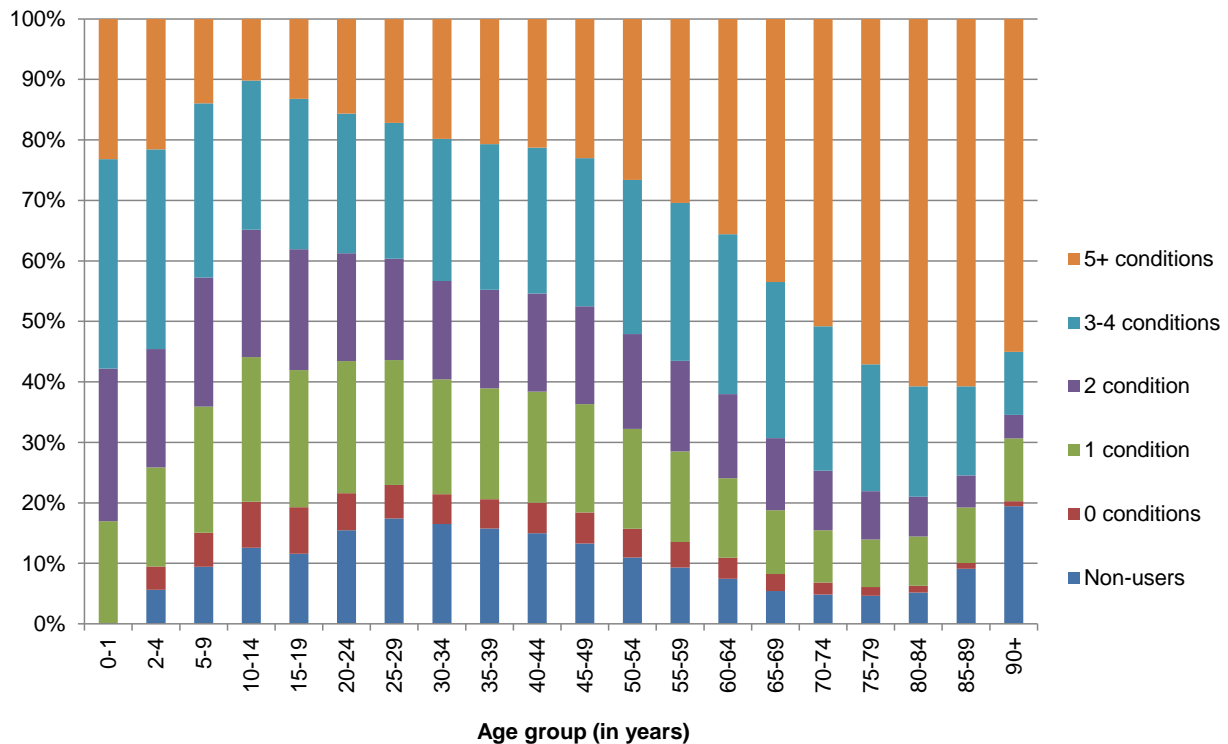
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

3. Health condition classification

Table 13 Number of non-users by 10-year age groups, by province

	British Columbia	Alberta	Ontario	All 3 provinces
Non-users during the concurrent period (2010-2011 and 2011-2012)	**499,575**	**479,443**	**1,676,656**	**2,655,674**
< 2 years old	0	0	0	0
2-9 years old	28,125	45,328	86,541	159,994
10-19 years old	64,957	74,448	186,809	326,214
20-29 years old	109,525	101,910	304,487	515,922
30-39 years old	85,854	90,185	329,196	505,235
40-49 years old	87,240	71,265	331,367	489,872
50-59 years old	68,898	53,210	216,665	338,773
60-69 years old	30,812	23,798	97,660	152,270
70-79 years old	11,514	9,842	43,882	65,238
80-89 years old	7,600	6,115	42,845	56,560
90+ years old	5,050	3,342	37,204	45,596

Suspected data quality issues with death data. This becomes particularly noticeable when older people are identified as non-users.

Sources

4. Health profile groups (HPGs)

The health profile grouping methodology leverages all of the methodology used in the assignment of health conditions, and then synthesizes an individual's clinical profile down to the most complex and clinically relevant health condition present.

As illustrated in Figure 6, once all of the health conditions have been assigned, a health profile group is then assigned to each individual.

Figure 6 Health condition and health profile group assignment
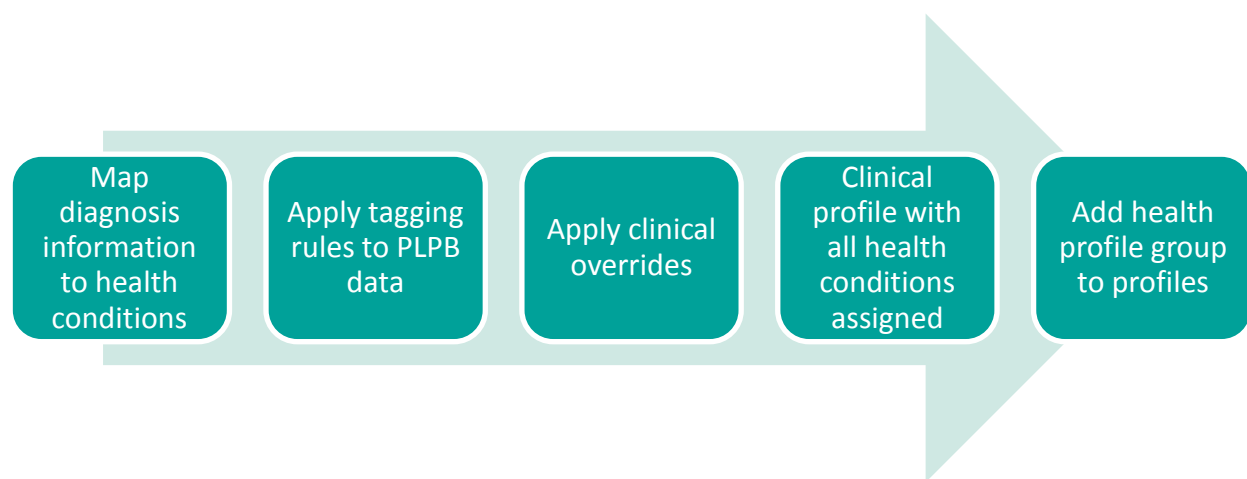
Map diagnosis information to health conditions → Apply tagging rules to PLPB data → Apply clinical overrides → Clinical profile with all health conditions assigned → Add health profile group to profiles

Sources
Canadian Institute for Health Information, population grouping methodology, 2016

The health profile grouping methodology is comprised of four key components.

1. **Branches.** Branches represent a synthesized clinical profile for each individual. The 226 health conditions were rolled up into 162 branches. Two additional branches were created for users without health conditions and non-users, for a total of 164 branches.

2. **A ranking.** A ranking was developed to order all of the branches from the most clinically complex and resource intensive to the least complex. This ranking is used to identify the single branch to assignto each individual.

3. **Presence of relevant comorbidities.** Where considerable volume and cost differentiations existed, major and moderate branches were split into two groups – one to identify individuals with no other major or moderate branches; and one to identify those with at least one other major or moderate branch (prior to applying the ranking to choose the most relevant branch).

48

4. Health status groups

4. **Assigning cost weights.** An average concurrent and prospective health profile group cost weight is assigned to each individual. The cost weights assigned to the health profile group represent the average cost of all individuals within a health profile group divided by the average cost of all individuals in the population.

To ensure that the branches, ranking and categories were clinically relevant, the methodology was developed in consultation with physician experts and the population grouping methodology expert group. The following sections below will provide details on each of the methodological components.

4.1 Assigning branches

The development of the health profile groups starts with rolling up the 226 health conditions into 162 branches. Branches were developed with the following considerations in mind:

- Group together one or more health conditions that are clinically similar and require similar cost resources.

- Delineate between severe and more moderate cases. For example, branches were created to identify individuals who had suffered from stroke with and without paralysis, as the clinical profile and associated costs would be very different between these two types of stroke patients.

- Provide enough differentiation between individuals with minor conditions, to ensure that analysis would also be meaningful for those looking to explore health care utilization in the community sectors as well as the hospital sector.

- Finally, branches were created to identify health system users with no health conditions and non-users.

In total, the 226 health conditions were rolled up to 162 branches. A one-to-one roll-up was applied to 122 health conditions. In these instances, each of the 122 health conditions is represented by a branch in the health profile group classification. A many-to-one roll-up was applied to 86 health conditions. In these instances, 2 or more health conditions were assigned to one branch. This added another 22 branches to the health profile group classification. The remaining health conditions were analyzed to delineate between severe and more moderate instances of health conditions to create 18 branches. Two branches were created, as indicated above, for users without health conditions and non-users, for a total of 164 branches. Appendix C, table C1 provides full details for branch assignment for all health conditions.

Once all of the health conditions have been rolled up to the appropriate branches, the health profile group hierarchy is then applied to identify the branch that best represents the clinical and cost profile for each individual and to assign that branch to the individual.

4.2 Ranking health profile groups

Each of the 164 branches is ranked from the most clinically complex and resource intensive to the least resource intensive. This ranking was developed by isolating the concurrent and prospective cost effect of the presence of each of the branches in the dataset, regardless of branch combinations. Taking both concurrent and prospective results into account, an average cost ranking was then established. Additionally, some clinical adjustments to the rankings were implemented and a preliminary overall branch ranking created.

Using this preliminary ranking and assignment of persons to branches, the average cost of persons within each branch was assessed, which resulted in some changes to the rankings as well as some further clinical adjustments.

4.2.1 Assignment of health profile categories

Each of the 164 branches is also linked to one of 16 categories, as follows:

- Palliative
- Major acute
- Major chronic
- Manor cancer
- Major mental health
- Major newborn
- Moderate acute
- Moderate chronic

- Other cancer
- Other mental health
- Minor acute
- Minor chronic
- Obstetrics
- Healthy newborn
- Users with no health conditions
- Non-users

Appendix C, table C2 contains the rankings for the branches, the category assignment and the associated volumes and average concurrent prospective costs for the foundation data.

4.3 Splitting branches by the presence of comorbidities

In many instances, the resources required to manage the health needs of individuals with multiple conditions tends to be greater than the resources needed to manage the health needs of individuals with only one condition. When developing health profile groups, analysis was undertaken to identify branches where the resources required to treat patients varied considerably between those with major or moderate comorbidities and those without. For each major and moderate branch (90 of the 164 branches), analysis was undertaken to determine if there was reasonable evidence to warrant splitting the branch into two – those without major/moderate comorbidities (i.e.: assigned to one major/moderate branch prior to applying the ranking) and those with major/moderate comorbidities (i.e. those assigned to 2 or more major/moderate branches prior to applying the ranking).

Decisions to split branches to identify comorbidities were based generally on the volume of cases (a minimum of 1,000 cases with and without comorbidities) and the difference between costs (required costs to be approximately 2+ times greater for those with comorbidities). Of the 90 branches categorized as major or moderate, 75 branches were further split to segment those with and without major or moderate comorbidities. Table 14 provides a few examples of branches where volumes and costs differentials warranted splitting the branch into two groups.

Table 14 Sample of branches split due to the presence of one or more major or moderate comorbidities, concurrent period (2010-2011 and 2011-2012), all 3 provinces

Branch	Volume		Average cost	
	No major/moderate comorbidities	With major/moderate comorbidities	No major/moderate comorbidities	With major/moderate comorbidities
N001 – Extremely low birth weight or immaturity	1,058	2,253	$18,258	$152,240
E030 - Heart failure with CAD/arrhythmia	8,978	80,398	$8,520	$26,096
J002 - Diabetes with PVD/other chronic vascular diagnosis	5,788	39,649	$4,694	$23,628
R009 - Leukemia/lymphoma	16,962	26,627	$3,935	$12,489

Sources
Canadian Institute for Health Information, population grouping methodology, 2016

15 of the 90 major and moderate branches were not split. For these branches, either the volume of cases was too low to warrant a health profile group or there was not enough difference in average costs between the two groups. Table 15 provides a few examples of branches where volumes and costs differentials did not warrant splitting the branch into two groups. For the first three branches presented in Table 15, there was not enough volume of cases without major or moderate comorbidities. While there was enough volume for the fourth branch, there wasn't enough difference between average costs to warrant splitting the branch into two groups.

Table 15 Sample of branches not split due to the presence of one or more major or moderate comorbidities, concurrent period (2010-2011 and 2011-2012), all 3 provinces

	Volume		Average cost	
Branch	**No major/moderate comorbidities**	**With major/moderate comorbidities**	**No major/moderate comorbidities**	**With major/moderate comorbidities**
D001 - Respiratory failure with heart failure	30	11,318	$19,458	$97,997
D003 - Pulmonary hypertension	334	10,321	$3,079	$41,292
S004 - Presence of ostomy	525	16,308	$7,331	$44,079
R020 - Uterine cancer	3,286	4,039	$6,724	$10,484

Sources
Canadian Institute for Health Information, population grouping methodology, 2016

The further segmentation of the 75 branches to differentiate between those with and without, major or moderate comorbidities results in 239 final health profile groups.

4.4 Relationship between branches, categories and groups

All components of the health profile group classification are connected. As illustrated in Figure 7, the 164 branches can be rolled up into one of 16 categories, or further segmented into the 239 health profile groups. This allow users to analyze the data at the level most meaningful to them.

Figure 7   Relationship between components of the health profile group classification
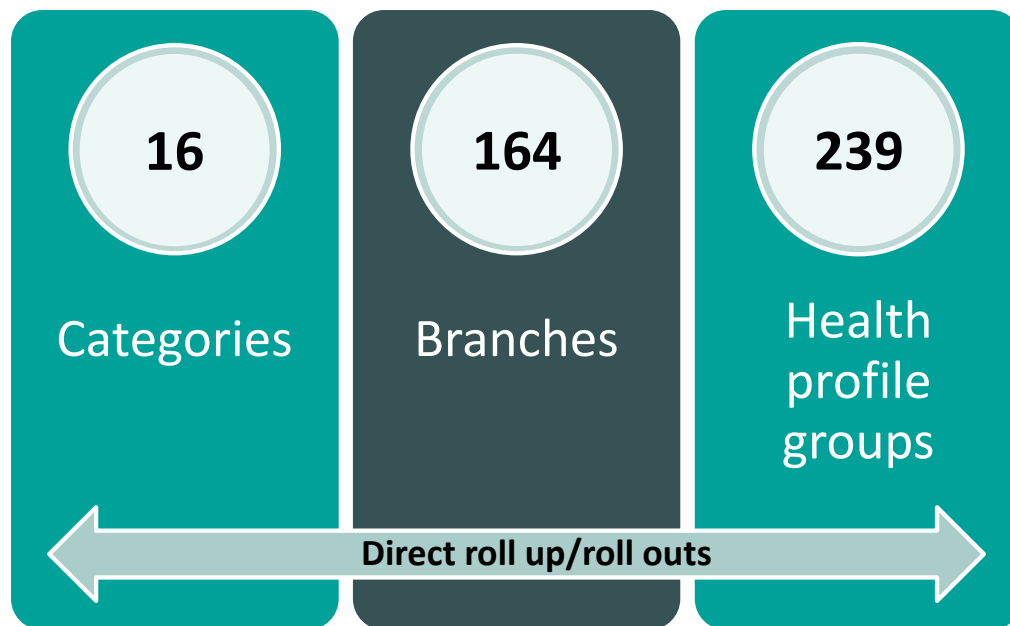
Sources
Canadian Institute for Health Information, population grouping methodology, 2016

5.Functional status classification

Functional status describes a person's ability to perform normal daily activities required to meet basic needs, fulfill usual roles, and maintain health and well-being. It can be influenced by biological or physiological impairment, symptoms, mood, and other factors.[32]

Functional status considers both functional capacity and functional performance. Functional capacity represents a person's maximum capacity to perform daily activities in the physical, psychological, social, and spiritual domains of life. Functional performance refers to the activities people actually do during the course of their daily lives.[32]

Functional status is useful information to incorporate in the classification and would ideally be added to the clinical profiles for special populations. By doing so, the methodology could differentiate health care need beyond what is achievable by looking at health conditions alone.

The functional status classification in the population grouping methodology is focused on describing residents in long-term care facilities and patients in complex continuing care beds (as reported in CCRS).

5.1 RAI-MDS 2.0 outcome scales

The RAI-MDS 2.0 instrument was developed for residents in long-term care facilities and patients in complex continuing care beds, and is used to assess people at the point of care, generating electronic reports that flag risks and inform care planning. It produces a variety of functional status outcome measures. Six of them are included in the functional status classification (Table 16). The functional status methodology uses the outcome measures reported for a resident on his or her last assessment during the concurrent period for all scales excluding the Pressure Ulcer Risk Scale (PURS). The functional status methodology for PURS uses outcome measures reported for all assessments during the concurrent period, so that changes in health status can be measured.

Table 16 RAI-MDS 2.0 outcome scales used in the functional status classification[33]

Outcome Scale	Description	RAI-MDS 2.0 assessment items	Score range
CHESS (changes in health, end-stage disease and signs and symptoms)	Detects frailty and health instability. Designed to identify residents at risk of serious decline.	• Decline in cognition (B6) • Decline in ADL (G9) • Dehydration (J1c) • Edema (J1g) • Shortness of breath (J1l) • Vomiting (J1o) • End-stage disease (J5c) • Weight loss (K3a) • Leaving food uneaten (K4c)	**0–5** Higher scores indicate higher levels of medical complexity and are associated with adverse outcomes, such as mortality, hospitalization, pain, caregiver stress and poor self-rated health.
Pain scale	Summarizes the presence and intensity of pain.	• Frequency of pain (J2a) • Intensity of pain (J2b)	**0–3** Higher scores indicate a more severe pain experience.
ADL* self-performance hierarchy scale (* Activities of Daily Living)	Reflects the disablement process by grouping ADL performance levels into discrete stages of loss (early loss: personal hygiene; middle loss: toileting and locomotion; late loss: eating).	• Personal hygiene (G1jA) • Toilet use (G1iA) • Locomotion (G1eA) • Eating (G1hA)	**0–6** Higher scores indicate greater decline (progressive loss) in ADL performance.
ABS Aggressive Behaviour Scale	Provides a measure of aggressive behaviour.	• Verbally abusive (E4b) • Physically abusive (E4c) • Socially inappropriate/disruptive behaviour (E4d) • Resists care (E4e)	**0–12** Higher scores indicate higher levels of aggressive behaviour.
interRAI PURS Pressure Ulcer Risk Scale	Differentiates risk for developing pressure ulcers.	• Bed mobility self-performance (G1aA) • Walk in room self-performance (G1cA) • Bowel incontinence (H1a) • Shortness of breath (J1l) • Daily pain (J2a) • Weight loss (K3) • History of resolved ulcer (M3)	**0–8** Higher scores indicate a higher relative risk for developing a pressure ulcer.
CPS Cognitive Performance Scale	Describes the cognitive status of a resident.	• Comatose (B1) • Short-term memory (B2a) • Cognition skills for daily decision-making (B4) • Expressive communication (C4) • Eating (G1hA)	**0–6** Higher scores indicate more severe cognitive impairment.

5.2 Functional status classification applied to foundation data

There are 205,633 persons, or 0.9% of the target population, who had one or more RAI-MDS 2.0 assessments completed during the concurrent period (9,326 of which were in both a CCC and LTC facility in Ontario). As illustrated in Table 17, the median age is 85 years and two-thirds of the population is female. The typical person has 11 health conditions after the overrides are applied.

Persons in Ontario's complex continuing care beds are more likely to be medically unstable (higher CHESS) experience pain (higher Pain Scale) and die, compared to residents in long-term care facilities. The long-term care population is otherwise similar in terms of functional status measures across the three provinces.

Table 17   Profile of CCRS residents with functional status information in the foundation data

	Complex continuing care* - Ontario only	Long-term care		
		British Columbia	Alberta	Ontario*
Persons with one or more RAI-MDS 2.0 assessments during the concurrent period	34,426	36,645	23,025	120,802
Median number of days in facility during concurrent period	36	512	413	471
Percent female	57%	66%	65%	69%
Median age (in years)	81	86	86	86
Median number of health conditions tagged over concurrent period	13	11	12	11
Percent died during the concurrent period	38%	34%	31%	33%
Median and interquartile range for:				
CHESS	2 (1-3)	1 (0-2)	1 (0-2)	1 (0-2)
Pain scale	1 (0-2)	0 (0-1)	0 (0-1)	0 (0-1)
ADL hierarchy scale	4 (2-5)	3 (2-5)	4 (3-5)	4 (3-5)
Aggressive behaviour scale	0 (0-1)	0 (0-1)	1 (0-3)	0 (0-3)
Pressure ulcer risk scale	3 (1-4)	2 (1-3)	3 (1-3)	3 (1-3)
Cognitive performance scale	2 (0-3)	3 (2-5)	3 (2-5)	3 (2-5)

Notes
* 9,326 Ontario residents are represented in both the Complex Continuing Care and a Long-Term Care sectors. This analysis categorized them based on the sector reported on the last assessment during the concurrent period.
Excludes 990 residents with overlapping episodes since this would result in days being double counted.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

6. Socioeconomic status classification

Socioeconomic status is a combined measure of a person's economic and social position in relation to others. Income, education and living arrangements are examples of information used to determine socioeconomic status.

Person-level data on socioeconomic status is not currently collected comprehensively across Canada. Hence, the population grouping methodology uses area-based socioeconomic measures, which originate from survey data, and attribute these measures to persons based on the information affiliated to the region they reside in. The two area-based socioeconomic measures provided in the beta release are derived using Statistics Canada survey data.

The population grouping methodology must associate each person to a geographic region in order to assign an area-based socioeconomic measure. The spatial dimension on which area-based socioeconomic measures are assigned are made as small as possible in order to ensure a high degree of homogeneity in the socioeconomic conditions attributed to each resident.[34] The smallest region available when using Statistics Canada survey data is dissemination area. Dissemination areas represent populations of approximately 400 to 700 people.

The population grouping methodology assigns a dissemination area to each person by taking his or her postal code, as recorded in the Registered Persons Data, using the Postal Code Conversion File (PCCF).[35] The methodology retains two dissemination area values for each postal code: one relates to those used in the 2006 Census and the other to those used in the 2011 Census.

Persons with invalid or missing postal code do not have a dissemination area nor an area-based socioeconomic status measure assigned. There are some cases where dissemination area can be derived but where area-based socioeconomic status measures are missing due to data suppression in survey data.

6.1 Deprivation index

The Institut national de santé publique du Québec (INSPQ) Deprivation Index is one of the area-based socioeconomic measures available in the population grouping methodology. The Deprivation Index includes two dimensions of socioeconomic status, social and material, and for each of these measures assigns a value of 1 to 5 with increasing values indicating higher socioeconomic status. These two dimensions are constructed from six items from the 2011 National Household Survey:

• Material dimension	• Social dimension:
– Persons without high school diploma	– Persons living alone
– Employment/population ratio	– Persons separated, divorced or widowed
– Average personal income	– Single-parent families

The population grouping methodology uses the 2011 Census dissemination area associated with a postal code when assigning Deprivation Index. The INSPQ methodology assigns each dissemination area to one of five groups, or quintiles, for each of the social and material factors. The quintiles represent different levels of socioeconomic deprivation. Residents assigned to the highest quintile reside in areas associated with the highest socioeconomic status areas and are least deprived. Each quintile represents 20% of the total population.

There are four versions of the Deprivation Index available. With the national version, the quintiles estimate the distribution as well as the extent of social and material inequalities throughout Canada. Each dissemination area in Canada is compared with each other, and the country as a whole is used as the reference region.[36]

The regional version estimates the variations in deprivation within the five Canadian regions: Atlantic, Quebec, Ontario, the Prairies and British Columbia. The three territories are excluded. The metropolitan version does the same for Canada's three most populated census metropolitan areas (CMA): Montreal, Toronto and Vancouver. Finally, the geographical zones version distinguishes four large geographical entities: the three largest CMAs (Toronto, Montreal and Vancouver); all the other CMAs combined (100,000 to one million people); all the census agglomerations combined (10,000 to 100,000 people); and small towns and rural regions (<10,000 people).[37]

6.2 Canadian marginalization index

The Canadian Marginalization Index (CAN-Marg) was developed by researchers at the Centre for Research on Inner City Health at St. Michael's Hospital in Toronto. CAN-Marg includes four dimensions of socioeconomic status: residential instability, material deprivation, ethnic concentration, and dependency. Each of these four dimensions is assigned values of 1 to 5 based on 18 items from the 2006 Long-Form Census, as described in Table 18. Higher values indicate higher socioeconomic status.

The Canadian Marginalization Index is derived based on the 2006 Census dissemination area associated with a postal code. CAN-Marg differentiates the area-based socioeconomic measures using quintiles that have been created by sorting the marginalization data into five groups, ranked from 1 (least marginalized) to 5 (most marginalized). Each group contains a fifth of the geographic units. For example, if an area has a value of 5 on the material deprivation scale, it means it is in the most deprived 20 percent of areas in Canada. The quintiles were created Canada-wide to enable comparability across the country.[38]

Table 18 Canadian marginalization index

Dimension	2006 long-form census survey items	
Residential instability	• Proportion living alone • Proportion of youth population aged 5-15 • Crowding: Average number of persons per dwelling • Proportion multi-unit housing	• Proportion of the population that is married/common-law • Proportion of dwellings that are owned • Proportion of residential mobility (same house as 5 years ago)
Material deprivation	• Proportion 25+ without certificate, diploma or degree • Proportion of lone-parent families • Proportion government transfer payment	• Proportion unemployment 15+ • Proportion below low income cut-off • Proportion of homes needing major repair
Dependency	• Proportion of seniors (65+) • Dependency ratio (0-14 + 65+)/ (15-64)	• Labour force participation rate (aged 15 and older)
Ethnic concentration	• Proportion of 5-year recent immigrants	• Proportion of visible minority

6.3 Quintile of annual income per person equivalent

PCCF also generates the Quintile of Annual Income per Person Equivalent (QAIPPE) based on 2006 Census data.[35] This income measure is also included in the population grouping methodology to enable further study of this specific dimension of socioeconomic status for risk adjustment. Values range from 1 to 5, with 5 representing the highest annual income quintile.

6.4 Socioeconomic status information in the foundation data

Figure 8 describes the population in the foundation data that have area-based socioeconomic status information available. One of the reasons why a person might not be assigned an area-based socioeconomic status measure is if there are data quality issues with his or her postal code in the RPDB that prevents assignment to a Census dissemination area. Another reason is that Census data can be suppressed for certain dissemination areas. Either of these reasons would result in an area-based socioeconomic status measure being unavailable (CAN-Marg, QAIPPE or INSPQ).

6. Socioeconomic status classification

Figure 8 Persons in the foundation data with area-based socioeconomic measures available
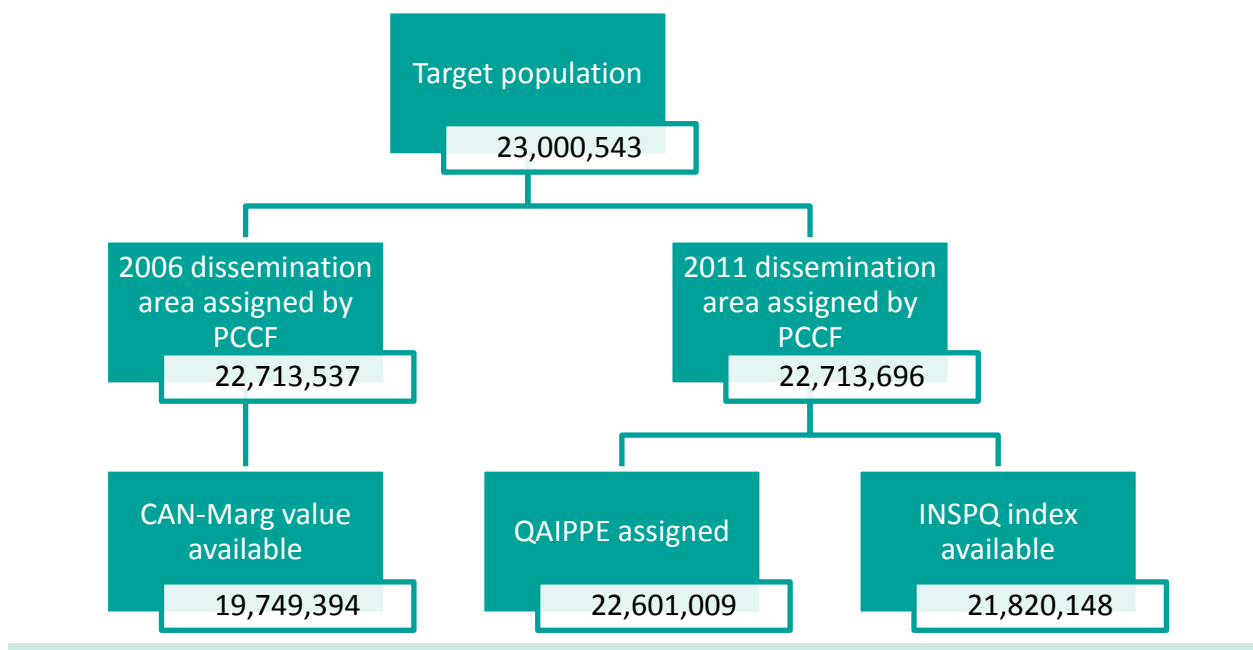
```
                    Target population
                       23,000,543

   2006 dissemination              2011 dissemination
   area assigned by                area assigned by
        PCCF                            PCCF
     22,713,537                      22,713,696

   CAN-Marg value          QAIPPE assigned        INSPQ index
     available                                     available
     19,749,394             22,601,009            21,820,148
```

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

Analysis of costs information in the foundation data shows that persons associated with lower socioeconomic status have higher costs, on average, particularly for the INSPQ social deprivation measure, CAN-Marg's dependency and instability measures, and the QAIPPE income measure. Persons residing in the most ethnically concentrated regions (CAN-Marg Ethnicity) tend to be lower cost.

6.5 Health region and rurality measure

To aid in potential segmentation analysis, both health regions and a measure of rurality are included in the population grouping methodology.

Statistics Canada CANSIM Table 109-5355[39] provides the following definition of a health region:

> *Health regions are legislated administrative areas defined by provincial ministries of health. These administrative areas represent geographic areas of responsibility for hospital boards or regional health authorities. Health regions, being provincial administrative areas, are subject to change.*

Health region boundaries presented in this publication correspond to the health regions presented in the latest file available from Statistics Canada used by CIHI to map postal codes to health regions (i.e., health regions as of 2015).

A measure of rurality, called the Statistical Area Classification, is an output from the PCCF that is retained by the population grouping methodology.[35] A Statistical Area Classification code is assigned to each person based on the 2011 Census dissemination area affiliated with his or her postal code.

7. Predictive indicators for health care costs

Population grouping methodologies commonly include predictive indicators that provide insight into the use of the health care system and the health of the population. Perhaps the most common indicator found in a population grouping methodology is an indicator of health care cost. CIHI's population grouping methodology includes two types of cost weight indicators – one based on the 226 health conditions and one based on the 239 health profile groups. For each of these indicators, cost weights are produced for the concurrent period and the prospective period. Each of the cost weights encompasses hospital inpatient, day surgery, and emergency room costs covered by hospital budgets, as well as publicly funded physician costs.

7.1 Data for indicator development

The foundation data (as described in the chapter 2) was used in the modeling of the predictive indicators. Table 19 provides some basic statistics on this foundation data for each indicator sub-population.

The modeling of the concurrent cost weights is based on all persons who are eligible for health care at any time during the concurrent period. The modeling of the prospective cost weights is based on only those persons who are eligible for health care on the last day of the concurrent period. Anyone not eligible for health care on the last day of the concurrent period is assigned a prospective cost weight of zero. The number of persons included in predictive modeling is slightly lower for the prospective models than for the concurrent models.

Table 19 Descriptive statistics on foundation data, by indicator sub-population, all 3 provinces

Period	Statistics	Non-users	Users without health conditions	Users with health conditions	Overall
			Indicator sub-population		
Concurrent (2010-2011 and 2011-2012)	Number eligible for health care	2,655,674	1,076,288	19,268,581	23,000,543
	Number with concurrent cost of $0	2,623,704	17,784	2,931	2,644,419
	Minimum concurrent cost*	$0	$0	$0	$0
	Average concurrent cost	$2	$140	$3,388	$2,845
	Maximum concurrent cost	$4,845	$45,180	$9,798,919	$9,798,919
Prospective (2012-2013)	Number eligible for health care	2,461,493	1,045,026	18,852,499	22,359,018
	Number with propective cost of $0	1,829,837	351,896	2,095,357	4,277,090
	Minimum propective cost	$0	$0	$0	$0
	Average propective cost	$212	$443	$1,703	$1,480
	Maximum propective cost	$3,973,380	$635,674	$4,157,177	$4,157,177

Notes

* Health system users, either with or without health conditions may have a minimum cost of $0, if all their costs come from sources not included in the cost foundation data. This includes PLPB records where we were unable to impute costs or CCRS and OMHRS records where cost data is not yet included in the foundation data.

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.

Patient-Level Physician Billing Data, Alberta Health Services.

Alberta Health Registration Files, Alberta Health.

Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.

B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

7.2 Health condition costs

7.2.1 Indicator sub-populations for cost weights based on the 226 health conditions

The predictive indicator methodology assigns the cost indicators to everyone eligible for health care in the jurisdiction, as identified from the RPD. For the development and the assignment of the cost weights this eligible population is segmented into three sub-populations: non-users, users without health conditions, and users with health conditions. Segmenting the population into the three sub-populations is based on clinical information from the concurrent period, as represented by the population grouping methodology's clinical classification.

To assign a person to one of the three sub-populations, the methodology looks at all the data sources fed into the software. If, on one or more occasion during the concurrent period, a person receives care from:

1. a physician who is paid either on a fee-for-service or alternative-payment-plan basis (i.e. is captured in the PLPB data), or

2. a hospital either as an inpatient (including mental health hospitalizations), day surgery, clinic or emergency department patient, or

3. long-term care or complex-continuing care centre,

then the population grouping methodology will consider the person to be a health system user. Otherwise it will consider the person to be a non-user.

A person considered to be a non-user by the population grouping methodology may have had contact with a part of the health care system that is not taken into account when assigning the health condition categories. Examples of areas of the health care system that are not used by the classification in assigning health condition categories include the following: hospital-based clinics, out-of-province care, publicly funded prescription drugs.

Users are further segmented into "users without health conditions" and "users with health conditions", based upon whether or not they have one or more of the 226 health conditions in the population grouping methodology.

The criteria for segmenting the population into the 3 sub-populations are summarized below in Table 20.

Table 20 Characteristic of the three indicator sub-populations

	Indicator sub-population		
Characteristics of each person in category	Non-users	Users without health conditions	Users with health conditions
Registered for health care for part or all of the concurrent period	Yes	Yes	Yes
During the concurrent period, received care: • In a hospital as an inpatient (including mental health hospitalizations) or day surgery patient, and/or • In a long-term care facility or a complex-continuing care bed* , and/or • From a physician that was paid for on a fee-for-service or Alternative Payment Plan basis	No	Yes	Yes
During the concurrent period, had one or more of the 226 health conditions	No	No	Yes

Notes
* Ontario residents only.
Source
Canadian Institute for Health Information, population grouping methodology, 2016

7.2.2 Model specification for the health condition cost weights

For the cost weight indicators based on the 226 health conditions, the methodology and the values it assigns are based on linear regression models. Ordinary least square estimation method was employed in estimating fitting these models. These models used cost as the response variable. The predictor variables are age, sex, the 226 health conditions, and the most influential 2-way health condition interactions.

The methodology for the concurrent cost weight and the prospective cost weight differs among the three indicator sub-populations. The statistical models used were developed separately for each of the three sub-populations described above and takes into account the characteristics of each sub-population.

For each indicator sub-population, the data was randomly split into two partitions; one partition consisted of 70% of the records and the other partition consisted of the remaining 30% of the records. The 70% partition was used in the derivation of the cost weights and the 30% partition was used to evaluate the goodness of fit of the weights and provide an assessment of how well they explain cost variations in "new" data.

Indicator sub-populations

Non-users

A person who did not use the health care system in the concurrent period is assigned a concurrent cost weight of zero. They are assigned a prospective cost weight based upon their age and sex. An exception to this is that any person who is not eligible for health care on the last day of the concurrent period is assigned a prospective cost weight of zero.

The prospective cost weight is determined based upon a regression model where the unit of observation in this model is the person (i.e. one record per person) and the response variable is cost (i.e. the sum of hospital inpatient, day surgery, and emergency room costs covered by hospital budgets and publicly funded physician costs, from the prospective period). The predictor variables in the prospective model are combinations of 19 age groups (based on age on the last day of the concurrent period) and sex (male or female).

Table 21   Age groups used by predictive models

Age groups			
0 years	15-19 years	40-44 years	65-69 years
1 year	20-24 years	45-49 years	70-74 years
2-4 years	25-29 years	50-54 years	75-79 years
5-9 years	30-34 years	55-59 years	80+ years
10-14 years	35-39 years	60-64 years	

Source
Canadian Institute for Health Information, population grouping methodology, 2016

The regression model can be specified as:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots \beta_i X_{ij} + \cdots + \beta_{38} X_{i38} \qquad \text{(Equation 2)}$$

where:

Y_i = cost for person *i* in the prospective period

X_{ij} = 1 if person *i* is in age-sex group *j*, 0 otherwise

β_j = the model parameter for predictor X_j

This model assigns a predicted cost to each person. These predicted costs are estimated as the average cost of each of the 38 age-sex categories. Ordinary least squares estimation method was used to calculate these averages, so as to allow the use of the Cook's Distance statistic[40] to identify and remove influential records. From the estimation data (i.e. the 70% partition) there were 19 cases set as influential records, due to having anomalous Cook's Distance values and extreme high cost value.

Users with no health conditions

A person who used the health care system in the concurrent period but does not have any of the 226 health conditions, is assigned a concurrent cost weight and a prospective cost weight based on their age and sex. The cost weights were derived using an approach similar to that used to model the prospective cost weight for the non-user group: two regression models (one for concurrent and one for prospective) using ordinary least squares estimation to calculate the cost weights as the average cost of the 38 age-sex categories.

From the estimation data (i.e. the 70% partition) there were 22 cases identified as influential observations based on the Cook's Distance statistics and these were all excluded from the

8. Predicting health system use for select services

concurrent modeling. In the prospective modeling, 77 cases were excluded based on Cook's Distance values.

Users with health conditions

A person who used the health care system during the concurrent period and has one or more of the 226 health conditions is assigned a concurrent cost weight and a prospective cost weight based on knowledge of which of the 226 health conditions that they had during the concurrent period. A person who is not eligible for health care on the last day of the concurrent period is assigned a prospective cost weight of zero.

These cost weights were derived using linear regression models. The predictor variables in the models are the 226 health conditions (with clinical overrides applied), and interactions for specific pairs of health conditions.

Each of the concurrent regression model and the prospective regression model can be specified as:

$$Y_i = \beta_1 X_{i1} + \cdots + \beta_j X_{ij} + \cdots + \beta_{225} X_{i225} + \theta_1 Z_{i1} + \cdots + \theta_k Z_{ik} + \cdots + \theta_M Z_{iM} \qquad \text{(Equation 3)}$$

where:

Y_i = cost for person i in the period

X_{ij} = 1 if person i has health condition j, 0 otherwise; for j = 1 to 226

Z_{ik} = 1 if person i has health condition interaction k, 0 otherwise; for k=1 to M

M = the number of interactions in the model: 291 for the concurrent model and 290 for the prospective model

β_j = the model parameter for predictor X_j

θ_k = the model parameter for predictor Z_k

Both the concurrent and prospective models contain two-way interactions between health conditions. Testing all possible two-way interactions was not possible, mainly due to the high computational requirements in doing so. And so, the number of interaction effects tested in each of the concurrent and prospective models was restricted to around 300. Separately for each of the concurrent and prospective models, the 300 interactions were identified and tested using the following steps:

1. A regression model with only main effects was fitted. Parameters were estimated using ordinary least squares, predicted cost values were obtained for each person, and residuals derived.

2. For each possible two-way interaction (i.e. each combination of two conditions that could be constructed using the 226 health conditions), the number of people with that pair of health

conditions was counted. The sum of squared error (SSE) was also calculated for the people with that pair of health conditions, using the residuals from Step 1 above.

3. Pairs of conditions with a count less than 764 (the median count) for concurrent model and 706 for prospective model were filtered out.

4. Of those pairs that remained, the 300 pairs with the largest SSE (and hence with the most potential to improve model prediction overall) were identified as candidates for testing in the regression model.

Separately for each of the concurrent and prospective models, two-way interactions for those 300 identified pairs were then tested in the regression models.

After the above steps for identifying the "top" interactions, additional interactions were identified and tested for several health conditions having negative main-effect parameter estimates. The objective in identifying and testing these additional interactions was to try and provide main-effect parameter estimates that were positive in value, particularly for the ones that were significant.

Health condition P42 – Influenza and I81 – Signs, Symptoms Skin had negative parameter estimates in the concurrent model. Two-way interactions involving P42 and I81 were investigated but this did not result in a positive main effect for P42 and I81. As an alternative solution, for estimation in the regression models, P42 – Influenza was combined with P43 – Other Viral Infection and I81 – Signs, Symptoms Skin was rolled up with I04 – Acne/Rosacea.

Based on a p-value of 0.05, some of the main-effect parameter estimates were also insignificant. Predictors with insignificant main effects were removed from the model and parameters of the remaining predictors were re-estimated. These insignificant predictors are contained in Tables 22 and 23 for the concurrent and prospective models respectively.

8. Predicting health system use for select services

Table 22   Initial concurrent model – health conditions with insignificant parameter estimates

Health condition	Parameter estimate	Standard error	p-value
C06 - Disorder of external ear	13.46	11.96	0.26
C07 - Other disease middle/inner ear & mastoid	51.91	44.21	0.24
C42 - Acute ENT, upper respiratory condition (incl. benign neoplasm, croup)	4.81	6.62	0.47
H10 - Other musculoskeletal & connective tissue condition	188.15	116.82	0.11
I04 - Acne/rosacea & i81 - signs, symptoms skin (rolled up together)	19.27	11.78	0.10
I42 - Skin infection (incl. cellulitis)	4.22	11.73	0.72
P42 – Influenza & P43 - Other viral infection (rolled up together)	8.15	10.97	0.46

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

Table 23   Initial prospective model – health conditions with insignificant parameter estimates

Health condition	Parameter estimate	Standard error	p-value
D43 - Acute infectious/parasitic respiratory disease (excl. pneumonia)	393.45	278.41	0.16
F46 - Disease of appendix	37.29	58.10	0.52
J03 - Goitre	64.47	39.62	0.10
J04 - Hyperthyroidism	77.17	43.03	0.07
L04 - Endometriosis/pelvic inflammatory disease	22.75	40.82	0.58
M42 - Unsuccessful/ectopic pregnancy	46.02	23.68	0.05

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.

From the re-estimation of the remaining predictors, all parameter estimates of the main health conditions were positive and statistically significant.

The fitting of the model provides a parameter estimate for each of the significant main effects (217 for the concurrent model and 220 for the prospective model) and the two-way interactions (291 for the concurrent model and 290 for the prospective model). Since age and sex did not improve the performance of the models they were excluded; section 7.2.4 explains this further.

For a person with a specific set of health conditions, their predicted cost is the sum of the main-effect parameter estimates and the interaction-effect parameter estimates, for the specific health conditions that person has.

While main-effect parameter estimates are all non-negative in value, an interaction-effect parameter estimate can be either positive or negative. Hence it is possible that for a person with a certain combination of health conditions, the predicted cost from the model can be negative in value. For people with one or more health conditions, negative predicted concurrent cost occurs in about 0.5%, and negative predicted prospective cost occurs in about 0.05% of people.

To prevent negative predictive cost weights being assigned in the grouping methodology software, the software contains logic that the predictive cost weight assigned to a person has to be equal to or greater than the cost weight that would apply if the person had only one of the conditions.

In the estimation process, influential observations (i.e. persons) were identified and removed from the estimation data (i.e. the 70% partition). In the concurrent model there were 322 cases set as influential observations based upon their Cook's Distance values. Also in the estimation data, there were 12,992 observations with a cost value of $0, 143 cases with a cost of over $2 million, and 4,632 cases with 25 or more health conditions. These influential and outlier observations were removed from the dataset during model fitting, prior to final parameter estimation.

In the prospective model, 163 observations were set as influential based on Cook's Distance values. Outlier observations were also identified: 54 cases had a cost of over $1 million, and 3,754 cases had 25 or more health conditions. These influential observations and outlier observations were removed from the dataset during model fitting, prior to final parameter estimation. Observations with $0 in the prospective period were not considered outliers as it is expected to have cases in this sub-population that incur no cost in the prospective year.

8. Predicting health system use for select services

Evaluation of Performance

Model fitting performance review

In order to evaluate the goodness of fit of the models, the following measures were used: bias (average residual), prediction error (i.e. mean absolute residual), and coefficient of multiple determination (R^2).

Bias is defined as:

$$Bias = \sum_i \left(Y_i - \hat{Y}_i \right)/n \qquad \text{(Equation 4)}$$

where:

Y_i = cost of person i

$\hat{Y}_i$ = predicted cost of person i

Prediction error (PE) is defined as:

$$PE = \sum_i \left| Y_i - \hat{Y}_i \right|/n \quad \text{(Equation 5)}$$

where:

Y_i = cost of person i

$\hat{Y}_i$ = predicted cost of person i

Coefficient of multiple determination (R^2) is defined as:

$$R^2 = \left(1 - \frac{MSE}{MST} \right) \times 100\% = \left(1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2/n}{\sum_i (Y_i - \bar{Y})^2/n} \right) \times 100\% \qquad \text{(Equation 6)}$$

where:

MSE = mean square error = $\sum_i \left(Y_i - \hat{Y}_i \right)^2/n$

MST = mean square total = $\sum_i (Y_i - \bar{Y})^2/n$

$\bar{Y}$ = overall average cost in the estimation data used in the fitting of the model. In the overall goodness of fit statistics (Tables 24 and 25), $\bar{Y}$ is calculated as the overall average using all three provinces data combined.

Concurrent models goodness of fit

Table 24 provides goodness-of-fit statistics for the concurrent models. 37 people in the foundation data have a total concurrent cost greater than $2M in the concurrent period. These 37 cases were excluded from the calculations in Table 24.

Table 24 shows a very reasonable goodness of fit to the data overall, as indicated by prediction error and R^2 numbers. There is no substantial change in the bias, prediction error, and R^2 between the estimation and validation datasets and so no evidence of overfitting to the estimation data.

Table 24 Concurrent models goodness of fit, all 3 provinces

Model	Par-tition	Volume	Average cost ($)	Average predicted cost ($)	Bias ($)	Prediction error	Mean square total	Mean square error	R^2 (%)
Users with no health conditions	Est.	752,769	140	140	0	103	41,621	39,935	4.1%
	Val.	323,519	140	140	0	103	40,155	38,455	4.2%
Users with health conditions	Est.	13,489,310	3,381	3,383	-1	2,147	169,188,577	88,488,978	47.7%
	Val.	5,779,234	3,385	3,382	3	2,151	172,641,254	90,871,063	47.4%
Overall*	Est.	14,242,079	3,210	3,211	-1	2,039	160,774,125	83,813,980	47.9%
	Val.	6,102,753	3,213	3,210	3	2,042	164,019,948	86,055,847	47.5%

Notes
*Non-users are not included in overall numbers.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

8. Predicting health system use for select services

Prospective models goodness of fit

Table 25 provides goodness of fit statistics for the prospective models. 74 cases, each with a total cost greater than $1M in the prospective period, were excluded from the calculations in these tables. Conclusions on the prospective goodness of fit statistics are similar to that of the concurrent goodness of fit statistics: no evidence of overfitting.

Table 25   Prospective models goodness of fit, all 3 provinces

Model	Par-tition	Volume	Average cost ($)	Average predicted cost ($)	Bias ($)	Prediction error	Mean square total	Mean square error	R² (%)
Non-users	Est.	1,723,539	212	214	-1	358	8,279,965	8,261,313	0.2%
	Val.	737,953	207	214	-7	352	6,319,235	6,301,270	0.3%
Users with no health conditions	Est.	731,509	442	441	1	605	10,172,789	10,054,088	1.2%
	Val.	313,517	445	441	5	608	13,212,534	13,091,669	0.9%
Users with health conditions	Est.	13,198,276	1,698	1,696	2	1,902	69,492,322	63,053,822	9.3%
	Val.	5,654,150	1,694	1,697	-3	1,899	69,342,104	62,966,320	9.2%
Overall	Est.	15,653,324	1,476	1,474	1	1,672	60,247,658	54,544,012	9.5%
	Val.	6,705,620	1,472	1,475	-3	1,668	60,049,094	54,398,478	9.4%

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

7.2.3   Rescaling to relative values

The parameter estimates from the regression models are dollar effects and the predicted values assigned are dollar amounts. For use in the grouping software, these parameter estimates are rescaled such that they and the resulting predictions are "relative" values. This relative indicator concept is used in many of CIHI's other case mix methodologies. Relative cost weights provide an estimate of a person's health care costs, relative to a reference, or anchor value (e.g. a person with a cost weight of 1.6 is estimated to cost 1.6 times as much as the anchor value).

In the population grouping methodology, the anchor point for the concurrent cost weight is $2,845, the two-year average cost for the target population in the foundation data (includes users and non-users) over 2010-2011 and 2011-2012. The anchor point for the prospective cost

77

weight is $1,480, the one-year average cost for the target population in the foundation data in 2012-2013.

The choice of the anchor point for the population grouping methodology is somewhat arbitrary, and was chosen for intuitive convenience (e.g. if a person has a concurrent cost weight of 1.6, then it is "estimated" that a person in the foundation data costs 1.6 times as much as the "average person").

The cost weights are based on the foundation data and as such they reflect variations in costs among people in 2010-2011 to 2012-2013, from Alberta, British Columbia and Ontario data combined. For persons with a certain clinical profile (e.g. combination of health conditions), the average cost in the foundation data may not be applicable to any one Canadian jurisdiction or to other years. An assumption in using the cost weights is that the average cost of persons with one clinical profile relative to the average cost of all persons (or to persons with another clinical profile) does remain the same across jurisdictions and across years. Hence the population grouping methodology's cost weights will still be relevant to individual jurisdictions and to other years. This is a standard assumption of cost weight methodologies both within CIHI and elsewhere.

Profiling based on cost weights

Part of the value of the cost weights is their usefulness in segmenting the population. Table 26 partitions the foundation data into deciles based on the concurrent cost weight and profiles each of these deciles based on cost and clinical complexity (i.e. number of health conditions). Each decile contains one-tenth of the target population, or about 2,300,050 persons. Users with cost > $2M are excluded from the compilation of the numbers in Table 26.

The top 10% of people (decile 10) cost on average 6.2 times (i.e. 17,564 ÷ 2,840) the average cost in the overall population and consume over 61% of total health care costs. They also have about 2.5 times (7.9 ÷ 3.2) the average number of health conditions in the overall population. The average concurrent cost weight value is 1.0 by definition. Table 26 also shows the distribution of the concurrent cost weights when assigned to foundation data. Table 27 provides similar insight based on the prospective cost weight.

As can be seen in row 1 of Table 26, there are some people that have a predicted cost of less than $0, due to negative parameter estimates for some regression-model interaction effects. As mentioned in Section 7.2.2, to prevent negative cost weights being assigned in the grouping methodology software, the software contains logic that the cost weight assigned to a person has to be equal to or greater than the cost weight that would apply if the person had only one of the conditions.

8. Predicting health system use for select services

Similarly, Table 27 partitions the foundation data into deciles based on the prospective cost weight. Persons not eligible for health care on the last day of the concurrent period and the 74 users with cost > $1M are excluded from the compilation of the numbers in Table 27.

Table 26 Profiling of population based on the concurrent cost weights, all 3 provinces

Decile	Volume	Average cost	Average predicted cost	Proportion of concurrent costs	Average number of health conditions	Average age (in years)	Minimum cost weight	Maximum cost weight
1	2,300,050	123	-8	0.43%	0.6	36.3	-4.38	0.00
2	2,300,051	124	24	0.44%	0.6	29.9	0.00	0.03
3	2,300,050	298	136	1.05%	1.0	33.3	0.03	0.06
4	2,300,051	440	235	1.55%	1.8	34.8	0.06	0.12
5	2,300,051	720	487	2.54%	2.7	35.6	0.12	0.23
6	2,300,050	1,056	859	3.72%	3.2	38.6	0.23	0.38
7	2,300,051	1,527	1,385	5.38%	3.8	37.5	0.38	0.62
8	2,300,050	2,356	2,330	8.29%	4.7	46.1	0.62	1.09
9	2,300,051	4,196	4,544	14.77%	5.6	48.7	1.09	2.26
10	2,300,051	17,564	18,413	61.84%	7.9	56.1	2.26	162.37
Overall	**23,000,506**	**2,840**	**2,845**	**100.00%**	**3.2**	**39.7**	**-4.38**	**162.37**

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

Table 27 Profiling of population based on the prospective cost weights, all 3 provinces

Decile	Volume	Average cost	Average predicted cost	Proportion of concurrent costs	Average number of health conditions	Average age (in years)	Minimum cost weight	Maximum cost weight
1	2,235,894	248	124	1.68%	0.4	27.6	-6.52	0.12
2	2,235,894	344	218	2.33%	0.9	29.8	0.12	0.18
3	2,235,895	464	317	3.15%	1.3	32.7	0.18	0.25
4	2,235,894	536	436	3.64%	1.8	32.8	0.25	0.34
5	2,235,895	653	581	4.43%	2.4	33.6	0.34	0.45
6	2,235,894	823	787	5.58%	3.2	34.8	0.45	0.62
7	2,235,894	1,114	1,098	7.55%	3.8	40.5	0.62	0.89
8	2,235,895	1,534	1,608	10.40%	4.5	46.9	0.89	1.34
9	2,235,894	2,380	2,598	16.14%	5.6	53.7	1.34	2.35
10	2,235,895	6,649	6,977	45.09%	8.2	61.0	2.35	65.66
Overall	**22,358,944**	**1,474**	**1,481**	**100.00%**	**3.2**	**39.3**	**-6.52**	**65.66**

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

7.2.4 Comments on the modeling of cost weights

Model specification and use of least squares estimation

The approach used in the modeling of the cost weights is similar to the approach used by other organizations that have developed similar cost models. These organizations include the German Federal Insurance Office[41], the U.S. Centers for Medicare and Medicaid[42], and the Nuffield Trust[43]. The approaches taken by these organizations have provided a good basis for CIHI to begin developing cost weight models for the population grouping methodology. There are some differences in CIHI's approach: CIHI's approach includes separating the population into 3 indicator sub-populations: non-users, users without health conditions, and users with health conditions.

8. Predicting health system use for select services

For users with health conditions, the concurrent and prospective models (i.e. Equation 3 in Section 7.2.2) specify that the 226 health conditions and the two-way interactions between health conditions are associated with cost in an additive manner. They do not take into account the wide variation in the number of health conditions and the morbidity complexity that exist in this sub-population, and the effect that these variations have on health care costs. These simple models and the use of least squares to estimate the model parameters do not restrict the resulting parameter estimates to be non-negative, do not restrict the predicted cost for a person to be non-negative, and result in biases (e.g. under-estimation of cost for people with one or two health conditions).

As discussed in section 7.2.2, with the introduction of the two-way interactions, ordinary least squares estimation did provide positive parameter estimates for the significant main effects in the models.

Separating the population into three indicator sub-populations

In the development of the cost weight models for the alpha and beta release of the population grouping methodology, fitting of just one concurrent model and one prospective model to the entire population (i.e. combining the three indicator sub-populations) was initially investigated. These initial models included the 38 age-sex categories and the health conditions as predictors. The results were that in that concurrent model, the majority of the age-sex parameter estimates were negative, which along with negative parameter estimates for some of the health conditions, resulted in negative predicted concurrent cost for a substantial portion (approximately 38%) of the population. For the prospective model, only one age-sex parameter (female age 2 to 4) was negative, resulting in negative predicted costs for only 0.5% of the population.

As mentioned already, least squares estimation chooses parameter estimates that minimize the overall mean square error and does not place any restriction for the parameter estimates to be non-negative. It also does not restrict the range of predicted cost values to be non-negative. These negative parameter estimates are not appealing from a subject matter perspective since they provide negative predicted costs for a substantial portion of people.

Due to these negative parameter estimates and resulting negative predicted values, along with the large estimation biases that resulted for each of the three sub-populations individually, it was concluded that a single regression model did not adequately fit the data and did not provide useful predicted costs and that it would be more appropriate to split the population into the three indicator sub-populations and fit regression models separately to each of the sub-populations.

Age and sex effects

Age and sex effects are used in the prospective model for non-users, and the concurrent model and the prospective model for users without health conditions. For the alpha and beta release of the population grouping methodology, age and sex effects were tested in the models for users with health conditions. However, for users with health conditions, including age-sex effects in the models did not make any substantial increase in predictive power. More important, the age-sex parameter estimates were negative. These negative parameter estimates result in negative predicted costs for a portion of the population.

In consideration of these results, age-sex effects are not included in the cost-weight models for users with health conditions.

Two-way interaction effects in predictive cost weights

Two-way interaction effects for a small number of pairs of health conditions are included in the predictive models for the users with health conditions sub-population. The incorporation of an interaction for a specific pair of health conditions recognizes that the cost of treating one of the conditions in the pair is different depending whether or not the other condition is present. The regression model estimate for an interaction parameter can be either negative or positive. A negative estimate may be an indication that both conditions are treated simultaneously and hence there are some cost economies. A positive estimate may be an indication that a condition (or at least the person as a whole) is more complicated to treat if a second condition is present.

Negative parameter estimates for health-condition main-effects did occur in some preliminary models during the modeling process. An explanation for a negative main-effect parameter estimate may be that the vast majority of people with that condition have other conditions as well (perhaps several other conditions). The cumulative effect of the parameter estimates from those other conditions may result in an overestimate of costs for people with multiple conditions. To reduce this over-prediction and help minimize the sum of squared error, the negative parameter estimate is the best estimate for that predictor given the other predictors that are in the model.

Introduction of interactions involving the condition with that negative parameter resulted in positive parameter estimates for all health conditions, except two: P42 – Influenza and I81 – Signs, Symptoms Skin. An explanation for this ability of interactions to turn main-effect parameter estimates positive may be that the interaction effects provide an alternative way for the models to deal with the over-prediction for the multiple condition cases. As as result , the final models do not contain any negative parameter estimates for any health-condition main effects.

Interpretation of parameter estimates

The goal of the cost weight models is to provide predictions of the concurrent and prospective costs and allow segmentation of the population based on expected cost. They are not meant to be descriptive models, i.e. the objective of these models is not to measure the association between cost and each individual predictor.

There may be a tendency for users to look at the parameter estimates of individual health conditions and interpret them as the marginal, or incremental, cost of a health condition. However, these models are much too simplistic for this purpose. Indeed, CIHI's goal in developing these models is not to provide marginal cost estimates.

A feasibly small set of interactions was tested for incorporation into the models to improve the predictive power. However, there are many other confounders and interactions of the 226 health conditions that are not incorporated into these models. While the current models will continue to be reviewed and refined in future versions of the population grouping methodology, it is unlikely that models will be developed to the point where parameter estimates can be interpreted as marginal costs.

7.3 Cost weights based on the 239 health profile groups

7.3.1 Assigning cost weights

To derive the cost weights base on the 239 health profile groups, the average cost of all individuals in each group is calculated using the foundation data.

These average costs, one for each health profile group, are then rescaled to a "relative value" by dividing each one by the overall average cost of all individuals within the foundation data (referred to as the anchor point in section 7.2.3). The anchor point is $2,845 for the concurrent period and $1,480 for the prospective period.

These rescaled values can then be interpreted as comparisons to the "average person". For example, if the concurrent cost weight for a specific health profile group is 1.6, then it is "estimated" that a person assigned to that health profile group costs 1.6 times as much as the "average" person in the population.

The cost weights are based on the foundation data and as such they reflect variations in costs among people in 2010-2011 to 2012-2013, from Alberta, British Columbia and Ontario data combined. The average cost in the foundation data may not be applicable to any one Canadian jurisdiction or to other years. An assumption in using the cost weights however is that the

average cost of persons with one clinical profile relative to the average cost of all persons (or to persons with another clinical profile) does remain the same across jurisdictions and across years. Hence the population grouping methodology's cost weights will still be relevant to individual jurisdictions and to other years. This is a standard assumption of cost weight methodologies both within CIHI and elsewhere.

7.3.2 Evaluation of performance

Goodness of fit for the health profile group models was analyzed using the same criteria outlined in section 7.2.2 for the health condition classification models - bias (average residual), prediction error (i.e. mean absolute residual), and coefficient of multiple determination (R^2) were evaluated.

Concurrent models goodness of fit

Table 28 provides goodness-of-fit statistics for the overall concurrent and prospective models. Similar to previous analysis of model fit for the health condition classification in section 7.2.2, the 37 people in the concurrent period and the 74 people in the prospective period with anomalously high costs were excluded from the calculations.

Table 28 shows a very reasonable goodness of fit to the data overall, as indicated by prediction error and R^2 numbers. There is no substantial change in the bias, prediction error, and R^2 between the estimation and validation datasets and so no evidence of overfitting to the estimation data.

8. Predicting health system use for select services

Table 28 Overall concurrent and prospective models goodness of fit using health profile groups, all 3 provinces

Model	Par-tition	Volume	Average cost ($)	Average predicted cost ($)	Bias ($)	Prediction error	Mean square total	Mean square error	R² (%)
Concurrent period*	Est.	14,242,079	3,210	3,210	0	2,364	160,774,125	110,365,125	31.4
	Val.	6,102,753	3,213	3,206	7	2,367	164,019,948	113,269,059	30.9
Prospective period	Est.	15,653,324	1,476	1,476	0	1,721	60,247,658	56,070,280	6.9
	Val.	6,705,620	1,472	1,477	-5	1,717	60,049,094	55,888,764	6.9

Notes
*Non-users are not included in overall numbers.
Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

8. Predictive indicators of use of select health system services

In addition to providing insight into the overall use of the health system via cost weights, CIHI's population grouping methodology includes three predictive indicators to offer insight on health system use for particular health sectors, namely visits to primary care physicians, visits to the emergency department and the probability of admission to a long-term care facility. Each of these models uses age, sex, and the 226 health conditions as the predictor variables. For all models developed, the data was randomly split into two partitions; one partition consisted of 70% of the records and the other partition consisted of the remaining 30% of the records. The 70% partition was used in the derivation of the model (*estimation dataset*) and the 30% partition was used to evaluate the goodness of fit of the predicted number of visits (*validation dataset*).

The following sections will provide details on the methodology and overall performance for these indicators.

Predictive outputs calculated by software

There are two types of outputs provided for each of these models: the raw outputs and rescaled outputs. The software calculates the predicted number of visits or probabilities based on the foundation data – this is referred to as the 'raw outputs'.

In addition, the software provides users with an opportunity to provide the additional data into the model, which allows outputs to be rescaled, using the average values for the geographic boundaries being analyzed (i.e. the average number of PHC visits, the average number of ED visits and the average probability of entering LTC for persons ages 65 and older). This produces outputs that reflect the population health needs of the geographic area being analyzed.

8.1 Predicting number of primary care visits

The population grouping methodology includes an indicator to predict the expected number of primary care visits for each individual in the prospective year for each of the three sub-populations: users with 1 or more health conditions, users with no health conditions and non-users. Future versions of the methodology will look to expand this indicator to include predicted number of visits for the concurrent period.

8.1.1 Defining a primary care visit

For the purposes of the primary-care-visit indicator, the definition of a primary care visit includes visits to family medicine physicians for select services.

Identifying family medicine physicians

Primary care can be provided by different types of health care providers, including, but not limited to: nurses, nurse practitioners, dieticians and physicians. Within the PLPB data, the coverage of health care providers other than physicians is not comprehensive. Therefore the indicator is focused on predicting primary care visits to a family medicine physician.

The specialty of the physician providing the service is included in all PLPB data. In some instances the PLPB data will include both the latest acquired certified specialty[xii] and the payment plan specialty. The payment plan specialty may be different than the latest acquired certified specialty because it represents the area in which the physician was paid for his or her services. To maintain consistency with the definition of family medicine physician used within CIHI's national physician database (NPDB)[13], family medicine physicians were identified using the payment plan specialty and included those providers who identified their specialty as: residency, general practice, family practice, community medicine/public health or emergency medicine.

Identifying a "visit"

One visit to a physician can result in multiple claims within the PLPB data. To ensure that each visit was only counted once, claims were summarized by encrypted health card number, physician identifier and service date. For example, all claims for individual 'John Q. Public ' to physician 'Jane Q. Physician' on date X would be counted as one visit. If 'John Q. Public ' visited a second physician on date X, that would count as a separate visit.

[xii] The latest acquired specialty must be designated by the Royal College of Physicians and Surgeons of Canada, the Collège des médecins du Québec or the College of Family Physicians of Canada.

Identifying select services

Each jurisdiction has their own payment schedule (or schedule of medical benefits) and their own fee-code terminology to identify services provided by physicians and the associated payments. CIHI classifies these fee codes into meaningful categories of service delivery using the National Grouping System (NGS)[13] , which are comparable across jurisdictions. There are 121 NGS groups, examples of which include assessments, consulations, surgery, delivery, etc. Appendix D lists the NGS groups and describes them in more detail.

NGS categories identifing the various services provided were assigned to each family physician visit in the PLPB data. The NGS category together with the location of the visit were used to identify primary care visits that contained services considered in scope for the primary-care-visit indicator, using the following steps:

1. First, visits that include a major assessment or other assessment were analysed. Assessments taking place in an office or unspecified setting were identified as potential primary care visits; whereas assessments taking place in a hospital setting or other location were excluded. Visits that contained a claim for an assessment that took place in an 'unspecified location' were included as potential primary care visits because some jurisdictions do not include the location of the claim, and all claims are 'unspecified'.

   A visit that included a claim for a major assessment or other assessment within an office or unspecified location were further investigated to understand what other claims were billed as part of that visit. If a visit that contained a claim for an assessment also contained a claim for a major surgery, then the visit was not included as a primary care visit, since the assumption is that the assessment was related to the major surgery and not primary care.

2. Second, visits to family medicine physicians where there was no assessment claim, but there was a claim for psychotherapy/counselling services, minor surgeries, obstetrical services, other therapeutic services, or diagnostic services were identified as potential primary care visits.

   Claims for these visits were also further investigated to understand what other claims were billed as part of that visit. If the visit also included a claim for a major surgery, the visit was not included as a primary care visit.

3. Third, visits with only claims for consultations, hospital care days, major surgeries, special calls, surgical assistance, anesthesia and/or radiology and laboratory services were not included as a primary care visit.

The logic for identifying primary care services for this indicator is described in Figure 9 below.

8. Predicting health system use for select services

Figure 9   Identifying primary care visits
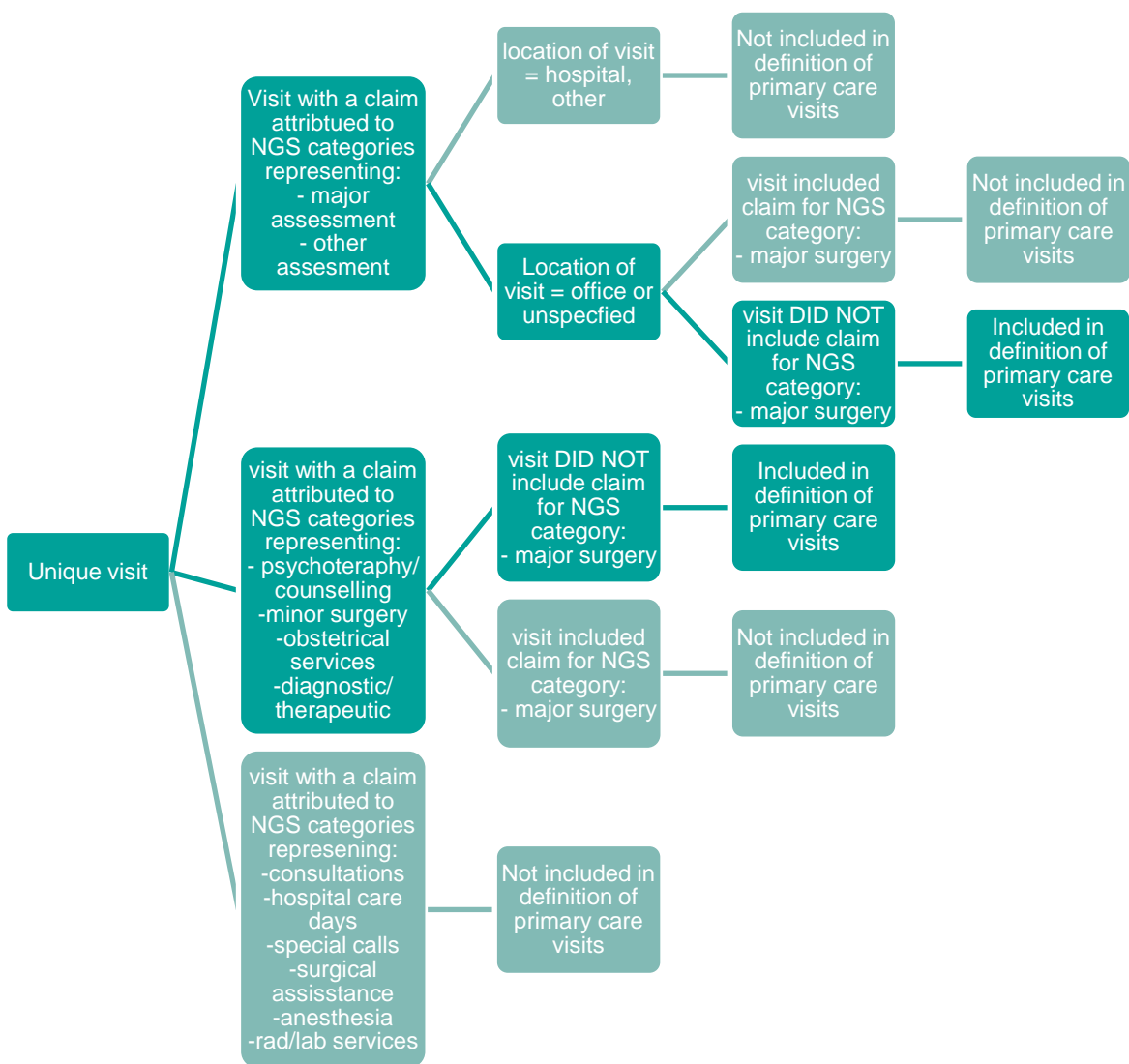
Sources
Canadian Institute for Health Information, population grouping methodology, 2016

8.1.2 Model specification – predicting primary care visits

The primary care visit indicator is based on linear regression models, with age, sex, the 226 health conditions and the most influential 2-way health condition interactions as predictor variables, and using ordinary least squares for model fitting. The statistical models were developed separately for each of the three sub-populations (non-users, users with no health conditions and users with health conditions).

Choosing the most suitable error distribution

Past experience with the cost weight indicator models for the population grouping methodology illustrated that assuming a normal distribution performed well for modeling costs. However, for the primary care visit indicator, other distributions more suitable for modeling of count data, were considered. To this end a poisson regression model was investigated and fitted. As a comparison, an alternate set of regression models that assumed a normal distribution of the response variable were also evaluated. The overall performance of the two models was similar, however the model assuming a normal distribution yielded a slightly higher R^2 value and lower bias (average residual) and prediction err0ed (i.e. mean absolute residual). In the end, the indicator was modeled assuming a normal distribution and fitted using ordinary least squares , based on performance and simplicity of interpretation of parameters.

Non-users and users without health conditions

The predicted number of primary care visits in the prospective period for both non-users and users without health conditions is based upon their age and sex. An exception to this is that any person who is not eligible for health care on the last day of the concurrent period is assigned a prospective number of primary care visits equal to zero.

The unit of observation in this model is a person (i.e. one record per person), the response variable is number of primary care visits, and the predictor variables are combinations of 19 age-sex groups.

8. Predicting health system use for select services

The regression model can be specified as:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \cdots \beta_i X_{ij} + \cdots + \beta_{38} X_{i38} \qquad \text{(Equation 7)}$$

where:

    Y_i = number of primary care visits for person *i* in the prospective period

    X_{ij} = 1 if person *i* is in age-sex group *j*, 0 otherwise

    β_j = the model parameter for predictor X_j

Cook's Distance statistic was used to identify and remove extreme influential records (outliers). From the estimation data (i.e. the 70% partition) there were 6 cases (in the non-user sub-population) set as outliers, due to having anomalous Cook's Distance values.

Users with health conditions

A person who used the health care system during the concurrent period and has one or more of the 226 health conditions is assigned a predicted number of primary care visits in the prospective period based on age, sex and health conditions identified in the concurrent period (i.e. presence or absence of each of the 226 conditions, with clinical overrides applied). The most influential 2-way interactions between health conditions were also included to account for the effect that one health condition has on the number of PC visits in the presence of another health condition. A person not eligible for health care on the last day of the concurrent period is assigned a prospective number of primary care visits equal to zero.

The final regression model, which uses age, sex, the 226 health conditions and the most influential 2-way health condition interactions as predictor variables, can be specified as:

$$Y_i = \sum_{j=1}^{38} \beta_j X_{ij} + \sum_{k=1}^{226} \gamma_k N_{ik} + \sum_{l=1}^{399} \theta_l Z_{il} \qquad \text{(Equation 8)}$$

where:

    Y_i = number of primary care visits for person *i* in the prospective period

    X_{ij} = 1 if person *i* is in age-sex group *j*, 0 otherwise

    β_j = the model parameter for predictor X_j

    N_{ik} = 1 if person *i* has health condition k, 0 otherwise; for k = 1 to 226

    γ_k = the model parameter for predictor N_k

    Z_{il} = 1 if person *i* has health condition interaction *l*, 0 otherwise; for *l* =1 to 399

    θ_l = the model parameter for predictor Z_l

In order to maintain a manageable number of interactions only two-way interactions (i.e. interaction between two health conditions) were considered for inclusion in the model.

Testing all possible two-way interactions was not possible, mainly due to the high computational requirements. And so, the number of interactions tested was restricted to 300. The 300 interactions were identified and tested using the following steps:

1. A regression model with only main effect variables was fitted. Parameters were estimated using least squares. Predicted number of primary care visits in the prospective year were obtained for each person, and residuals derived.

2. The majority of users with health conditions have between 1 and 15 health conditions, therefore two-way interactions that occur in individuals with 2 to 15 conditions were identified. The number of people with that pair of health conditions was counted. The sum of squared error (SSE) was also calculated for the people with that pair of health conditions, using the residuals from Step 1 above.

3. Pairs of conditions with a count less than 715 (the median count) were filtered out.

4. Of those pairs that remained, the 300 pairs with the largest SSE (and hence with the most potential to improve model prediction overall) were identified as candidates for testing in the regression model.

After the above steps for identifying the "top 300" interactions, 47 health conditions had negative parameter estimates. After consultation with clinical specialists we identified the health conditions most likely to require follow up or monitoring by a family physician in the prospective year. These conditions are expected to have a positive rather than negative effect. Therefore, additional interactions were identified and tested to turn these negative health condition parameter estimates into positive in values.

Based on a p-value of 0.05, some of the main-effect parameter estimates were also insignificant. Predictors with insignificant main effects were removed from the model and parameters of the remaining predictors were re-estimated.

The fitting of the model provides a parameter estimate for each of the age-sex groups, the 203 significant health conditions and the 399 healthcondition two-way interactions. Each individual's predicted number of primary care visits in the prospective period is the sum of the parameter estimates for their age and sex group, their specific health conditions and any applicable 2-way health condition interactions.

The final model does include some health condition main-effect parameters and interaction-effect parameters with negative estimates, thus the model could predict a negative number of primary care visits for an individual with a certain combination of health conditions. This occurs in 0.07% of persons, and for these persons, the software includes logic that assigns these individuals a predicted number of primary care visits equal to zero.

In the estimation process, influential observations (i.e. persons) were identified and removed from the estimation data (i.e. the 70% partition). There were 4 cases set as influential observations based upon their Cook's Distance values. These influential observations were removed from the dataset during model fitting, prior to final parameter estimation. Also in the estimation data, there were 3,754 observations with 25 or more health conditions; these outlier observations were also removed from the dataset during model fitting,.

8.1.3 Evaluation of performance – predicting primary care visits

Goodness of fit was analyzed using the same criteria outlined in section 7.2.2 for the health condition classification models - bias (average residual), prediction error (i.e. mean absolute residual), and coefficient of multiple determination (R^2) were evaluated.

Prospective model goodness of fit

Table 29 provides goodness-of-fit statistics for the model predicting prospective number of primary care visits. Table 29 shows a very reasonable goodness of fit to the data overall, as indicated by prediction error and R^2 numbers. There is no substantial change in the bias, prediction error, and R^2 between the estimation and validation datasets and so no evidence of overfitting to the estimation data.

Table 29 Predicting number of primary care visits – prospective period, goodness of fit, all 3 provinces

Model	Par-tition	Volume	Average number of PHC visits	Average predicted number of PHC visits	Bias	Prediction error	Mean square total	Mean square error	R^2 (%)
Non-users	Est.	1,723,540	0.64	0.64	0	1.01	4.18	4.15	0.59
	Val.	737,953	0.64	0.64	0	1.01	3.92	3.90	0.59
Users with no health conditions	Est.	731,509	1.71	1.71	0	1.76	7.75	7.46	3.74
	Val.	313,517	1.71	1.71	0	1.76	7.92	7.63	3.63
Users with health conditions	Est.	13,198,330	5.05	5.05	0	3.40	46.29	35.50	23.31
	Val.	5,654,169	5.05	5.05	0	3.40	46.57	35.80	23.14
Overall	Est.	15,653,379	4.41	4.41	0	3.05	42.09	30.45	27.66
	Val.	6,705,639	4.41	4.41	0	3.05	42.32	30.68	27.49

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.

Patient-Level Physician Billing Data, Alberta Health Services.

Alberta Health Registration Files, Alberta Health.

Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.

B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

We also looked at the overall number of predicted primary care visits by number of health conditions to see how well the predicted values aligned with the data. As shown in Table 30, the predicted average number of visits aligns very closely with the actual average number of visits for all health condition count cohorts.

8. Predicting health system use for select services

Table 30 Average number of primary care visits by health condition count – prospective period, goodness of fit, all 3 provinces

Number of health conditions	Population count	Actual average number of primary care visits	Predicted average number of primary care visits
1	3,801,101	2.4	2.3
2	3,653,670	3.2	3.2
3	3,155,942	4.1	4.1
4	2,475,428	5.1	5.2
5	1,812,471	6.2	6.2
6	1,271,204	7.3	7.4
7	865,239	8.5	8.5
8	581,106	9.7	9.6
9	389,078	10.8	10.7
10 or more	847,260	13.1	13.0
Total	**18,852,499**	**5.1**	**5.1**

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.

Patient-Level Physician Billing Data, Alberta Health Services.

Alberta Health Registration Files, Alberta Health.

Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.

B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

8.2 Predicting number of visits to an emergency department

The population grouping methodology includes an indicator to predict the number of expected visits to an emergency department for each individual in the concurrent and prospective time periods for each of the three sub-populations: users with one or more health conditions, users with no health conditions and non-users.

8.2.1   Defining an emergency department visit

Emergency department visits were identified in the foundation data by the presence of an ED abstract in the National Ambulatory Care Reporting System (NACRS ambulatory care type code = 10, 11 or 12). Of the three jurisdictions in the foundation data, NACRS data is only available for Alberta and Ontario. To this end, modeling concurrent and prospective number of ED visits is built on data from these two jurisdictions which includes 15,146,734 persons.

8.2.2   Model specification – predicting ED visits

Number of ED visits in the prospective period are assigned using a 2 step model:

4. Step 1 involves a logistic regression to calculate the probability of visiting the ED. This model was built using all individuals within the dataset.

5. Step 2 involves a least squares linear regression models to predict the number of visits to the ED. This model was developed using a subsample of the population who had visited the ED at least once during the respective research period. For example, for the prospective model, the sample would include those who visited the ED in the prospective period; while the concurrent model was built using those who visited the ED in the concurrent period.

The outputs from the two stages are then multiplied together to calculate the predicted number of ED visits for each individual.

For both steps, the predictor variables included age, sex and the 226 health conditions. The statistical models used were developed separately for each of the three sub-populations (non-users, users with no health conditions and users with health conditions).

Choosing the most appropriate regression technique

Approximately 25% of individuals visit the ED in the prospective period. Given the high proportion of non-events (i.e. individuals not visiting the ED), consideration was given to the best way to develop a model that can identify events and best predict the number of events. To

this end, both a 2 step model (as described above) and a 1 step least squares linear regression model were evaluated. While the prediction error and R^2 for the two models was similar, the 1 step model resulted in a high proportion of negative predicted values. Just over 10% of individuals were assigned a negative predicted value in the concurrent period using the 1 step model, compared to 0.04% using the 2 step model. In the prospective period, the 1 step model predicted negative visit counts for 0.69%, compared to 0.00% for the 2 step model.

Based on the overall performance of the two models, a 2 step model was identified as the preferred option for modeling predicted number of ED visits in the concurrent and prospective periods.

Non-users and users without health conditions

Non-users (a person who did not use the health care system in the concurrent period) is assigned a predicted number of ED visits equal to zero for the concurrent period, and a prospective predicted number of ED visits based upon their age and sex.

Users without health conditions are assigned a concurrent and prospective predicted number of ED visits based upon their age and sex.

An exception to this is that any person who is not eligible for health care on the last day of the concurrent period is assigned a prospective number of ED visits equal to zero.

The unit of observation in this model is the person (i.e. one record per person), the response variable is number of ED visits, and the predictor variables are combinations of 38 age-sex groups.

The model for the concurrent and prospective periods can be specified as:

$$Y_i = \frac{\prod_{j=1}^{38} e^{\beta_j X_{ij}}}{1 + \left(\prod_{j=1}^{38} e^{\beta_j X_{ij}}\right)} * \sum_{j=1}^{38} \beta\beta_m XX_{im} \qquad \text{(Equation 9)}$$

where:

Y_i = number of ED visits for person i in the prospective period

X_{ij} = 1 if person i is in age-sex group j, 0 otherwise

β_j = the model parameter for predictor X_j

XX_{im} = 1 if person i is in age-sex group m, 0 otherwise

$\beta\beta_m$ = the model parameter for predictor XX_m

This model assigns a predicted number of ED to each person, which are estimates of the average number of ED visits of each of the 38 age-sex categories.

In step 2 of the model, least squares estimation method was used so as to allow the use of the Cook's Distance statistic to identify and remove influential records. Influential records identified and removed from the estimation data included 7 cases (all from the users with no health conditions sub-population) in the model for the concurrent period and 31 cases (17 from the non-user sub-population and 14 from the users with no health conditions sub-population) in the model for the prospective period.

Users with health conditions

A person who used the health care system during the concurrent period and has one or more of the 226 health conditions is assigned a concurrent and prospective predicted number of ED visits based on their age-sex group and knowledge of which of the 226 health conditions that they had during the concurrent period. A person who is not eligible for health care on the last day of the concurrent period is assigned a prospective number of ED visits equal to zero.

The predictor variables in the models are the 38 age-sex groups and the 226 health conditions, with clinical overrides applied.

8. Predicting health system use for select services

The model for the concurrent and prospective periods can be specified as:

$$Y_i = \frac{\prod_{j=1}^{38} e^{\beta_j X_{ij}} * \prod_{k=1}^{226} e^{\Upsilon_k N_{ik}}}{1 + \left(\prod_{j=1}^{38} e^{\beta_j X_{ij}} * \prod_{k=1}^{226} e^{\Upsilon_k N_{ik}}\right)} * \sum_{j=1}^{38} \beta\beta_m XX_{im} + \sum_{n=1}^{226} \Upsilon\Upsilon_n NN_{in} \qquad \text{(Equation 10)}$$

where:

Y_i = number of ED visits for person i in the prospective period

X_{ij} = 1 if person i is in age-sex group j, 0 otherwise

β_j = the model parameter for predictor X_j

N_{ik} = 1 if person i has health condition k, 0 otherwise; for k = 1 to 226

Υ_k = the model parameter for predictor N_k

XX_{im} = 1 if person i is in age-sex group m, 0 otherwise

$\beta\beta_m$ = the model parameter for predictor XX_m

NN_{in} = 1 if person i has health condition n, 0 otherwise; for n = 1 to 226

$\Upsilon\Upsilon_n$ = the model parameter for predictor NN_n

Based on a p-value of 0.05, 26 of the main-effect health condition parameter estimates were insignificant. Predictors with insignificant main effects were removed from the model and parameters of the remaining predictors were re-estimated.

The final model does include some health condition main-effect parameters with negative estimates, thus the model could predict a negative number of ED visits for an individual with a certain combination of health conditions. This occurs in less than 0.00% of cases. For these cases, the software includes logic that assigns these individuals a predicted number of ED visits equal to zero.

In the estimation process, influential observations (i.e. persons) were identified and removed from the estimation data (i.e. the 70% partition).

- For the model predicting ED visits in the concurrent period, there were 49 cases set as influential observations based upon their Cook's Distance values and 9,905 observations with 25 or more health conditions.

- For the model predicting ED visits in the prospective period, there were 38 cases set as influential observations based upon their Cook's Distance values and 9,905 observations with 25 or more health conditions.

These influential observations were removed from the dataset during model fitting, prior to final parameter estimation.

8.2.3 Evaluation of performance – predicting ED visits

Goodness of fit was analyzed using the same criteria outlined in section 7.2.2 for the health condition classification models - bias (average residual), prediction error (i.e. mean absolute residual), and coefficient of multiple determination (R^2) were evaluated.

Tables 31 and 32 provides goodness-of-fit statistics for the model predicting concurrent and prospective number of ED visits. Tables 31 and 32 show a very reasonable goodness of fit to the data overall, as indicated by prediction error and R^2 numbers. There is no substantial change in the bias, prediction error, and R^2 between the estimation and validation datasets and so no evidence of overfitting to the estimation data.

Table 31 Predicting number of emergency department visits – concurrent period, goodness of fit, AB and ON

Model	Par-tition	Volume	Average number of ED visits	Average predicted number of ED visits	Bias	Prediction error	Mean square total	Mean square error	R^2 (%)
Users with no health conditions	Est.	626,619	0.09	0.09	0.00	0.16	0.11	0.11	1.2%
	Val.	268,401	0.09	0.09	0.00	0.16	0.12	0.12	1.1%
Users with health conditions	Est.	10,602,994	0.97	0.97	0.00	0.91	4.92	3.70	24.9%
	Val.	4,543,740	0.97	0.97	0.00	0.91	4.82	3.59	25.4%
Overall	Est.	11,229,613	0.92	0.92	0.00	0.87	4.70	3.50	25.5%
	Val.	4,812,141	0.92	0.92	0.00	0.87	4.60	3.40	26.0%

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

Table 32 Predicting number of emergency department visits – prospective period, goodness of fit, AB and ON

Model	Par-tition	Volume	Average number of ED visits	Average predicted number of ED visits	Bias	Prediction error	Mean square total	Mean square error	R^2 (%)
Non-users	Est.	1,509,358	0.10	0.10	0.00	0.18	0.21	0.21	0.5%
	Val.	646,741	0.10	0.10	0.00	0.18	0.21	0.20	0.5%
Users with no health conditions	Est.	626,619	0.19	0.19	0.00	0.33	0.40	0.40	0.2%
	Val.	268,401	0.19	0.19	0.00	0.33	0.39	0.39	0.2%
Users with health conditions	Est.	10,602,994	0.47	0.47	0.00	0.65	1.68	1.55	8.0%
	Val.	4,543,740	0.47	0.47	0.00	0.65	1.68	1.55	8.1%
Overall	Est.	12,738,971	0.41	0.41	0.00	0.58	1.46	1.33	8.8%
	Val.	5,458,882	0.41	0.41	0.00	0.58	1.46	1.33	8.8%

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

8.3 Predicting probability of admission to long-term care for persons ages 65 and older

The population grouping methodology includes an indicator to predict the probability of persons ages 65 and older entering long-term care in the prospective period. The purpose of this model is not to classify individuals as 'likely' or 'unlikely' to be admitted to long-term care in the prospective year; but to identify a cohort of individuals who are more likely than others to enter long-term care.

8.3.1 Defining admission to long-term care

Admission to long-term care was identified in the foundation data by the presence of a RAI-MDS 2.0 form completed for a LTC facility in CCRS. Those admitted to a CCC bed in Ontario were excluded from this definition. Additionally, admissions to LTC were only predicted for individuals who were <u>not</u> in LTC at any point in the concurrent period, and who were eligible for health care on the last day of the concurrent period.

Model focused on those ages 65 or older

Analysis of the those in LTC in the foundation data indicated that 94% of LTC residents were 65 years or older. To this end, models were tested based on the entire age population and those ages 65 and older. Models focusing on those ages 65 and older outperformed those based on all age groups, both at predicting admissions at the individual level and on identifying a cohort most likely to enter LTC in the prospective period. To this end, CIHI's population grouping methodology model on predicting admissions to LTC has been developed for those ages 65 or older at the end of the concurrent period.

8.3.2 Model specification – probability of admission to LTC

A logistic regression model was used to predict the probability of entering LTC for those ages 65 and older.

Modeling rare events

1.17% of individuals (36,031 out of 3,092,011) entered LTC in the prospective period. Due to the small number of 'events' (i.e. admission to LTC), model performance was optimized by creating one model for all sub-populations together, and using the presence of one or more health conditions as a predictor variable.

As discussed above, a very small percentage of persons ages 65 and older will enter LTC in a given year (in this case, the prospective year). Traditional statistical techniques can often

8. Predicting health system use for select services

underestimate the probability of rare events. To this end, sampling techniques can be applied, such as sampling all 'events' and only a small proportion of 'non-event's to create a more optimal dataset for logistic regression analysis. This is known as 'choice-based' or 'endogenous stratified' sampling[44].

As illustrated in Figure 10, we applied a 1:3 ratio of 'events' to 'non-events' to the estimation dataset (70% partition). To this end, all 'events' (i.e. admissions to LTC in prospective period) and 3% of non-events, which were randomly selected from the 70% partition, resulting in 100,592 cases, which were used as the final estimation dataset. The estimation dataset based on 100,592 cases was weighted to be representative of the full partition dataset (the original 2,164,227 cases).

Figure 10 Applying a choice-based sampling methodology to optimize logistic regression model for predicting probability of admission to LTC
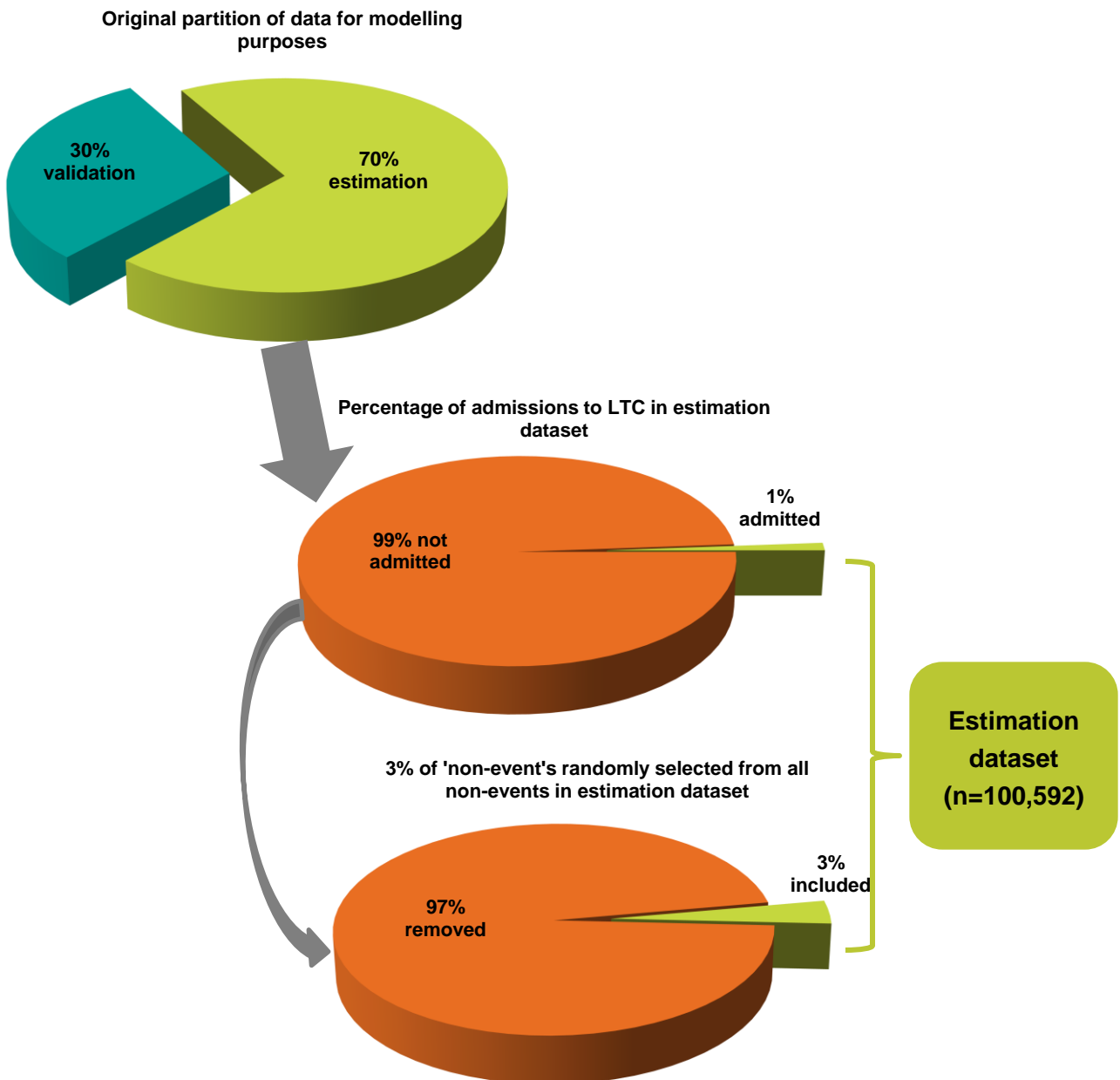
Original partition of data for modelling purposes

30%
validation

70%
estimation

Percentage of admissions to LTC in estimation dataset

1%
admitted

99% not
admitted

3% of 'non-event's randomly selected from all non-events in estimation dataset

3%
included

97%
removed

Estimation dataset (n=100,592)

Source
Canadian Institute for Health Information, population grouping methodology, 2016

8. Predicting health system use for select services

Predictor variables included in probability of admission to LTC model

Individuals ages 65 and older who were not in LTC during the concurrent period are assigned a probability of being admitted to LTC in the prospective period. Predicted probability of admission to LTC is based on the individuals age-sex group, whether or not they have 1 or more health conditions, the health conditions identified in the concurrent period and associated 2-way health condition interactions.

Persons not eligible for health care on the last day of the concurrent period are not assigned a probability of entering LTC in the prospective period. The model can be specified as:

$$Y_i = \frac{\prod_{j=1}^{38} e^{\beta_j X_{ij}} * \prod_{k=1}^{5} e^{\Upsilon_k N_{ik}} * \prod_{z=1}^{2} e^{\theta_l Z_l} * e^{\Phi}}{1 + \left(\prod_{j=1}^{38} e^{\beta_j X_{ij}} * \prod_{k=1}^{226} e^{\Upsilon_k N_{ik}}\right)}$$ (Equation 11)

where:

Y_i = probability of admission to LTC for person *i* in the prospective period

X_{ij} = 1 if person *i* is in age-sex group *j*, 0 otherwise

β_j = the model parameter for predictor X_j

N_{ik} = 1 if person *i* has health condition k, 0 otherwise; for k = 1 to 5

Υ_k = the model parameter for predictor N_k

Z_{il} = 1 if person *i* has health condition interaction *l*, 0 otherwise; for *l* =1 to 2

θ_l = the model parameter for predictor Z_l

Φ = the model parameter for those with 1 or more health conditions

Health conditions and associated interactions included – probability of admission to LTC model

Based on a p-value of 0.05, the majority of health conditions were insignificant. Health conditions with insignificant estimates were removed, leaving 5 health conditions:

- A02 – Hereditary/degenerative condition of the nervous system
- A04 – Parkinson's disease/parkinsonism
- Q01 – Dementia (incl. Alzheimer's)
- Q03 – Mental disorder resulting from brain injury or other illness
- Q81 – Delirium

All possible 2-way interactions between significant health conditions were tested – two of which were significant:

- Q01 – Dementia (incl. Alzheimer's) & A02 – Hereditary/degenerative condition of the nervous system
- Q01 – Dementia (incl. Alzheimer's) & A04 – Parkinson's disease/parkinsonism

8.3.3 Evaluation of performance – probability of admission to LTC

A standard output on model performance for logistic regression models is the Hosmer-Lemeshow goodness-of-fit statistic. This statistic is designed to evaluate how well the model correctly predicts admissions to long-term care at the individual level. The Hosmer-Lemeshow goodness-of-fit statistic is represented by:

$$X_{HL}(g-2) \ = \ \sum_{i=1}^{g} \frac{(O_i - N_i \pi_i)^2}{N_i \pi_i (1 - \pi_i)}$$ (Equation 12)

where:

O = number of observed events at level i

N = number of observations events at level i

π = average probability of event at level i

g = total number of levels (ventiles)

This statistic measures the lack of fit, so relatively low Chi Square values and a high p value are indicators of good fit. The Hosmer-Lemeshow goodness-of-fit statistics for our model indicate that it did not perform well at the individual level. The Chi Square value from this model was 48,142.8 and the p <= 0.000. That is to say, when you sort the probabilities calculated for each individual, those predicted to be admitted to LTC are the 1.1653% of individuals with the highest probabilities[xiii]. As illustrated in Table 33, of those with the highest 1.1653% of predicted probabilities, only 18% (6,540 out of 36,031) were actually in LTC in the prospective period.

[xiii] The top 1.1653% was chosen to match the actual percentage of individuals in LTC in the dataset (as identified in section 8.3.2.

Table 33 Comparison between observed and predicted admissions to LTC –
prospective period, all 3 provinces

		Observed LTC admissions		
		No	Yes	Total
Those predicted to be admitted to LTC (highest 1.1653% of probabilities)	**No**	3,026,489	29,491	3,055,980
	Yes	29,491	**6,540**	36,031
	Total	3,055,980	36,031	3,092,011

Sources
Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services.

As mentioned at the onset of this section, the purpose of this model is not to classify people as 'likely' or 'unlikely' to enter LTC in the prospective year, but to identify a cohort of individuals most likely to enter LTC. To this end, individuals with the highest probability of entering LTC could be targeted and programs could be directed to their specific health care needs.

One way to determine how well the model is able to identify those most likely to enter LTC is to sort individuals into ventiles to determine to what extent those in the top ventile represent the population actually entering LTC in the prospective year. As illustrated in Table 34, the top ventile (the 5% with the highest predicted probabilities of entering LTC in the prospective period) includes 49% (17,622 ÷ 36,031) of persons who actually entered LTC in the prospective year.

Additionally, the top ventile includes a considerably higher proportion of prospective LTC admissions than what is present in the total population. If health care decision makers, planners or providers want to target those most likely to be admitted to LTC, they would reach a considerably higher proportion of them if they focused on the top ventile (11.4% compared to 1.17% in the general population).

Table 34 Predicted probabilities of admission into LTC, sorted into ventiles – prospective period, all 3 provinces

Ventile	Number of cases	Number in LTC	Predicted number in LTC	Percentage in LTC	Percentage predicted in LTC
1	154,600	140	30	0.1%	0.0%
2	154,600	171	77	0.1%	0.0%
3	154,600	145	77	0.1%	0.0%
4	154,600	148	77	0.1%	0.0%
5	154,600	158	87	0.1%	0.1%
6	154,600	156	87	0.1%	0.1%
7	154,600	248	96	0.2%	0.1%
8	154,600	293	168	0.2%	0.1%
9	154,600	319	169	0.2%	0.1%
10	154,600	362	185	0.2%	0.1%
11	154,600	339	185	0.2%	0.1%
12	154,600	617	277	0.4%	0.2%
13	154,600	654	334	0.4%	0.2%
14	154,600	860	385	0.6%	0.2%
15	154,600	932	399	0.6%	0.3%
16	154,600	1,536	578	1.0%	0.4%
17	154,600	1,945	718	1.3%	0.5%
18	154,600	3,422	929	2.2%	0.6%
19	154,600	5,964	1,442	3.9%	0.9%
20	154,611	17,622	6,517	11.4%	4.2%
Overall	**3,092,011**	**36,031**	**12,814**	**1.17%**	**0.41%**

Sources

Discharge Abstract Database / National Ambulatory Care Reporting System / Continuing Care Reporting System / Ontario Mental Health Reporting System, Canadian Institute for Health Information.
Patient-Level Physician Billing Data, Alberta Health Services.
Alberta Health Registration Files, Alberta Health.
Registered Persons Data / Patient-Level Physician Billing Data, Ontario Ministry of Health and Long-Term Care.
B.C. Health Ideas Client Roster / Patient-Level Physician Billing Data, B.C. Ministry of Health Services

9. Future iterations of the population grouping methodology

9.1 Additional data

Currently, CIHI does not have complete clinical and cost data for all health sectors. Over the time, with future releases, CIHI will work to add in missing cost data to accompany existing clinical data: (i.e.: Ontario mental health inpatient stays, British Columbia emergency department and hospital clinic visits and LTC and CCC stays for all three jurisdictions).

In addition to adding in missing cost data for existing data sources, CIHI will look at the feasibility of including functional status information from the OMHRS data, as this could differentiate health care need beyond what is achievable by looking at health conditions alone.

Additionally, over time, CIHI will work towards including other sources of clinical and cost data, such as home care, rehabilitation and prescription drug data; as well as data from other jurisdictions.

9.2 Additional predictive indicators

The predictive indicators included in version 1.0 of the population grouping software represent the indicators identified by CIHI and our expert panel as being interesting indicators for health system planning and decisions making. However, it is by no means an exhaustive list of useful indicators for monitoring population health, disease profiling and predicting health care utilization. Over time, CIHI will incorporate additional predictive indicators to provide health system planners and policy-makers with additional real-time evidence to support decision-making.

Appendix A – Health conditions, PLPB tagging rules and clinical overrides

Table A1   Health conditions, PLPB tagging rules and clinical overrides

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
A01:Congenital malformation of the nervous system (incl. spina bifida)	2	A84, S81
A02:Hereditary/degenerative condition of the nervous system	2	A84, S81
A03:Muscular dystrophy & other myopathy	2	A84, H81, S81
A04:Parkinson's disease/parkinsonism	2	A02, A84, H81, S81
A05:Multiple sclerosis & other demyelinating disease of cns	2	A02, A82, A84, H81, S81
A06:Cerebral palsy	2	A02, A07, A84, H81, Q10, S81
A07:Paralytic syndrome/spinal cord injury	2	A08, A42, A84, H06, H42, H81, H82, S81
A08:Disorder of the peripheral nervous system (incl. Carpal tunnel, bells palsy)	2	A84, S81
A09:Epilepsy	2	A82, A84, S81
A10:Other cerebral and spinal disease/disorder	2	A82, A84, H42, S81
A11:Other disease of the nervous system	2	A84, S81
A41:Stroke	0	A42, A43, A84, S81
A42:Cerebrovascular disorder (excl. stroke)	1	A43, A84, S81
A43:Transient ischemic attack	1	A84, S81
A44:Intracranial injury	2	A08, A46, A81, A82, A83, A84, C08, H81, Q81, S45, S81
A45:Skull fracture	2	A08, A46, A81, A82, A83, A84, C08, H81, Q81, S45, S81
A46:Concussion	1	A83, A84, C08, H81, Q81, S45, S81
A47:Infection/inflammation nervous system	1	A82, A83, A84, P43, P44, P45, S81

Appendix A

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
A48:Nerve injury	1	A84, S48, S81
A81:Coma unspecified	2	A84, Q81, S81
A82:Seizure	2	A84, S81
A83:Migraine/other headache	2	A84, S81
A84:Neurological, sensory, cognitive signs & symptoms	1	S81
B01:Cataract/lens disorder	2	B05, B07, S81
B02:Retinopathy	2	B05, B07, B41, S81
B03:Strabismus	1	B05, B07, S81
B04:Disorder of the eyelids & lacrimal system	1	B07, S81
B05:Vision impairment (incl. blindness)	2	
B06:Glaucoma	2	B05, B07, S81
B07:Other eye diseases & disorders	2	S81
B41:Retinal break/detachment	2	B05, S81
B42:Infection/inflammation eye	1	B07, S81
B43:Injury of eye and periocular area	1	S81
C01:Congenital/acquired malformation of ear/nose/throat	2	
C02:Disease of oral cavity, salivary gland, or jaw	2	
C03:Dental caries/gingivitis	1	S81
C04:Chronic upper respiratory condition (incl. tonsillitis)	1	C42, S81
C05:Hearing loss (incl. deafness)	2	
C06:Disorder of external ear	1	C41, S81
C07:Other disease middle/inner ear & mastoid	2	S81
C08:Meniere's/dizziness/vertigo	1	C42
C41:Otitis media	1	C05, C07, C42
C42:Acute ent, upper respiratory condition (incl. benign neoplasm, croup)	1	S81

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
D01:Pleural disease (incl. pleural effusion)	1	D44, D81, S81
D02:Congenital disorder of the respiratory system	2	D44, D81, S81
D03:Chronic obstructive pulmonary disease	2	D06, D43, D44, D81, S81
D04:Pulmonary hypertension	2	E10, D81, S81
D05:Other chronic lung disease	2	D81, S81
D06:Asthma	2	C42, D81, S81
D41:Respiratory failure	0	C42, D81, S81
D42:Pneumonia	1	C42, D43, D44, D81, S81
D43:Acute infectious/parasitic respiratory disease (excl. pneumonia)	1	C42, D81, S81
D44:Acute and other respiratory diseases & disorders	1	D81, S81
D81:Signs, symptoms respiratory system	1	S81
E01:Heart failure	2	E03, E06, E10, E82, S81
E02:Malformation of the cardiovascular system	2	E06, E82, S81
E03:Cardiac valve disease	2	E06, E82, S81
E04:Coronary artery disease	2	E06, E82, J09, S81
E05:Arrhythmia	2	E06, E82, S81
E06:Other heart disease	2	E82, S81
E07:Peripheral venous disease/ phlebitis/thrombophlebitis/DVT	2	E08, E09, E82, S81
E08:Uncomplicated varicose veins	2	E82, S81
E09:Other vascular system disease	2	E82, S81
E10:Hypertension	2	E82, S81
E11:Aortic aneurysm	2	E82, S81
E12:Peripheral artery disease	2	E82, S81
E41:Acute myocardial infarction/shock/arrest	0	E04, E05, E06, E10, E43, E82, J09, S81
E43:Unstable angina	2	E04, E06, E10, E82, S81

Appendix A

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
E82:Signs, symptoms cardiovascular system	1	S81
F01:Congenital malformation of the digestive & hepatobiliary system	2	F81, F82, S81
F02:Gastroesophageal reflux disease	2	F81, F82, S81
F03:Other disease of the esophagus	2	F02, F81, F82, S81
F04:Gastritis & duodenitis	1	F02, F81, F82, S81
F05:Hernia	2	F81, F82, S81
F06:Inflammatory bowel (incl. crohn's, ulcerative colitis)	2	F08, F09, F11, F81, F82, S81
F07:Vascular disorder of intestine (incl. ischemic bowel)	2	F81, F82, S81
F08:Diverticulitis	2	F81, F82, S81
F09:Hemorrhoids	2	F11, F42, F81, S81
F10:Benign neoplasm/polyp colon, rectum, anus	2	F09, F42, F81, S81
F11:Other rectal/anal disorder	2	F81, S81
F12:Other disease/disorder of the digestive system	2	F81, F82, S81
F41:Acute gastrointestinal infection	1	F81, F82, S81
F42:Unspecified gastrointestinal hemorrhage	1	O02, F81, F82, S81
F43:Gastrointestinal obstruction	1	F05, F11, F81, F82, S81
F44:Peptic ulcer	2	F04, F81, F82, S81
F45:Peritoneal disease (incl. peritonitis)	1	F81, F82, S81
F46:Disease of appendix	1	F81, F82, S81
F81:Signs, symptoms digestive & hepatobiliary system	1	S81
F82:Abdominal pain	1	S81
G01:Cirrhosis	2	F82, G02, G04, S81
G02:Chronic liver disease (incl. hepatitis)	2	F82, G04, S81
G03:Pancreatic disease	2	G05, F81, F82, H82, S81
G04:Acute liver disease & other hepatic disorders	1	F82, S81

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
G05:Diseases of gallbladder & biliary tract	2	F82, S81
H01:Rheumatoid & other inflammatory arthropathy (excl. gout)	2	H02, H10, H81, H82, S81
H02:Osteoarthritis	2	H10, H81, H82, S81
H03:Crystal arthropathy (incl. gout)	1	H81, H82, S81
H04:Osteoporosis	2	H10, H81, H82, S81
H05:Other bone disease	2	H81, H82, S81
H06:Vertebral/disk & other disease of back	2	H81, H82, S81
H07:Systemic connective tissue disorder (incl. lupus, scleroderma)	2	H10, H81, H82, S81
H08:Musculoskeletal malformation (excl. spine)	2	H10, H81, S81
H09:Myositis and soft tissue disorder (incl. muscle inflammation)	1	H81, H82, S81
H10:Other musculoskeletal & connective tissue condition	2	H81, H82, S81
H41:Traumatic amputation arm/hand/leg/foot	1	H44, H81, S45, S48, S81
H42:Fracture/dislocation vertebrae, pelvis	1	H44, H81, H82, S45, S48, S81
H43:Fracture femur	1	H44, H81, S45, S48, S81
H44:Other fracture/dislocation	1	H81, S45, S48, S81
H45:Benign/unspecified musculoskeletal neoplasm	1	H81, S81
H46:Joint/tendon disorder and injury (incl. pain, sprain, strain)	1	H81, S45, S81
H47:Musculoskeletal infections	1	H81, S81
H81:Neuromuscular signs & symptoms	1	S81
H82:Back pain	1	S81
I01:Autoimmune skin disorder	2	I03, I04, I09, I81, S81
I02:Papulosquamous disorder/ psoriasis	2	I03, I04, I09, I81, S81
I03:Eczema/dermatitis/hives	2	I04, I09, I81, S81

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
I04:Acne/rosacea	2	I09, I81, S81
I05:Skin ulcer (incl. decubitus)	2	I09, I81, S81
I06:Benign skin neoplasm	1	I09, I81, S81
I07:Benign disorder of breast	1	I09, I81, S81
I08:Disease of hair/nail/sweat glands	1	I09, I81, S81
I09:Other condition of skin/subcutaneous tissue	1	I81, S81
I41:Serious burn/frostbite	1	I43, I81, S81
I42:Skin infection (incl. cellulitis)	1	I81, S81
I43:Superficial skin injury/contusion/non-serious burn	1	I81, S81
I81:Signs, symptoms skin	1	S81
J01:Cystic fibrosis	2	D05, D43, D44, D81, F81, S81
J02:Diabetes mellitus	2	J12, K81, S81
J03:Goitre	2	J04, J05, S81
J04:Hyperthyroidism	2	S81
J05:Hypothyroidism	2	J10, S81
J06:Adrenal disorder	2	S81
J07:Other disease of the endocrine system	2	J13, S81
J08:Malnutrition & vitamin deficiency	2	J13, S81
J09:Hypercholesterolaemia and other dyslipidemia	2	S81
J10:Obesity	1	
J11:Disorder of electrolyte acid base balance	1	S81
J12:Hypoglycemia	1	J13, S81
J13:Other disorders of metabolism	1	J10, S81
K01:Chronic kidney disease/failure	2	H82, K04, K41, K81, S81
K02:Benign/unspecified urinary neoplasm	1	K81, S81
K03:Other disease/disorder bladder & urethra	1	K81, S81

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
K04:Other disease of the urinary system	1	K81, S81
K05:Prostate disorder	2	K43, K81, S81
K06:Other disease of the male reproductive system	2	K81, S81
K07:Urinary incontinence (stress)	2	K81, S81
K41:Acute/other kidney disease/ failure	1	H82, K81, S81
K42:Urinary tract infection/cystitis	1	F82, K81, S81
K43:Urinary obstruction & retention	1	K81, S81
K81:Signs, symptoms urinary	1	S81
L01:Female infertility	2	L02, L04, L05, S81
L02:Menstruation disorder (incl. menopause)	2	F82, S81
L03:Congenital/other female reproductive condition	2	S81
L04:Endometriosis/pelvic inflammatory disease	2	F82, H42, L05, S81
L05:Other uterine, fallopian, ovarian condition (incl. prolapse, fistula)	2	K07, K81, S81
M41:Antepartum condition	2	H82, S81
M42:Unsuccessful/ectopic pregnancy	1	F82, S81
M43:Completed pregnancy	1	S81, M41
N41:Extremely low birth weight or immaturity	1	N42, N43, N44, S81
N42:Serious perinatal condition	1	N43, N44, S81
N43:Complicated neonatal condition	1	N44, S81
N44:Healthy newborn	8	
O01:Disease of white blood cells (incl. neutropenia)	2	O07, S81
O02:Anemia disorder (excl. nutritional)	2	O07, S81
O03:Coagulation & hemorrhagic disorder	2	O02, O07, S81
O04:Spleen disorder	2	F82, O03, S81
O05:Lymphatic system disorder (excl. spleen)	2	S81
O06:Disorder of immune mechanism	2	S81

Appendix A

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
O07:Other disease of blood & blood forming organs	2	S81
P01:Human immunodeficiency virus (HIV) infection	2	P43, P44, P45, O01, O06, O07, S81
P02:Tuberculosis disease	2	D81, P44, P45, S81
P03:Sexually transmitted disease (excl. HIV)	1	S81
P41:Sepsis	0	P43, P44, P45, S81
P42:Influenza	1	S81
P43:Other viral infection	1	S81
P44:Other bacterial infection	1	S81
P45:Other & unspecified infection	1	S81
Q01:Dementia (incl. alzheimer's)	2	A84, Q03, Q14, Q81, Q82, S81
Q02:Delusional disorder (incl. schizophrenia)	2	A84, Q04, Q05, Q06, Q11, Q14, Q82, S81
Q03:Mental disorder resulting from brain injury or other illness	2	A41, A42, A44, A45, A46, A84, Q14, Q82, S02, S45, S81
Q04:Depression	2	Q11, Q12, Q14, Q82, S81
Q05:Bipolar/manic mood disorder	2	Q04, Q06, Q11, Q12, Q14, Q82, S81
Q06:Personality disorder	2	Q11, Q12, Q13, Q14, Q82, S81
Q07:Drug/alcohol abuse/dependence	2	S44, Q14, Q82, S81
Q08:Eating disorder	2	F81, F82, J08, J10, J11, J13, Q11, Q14, S81
Q09:Intellectual disorder/delay	2	Q14, Q82, S81
Q10:Developmental disorder (incl. autism spectrum)	2	Q11, Q13, Q14, Q82, S81
Q11:Neurotic/anxiety/obsessive compulsive disorder	2	Q14, Q82, S81
Q12:Adjustment reaction/stress	1	Q14, Q82, S81
Q13:Emotional and behavioural disorder with onset generally in childhood	2	Q14, Q82, S81
Q14:Other/unspecified mental disorder/condition	2	Q82, S81
Q81:Delirium	1	A84, Q82, S81
Q82:Mental health signs & symptoms	1	S81

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
R01a:Brain cancer	2	R17, A84, S81
R02c:Oral/ear/nose/throat cancer	2	C02, C42, R17, S81
R03d:Lung cancer	2	C42, D01, D42, D43, D44, D81, H82, R17, S81
R04f:Colorectal cancer	2	F10, F42, F81, F82, R17, S81
R05f:Other digestive & hepatobiliary cancer	2	F10, F81, F82, R17, S81
R06h:Musculoskeletal cancer	2	H45, H81, R17, S81
R07k:Renal cancer	2	H82, K02, K81, R17, S81
R08i:Breast cancer	2	I07, R17, S81
R09i:Skin cancer	2	I06, I81, R17, S81
R10j:Thyroid cancer	2	J04, J05, R17, S81
R11k:Prostate cancer	2	K05, K81, R17, S81
R12k:Bladder cancer	2	H82, K02, K03, K81, R17, S81
R13l:Ovarian cancer	2	F82, L04, L05, R17, S81
R14l:Uterine cancer	2	L02, L04, L05, R17, S81
R15l:Cervical cancer	2	L02, R17, S81
R16o:Leukemia/lymphoma	2	O01, O02, O07, R17, S81
R17:Other & unspecified primary cancer	2	S81
R18:Metastatic cancer	2	R01A, R02C, R03D, R04F, R05F, R06H, R07F, R08I, R09I, R10J, R11K, R12K, R13L, R14L, R15L, R16O, R17, S81
S01:Palliative state	0	
S02:Chromosomal/other/unspecified congenita syndrome/disorder	2	Q09, Q10, S81
S03:Transplant recipient	2	S81
S04:Presence of ostomy	2	F81, S81
S41:Transplant complication	1	S03, S46, S81
S42:Orthopedic complication	2	S46, S81

Appendix A

Health condition (index health condition)	PLPB tagging rule	Clinical overrides (conditions un-tagged when coexisting with index health condition)
S43:Ostomy complication	1	F81, S04, S46, S81
S44:Poisoning/adverse effects	1	F82, S81
S45:Other trauma/injury (not elsewhere classified)	1	I43, S81
S46:Other complication of medical or surgical care	1	S81
S47:Shock	1	S81
S48:Open wound	1	I43, S81
S49:Internal trauma	1	I43, S81
S50:Reaction to transfusion, medication or anaesthesia	1	S81
S81:General signs, symptoms (not elsewhere classified)	1	

Appendix B – Health conditions: volumes and average costs, all three provinces

Table B1   Health conditions, volumes and average costs for the concurrent period, all three provinces

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
A01:Congenital malformation of the nervous system (incl. spina bifida)	9,172	$37,825	$3,627	$2,693	$31,505
A02:Hereditary/degenerative condition of the nervous system	43,978	$32,545	$3,545	$2,492	$26,508
A03:Muscular dystrophy & other myopathy	12,985	$25,931	$3,706	$1,991	$20,235
A04:Parkinson's disease/parkinsonism	46,339	$16,780	$3,505	$1,119	$12,156
A05:Multiple sclerosis & other demyelinating disease of cns	41,849	$9,781	$2,226	$1,104	$6,451
A06:Cerebral palsy	12,380	$18,905	$2,429	$1,434	$15,042
A07:Paralytic syndrome/spinal cord injury	35,908	$46,396	$4,811	$2,882	$38,703
A08:Disorder of the peripheral nervous system (incl. Carpal tunnel, bells palsy)	161,134	$12,471	$2,891	$1,372	$8,208
A09:Epilepsy	99,642	$15,094	$2,521	$1,334	$11,239
A10:Other cerebral and spinal disease/disorder	37,828	$35,690	$4,199	$3,046	$28,444
A11:Other disease of the nervous system	71,128	$26,113	$4,332	$1,769	$20,013
A41:Stroke	80,454	$30,886	$4,475	$1,828	$24,583
A42:Cerebrovascular disorder (excl. stroke)	49,716	$18,831	$3,547	$1,890	$13,395
A43:Transient ischemic attack	83,270	$11,821	$2,988	$1,202	$7,631

Appendix B

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
A44:Intracranial injury	41,922	$19,785	$3,049	$1,382	$15,355
A45:Skull fracture	5,002	$31,921	$3,133	$2,695	$26,093
A46:Concussion	123,330	$4,270	$1,127	$573	$2,570
A47:Infection/inflammation nervous system	28,995	$26,609	$2,927	$2,058	$21,624
A48:Nerve injury	21,996	$11,752	$1,847	$1,699	$8,207
A81:Coma unspecified	7,686	$39,359	$3,396	$3,772	$32,191
A82:Seizure	17,244	$25,440	$3,708	$1,998	$19,734
A83:Migraine/other headache	211,665	$7,870	$2,375	$813	$4,682
A84:Neurological, sensory, cognitive signs & symptoms	230,257	$7,609	$1,892	$927	$4,790
B01:Cataract/lens disorder	708,417	$8,123	$2,713	$747	$4,663
B02:Retinopathy	312,240	$10,131	$2,889	$956	$6,285
B03:Strabismus	172,236	$5,140	$1,336	$476	$3,328
B04:Disorder of the eyelids & lacrimal system	499,079	$4,794	$1,719	$511	$2,564
B05:Vision impairment (incl. blindness)	657,419	$2,627	$1,002	$214	$1,411
B06:Glaucoma	492,007	$6,550	$2,389	$632	$3,530
B07:Other eye diseases & disorders	166,398	$6,277	$2,049	$536	$3,691
B41:Retinal break/detachment	36,695	$5,955	$2,427	$576	$2,952
B42:Infection/inflammation eye	1,365,612	$3,689	$1,386	$355	$1,948
B43:Injury of eye and periocular area	201,705	$3,563	$1,215	$357	$1,991
C01:Congenital/acquired malformation of ear/nose/throat	29,931	$13,006	$2,108	$942	$9,956
C02:Disease of oral cavity, salivary gland, or jaw	119,019	$10,666	$2,191	$987	$7,489
C03:Dental caries/gingivitis	311,861	$7,151	$1,709	$651	$4,791

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
C04:Chronic upper respiratory condition (incl. tonsillitis)	369,261	$4,736	$1,574	$622	$2,539
C05:Hearing loss (incl. deafness)	82,701	$8,807	$2,403	$697	$5,707
C06:Disorder of external ear	738,959	$4,058	$1,343	$479	$2,236
C07:Other disease middle/inner ear & mastoid	150,329	$4,981	$1,923	$419	$2,640
C08:Meniere's/dizziness/vertigo	161,611	$7,662	$2,267	$780	$4,615
C41:Otitis media	1,242,809	$2,684	$1,002	$253	$1,429
C42:Acute ent, upper respiratory condition (incl. benign neoplasm, croup)	5,885,750	$2,499	$1,045	$274	$1,179
D01:Pleural disease (incl. pleural effusion)	56,447	$43,728	$4,537	$3,577	$35,614
D02:Congenital disorder of the respiratory system	2,931	$52,472	$3,045	$3,853	$45,574
D03:Chronic obstructive pulmonary disease	299,631	$19,308	$3,309	$1,627	$14,372
D04:Pulmonary hypertension	13,742	$54,325	$5,171	$4,869	$44,285
D05:Other chronic lung disease	39,748	$24,046	$3,717	$2,238	$18,091
D06:Asthma	493,737	$4,570	$1,594	$461	$2,516
D41:Respiratory failure	30,916	$89,697	$6,981	$5,488	$77,228
D42:Pneumonia	679,368	$14,680	$2,363	$1,113	$11,205
D43:Acute infectious/parasitic respiratory disease (excl. pneumonia)	7,497	$15,565	$4,201	$930	$10,434
D44:Acute and other respiratory diseases & disorders	2,101,824	$3,518	$1,363	$367	$1,788
D81:Signs, symptoms respiratory system	1,331,922	$4,130	$1,489	$464	$2,176
E01:Heart failure	225,716	$30,109	$4,491	$2,346	$23,272
E02:Malformation of the cardiovascular system	35,705	$35,217	$3,524	$2,531	$29,162

Appendix B

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
E03:Cardiac valve disease	30,985	$22,366	$3,952	$2,503	$15,911
E04:Coronary artery disease	295,599	$16,148	$3,569	$1,529	$11,050
E05:Arrhythmia	426,202	$17,481	$3,463	$1,577	$12,441
E06:Other heart disease	246,139	$8,271	$2,733	$723	$4,814
E07:Peripheral venous disease/ phlebitis/thrombophlebitis/DVT	76,023	$28,239	$4,620	$1,780	$21,838
E08:Uncomplicated varicose veins	99,810	$5,143	$2,028	$484	$2,631
E09:Other vascular system disease	179,820	$27,332	$4,293	$2,342	$20,698
E10:Hypertension	2,278,188	$6,371	$2,033	$629	$3,709
E11:Aortic aneurysm	15,426	$41,204	$5,517	$4,145	$31,542
E12:Peripheral artery disease	31,805	$46,215	$5,956	$3,903	$36,356
E41:Acute myocardial infarction/shock/arrest	90,053	$39,379	$4,998	$2,923	$31,458
E43:Unstable angina	38,128	$24,406	$3,997	$2,571	$17,838
E82:Signs, symptoms cardiovascular system	1,565,011	$3,827	$1,529	$439	$1,859
F01:Congenital malformation of the digestive & hepatobiliary system	8,382	$34,985	$2,838	$2,686	$29,461
F02:Gastroesophageal reflux disease	61,702	$22,600	$3,357	$1,935	$17,308
F03:Other disease of the esophagus	174,702	$12,348	$2,669	$1,177	$8,502
F04:Gastritis & duodenitis	683,544	$6,432	$1,886	$733	$3,813
F05:Hernia	238,802	$9,933	$2,419	$1,017	$6,498
F06:Inflammatory bowel (incl. crohn's, ulcerative colitis)	119,457	$12,818	$2,553	$1,163	$9,102
F07:Vascular disorder of intestine (incl. ischemic bowel)	13,367	$42,030	$5,032	$3,268	$33,730
F08:Diverticulitis	209,126	$10,085	$2,594	$1,028	$6,462

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
F09:Hemorrhoids	173,461	$5,955	$1,986	$660	$3,309
F10:Benign neoplasm/polyp colon, rectum, anus	286,696	$7,115	$2,287	$793	$4,035
F11:Other rectal/anal disorder	246,595	$8,583	$2,417	$776	$5,389
F12:Other disease/disorder of the digestive system	170,658	$21,232	$3,672	$1,565	$15,996
F41:Acute gastrointestinal infection	1,153,896	$6,792	$1,622	$578	$4,592
F42:Unspecified gastrointestinal hemorrhage	78,501	$23,821	$2,764	$2,535	$18,522
F43:Gastrointestinal obstruction	81,136	$32,260	$4,116	$2,697	$25,448
F44:Peptic ulcer	77,803	$20,466	$3,427	$1,709	$15,330
F45:Peritoneal disease (incl. peritonitis)	36,543	$42,040	$4,857	$3,914	$33,269
F46:Disease of appendix	64,257	$10,196	$1,994	$1,205	$6,997
F81:Signs, symptoms digestive & hepatobiliary system	3,060,448	$4,153	$1,499	$408	$2,245
F82:Abdominal pain	169,037	$4,074	$1,226	$672	$2,176
G01:Cirrhosis	35,996	$23,465	$3,659	$2,172	$17,634
G02:Chronic liver disease (incl. hepatitis)	25,633	$29,570	$3,945	$2,453	$23,173
G03:Pancreatic disease	36,512	$27,146	$3,577	$2,638	$20,932
G04:Acute liver disease & other hepatic disorders	233,661	$7,154	$2,024	$683	$4,448
G05:Diseases of gallbladder & biliary tract	124,739	$14,444	$2,840	$1,499	$10,106
H01:Rheumatoid & other inflammatory arthropathy (excl. gout)	160,634	$12,251	$2,978	$1,159	$8,114
H02:Osteoarthritis	779,886	$10,451	$2,951	$836	$6,665
H03:Crystal arthropathy (incl. gout)	224,498	$7,452	$1,867	$759	$4,827

Appendix B

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
H04:Osteoporosis	85,595	$23,503	$4,091	$1,493	$17,919
H05:Other bone disease	139,081	$11,647	$2,662	$1,009	$7,976
H06:Vertebral/disk & other disease of back	797,636	$6,578	$2,218	$731	$3,629
H07:Systemic connective tissue disorder (incl. lupus, scleroderma)	58,386	$13,777	$3,185	$1,454	$9,139
H08:Musculoskeletal malformation (excl. spine)	280,665	$9,977	$2,706	$766	$6,506
H09:Myositis and soft tissue disorder (incl. muscle inflammation)	930,073	$7,700	$1,828	$908	$4,964
H10:Other musculoskeletal & connective tissue condition	26,736	$7,459	$2,257	$920	$4,282
H41:Traumatic amputation arm/hand/leg/foot	4,423	$33,687	$1,821	$3,950	$27,917
H42:Fracture/dislocation vertebrae, pelvis	69,607	$18,593	$2,823	$1,521	$14,249
H43:Fracture femur	77,172	$30,752	$4,275	$2,200	$24,277
H44:Other fracture/dislocation	840,684	$5,637	$1,503	$648	$3,486
H45:Benign/unspecified musculoskeletal neoplasm	56,995	$6,021	$1,689	$812	$3,520
H46:Joint/tendon disorder and injury (incl. pain, sprain, strain)	4,734,645	$4,534	$1,519	$485	$2,530
H47:musculoskeletal infections	35,966	$41,540	$4,835	$3,543	$33,162
H81:neuromuscular signs & symptoms	1,322,416	$3,249	$1,222	$321	$1,706
H82:Back pain	199,059	$3,402	$1,119	$505	$1,778
I01:Autoimmune skin disorder	1,341	$36,188	$4,564	$2,695	$28,929
I02:Papulosquamous disorder/ psoriasis	78,708	$4,983	$1,801	$539	$2,643
I03:Eczema/dermatitis/hives	853,451	$4,511	$1,565	$368	$2,578

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
I04:Acne/rosacea	247,210	$2,169	$1,026	$210	$933
I05:Skin ulcer (incl. decubitus)	41,441	$51,379	$5,627	$2,992	$42,760
I06:Benign skin neoplasm	962,515	$3,691	$1,517	$412	$1,761
I07:Benign disorder of breast	668,315	$3,632	$1,524	$474	$1,634
I08:Disease of hair/nail/sweat glands	407,036	$5,458	$1,565	$606	$3,287
I09:Other condition of skin/subcutaneous tissue	1,272,694	$4,125	$1,446	$411	$2,268
I41:Serious burn/frostbite	2,975	$40,209	$2,750	$3,010	$34,449
I42:Skin infection (incl. cellulitis)	2,117,408	$5,926	$1,527	$617	$3,782
I43:Superficial skin injury/contusion/non-serious burn	1,038,284	$5,167	$1,519	$459	$3,189
I81:Signs, symptoms skin	492,522	$3,514	$961	$511	$2,042
J01:Cystic fibrosis	1,808	$38,001	$2,782	$3,385	$31,834
J02:Diabetes mellitus	1,380,843	$8,678	$2,426	$777	$5,474
J03:Goitre	56,081	$6,519	$2,455	$723	$3,341
J04:Hyperthyroidism	47,086	$7,545	$2,311	$667	$4,567
J05:Hypothyroidism	361,584	$8,694	$2,153	$772	$5,768
J06:Adrenal disorder	10,176	$27,551	$3,833	$2,350	$21,368
J07:Other disease of the endocrine system	143,234	$11,496	$2,714	$1,081	$7,700
J08:Malnutrition & vitamin deficiency	338,069	$13,434	$2,758	$1,077	$9,598
J09:Hypercholesterolaemia and other dyslipidemia	556,243	$5,069	$1,725	$483	$2,861
J10:Obesity	433,776	$5,783	$1,795	$582	$3,406
J11:Disorder of electrolyte acid base balance	216,641	$34,030	$3,881	$2,863	$27,285
J12:Hypoglycemia	6,024	$17,571	$2,023	$1,681	$13,867

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
J13:Other disorders of metabolism	129,906	$15,129	$2,833	$1,052	$11,245
K01:Chronic kidney disease/failure	178,332	$25,441	$4,869	$2,169	$18,403
K02:Benign/unspecified urinary neoplasm	22,256	$9,681	$2,649	$907	$6,125
K03:Other disease/disorder bladder & urethra	1,261,225	$7,028	$1,976	$641	$4,411
K04:Other disease of the urinary system	22,343	$11,644	$1,882	$1,145	$8,617
K05:Prostate disorder	247,242	$9,082	$2,465	$835	$5,782
K06:Other disease of the male reproductive system	169,788	$6,338	$1,578	$602	$4,158
K07:Urinary incontinence (stress)	138,322	$18,078	$3,768	$1,028	$13,282
K41:Acute/other kidney disease/ failure	276,683	$18,073	$2,844	$1,598	$13,631
K42:Urinary tract infection/cystitis	1,240,338	$9,292	$2,061	$807	$6,424
K43:Urinary obstruction & retention	249,498	$12,317	$2,443	$1,335	$8,540
K81:Signs, symptoms urinary	366,372	$3,309	$1,271	$426	$1,612
L01:Female infertility	110,839	$5,333	$2,714	$521	$2,097
L02:Menstruation disorder (incl. menopause)	673,557	$4,121	$1,715	$493	$1,912
L03:Congenital/other female reproductive condition	611,955	$4,405	$1,757	$502	$2,146
L04:Endometriosis/pelvic inflammatory disease	52,479	$9,977	$2,454	$1,193	$6,330
L05:Other uterine, fallopian, ovarian condition (incl. prolapse, fistula)	159,586	$7,300	$2,340	$799	$4,161
M41:Antepartum condition	144,396	$3,160	$1,535	$394	$1,232
M42:Unsuccessful/ectopic pregnancy	170,233	$4,899	$1,767	$615	$2,518
M43:Completed pregnancy	441,085	$7,542	$2,413	$1,013	$4,116

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
N41:Extremely low birth weight or immaturity	3,371	$113,437	$4,651	$3,112	$105,673
N42:Serious perinatal condition	46,009	$18,117	$1,484	$860	$15,773
N43:Complicated neonatal condition	65,740	$7,769	$923	$535	$6,312
N44:Healthy newborn	370,916	$2,222	$571	$163	$1,488
O01:Disease of white blood cells (incl. neutropenia)	17,066	$33,441	$4,076	$2,748	$26,617
O02:Anemia disorder (excl. nutritional)	177,473	$25,259	$4,169	$1,838	$19,251
O03:Coagulation & hemorrhagic disorder	98,513	$28,138	$3,993	$2,390	$21,756
O04:Spleen disorder	1,645	$61,224	$4,741	$5,357	$51,126
O05:Lymphatic system disorder (excl. spleen)	27,627	$12,693	$2,241	$1,128	$9,323
O06:Disorder of immune mechanism	10,596	$21,933	$2,988	$2,089	$16,856
O07:Other disease of blood & blood forming organs	79,368	$9,695	$2,581	$944	$6,171
P01:Human immunodeficiency virus (HIV) infection	21,332	$9,206	$2,212	$1,034	$5,960
P02:tuberculosis disease	40,111	$7,767	$1,858	$707	$5,202
P03:sexually transmitted disease (excl. HIV)	235,279	$2,821	$1,181	$253	$1,387
P41:Sepsis	46,148	$70,336	$6,090	$4,256	$59,990
P42:Influenza	556,918	$4,078	$1,495	$370	$2,213
P43:Other viral infection	1,493,156	$3,145	$1,191	$273	$1,682
P44:Other bacterial infection	196,446	$27,003	$3,526	$1,997	$21,480
P45:Other & unspecified infection	406,068	$8,077	$1,805	$667	$5,605
Q01:Dementia (incl. alzheimer's)	242,489	$21,176	$3,762	$1,198	$16,216
Q02:Delusional disorder (incl. schizophrenia)	127,773	$18,433	$3,486	$1,747	$13,200

Appendix B

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
Q03:Mental disorder resulting from brain injury or other illness	13,911	$43,164	$3,677	$3,998	$35,489
Q04:Depression	714,456	$7,897	$2,152	$822	$4,924
Q05:Bipolar/manic mood disorder	173,581	$12,093	$3,267	$1,333	$7,494
Q06:Personality disorder	25,480	$10,638	$3,136	$1,022	$6,480
Q07:Drug/alcohol abuse/dependence	284,722	$14,079	$3,086	$1,303	$9,690
Q08:Eating disorder	8,504	$21,127	$4,889	$2,247	$13,990
Q09:Intellectual disorder/delay	12,106	$18,006	$3,222	$1,587	$13,197
Q10:Developmental disorder (incl. autism spectrum)	59,443	$7,802	$1,598	$648	$5,557
Q11:Neurotic/anxiety/obsessive compulsive disorder	1,064,722	$4,956	$2,084	$391	$2,480
Q12:Adjustment reaction/stress	321,601	$5,491	$1,636	$671	$3,185
Q13:Emotional and behavioural disorder with onset generally in childhood	176,780	$4,292	$1,372	$475	$2,446
Q14:Other/unspecified mental disorder/condition	34,016	$4,430	$1,610	$485	$2,335
Q81:Delirium	29,142	$53,405	$4,888	$4,902	$43,615
Q82:Mental health signs & symptoms	374,250	$4,076	$1,769	$349	$1,959
R01a:Brain cancer	8,700	$28,638	$4,832	$2,574	$21,232
R02c:Oral/ear/nose/throat cancer	17,323	$13,725	$3,235	$1,293	$9,197
R03d:Lung cancer	26,648	$19,076	$3,800	$2,051	$13,226
R04f:Colorectal cancer	49,656	$16,198	$3,426	$1,614	$11,158
R05f:Other digestive & hepatobiliary cancer	25,736	$21,208	$3,917	$2,048	$15,242
R06h:Musculoskeletal cancer	7,516	$18,233	$3,408	$1,927	$12,898
R07k:Renal cancer	15,387	$17,073	$3,782	$1,969	$11,322

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
R08i:Breast cancer	90,004	$7,930	$2,731	$1,113	$4,086
R09i:Skin cancer	250,485	$6,723	$2,476	$629	$3,617
R10j:Thyroid cancer	23,944	$6,768	$2,639	$728	$3,401
R11k:Prostate cancer	102,081	$8,965	$2,703	$913	$5,349
R12k:Bladder cancer	38,386	$13,775	$3,300	$1,357	$9,118
R13l:Ovarian cancer	2,130	$18,472	$3,243	$2,284	$12,945
R14l:Uterine cancer	9,440	$12,304	$3,253	$1,262	$7,790
R15l:Cervical cancer	10,703	$7,069	$1,997	$883	$4,189
R16o:Leukemia/lymphoma	56,824	$18,225	$3,262	$1,650	$13,314
R17:Other & unspecified primary cancer	46,558	$11,686	$2,872	$876	$7,938
R18:Metastatic cancer	77,522	$26,003	$4,853	$2,580	$18,569
S01:Palliative state	114,769	$41,365	$4,241	$2,504	$34,620
S02:Chromosomal/other/unspecified congenita syndrome/disorder	21,708	$20,748	$2,682	$1,539	$16,526
S03:Transplant recipient	10,183	$42,187	$5,743	$4,668	$31,777
S04:Presence of ostomy	18,640	$55,078	$5,751	$3,750	$45,577
S41:Transplant complication	3,086	$88,442	$8,111	$8,365	$71,965
S42:Orthopedic complication	13,145	$33,690	$4,837	$3,048	$25,804
S43:Ostomy complication	7,953	$92,783	$7,177	$5,633	$79,973
S44:Poisoning/adverse effects	555,546	$8,145	$2,239	$640	$5,266
S45:Other trauma/injury (not elsewhere classified)	443,552	$4,757	$1,351	$421	$2,985
S46:Other complication of medical or surgical care	176,724	$34,681	$4,445	$3,277	$26,959
S47:Shock	6,682	$80,672	$3,568	$9,963	$67,141

Appendix B

Health condition (index health condition)	Number of persons during concurrent period	Average cost during the concurrent period, in Dollars			
		All costs	Physician costs, non-hospital care	Physician costs, hospital care	Hospital inpatient, day surgery, ED
S48:Open wound	780,566	$4,358	$1,212	$498	$2,648
S49:Internal trauma	24,156	$33,825	$3,406	$3,419	$26,999
S50:Reaction to transfusion, medication or anaesthesia	129,542	$8,478	$1,503	$1,221	$5,755
S81:General signs, symptoms (not elsewhere classified)	339,550	$297	$181	$26	$91

Appendix C, Table C1 – Health profile grouping methodology

Table C1 Health profile grouping methodology: mapping branches to health conditions

Branch	Branch description	Health condition Code	Health condition description
A001	Paralytic syndrome with condition other than stroke	A07	Paralytic syndrome/spinal cord injury
A003	Paralytic syndrome with stroke	A41+A07	A41-Stroke, A07-Paralytic syndrome
A004	Congenital malformation of the nervous system (incl. spina bifida)	A01	Congenital malformation of the nervous system (incl. spina bifida)
A006	Stroke without paralytic syndrome	A41	Stroke
A008	Other cerebral and spinal disorder (incl benign neoplasms)	A10	Other cerebral and spinal disease/disorder
A010	Parkinson's disease/parkinsonism	A04	Parkinson's disease/parkinsonism
A012	Cerebral palsy	A06	Cerebral palsy
A014	Major chronic nervous system	A02	Hereditary/degenerative condition of the nervous system
		A03	Muscular dystrophy & other myopathy
		A11	Other disease of the nervous system
A030	Cerebrovascular disorder (excl. stroke)	A42	Cerebrovascular disorder (excl. stroke)
A032	Seizure	A82	Seizure
A033	Disorder of the peripheral nervous system (incl. carpal tunnel, bells palsy)	A08	Disorder of the peripheral nervous system (incl. carpal tunnel, bells palsy)
A035	Epilepsy	A09	Epilepsy
A037	Multiple sclerosis & other demyelinating disease of central nervous system	A05	Multiple sclerosis & other demyelinating disease of cns

Appendix C

Branch	Branch description	Health condition Code	Health condition description
A039	Transient ischemic attack	A43	Transient ischemic attack
A050	Neurological, sensory, cognitive signs & symptoms (incl insomnia)	A84	Neurological, sensory, cognitive signs & symptoms
A051	Migraine/other headache	A83	Migraine/other headache
B030	Retinopathy	B02	Retinopathy
B032	Cataract/lens disorder	B01	Cataract/lens disorder
B050	Retinal break/detachment	B41	Retinal break/detachment
B051	Other eye diseases & disorders	B07	Other eye diseases & disorders
B052	Chronic vision impairment	B03	Strabismus
		B05	Vision impairment (incl. blindness)
		B06	Glaucoma
B053	Minor acute eye (incl conjunctivities)	B04	Disorder of the eyelids & lacrimal system
		B42	Infection/inflammation eye
		B43	Injury of eye and periocular area
C054	Disease of oral cavity, salivary gland, or jaw	C02	Disease of oral cavity, salivary gland, or jaw
C055	Chronic upper respiratory condition (incl. tonsillitis)	C04	Chronic upper respiratory condition (incl. tonsillitis)
C056	Disorder of external ear	C06	Disorder of external ear
C057	Otitis media	C41	Otitis media
C058	Acute ent, upper respiratory condition (incl. colds, croup)	C42	Acute ent, upper respiratory condition (incl. benign neoplasm, croup)
D001	Respiratory failure with heart failure	D41+E01	D41-Respiratory failure, E01-Heart failure
D002	Respiratory failure without heart failure	D41	Respiratory failure
D003	Pulmonary hypertension	D04	Pulmonary hypertension

Branch	Branch description	Health condition Code	Health condition description
D004	Pneumonia with copd	D42+D03	D42-Pneumonia, D03-Chronic obstructive pulmonary disease
		D42+D05	D42-pneumonia, D05-Other chronic lung disease
D030	COPD without pneumonia	D03	Chronic obstructive pulmonary disease
		D05	Other chronic lung disease
D050	Pneumonia without COPD	D42	Pneumonia
D051	Asthma	D06	Asthma
D052	Minor acute respiratory (incl cough, bronchitis)	D43	Acute infectious/parasitic respiratory disease (excl. pneumonia)
		D44	Acute and other respiratory diseases & disorders
		D81	Signs, symptoms respiratory system
E001	Acute myocardial infarction with heart failure	E41+E01	E41-Acute myocardial infarction, E01-Heart failure
E003	Heart failure with CAD/arrhythmia	E04+E01	E04-Coronary artery disease, E01-Heart failure
		E05+E01	E05-Arrhythmia, E01-Heart failure
		E06+E01	E06-Other heart disease, E01-Heart failure
E005	Acute myocardial infarction without heart failure	E41	Acute myocardial infarction/shock/arrest
E007	Heart failure without AMI, respiratory failure or CAD	E01	Heart failure
E009	Cardiac valve disease	E03	Cardiac valve disease
E011	Unstable angina	E43	Unstable angina
E013	PVD/other chronic vascular dx (without diabetes mellitus/hypoglycemia)	E07	Peripheral venous disease/phlebitis/thrombophlebitis/dvt
		E09	Other vascular system disease
		E11	Aortic aneurysm
		E12	Peripheral artery disease

Appendix C

Branch	Branch description	Health condition Code	Health condition description
E030	CAD/arrhythmia without heart failure	E04	Coronary artery disease
		E05	Arrhythmia
		E06	Other heart disease
E050	Hypertension	E10	Hypertension
E051	Signs, symptoms cardiovascular system (incl chest pain, palpitations)	E82	Signs, symptoms cardiovascular system
F001	Major acute GI/peritonitis	F43	Gastrointestinal obstruction
		F45	Peritoneal disease (incl. peritonitis)
F030	Unspecified gastrointestinal hemorrhage	F42	Unspecified gastrointestinal hemorrhage
F032	Disease of appendix	F46	Disease of appendix
F034	Inflammatory bowel (incl. crohn's, ulcerative colitis)	F06	Inflammatory bowel (incl. crohn's, ulcerative colitis)
F036	Peptic ulcer	F44	Peptic ulcer
F050	Hernia	F05	Hernia (excl. diaphragmatic)
F051	Gastroesophageal reflux disease	F02	Gastroesophageal reflux disease
F052	Other disease/disorder of the digestive system	F07	Vascular disorder of intestine (incl. ischemic bowel)
		F12	Other disease/disorder of the digestive system
F053	Other disease of the esophagus (incl esophagitis)	F03	Other disease of the esophagus
F054	Diverticulitis	F08	Diverticulitis
F055	Benign neoplasms and other rectal/anal disorders	F09	Hemorrhoids
		F10	Benign neoplasm/polyp colon, rectum, anus
		F11	Other rectal/anal disorder

Branch	Branch description	Health condition Code	Health condition description
F056	Minor GI acute (incl. GI symptom, ill-defined infection)	F41	Acute gastrointestinal infection
		F81	Signs, symptoms digestive & hepatobiliary system
		F82	Abdominal pain
F057	Gastritis & duodenitis	F04	Gastritis & duodenitis
G001	Pancreatic disease	G03	Pancreatic disease
G003	Cirrhosis	G01	Cirrhosis
G030	Diseases of gallbladder & biliary tract	G05	Diseases of gallbladder & biliary tract
G032	Chronic liver disease (incl. hepatitis)	G02	Chronic liver disease (incl. hepatitis)
H001	Musculoskeletal infections	H47	Musculoskeletal infections
H003	Major acute msk (incl fracture femur/spine, limb amputations)	H41	Traumatic amputation arm/hand/leg/foot
		H42	Fracture/dislocation vertebrae, pelvis
		H43	Fracture femur
H030	Rheumatoid/osteoarthritis& other inflammatory arthropathy (excl. gout)	H01	Rheumatoid & other inflammatory arthropathy (excl. gout)
		H02	Osteoarthritis
H032	Osteoporosis	H04	Osteoporosis
H034	Systemic connective tissue disorder (incl. lupus, scleroderma)	H07	Systemic connective tissue disorder (incl. lupus, scleroderma)
H050	Other bone disease	H05	Other bone disease
H051	Musculoskeletal malformation (excl. spine)	H08	Musculoskeletal malformation (excl. spine)
H052	Vertebral/disk & other disease of back	H06	Vertebral/disk & other disease of back
H053	Other fracture/dislocation	H44	Other fracture/dislocation
H054	Back pain	H82	Back pain
H055	Crystal arthropathy (incl. gout)	H03	Crystal arthropathy (incl. gout)

Appendix C

Branch	Branch description	Health condition Code	Health condition description
H056	Myositis and soft tissue disorder (incl. muscle inflammation)	H09	Myositis and soft tissue disorder (incl. muscle inflammation)
H057	Joint/tendon disorder and injury (incl. pain, sprain, strain)	H46	Joint/tendon disorder and injury (incl. pain, sprain, strain)
H058	Neuromuscular signs & symptoms	H81	Neuromuscular signs & symptoms
I001	Skin ulcer (incl. decubitus)	I05	Skin ulcer (incl. decubitus)
I050	Minor acute skin (incl dermatitis, cellulitis)	I03	Eczema/dermatitis/hives
		I04	Acne/rosacea
		I06	Benign skin neoplasm
		I07	Benign disorder of breast
		I08	Disease of hair/nail/sweat glands
		I09	Other condition of skin/subcutaneous tissue
		I42	Skin infection (incl. cellulitis)
		I43	Superficial skin injury/contusion/non-serious burn
		I81	Signs, symptoms skin
J001	Cystic fibrosis	J01	Cystic fibrosis
J002	Diabetes/hypoglycemia with PVD/other chronic vasc dx	J02+E07	J02 - Diabetes mellitus, Peripheral venous disease/phlebitis/thrombophlebitis/ DVT
		J02+E09	J02 - Diabetes mellitus, Other vascular system disease
		J02+E11	J02 - Diabetes mellitus, Aortic aneurysm
		J02+E12	J02 - Diabetes mellitus, Peripheral artery disease
		J12+E07	J12- Hypoglycemia, E07 - Peripheral venous disease/phlebitis/thrombophlebitis/DVT
		J12+E09	J12- Hypoglycemia, E09 - Other vascular system disease
		J12+E11	J12- Hypoglycemia, E11 - Aortic aneurysm
		J12+E12	J12- Hypoglycemia, E12 - Peripheral artery disease

Branch	Branch description	Health condition Code	Health condition description
J004	Diabetes/hypoglycemia with chronic kidney disease/failure	J02+K01	J02 - Diabetes mellitus, K01 - Chronic kidney disease/failure
		J12+K01	J12- Hypoglycemia, K01 - Chronic kidney disease/failure
J030	Adrenal/electrolyte/metabolism disorders	J06	Adrenal disorder
		J11	Disorder of electrolyte acid base balance
J032	Diabetes/hypoglycemia without chronic kidney disease or PVD/other chronic vasc dx	J02	Diabetes mellitus
		J12	Hypoglycemia
J050	Other disease of the endocrine system	J07	Other disease of the endocrine system
J051	Other disorders of metabolism	J13	Other disorders of metabolism
J052	Malnutrition & vitamin deficiency	J08	Malnutrition & vitamin deficiency
J053	Hypothyroidism	J05	Hypothyroidism
J054	Obesity	J10	Obesity
J055	Hypercholesterolaemia and other dyslipidemia	J09	Hypercholesterolaemia and other dyslipidemia
K001	Chronic kidney disease/failure (without diabetes mellitus/hypoglycemia)	K01	Chronic kidney disease/failure
K030	Urinary incontinence (stress)	K07	Urinary incontinence (stress)
K032	Acute/other kidney disease/failure	K41	Acute/other kidney disease/failure
K050	Prostate disorder	K05	Prostate disorder
K051	Minor acute urinary (incl uti, cystitis)	K42	Urinary tract infection/cystitis
		K43	Urinary obstruction & retention
		K81	Signs, symptoms urinary
K052	Other disease of the male reproductive system	K06	Other disease of the male reproductive system
K053	Other disease/disorder bladder & urethra	K03	Other disease/disorder bladder & urethra

Branch	Branch description	Health condition Code	Health condition description
L030	Endometriosis/pelvic inflammatory disease	L04	Endometriosis/pelvic inflammatory disease
L032	Female infertility	L01	Female infertility
L034	Other uterine, fallopian, ovarian condition (incl. prolapse, fistula)	L05	Other uterine, fallopian, ovarian condition (incl. prolapse, fistula)
L050	Menstruation disorder (incl. menopause)	L02	Menstruation disorder (incl. menopause)
L051	Congenital/other female reproductive condition	L03	Congenital/other female reproductive condition
M030	Obstetrics	M41	Antepartum condition
		M43	Completed pregnancy
M050	Unsuccessful/ectopic pregnancy	M42	Unsuccessful/ectopic pregnancy
N001	Extremely low birth weight or immaturity	N41	Extremely low birth weight or immaturity
N003	Serious perinatal condition	N42	Serious perinatal condition
N005	Newborn/neonate with other complicating diagnosis	N43	Newborn/neonate with other complicating diagnosis
N070	Healthy newborn	N44	Healthy newborn
O001	disease of white blood cells (incl. neutropenia)	O01	Disease of white blood cells (incl. neutropenia)
O030	Major chronic blood disorder (incl coagulation disorders, anemia)	O02	Anemia disorder (excl. nutritional)
		O03	Coagulation & hemorrhagic disorder
O050	Other disease of blood & blood forming organs	O07	Other disease of blood & blood forming organs
O051	Lymphatic system disorder (excl. spleen)	O05	Lymphatic system disorder (excl. spleen)
P001	Sepsis	P41	Sepsis
P030	Human immunodeficiency virus (hiv)	P01	Human immunodeficiency virus (HIV) infection

Branch	Branch description	Health condition Code	Health condition description
	infection		
P050	Other bacterial infection	P44	Other bacterial infection
P051	Influenza	P42	Influenza
P052	Sexually transmitted disease (excl. HIV)	P03	Sexually transmitted disease (excl. HIV)
P053	Other & unspecified infection	P45	Other & unspecified infection
P054	Other viral infection	P43	Other viral infection
Q001	Mental disorder resulting from brain injury or other illness	Q03	Mental disorder resulting from brain injury or other illness
Q003	Delirium	Q81	Delirium
Q004	Eating disorder	Q08	Eating disorder
Q006	Dementia (incl. alzheimer's)	Q01	Dementia (incl. alzheimer's)
Q008	Delusional disorder (incl. schizophrenia)	Q02	Delusional disorder (incl. schizophrenia)
Q010	Other major mental health (incl bipolar disorder)	Q05	Bipolar/manic mood disorder
		Q09	Intellectual disorder/delay
Q070	Personality disorder	Q06	Personality disorder
Q072	Drug/alcohol abuse/dependence	Q07	Drug/alcohol abuse/dependence
Q074	Depression	Q04	Depression
Q075	Developmental disorder (incl. autism spectrum)	Q10	Developmental disorder (incl. autism spectrum)
Q076	Neurotic/anxiety/obsessive compulsive disorder	Q11	Neurotic/anxiety/obsessive compulsive disorder
Q077	Behavioural & emotional childhood/adolescent disorder	Q13	Emotional and behavioural disorder with onset generally in childhood
Q078	Other mental health (incl adjustment reaction)	Q12	Adjustment reaction/stress
		Q14	Other/unspecified mental disorder/condition

Appendix C

Branch	Branch description	Health condition Code	Health condition description
Q079	Mental health signs & symptoms	Q82	Mental health signs & symptoms
R001	Metastatic cancer	R18	Metastatic cancer
R003	Brain cancer	R01A	Brain cancer
R005	Other digestive & hepatobiliary cancer	R05F	Other digestive & hepatobiliary cancer
R007	Lung cancer	R03D	Lung cancer
R009	Leukemia/lymphoma	R16O	Leukemia/lymphoma
R011	Ovarian cancer	R13L	Ovarian cancer
R012	Bladder cancer	R12K	Bladder cancer
R014	Musculoskeletal cancer	R06H	Musculoskeletal cancer
R016	Renal cancer	R07K	Renal cancer
R018	Colorectal cancer	R04F	Colorectal cancer
R020	Uterine cancer	R14L	Uterine cancer
R021	Oral/ear/nose/throat cancer	R02C	Oral/ear/nose/throat cancer
R070	Breast cancer	R08I	Breast cancer
R072	Prostate cancer	R11K	Prostate cancer
R074	Skin/cervical/thyroid/other & unspecified cancers	R09I	Skin cancer
		R10J	Thyroid cancer
		R15L	Cervical cancer
		R17	Other & unspecified primary cancer
S001	Palliative state (acute)	S01	Palliative state
S002	Transplant complication	S41	Transplant complication
S003	Ostomy complication	S43	Ostomy complication
S004	Presence of ostomy	S04	Presence of Ostomy
S005	Transplant recipient	S03	Transplant recipient

Branch	Branch description	Health condition Code	Health condition description
S030	Chromosomal/other/unspecified congenital syndrome/disorder	S02	Chromosomal/other/unspecified congenital syndrome/disorder
S050	Other complication of medical or surgical care	S46	Other complication of medical or surgical care
S051	Reaction to transfusion, medication or anaesthesia	S50	Reaction to transfusion, medication or anaesthesia
S052	Drug/alcohol poisoning (not elsewhere classified)	S44	Poisoning/adverse effects
S053	Other trauma/injury (not elsewhere classified)	S45	Other trauma/injury (not elsewhere classified)
S054	Open wound	S48	Open wound
X001	Other major acute (incl shock, coma)	A44	Intracranial injury
		A45	Skull fracture
		A47	Infection/inflammation nervous system
		A81	Coma unspecified
		D01	Pleural disease (incl. pleural effusion)
		I41	Serious burn/frostbite
		O04	Spleen disorder
		S42	Orthopedic complication
		S47	Shock
		S49	Internal trauma
X003	Other major chronic (incl congential anomaly heart)	D02	Congenital disorder of the respiratory system
		E02	Malformation of the cardiovascular system
		F01	Congenital malformation of the digestive & hepatobiliary system
		I01	Autoimmune skin disorder
X030	Other moderate chronic	C01	Congenital/acquired malformation of

Branch	Branch description	Health condition Code	Health condition description
			ear/nose/throat
		C05	Hearing loss (incl. deafness)
		H10	Other musculoskeletal & connective tissue condition
		I02	Papulosquamous disorder/psoriasis
		J03	Goitre
		J04	Hyperthyroidism
		K04	Other disease of the urinary system
		O06	Disorder of immune mechanism
		P02	Tuberculosis disease
X050	**Other minor acute (incl general symptoms)**	A46	Concussion
		A48	Nerve injury
		C03	Dental caries/gingivitis
		C07	Other disease middle/inner ear & mastoid
		C08	Meniere's/dizziness/vertigo
		E08	Uncomplicated varicose veins
		G04	Acute liver disease & other hepatic disorders
		H45	Benign/unspecified musculoskeletal neoplasm
		K02	Benign/unspecified urinary neoplasm
		S81	General signs, symptoms (not elsewhere classified)
Z000	**Users without health conditions**	n/a	Users without health conditions
Z098	**Non-users**	n/a	Non-users

Appendix C, Table C2 – Health profile groups: volumes and average costs, all three provinces

Table C2 Health profile grouping methodology: mapping branches to health conditions

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
1	S001	Palliative state (acute)	114,764	$41,142	$10,813
2	D001A	Respiratory failure with heart failure	11,348	$97,789	$24,640
3	S002A	Transplant complication	3,009	$84,369	$21,590
4	S003A	Ostomy complication	7,631	$82,777	$18,314
5	N001	Extremely low birth weight or immaturity without sig comorbidities	1,058	$18,258	$4,458
5	N002	Extremely low birth weight or immaturity with sig comorbidities	2,253	$152,240	$12,128
6	D002A	Respiratory failure without heart failure	18,682	$76,759	$16,501
7	E001C	Acute myocardial infarction with heart failure without sig comorbidities	1,365	$21,034	$6,515
7	E002C	Acute myocardial infarction with heart failure with sig comorbidities	23,461	$50,475	$15,558
8	D003C	Pulmonary hypertension	10,655	$40,094	$17,032
9	S004C	Presence of ostomy	16,833	$42,933	$12,154
10	Q001	Mental disorder resulting from brain injury or other illness without sig comorbidities	1,494	$3,654	$3,190
10	Q002	Mental disorder resulting from brain injury or other illness with sig comorbidities	11,314	$39,243	$12,630
11	A001C	Paralytic syndrome with condition other than stroke without sig comorbidities	1,852	$7,168	$2,545

Appendix C

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
11	A002C	Paralytic syndrome with condition other than stroke with sig comorbidities	11,572	$40,380	$12,160
12	Q003A	Delirium	22,221	$38,499	$14,385
13	I001C	Skin ulcer (incl. decubitus) without sig comorbidities	2,018	$3,521	$3,743
13	I002C	Skin ulcer (incl. decubitus) with sig comorbidities	30,335	$34,871	$13,440
14	J001C	Cystic fibrosis	1,606	$24,026	$13,734
15	A003C	Paralytic syndrome with stroke	15,525	$31,814	$7,946
16	S005C	Transplant recipient without sig comorbidities	1,025	$2,801	$3,739
16	S006C	Transplant recipient with sig comorbidities	7,703	$32,324	$12,540
17	E003C	Heart failure with CAD/arrhythmia without sig comorbidities	8,978	$8,520	$6,177
17	E004C	Heart failure with CAD/arrhythmia with sig comorbidities	80,398	$26,096	$12,496
18	R001	Metastatic cancer without sig comorbidities	17,794	$9,531	$5,546
18	R002	Metastatic cancer with sig comorbidities	49,453	$23,721	$10,744
19	R003	Brain cancer without sig comorbidities	1,474	$4,185	$4,040
19	R004	Brain cancer with sig comorbidities	6,264	$23,411	$10,132
20	J002C	Diabetes/hypoglycemia with PVD/oth chronic vasc dx without sig comorbidities	5,788	$4,694	$5,042
20	J003C	Diabetes/hypoglycemia with PVD/oth chronic vasc dx with sig comorbidities	39,649	$23,628	$10,923
21	H001A	Musculoskeletal infections without sig comorbidities	6,187	$3,980	$1,751
21	H002A	Musculoskeletal infections with sig comorbidities	18,493	$26,996	$8,960
22	R005	Other digestive & hepatobiliary cancer without sig comorbidities	4,831	$3,540	$3,604
22	R006	Other digestive & hepatobiliary cancer with sig comorbidities	17,591	$17,346	$9,356

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
23	P001A	Sepsis without sig comorbidities	1,431	$7,755	$1,393
23	P002A	Sepsis with sig comorbidities	18,998	$34,081	$9,022
24	Q004	Eating disorder without sig comorbidities	2,319	$13,057	$4,378
24	Q005	Eating disorder with sig comorbidities	5,638	$20,670	$7,309
25	Q006	Dementia (incl. alzheimer's) without sig comorbidities	19,539	$3,034	$5,761
25	Q007	Eementia (incl. alzheimer's) with sig comorbidities	163,309	$15,884	$7,840
26	R007	Lung cancer without sig comorbidities	4,975	$5,339	$5,783
26	R008	Lung cancer with sig comorbidities	16,215	$14,708	$9,065
27	A004C	Congenital malformation nervous syst (incl. spina bifida) without sig comorbidities	2,122	$3,455	$2,219
27	A005C	Congenital malformation nervous syst (incl. spina bifida) with sig comorbidities	4,669	$24,608	$7,845
28	E005A	Acute myocardial infarction without heart failure without sig comorbidities	13,817	$15,953	$2,502
28	E006A	Acute myocardial infarction without heart failure with sig comorbidities	32,600	$24,311	$5,927
29	A006A	Stroke without paralytic syndrome without sig comorbidities	2,928	$11,574	$3,356
29	A007A	Stroke without paralytic syndrome with sig comorbidities	21,039	$20,820	$6,641
30	D004A	Pneumonia with COPD without sig comorbidities	8,942	$5,055	$4,482
30	D005A	Pneumonia with COPD with sig comorbidities	33,886	$16,472	$10,550
31	J004C	Diabetes/hypoglycemia with chronic kidney disease/failure without sig comorbidities	10,078	$3,696	$5,396
31	J005C	Diabetes/hypoglycemia with chronic kidney disease/failure with sig comorbidities	24,598	$12,389	$10,562
32	E007C	Heart failure without AMI, respiratory failure or CAD without sig comorbidities	11,843	$3,099	$4,954

Appendix C

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
32	E008C	Heart failure without AMI, respiratory failure or CAD with sig comorbidities	38,496	$10,992	$8,868
33	E009C	Cardiac valve disease without sig comorbidities	4,741	$2,942	$4,212
33	E010C	Cardiac valve disease with sig comorbidities	17,962	$16,419	$7,772
34	R009	Leukemia/lymphoma without sig comorbidities	16,962	$3,935	$3,898
34	R010	Leukemia/lymphoma with sig comorbidities	26,627	$12,489	$7,885
35	E011A	Unstable angina without sig comorbidities	6,601	$9,760	$2,510
35	E012A	Unstable angina with sig comorbidities	17,518	$17,520	$5,355
36	G001C	Pancreatic disease without sig comorb	6,812	$7,066	$2,345
36	G002C	Pancreatic disease with sig comorbidities	17,888	$16,020	$6,197
37	R011	Ovarian cancer	1,822	$14,848	$5,304
38	F001A	Major acute GI/peritonitis without sig comorbidities	15,595	$4,585	$2,101
38	F002A	Major acute GI/peritonitis with sig comorbidities	50,512	$16,395	$5,675
39	Q008	Delusional disorder (incl. schizophrenia) without sig comorbidities	50,320	$5,909	$3,415
39	Q009	Delusional disorder (incl. schizophrenia) with sig comorbidities	47,574	$15,728	$6,661
40	R012	Bladder cancer without sig comorbidities	8,760	$4,667	$3,778
40	R013	Bladder cancer with sig comorbidities	19,977	$9,936	$6,764
41	R014	Musculoskeletal cancer without sig comorbidities	2,314	$4,075	$2,675
41	R015	Musculoskeletal cancer with sig comorbidities	2,961	$12,727	$6,012
42	G003C	Cirrhosis without sig comorbidities	7,758	$2,373	$2,793
42	G004C	Cirrhosis with sig comorbidities	14,383	$10,343	$7,955
43	A008A	Other cerebral/spinal disorder (incl. benign neoplasms) without sig comorbidities	7,536	$4,285	$2,361
43	A009A	Other cerebral/spinal disorder (incl. benign neoplasms) with sig comorbidities	12,970	$13,634	$5,264

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
44	E013C	PVD/other chronic vascular dx (without diabetes/hypoglyc) without sig comorbidities	36,273	$2,987	$2,483
44	E014C	PVD/other chronic vascular dx (without diabetes/hypoglyc) with sig comorbidities	79,220	$11,626	$5,580
45	A010C	Parkinson's disease/parkinsonism without sig comorbidities	7,995	$2,350	$3,958
45	A011C	Parkinson's disease/parkinsonism with sig comorbidities	14,567	$7,202	$6,625
46	X001A	Other major acute (incl. shock, coma) without sig comorbidities	52,238	$3,939	$1,324
46	X002A	Other major acute (incl. shock, coma) with sig comorbidities	47,549	$14,688	$4,589
47	N003	Serious perinatal condition without sig comorbidities	33,733	$8,524	$1,132
47	N004	Serious perinatal condition with sig comorbidities	7,100	$28,516	$4,312
48	K001C	Chronic kidney disease/failure (without diabetes/hypoglycemia) without sig comorbidities	22,717	$3,141	$3,702
48	K002C	Chronic kidney disease/failure (without diabetes/hypoglycemia) with sig comorbidities	30,129	$8,210	$6,485
49	R016	Renal cancer without sig comorbidities	2,753	$4,660	$2,719
49	R017	Renal cancer with sig comorbidities	5,996	$10,360	$4,752
50	R018	Colorectal cancer without sig comorbidities	12,613	$4,097	$3,371
50	R019	Colorectal cancer with sig comorbidities	17,537	$9,054	$5,158
51	H003A	Major acute MSK (incl. fract femur/spine, limb amput) without sig comorbidities	37,651	$3,936	$1,528
51	H004A	Major acute MSK (incl. fract femur/spine, limb amput) with sig comorbidities	35,314	$13,493	$5,503
52	R020	Uterine cancer	7,325	$8,797	$3,185
53	R021	Oral/ear/nose/throat cancer without sig comorbidities	6,064	$3,755	$3,235

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
53	R022	Oral/ear/nose/throat cancer with sig comorbidities	5,900	$8,138	$5,367
54	A012C	Cerebral palsy without sig comorbidities	4,923	$3,182	$2,897
54	A013C	Cerebral palsy with sig comorbidities	3,480	$9,420	$4,727
55	X003C	Other major chronic (incl. congential anomaly heart) without sig comorbidities	13,735	$4,035	$2,038
55	X004C	Other major chronic (incl. congential anomaly heart) with sig comorbidities	12,148	$11,407	$4,400
56	O001A	Disease of white blood cells (incl. neutropenia) without sig comorbidities	3,870	$3,033	$1,389
56	O002A	Disease of white blood cells (incl. neutropenia) with sig comorbidities	4,379	$10,921	$3,983
57	A014C	Major chronic nervous system without sig comorbidities	24,673	$2,890	$2,287
57	A015C	Major chronic nervous system with sig comorbidities	32,413	$8,435	$4,750
58	Q010	Other major mental health (incl. bipolar disorder)	81,462	$3,873	$2,297
58	Q011	Other major mental health (incl. bipolar disorder) with sig comorbidities	60,899	$8,729	$4,384
59	A030A	Cerebrovascular disorder (excl. stroke)	7,859	$2,699	$2,597
59	A031A	Cerebrovascular disorder (excl. stroke) with sig comorbidities	14,485	$7,142	$5,063
60	O030C	Chronic blood disorder (incl. coagulation disorders, anemia)	38,864	$2,569	$1,887
60	O031C	Chronic blood disorder (incl. coagulation disorders, anemia) with sig comorbidities	72,274	$8,231	$4,312
61	J030A	Adrenal/electrolyte/metabolism disorders	33,614	$3,594	$1,723
61	J031A	Adrenal/electrolyte/metabolism disorders with sig comorbidities	40,353	$9,812	$5,070
62	D030C	COPD without pneumonia	60,001	$2,209	$2,722

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
62	D031C	COPD without pneumonia with sig comorbidities	67,197	$5,638	$4,658
63	R070	Breast cancer without sig comorbidities	39,569	$4,426	$2,414
63	R071	Breast cancer with sig comorbidities	30,954	$7,115	$3,548
64	S030C	Chromosomal/other/unspec congenital syndrome/dis without sig comorbidities	7,920	$3,134	$1,754
64	S031C	Chromosomal/other /unspec congenital syndrome/dis with sig comorbidities	4,905	$7,929	$2,984
65	E030C	CAD/arrhythmia without heart failure without sig comorbidities	246,041	$3,146	$2,297
65	E031C	CAD/arrhythmia without heart failure with sig comorbidities	228,984	$6,132	$4,025
66	F030A	Unspecified gastrointestinal hemorrhage without sig comorbidities	19,437	$2,366	$1,672
66	F031A	Unspecified gastrointestinal hemorrhage with sig comorbidities	16,375	$6,699	$3,700
67	Q070	Personality disorder without sig comorbidities	12,482	$3,789	$2,046
67	Q071	Personality disorder with sig comorbidities	6,942	$7,696	$3,388
68	R072	Prostate cancer without sig comorbidities	32,463	$3,366	$2,521
68	R073	Prostate cancer with sig comorbidities	26,510	$5,312	$3,730
69	M030	Obstetrics without sig comorbidities	432,495	$5,373	$2,030
69	M031	Obstetrics with sig comorbidities	107,619	$8,240	$2,784
70	G030A	Diseases of gallbladder & biliary tract without sig comorbidities	45,510	$5,009	$1,523
70	G031A	Diseases of gallbladder & biliary tract with sig comorbidities	25,611	$7,849	$2,801
71	H030C	Rheum/osteoarthritis & oth inflamm arthrop (excl. gout) without sig comorbidities	323,433	$3,262	$2,455
71	H031C	Rheum/osteoarthritis & oth inflamm arthrop (excl. gout) with sig comorbidities	185,081	$5,466	$3,698
72	Q072	Drug/alcohol abuse/dependence without sig	119,394	$3,314	$2,310

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
		comorbidities			
72	Q073	Drug/alcohol abuse/dependence with sig comorbidities	23,466	$5,985	$3,566
73	L030C	Endometriosis/pelvic inflammatory disease without sig comorbidities	26,358	$5,263	$1,536
73	L031C	Endometriosis/pelvic inflammatory disease with sig comorbidities	6,915	$8,787	$2,188
74	A032A	Seizure	3,402	$4,101	$1,847
75	F032A	Disease of appendix without sig comorbidities	35,235	$5,462	$889
75	F033A	Disease of appendix with sig comorbidities	7,891	$8,389	$1,773
76	K030C	Urinary incontinence (stress) without sig comorbidities	18,989	$3,475	$1,775
76	K031C	Urinary incontinence (stress) with sig comorbidities	7,388	$5,496	$3,069
77	H032C	Osteoporosis without sig comorbidities	6,653	$2,240	$1,942
77	H033C	Osteoporosis with sig comorbidities	3,228	$5,090	$3,322
78	A033A	Disorder peripheral nerv syst (incl. Carpal tunnel, bells palsy) without sig comorbidities	57,931	$3,324	$1,702
78	A034A	Disorder peripheral nerv syst (incl. Carpal tunnel, bells palsy) with sig comorbidities	22,094	$5,280	$3,036
79	F034C	Inflammatory bowel (incl. crohn's, ulcerative colitis) without sig comorbidities	51,455	$2,955	$1,927
79	F035C	Inflammatory bowel (incl. crohn's, ulcerative colitis) with sig comorbidities	12,913	$5,152	$3,044
80	P030C	Human immunodeficiency virus (HIV) infection without sig comorbidities	11,574	$1,943	$1,765
80	P031C	Human immunodeficiency virus (HIV) infection with sig comorbidities	1,763	$3,962	$2,527
81	A035C	Epilepsy without sig comorbidities	45,675	$2,404	$1,717
81	A036C	Epilepsy with sig comorbidities	5,912	$5,133	$3,430
82	B030C	Retinopathy without sig comorbidities	83,306	$1,956	$1,935

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
82	B031C	Retinopathy with sig comorbidities	64,431	$4,141	$3,121
83	L032C	Female infertility without sig comorbidities	58,126	$2,699	$2,279
83	L033C	Female infertility with sig comorbidities	4,724	$4,382	$2,718
84	H034C	Syst connective tissue dis (incl. lupus, scleroderma) without sig comorbidities	15,003	$2,317	$1,983
84	H035C	Syst connective tissue dis (incl. lupus, scleroderma) with sig comorbidities	4,625	$4,130	$3,353
85	A037C	Multiple sclerosis & oth demyelinating disease of cns without sig comorbidities	19,475	$2,178	$1,960
85	A038C	Multiple sclerosis & oth demyelinating disease of cns with sig comorbidities	2,773	$4,592	$3,236
86	S050A	Other complication of medical or surgical care	37,902	$6,002	$1,707
87	N005	Newborn/neonate with other complicating diagnosis	56,321	$5,191	$834
88	F036A	Peptic ulcer	18,397	$2,517	$1,476
88	F037A	Peptic ulcer with sig comorbidities	7,313	$4,310	$2,596
89	G032C	Chronic liver disease (incl. hepatitis)	4,383	$3,552	$1,906
90	B032C	Cataract/lens disorder	299,091	$2,819	$2,278
91	L034C	Other uterine, fallopian, ovarian condition (incl. prolapse, fistula)	90,665	$3,974	$1,734
92	A039A	Transient ischemic attack	22,306	$2,750	$2,272
93	K032A	Acute/other kidney disease/failure without sig comorbidities	91,512	$2,327	$1,474
93	K033A	Acute/other kidney disease/failure with sig comorbidities	13,838	$3,904	$2,992
94	F050A	Hernia	111,167	$3,496	$1,501
95	J032A	Diab/hypoglyc without chron kidney dis/PVD/chronic vasc dx without sig com	595,920	$1,672	$1,882
95	J033A	Diab/hypoglyc without chron kidney dis/PVD/chronic vasc dx w sig com	12,446	$2,896	$2,562

Appendix C

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
96	X030C	Other moderate chronic	195,874	$2,010	$1,218
97	F051A	Gastroesophageal reflux disease	12,466	$2,989	$1,347
98	F052A	Other disease/disorder of the digestive system	32,684	$2,839	$1,444
99	K050C	Prostate disorder	85,376	$1,929	$1,734
100	F053C	Other disease of the esophagus (incl. esophagitis)	50,166	$2,235	$1,409
101	C054A	Disease of oral cavity, salivary gland, or jaw	59,402	$2,779	$1,098
102	R074	Skin/cervical/thyroid/other & unspecified cancers	114,569	$1,997	$1,591
103	H050A	Other bone disease	46,114	$2,216	$1,369
104	O050C	Other disease of blood & blood forming organs	26,092	$1,808	$1,410
105	F054C	Diverticulitis	53,036	$2,549	$1,529
106	Q074	Depression	336,801	$1,794	$1,392
107	J050C	Other disease of the endocrine system	46,518	$1,708	$1,169
108	H051C	Musculoskeletal malformation (excl. spine)	105,872	$2,047	$1,176
109	O051A	Lymphatic system disorder (excl. spleen)	12,128	$2,017	$977
110	Q075	Developmental disorder (incl. autism spectrum)	40,944	$1,981	$976
111	A050A	Neurological, sensory, cognitive signs & symptoms (incl. insomnia)	94,352	$1,536	$1,314
112	B050A	Retinal break/detachment	13,934	$1,894	$1,225
113	E050C	Hypertension	757,155	$1,208	$1,406
114	M050A	Unsuccessful/ectopic pregnancy	61,678	$1,838	$1,173
115	Q076	Neurotic/anxiety/obsessive compulsive disorder	503,460	$1,571	$1,152
116	P050A	Other bacterial infection	35,847	$1,832	$810
117	J051C	Other disorders of metabolism	32,432	$1,275	$946
118	J052C	Malnutrition & vitamin deficiency	81,970	$1,246	$948
119	H052C	Vertebral/disk & other disease of back	243,915	$1,391	$1,099

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
120	F055A	Benign neoplasms and other rectal/anal disorders	142,789	$1,827	$948
121	H053A	Other fracture/dislocation	425,952	$1,602	$725
122	Q077	Behavioural & emotional childhood/adolescent disorder	103,156	$1,221	$843
123	C055A	Chronic upper respiratory condition (incl. tonsillitis)	164,199	$1,642	$832
124	N070	Healthy newborn	316,323	$1,731	$720
125	S051A	Reaction to transfusion, medication or anaesthesia	46,816	$973	$700
126	D050A	Pneumonia without copd	201,379	$1,177	$748
127	K051A	Minor acute urinary (incl. UTI, cystitis)	561,252	$970	$819
128	E051A	Signs, symptoms cardiovascular system (incl. chest pain, palpitations)	555,908	$1,000	$854
129	Q078	Other mental health (incl. adjustment reaction)	91,163	$845	$827
130	B051C	Other eye diseases & disorders	38,043	$823	$810
131	A051A	Migraine/other headache	45,125	$1,048	$852
132	L050C	Menstruation disorder (incl. menopause)	162,807	$1,052	$930
133	K052C	Other disease of the male reproductive system	54,999	$1,230	$571
134	J053C	Hypothyroidism	71,395	$754	$840
135	D051C	Asthma	171,854	$967	$678
136	H054A	Back pain	62,446	$664	$755
137	D052A	Minor acute respiratory (incl. cough, bronchitis)	1,059,378	$631	$592
138	F056A	Minor gi acute (incl. GI symptom, ill-defined infection)	990,244	$665	$606
139	Q079	Mental health signs & symptoms	69,284	$705	$640
140	H055A	Crystal arthropathy (incl. gout)	33,734	$481	$741
141	H056A	Myositis and soft tissue disorder (incl. muscle inflammation)	133,254	$647	$688

Appendix C

Rank	Health profile Group	Description	Number of persons during concurrent period	Average cost, in Dollars	
				Concurrent period	Prospective period
142	S052A	Drug/alcohol poisoning (not elsewhere classified)	76,499	$680	$590
143	K053C	Other disease/disorder bladder & urethra	137,720	$557	$638
144	L051A	Congenital/other female reproductive condition	68,205	$701	$671
145	J054C	Obesity	64,403	$489	$591
146	F057C	Gastritis & duodenitis	77,333	$641	$655
147	H057A	Joint/tendon disorder and injury (incl. pain, sprain, strain)	803,508	$551	$576
148	J055C	Hypercholesterolaemia and other dyslipidemia	51,987	$482	$660
149	P051A	Influenza	67,752	$352	$462
150	X050A	Other minor acute (incl. general symptoms)	449,412	$362	$495
151	S053A	Other trauma/injury (not elsewhere classified)	97,260	$476	$474
152	H058A	Neuromuscular signs & symptoms	266,032	$362	$517
153	S054A	Open wound	143,532	$426	$424
154	I050A	Minor acute skin (incl. dermatitis, cellulitis)	895,123	$358	$456
155	B052C	Chronic vision impairment	157,480	$321	$404
156	C056A	Disorder of external ear	67,105	$289	$430
157	P052A	Sexually transmitted disease (excl. HIV)	22,840	$251	$399
158	P053A	Other & unspecified infection	22,621	$254	$366
159	B053A	Minor acute eye (incl. conjunctivities)	166,870	$285	$408
160	C057A	Otitis media	124,851	$296	$352
161	C058A	Acute ent, upper respiratory condition (incl. colds, croup)	610,995	$209	$367
162	P054	Other viral infection	69,478	$199	$334
163	Z000	Users without health conditions	1,076,288	$140	$2
164	Z098	Non-users	2,655,674	$443	$212

Appendix D – National Grouping System

Categories and strata

1. Consultations

Major, initial, ordinary, minor, repeat, regional and operative consultations performed in offices, hospitals, chronic care and convalescent hospitals and nursing homes, as well as psychiatric and obstetrical consultations where no special call is involved

- Major consultations
- Other consultations

2. Major assessments

General and specific assessments, reassessments, initial visits with a complete exam, new condition seen for first time and including complete history and exam, complete specific exam depending upon the physician specialty, annual exams, newborn/premature care and special eye exams performed in any location where no special call is involved

- Office
- Hospital inpatient: Newborn
- Hospital inpatient: Other
- Hospital outpatient
- Hospital unspecified
- Unspecified
- Special eye

3. Other assessments

Partial or minor assessments, regional exams, first or subsequent or repeat exams, ordinary, pre- and post-natal care, well-baby care in any location, chronic and convalescent care, outpatient visits and other visits when physician is in the hospital, intermediate and minor assessments, partial assessments, follow-up exams and regional exams, additional patients seen during a special call and detention

- Office
- Hospital inpatient

- Hospital outpatient
- Hospital unspecified
- Unspecified — Location
- Special calls — Add
- Detention

4. Hospital care days

Regular visits up to 28/30/31 or 35/42 days, more than 28/30/31 or 35/42 days, inpatient supportive care, continuing care, concurrent care, directive care, convalescent care, palliative care and daily management

- Up to 28/30/31 or 35/42 days
- More than 28/30/31 or 35/42 days
- Other

5. Special calls

Visits at night, on Saturdays, Sundays and holidays, requiring travel to offices, homes, nursing homes, outpatient and emergency departments; also includes special visits, consultations, specific assessments and reassessments, general reassessments, ordinary home visits, home summary and specific exams

- Out-of-hours/emergency
- Other regular hours

6. Psychotherapy/counselling

Individual psychotherapy, hypnotherapy, narco-analysis, diagnostic/therapeutic interviews, group and family psychotherapy and interviews, interviews for physical medicine and counselling for drugs, family, genetic, marriage and contraception, and case conferences on behalf of patients with allied workers, teachers, clergy, etc.

- Individual psychotherapy
- Group/family psychotherapy
- Counselling

7. Major surgery

Based on the 1988 Ontario Schedule of Benefits, these procedures have a fee of more than $75.

Mastectomy and breast surgery

Simple, radical or modified radical; unilateral or bilateral; female or male; partial mastectomy or wedge resection; breast augmentation, reduction and mammoplasty

Breast tumour excision/biopsy

Tumour or tissue for biopsy and/or treatment

Other integumentary system

All other major surgery procedures performed on the integumentary system not listed above; for example, excisions, lesions, tumours, cysts, burn and skin grafts, pedicle and free island flaps, plastic planing, plastic surgery procedures

Fractures

Bone and joint fractures, including dislocations

Disc surgery

Procedures for disc removal and fusion, including discectomy

Arthroplasty — Hip

Total hip replacement, unipolar and bipolar arthroplasty and revisions

Arthroplasty — Knee

Knee arthroplasty and revisions

Other musculoskeletal system

All other major surgery performed on the musculoskeletal system not listed above, including bone grafts, arthrodesis, amputation, arthrotomy, bone, joint, muscle and tendon excision, reconstruction, orbito-cranial surgery and instrumentation

Sub-mucous resection

Septoplasty and resection

Rhinoplasty

Correction of nasal deformity

Other respiratory system

All other major surgery performed on the respiratory system not listed above, including excisions and repairs

Coronary artery bypass/repair

Coronary artery repair single, double, triple (or more) bypass; coronary endarterectomy

Coronary angioplasty

Percutaneous transluminal coronary angioplasty

Cardiac electrophysiology

Electrophysiological study (partial or full); HIS bundle; atrial pacing; catheter ablation; and insertion of pacemaker, defibrillator and VAD devices, including insertion of permanent endocardial electrodes, implantation of pack, permanent replacement and repair

Other heart/pericardium

All other major surgery procedures performed on the heart and pericardium system not listed above, including valves, septa of heart and heart transplant

Varicose veins/vein repair

Ligation and stripping of varicose veins but excluding injection of varicose veins; vein repair, venous anastomosis and suture of veins

Carotid endarterectomy

Endarterectomy, body tumour and bypass graft of the carotid artery

Other cardiovascular

All other major surgery performed on the cardiovascular system not listed above

Appendectomy

Excision of the appendix

Laparotomy

Any laparotomy performed as a surgical procedure

Gallbladder and biliary tract

All cholecystectomies and any additional payments made for other procedures performed at the same time; gallbladder removal, including cholecystostomy, cholecystotomy and all operations

on the gallbladder, bile duct and biliary tract; includes lithotripsy/litholapaxy performed on the biliary tract and/or bile duct

Tonsillectomy

Both adult and child tonsillectomies and payments for adenoidectomies performed at the same time

Hernias

All forms of hernia repair surgery

Colon and intestines (colectomy)

Total and hemi-colectomies and payments for other procedures performed at the same time; surgical removal of all or part of the large intestine and other intestine surgery, entero-enterostomy, enterostomy, enterotomy or cecostomy; colostomy and colo-colotomy; colon surgeries not including colonoscopy (see NGS 103); small intestine, small bowel surgeries

Rectum/proctotomy/anus and hemorrhoidectomy

Hemorrhoidectomies by cryotherapy and banding are excluded; includes proctotomy and other anus, rectum, rectal prolapse and peri-rectal tissue operations

Other digestive

All other major surgery performed on the digestive system not listed above; includes operations on lips, mouth, throat (excluding tonsillectomy), esophagus, stomach, pancreas and liver and gastrectomy, splenectomy, vagotomy, liver transplant and hepatectomy

Urinary system

All urinary system operations, including renal transplant, nephrotomy, nephrectomy, bladder, ureter and stress incontinence operations (male or female); includes all forms of lithotripsy/ litholapaxy performed for renal/upper ureteral stone removal

Prostate surgery (male)

All forms of prostate surgery, such as perineal, suprapubic, retropubic, transpubic and transurethral resection

Vasectomy (male)

Unilateral or bilateral ligation

Other male genital system

All other major surgery performed on the male genital system not including the prostate and vasectomies

Prolapse in female genital system

All forms of prolapse repair surgery in female genital system; includes cystocele and/or rectocele repair, prolapse uterus and LeFort's operation

Hysterectomy (female)

Total, subtotal, abdominal or vaginal or radical hysterectomies

Sterilization (female)

Tubal occlusion/interruption/removal by any method or approach

Other female

All other major surgery performed on the female genital system not listed above

Cataract surgery

All forms of cataract surgery, dislocated lens extraction and insertion of intra-ocular lens when paid in addition to the above

Light coagulation

Photocoagulation and cryopexy

Tympanoplasty

Tympanoplasty, myringoplasty, tympanomastoidectomy and payments for other procedures performed at the same time

Other eye/ear

All other major surgery performed on the organs of special senses not listed above

Other major surgery

All other major surgery not listed above, such as operations on the nervous system, on the hemic and lymphatic systems and on the endocrine system

8. Minor surgery

Incision, abscess, etc.

Incision of abscesses or hematomas under local or general anesthesia

Removal of foreign body

Foreign body removal under local or general anesthesia

Excision, tumour, etc.

Excision of verruca, papilloma, keratosis, pyogenic granuloma, moles, etc.

Suture wound

Repair, debridement and dressing

Excision of nail

Excision and/or destruction of fingernail or toenail

Chalazion

Single or multiple under local or general anesthesia

Myringotomy

Unilateral myringotomy with insertion of ventilation tubes

Minor fractures

Fractures with a fee less than $75

Other minor surgery

All other minor surgery not listed above

9. Sugical assistance

All services and payments for surgical assistance

10.Anesthesia

Nerve blocks

All forms of nerve blocks

Other anesthesia

All services and payments for anesthesia, excluding nerve blocks

11.Obstetrical services

Service at delivery

Attendance at delivery or Caesarean section, repair of third-degree/vaginal/cervical laceration, removal of retained placenta, scalp sampling, fetal monitoring and induction of labour

Delivery (excluding Caesarean section)

Delivery and multiple births, excluding Caesarean sections

Caesarean section

The procedure only

Therapeutic abortion

Therapeutic abortions only

Other obstetrical services

Fetoscopy, stress test, hypertension, fetal transfusion, toxemia of pregnancy, oxytocin challenge test, abortions (missed, threatened, without dilatation and curettage, incomplete, menstrual extraction and spontaneous), amniocentesis, ectopic pregnancy, suture for incomplete cervix during pregnancy, uterine inversion and emergency removal of sutures

12.Diagnostic/therapeutic services

ICU/resuscitation

Intensive care and resuscitation services

Allergy/hyposensitization

Tests for allergies and hyposensitization

Injection/aspiration of joint

Injection and/or aspiration of joints

Electrocardiogram

Services and payments for the professional component of electrocardiograms and the payments for the technical component

Esophagoscopy/gastroscopy

Services and payments for both these endoscopies as well as payments for procedures performed at the same time as the endoscopy

Laryngoscopy/bronchoscopy

Services and payments for both these endoscopies as well as payments for procedures performed at the same time as the endoscopy

Colonoscopy

Services and payments for this endoscopy as well as payments for procedures performed at the same time as the colonoscopy

Cystoscopy

Services and payments for this endoscopy as well as payments for procedures performed at the same time as the cystoscopy

Sigmoidoscopy

Services and payments for this endoscopy as well as payments for procedures performed at the same time as the sigmoidoscopy

Other endoscopy

All other endoscopies not listed above

Coronary angiography

Coronary angiography only

Procedures associated with imaging

Therapeutic radiology and radioisotopes

Dilatation and curettage

Dilatation and curettage and payments for procedures performed at the same time

Electroencephalography

Services and payments for the professional component of electrocardiograms and payments for the technical component

Cryotheraphy

Any cryotherapy identified in the payment schedules

Cardiac catheterization

Left heart, right heart and selective coronary catheterization

Biopsy

All non-surgical biopsies

Other diagnostic/therapeutic services

All other diagnostic/therapeutic services not listed above

13.Special services

Injections and immunizations

Injections (subcutaneous, intramuscular and for varicose veins) and immunizations regularly performed by nursing personnel. In some jurisdictions these are treated as separate services, while in others they are included in visit fees. Not included are intravascular injections performed by physicians and lumbar myelograms, which are included in diagnostic/ therapeutic services.

Papanicolaou smear

As with injections, this procedure is not always considered to be a separate service and is sometimes included in a visit fee

Insertion of intrauterine device

As with injections, this procedure is not always considered to be a separate service and is sometimes included in a visit fee

14.Miscellaneous services

Other identified

These are services that are not listed as fee items by a majority of the provinces and territories. Examples include sessional and standby fees (where identified by a fee code), mileage, telephone consultations, sexual assault exam for investigation and/or confirmation of alleged sexual assault and other services.

Unidentified

Most provinces and territories have coding errors or list codes that are unidentifiable. These include all services that are unidentified or are identified but paid for by social services, the attorney general, workers' compensation, etc. Examples include services related to impaired driving, to rape victims and as a result of injuries sustained at work.

References

1. Canadian Institute for Health Information. *National health expenditure trends 1975–2015, complete data tables.* Accessed November 9, 2016.

2. Canadian Institute for Health Information. *Inpatient hospitalizations, surgeries, newborns and childbirth indicators, 2014-2015,* March 2016.

3. Canadian Institute for Health Information. *Health care cost drivers: the facts*, November, 2011.

4. Canadian Institute for Health Information. *Highlights of 2011-2012 emergency department visits and inpatinet hospitalizations,* June 2013.

5. Canadian Institute for Health Information. MS Excel File titled CCRS Quick Stats, 2015–2016.xls. Accessed November 14, 2016 at https://www.cihi.ca/en/quick-stats.

6. Canadian Association of Retired Persons. Nothing specific for seniors in Ontario budget – could have done better in an election year, March 30, 2011. Accessed November 14, 2016.

7. Ontario Ministry of Health and Long-Term Care. Ontario Protecting Quality in Long-Term Care Homes, Health Bulletin May 30, 2012. Accessed November 14, 2016.

8. Canadian Life and Health Insurance Association Inc. CLHIA report on long-term care policy, improving the accessibility, quality and sustainability of long-term care in Canada. Accessed November 14, 2016.

9. Alberta Health. Alberta Health Care Insurance Plan Statistical Supplement 2012−2013. Accessed November 14, 2016.

10. Statistics Canada. Population by year, by province and territory (Number) (CANSIM, table 051-0001). Accessed November 14, 2016.

11. Statistics Canada. Components of population growth, Canada, provinces and territories (CANSIM, table 051-0004). Accessed November 14, 2016.

12. Statistics Canada. Migration: Interprovincial, 2009/2010 and 2010/2011 (Table 3) and Migration: Interprovincial, 2011/2012 (Table 2). Accessed November 14, 2016.

13. Canadian Institute for Health Information. *Physicians in Canada, 2015.* August 23, 2016.

14. Morris JN, Hawes C, Mor V, Phillips C, Fries BE, Nonemaker S, Murphy K, *Resident Assessment Instrument (RAI) RAI-MDS 2.0 user's manual, Canadian version*, Washington DC: interRAI 2010.

15. Germany's Federal Office of Administration [Bundesversicherungsamt]. MS Excel File titled 01_Anlage_1_ICD_Zuordnung_und_Gueltigkeitskriterien_AJ2013_Festlegung.xls. Assessed on November 14, 2016 at <http://www.bundesversicherungsamt.de/fileadmin/redaktion/Risikostrukturausgleich/Festlegungen/AJ_2013/Festlegungen_31_Abs.zip>.

16. Health Canada. Diseases and conditions. Accessed November 14, 2016.

17. Public Health Agency of Canada. Chronic diseases. Accessed November 14, 2016.

18. World Health Organization. WHO fact sheet. Accessed January 30, 2014.

19. Centers for Disease Control and Prevention. Diseases & conditions. Accessed November 14, 2016.

20. Department of Health of New York State. Chronic diseases and conditions. Accessed November 14, 2016.

21. Pope GC, Ellis RP, Ash AS, Ayanian JZ, Bates DW, et al. *Diagnostic cost group hierarchical condition category models for Medicare risk adjustment, final report*. December 2000.

22. Sauver JL, Warner DO, Yawn BP, Jacobson DJ, McGree BS, et al. *Why patients visit their doctors: assessing the most prevalent conditions in a defined American population*. Mayo Foundation for Medical Education and Research, Mayo Clin Proc. 2013;88(1):56-67

23. World Health Organization. Chronic diseases. Accessed January 30, 2014.

24. The Center for Managing Chronic Disease. What is chronic disease? Accessed January 30, 2014.

25. Kaiser Permanente. Defining a medical emergency. Accessed January 30, 2014.

26. Lix L, Yogendran M, Burchill C, Metge C, McKeen N, Moore D, Bond R. *Defining and validating chronic diseases: an administrative data approach.* Winnipeg, Manitoba Centre for Health Policy, July 2006.

27. Public Health Agency of Canada. Canadian Chronic Disease Surveillance System Conditions (MS Excel file titled ccdssv2013-2ng.xlsx). Accessed April 1, 2015.

References

28. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JH, Pope JE. *Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using medical records of rheumatologists.* Arthritis Care & Research. 2013;65(10):1582-1591.

29. Quan H, Khan N, Hemmelgarn BR, Tu K, Chen G, Campbell N, et al. *Validation of a case definition to define hypertension using administrative data*. Hypertension, Journal of the American Heart Association, pp 1243-48. December 2009.

30. Robitaille C, Bancej C, Dai S, Tu K, Rasali D, Blais C, et al. *Surveillance of ischemic heart disease should include physician billing claims: population-based evidence from administrative health data across seven Canadian provinces*. BMC Cardiovascular Disorders, 2013, 13:88.

31. Lix LM, Yogendran MS, Shaw SY, Burchill C, Metge C, Bond R. *Population-based data sources for chronic disease surveillance*. Chronic Diseases in Canada. Vol 29, No 1, 2008.

32. American Thoracic Society. Quality of life resource: functional status. Accessed October 15, 2015.

33. Canadian Institute for Health Information, *Job Aid: RAI-MDS 2.0 Outcome Scales.* January 2010.

34. Canadian Institute for Health Information. The INSPQ Deprivation Index for Health in Canada: Applications for Research, Policy and Practice. Workshop Proceedings Report. Accessed October 8, 2015.

35. Statistics Canada. Postal Code Conversion File (PCCF), Reference Guide: Overview. Accessed November 4, 2015.

36. Institut national de sante publique. Deprivation index, Canada, 2006. Accessed October 8, 2015.

37. Institut national de sante publique. Methodology guide "The material and social deprivation index: a summary". Accessed October 8, 2015.

38. St. Michael's Hospital, Centre for Research on Inner City Health. Canadian Marginalization Index user guide, version 1.0. Accessed October 13, 2015.

39. Statistics Canada. Table 109-5355 — Estimates of population (2011 Census and administrative data), by age group and sex for July 1st, Canada, provinces, territories, health regions (2015 boundaries) and peer groups, annual (number). CANSIM database. Accessed April 1, 2016.

40. Neter J, Wasserman W, Kutner MH. *Applied linear statistical models - regression, analysis of variance and experimental designs*. pp 407-409. Richard D. Irwin, Inc., Illinois, USA, 1985.

41. Buchner F, Göpffarth D, Wasem J. *The new risk adjustment formula in Germany – implementation and first experiences*. White paper. May 2012.

42. Pope GC, Kautter J, Ingber MJ, Freeman S, Sekar R, Newhart C. *Evaluation of the CMS-HCC risk adjustment model, final report*. RTI International white paper. March 2011.

43. The Nuffield Trust. *Developing a person-based resource allocation formula for allocations to general practices in England*. London UK. 2010.

44. King G, Zeng L. *Logistic Regression in Rare Events Data.* P 137-163, February 16, 2001. The Society for Political Methodology publication.

cihi.ca

14350-1116