


BMJ Open Identifying factors associated with high use of acute care in Canada: protocol of a population-based retrospective cohort study

Mengmeng Zhang ,¹ Jinhui Ma,¹ Feng Xie,^{1,2} Lehana Thabane^{1,3}

To cite: Zhang M, Ma J, Xie F, *et al.* Identifying factors associated with high use of acute care in Canada: protocol of a population-based retrospective cohort study. *BMJ Open* 2020;**10**:e038008. doi:10.1136/bmjopen-2020-038008

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-038008>).

Received 24 February 2020
Revised 20 August 2020
Accepted 25 August 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

²Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Ontario, Canada

³Biostatistics Unit/FSORC, Saint Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada

Correspondence to

Dr Lehana Thabane;
thabani@mcmaster.ca

ABSTRACT

Introduction High-cost users (HCUs) account for a small proportion of the population but use a disproportionately large share of healthcare resources. Although HCUs exist in all healthcare types, acute care is the most expensive type of service and the most significant contributor to expenditures among HCUs. This study aims to determine demographic, socioeconomic and clinical factors associated with being HCUs in adult patients (≥ 18 years) receiving acute care in Canada.

Methods and analysis This is a population-based analysis using a national linked dataset. Adult patients who had at least one interaction with acute care facilities each year from 2011 to 2014 were captured in the dataset, and those living in institutions or other collective residences were not covered. The primary outcome is HCU of acute care (yes/no), which is defined as whether a patient is within the top 10% of the highest acute care cost users in his/her province. Multilevel logistic regression will be used to identify factors associated with HCU and to examine the provincial variations of these identified risk factors. Sensitivity analyses investigating the influences of different high user definitions and missing data on the study results will also be performed.

Ethics and dissemination All researchers will follow the codes and rules set by Statistics Canada and the Research Data Centre and give priority to the confidentiality of the data during and after this study. The study findings will be published in peer-review journals and disseminated at academic conferences.

INTRODUCTION

Acute care is a type of short-term care for patients who are sick or injured or in the process of recovery from treatment. It involves emergency medicine, trauma care, prehospital emergency care, acute care surgery, critical care, urgent care and inpatient care.^{1 2} Acute care is crucial to the prevention of death and disability but meanwhile the costliest healthcare type in developed countries including Canada.²⁻⁴ Recent studies revealed that acute care accounted for the largest share of healthcare expenditures (28.3%) in Canada in 2018 and was expected

Strengths and limitations of this study

- The linked dataset used in our study allows us to analyse a broad range of demographic, socioeconomic and clinical factors that are potentially associated with high use of healthcare resources.
- This study will be conducted using national data, allowing us to compare similarities and differences of the acute care high-cost user characteristics across different provinces in Canada.
- The variations of patient characteristics across different definitions on high system users will be investigated through sensitivity analyses.
- Populations who did not participate in the surveys (eg, people living in institutions or other collective residences) are not included in our analysis and thus might be under-represented.
- The information for each patient is retrieved from different databases with different data collection time, and there is a time lag between the demographic and socioeconomic information collected within the Census 2006/National Household Survey 2011 and the status of patients being high system users or non-high system users from 2011 to 2014.

to increase by 2% in 2019.^{5 6} The substantial and growing demand for acute care services places pressure on healthcare systems and calls for further research to understand the distribution and determinants of healthcare costs and strategies to reduce it.

Research shows that a small proportion of patients consume a disproportionately large amount of resources. A recently published systematic review demonstrated that roughly 68%, 55% and 24% of healthcare costs were, respectively, spent on 10%, 5% and 1% patients in developed countries or regions.⁷ In Ontario, 61% of hospital and community care expenditures were consumed by only 5% of patients.⁴ As for the utilisation of physician services, 30% of physician services were used by 5% of patients in British Columbia.⁸ These high system users (HSUs) are commonly

defined using metrics such as cumulative costs, length of stay, frequency of hospitalisations and frequency of emergency department (ED) visits.⁹ Research has demonstrated that HSUs were about 12 times more likely to have hospital admissions (69.2% vs 5.4%) and eight times more likely to die (13.0% vs 1.7%) than non-high users within 2 years following their index physician visits.⁸ Also, an increased number of ED visits among patients with mental disorders is associated with higher mortality within 2 years following their index ED visits.¹⁰ In the context of rising healthcare costs and limited resources, the poorer health outcomes and higher mortality rates in HSUs make understanding the consumption of healthcare resources by HSUs a critical step to improve the efficiency and sustainability of healthcare systems.^{4 11}

According to previous studies, acute care is the largest source and driver of spending among HSUs.¹² A study estimated that acute care accounted for 62% of the high-cost user (HCU) costs in Ontario.⁴ Meanwhile, compared with non-HCUs, HCUs were found to be more likely to use acute care.^{12–14} A range of contributors to high acute care costs has been identified in previous studies conducted in Canada, which include older age, socioeconomic disadvantages (eg, personal or family low-income status) and medical complexity (eg, a higher level of comorbidities).^{7 13–17} However, few studies have adopted a national perspective and addressed the provincial variations in characteristics of HCUs of acute care. Prior studies have shown that age, sex and socioeconomic distributions of populations and delivery of healthcare services vary across the provinces of Canada.^{18 19} Large provincial variations exist regarding patients' access and experience of healthcare services, and the performance and quality of the healthcare system.²⁰ These differences might result in variations in HCU characteristics. Furthermore, most studies were conducted based on healthcare administrative databases or health surveys, which limited their capacity to address individual socioeconomic characteristics that could be heterogeneous among clinically similar patients and amenable to interventions.^{13–17} Thus, there is a lack of evidence on risk factors, especially socioeconomic factors, associated with being HCUs of acute care at the national level, and it is unknown how these factors may vary across provinces. This study aims to bridge these knowledge gaps and to identify socioeconomic, demographic and clinical factors associated with being HCUs of acute care in adult patients (≥ 18 years) in Canada.

METHODS AND DATA ANALYSIS

Data sources and study population

We will perform a retrospective cohort study using a national linked dataset on high users, the HSUs linked to T1 Family File-Census of the Population Long-Form-National Household Survey (HSUS-TIFF-NHS, hereinafter referred as 'the linked dataset'), released by Statistics Canada and the Canadian Institute for Health Information (CIHI).¹⁷ The confidential Master Data

File for the linked dataset will be used to address our research question.²¹ The original cohorts in the linked dataset were generated anonymously from CIHI's internal datasets: Discharge Abstract Database (DAD), National Ambulatory Care Reporting System (NACRS) and Ontario Mental Health Reporting System (OMHRS). Only patients captured by DAD at least once between the fiscal years of 2011/2012 and 2014/2015 are included in the linked dataset. Those who have interactions with only NACRS or OMHRS but not DAD are not included. Using encrypted patient identity numbers, anonymised patient records from the TIFF, the 2006 Census of Population Long-Form (2006 Census 2B), and the 2011 NHS will be linked to all the cohorts at the individual level.²² The TIFF database includes all individuals who have filed a T1 tax return.²³ The cross-sectional 2006 Census 2B enumerated the entire Canadian population, including Canadian citizens, landed immigrants and non-permanent residents (people who hold a work permit or study permit) and their families living in Canada.²⁴ In this Long-form Census, 20% of private dwellings in Canada were selected. Canadian citizens living temporarily outside Canada, full-time members of the Canadian Forces stationed outside Canada and people who live in collective dwellings (eg, hospitals, nursing homes and hotels) were not included. The cross-sectional NHS provides information about all persons living in Canada except for foreign residents and those excluded from the long-form Census.²⁵ These three databases complement each other to provide sociodemographic and socioeconomic information of the included population. The DAD database contains demographic, administrative and clinical data on all discharges from acute inpatient facilities.²⁶ Thus, the linked dataset will provide combined information about the patients' hospitalisations as well as their demographic and socioeconomic characteristics.²¹

The linked dataset involves eight subset cohorts to reflect the definitions of HSUs regarding acute care cost, the total length of stay, the number of hospitalisations and the number of ED visits each year in adults (≥ 18 years) and children.²² The cut-off value of 10% is used to define HSUs across all the above cohorts. The acute care costs in the linked dataset are calculated using provincial values of Cost of a Standard Hospital Stay (CSHS) in conjunction with Resource Intensity Weights (RIW).²² The CSHS is the average full cost of treating an average acute inpatient in a hospital which measures the cost efficiency of the hospital's acute care services.²² The RIW estimates each patient's relative cost weight compared with the average acute inpatient to measure the intensity of resource use. The total length of stay, number of hospitalisations and number of ED visits are annually cumulative values for each patient. Our study will focus on the adult acute care cost cohort, which covered adult patients (≥ 18 years) who have been hospitalised in an acute care facility from the fiscal year 2011/2012–2014/2015. In this cohort, HCUs are defined as the top 10% of each province's highest acute care cost adult users (≥ 18 years old) every year

from the fiscal year 2011/2012–2014/2015.²² Non-HCUs are patients randomly selected from the remaining 90% of adult patients in the same province each year with a sample ratio of 4:1 in the linked dataset. The rest of the adult patients in this acute care cost cohort are neither HCUs nor non-HCUs. No matching between the HCU group and the non-HCU group was performed in the linked dataset.

The definitions and selections of HCUs and non-HCUs in our analysis will be consistent with the methods used in the linked dataset for the following reasons. First, the raw cost data of the linked dataset is unavailable, and patients' status of being HCU, non-HCU or neither has already been classified in the dataset provided to us. Second, although not all remaining 90% of patients are included

as comparators, the sampling ratio of 4:1 between the HCU group and the non-HCU group is adequate for the power of our study and efficient for our analysis.^{27 28} Third, given the objective of our study being risk factor identification rather than effect or hazard measurement, matching is not necessary.²⁸

Selected variables

For the primary purpose of this study, variables will be selected according to the findings of previous studies and identified from the linked dataset. The descriptions of the data sources in our study are present in [table 1](#), while the dependent variable and the potential independent variables in our study are listed in [table 2](#).

Table 1 Descriptions of data sources to be used in the analyses

Data sources	Content and features	Covered populations	Selected variables	Limitations
HSUS 2011–2014 ²¹	It provides encrypted patient numbers that can be used to link with the other databases. It also provides demographic characteristics of patients, the status of being HSU, non-HSU or neither regarding acute care cost, the length of stay, the frequency of hospitalisations and ED visits in all provinces and territories of Canada except Quebec.	Patients who have had at least one interaction with DAD within each fiscal year.	Dependent variables: patient's status of being HSU or non-HSU regarding acute care cost, length of stay, frequency of hospitalisations and frequency of ED visits; Independent variables: age, sex, province.	<ul style="list-style-type: none"> ▶ The analysis is limited to the populations who participated in the surveys, and those not covered by the surveys might be under-represented. ▶ Since there is a time lag between the information collected in earlier Census 2006 and NHS 2011 and the information collected in subsequent HSUS and DAD, there might be some inaccurate data included in the analyses. However, variables related to identity and sex should not change over time. Thus, the order of using information for each variable from these three databases in our study will be T1FF 2011–2014, NHS 2011 and then Census 2006 so that the information used in the analyses is as up to date as possible.
DAD 2011–2014 ²⁶	It provides administrative, clinical and demographic information on hospital discharges from acute care facilities in all provinces and territories of Canada except Quebec.	Patients who have had interactions with acute care facilities.	Clinical factors: admission category, diagnosis code using the International Classification of Diseases, 10th Revision, with Canadian Enhancements (ICD-10-CA, which will be used to generate Elixhauser comorbidity index and score)	
T1FF 2011–2014 ²³	It provides tax-filers' demographic and socioeconomic information, including income, spendings, savings and pension plans and funds in all provinces and territories of Canada.	All persons who completed a T1 tax return for the year of reference or who received child benefits, their non-filing spouses, their non-filing and filing children who reported the same address as their parent.	Demographic factors: marital status; Socioeconomic factors: occupation classification, the after-tax low-income status of a census family, income adequacy deciles among Canadian residents	
Census 2006 - 2B ²⁴	It is the long form of Census 2006, and it provides information regarding participants' demographic and socioeconomic characteristics, including education, ethnicity, mobility, income and employment in all provinces and territories of Canada.	20% of all the occupied private dwellings in Canada. Persons living in collective dwellings or institutions were not covered.	Demographic factors: rurality of residence, marital status, immigrant status, visible minority; Socioeconomic factors: work activity during the reference year, occupation classification, the after-tax low-income status of a census family, the highest level of education - respondent	
NHS 2011 ²⁵	It replaced the mandatory long-form Census in 2011, and it is voluntary. It provides similar information as the long form of Census, which includes the demographic and socioeconomic characteristics of people living in all provinces and territories of Canada.	All persons who usually live in Canada including those living on Indian reserves and in other Indian settlements, permanent residents, non-permanent residents such as refugee claimants, holders of work or study permits, and members of their families living with them. Foreign residents and people living in collective dwellings are excluded.	Demographic factors: rurality of residence, marital status, immigrant status, visible minority; Socioeconomic factors: work activity during the reference year, occupation classification, the after-tax low-income status of a census family, the highest level of education respondent	

DAD, Discharge Abstract Database; ED, emergency department; HSU, high system user; HSUS, High System Users database; NHS, National Household Survey; T1FF, T1 Family File.

Table 2 Dependent and independent variables selected for the study

Variable names	Variable type	Descriptions	Rationales	Data sources
Dependent variables (primary analysis)				
HCU status	Categorical	It is the status of the adult being HCU or non-HCU, as defined by the cut-off value of 10%.	The primary objective of this study is to identify risk factors associated with being HCU in adult patients receiving acute care.	HSUS 2011–2014
Dependent variables (sensitivity analysis)				
HSU status defined by the total length of stay	Categorical	It is the status of the adult being HSU or non-HSU regarding each year's sum of different lengths of stay, as defined by the cut-off value of 10%.	There are different methods to define high users of healthcare resources. Previous studies have demonstrated different definitions capture significantly different groups of people, and the inpatient system burdens they caused are different. ^{9 43}	HSUS 2011–2014
HSU status defined by the frequency of hospitalisations	Categorical	It is the status of the adult being HSU or non-HSU regarding each year's accumulative number of hospitalisations, as defined by the cut-off value of 10%.		
HSU status defined by the frequency of ED visits	Categorical	It is the status of the adult being HSU or non-HSU regarding each year's accumulative number of ED visits, as defined by the cut-off value of 10%.		
Clinical factors				
Admission category	Categorical	It is the classification of a patient on his/her admission to a healthcare facility. It could be elective, urgent/emergent, newborn, stillborn or cadaveric donor.	Previous studies have found that patients admitted to acute care facilities in emergency tend to be high users. ¹³	DAD 2011–2014
Elixhauser comorbidity index	Categorical	It is a comorbidity index with 31 categories to classify the diagnosis of patients based on ICD-10.	Previous studies have shown that certain conditions would contribute to the high use of healthcare resources and that an increase in comorbidity scores and complications are associated with increased risk of being HCUs. ^{9 13 44}	DAD 2011–2014
Elixhauser comorbidity score	Continuous	It is the score derived from algorithms assigning weights to different categories of the Elixhauser comorbidity index.		
Demographic factors				
Age	Continuous	It is how old the patient is at the time of admission and in DAD.	Multiple studies showed that older age is associated with increased risk of being HCUs. ^{6 8 12}	HSUS 2011–2014
Sex	Categorical	It is the sex of the patient as being male or female.	Some studies demonstrated that males tended to have a higher risk of being HCUs than females. However, other studies showed that females had a higher risk of being HCUs than males. ⁶	HSUS 2011–2014
Rurality of residence	Categorical	It is the rurality of the patient's residence. Even though there are four categories in NHS 2011 indicating a patient's residence in rural areas, small, medium or large urban population centres, there are only two categories in Census 2006 as rural or urban areas. We will categorise patients' residence to two categories: rural or urban areas.	The information will be used to assess the rurality of a patient's residence. Studies showed that increased rurality of a person's residence was associated with increased risk of being HCUs. ^{6 8}	NHS 2011 and Census 2006
Province	Categorical	It is the provincial/territorial government from which the Healthcare Number was issued. There are ten provinces or territories included in our study: Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Northern Canada, Nova Scotia, Ontario, Prince Edward Island and Saskatchewan.	This variable will be used to identify the province in which a person lives or works.	HSUS 2011–2014
Marital status	Categorical	It refers to whether or not a person is living in a common-law union or the legal marital status. There are two categories in our study: married or in common law and others.	Studies have shown that being married is associated with decreased risk of being HCUs. ^{6 12}	T1FF 2011–2014, NHS 2011 and Census 2006
Immigrant status	Categorical	It refers to whether the person is a Canadian-born or a non-Canadian-born.	Being an immigrant might have an impact on the risk of being HCUs. ⁶	NHS 2011 and Census 2006
Visible minority	Categorical	It refers to whether a patient is a visible minority or non-visible minority.	Multiple studies showed that ethnicity is associated with increased cost. ^{6 8 12}	NHS 2011 and Census 2006
Socioeconomic factors				
Work activity during the reference year	Categorical	It refers to whether or not a person worked during the reference year (only for persons older than 15 years old).	A study showed that a person's employment status is associated with the risk of being HCUs. ⁶	NHS 2011 and Census 2006

Continued

Table 2 Continued

Variable names	Variable type	Descriptions	Rationales	Data sources
Occupation classification	Categorical	It refers to the occupation type of patients, according to the National Occupational Classification. There are ten broad categories.	Previous studies have demonstrated the unequal distribution of health and mortality across different occupation types, which might contribute to the different levels of healthcare resource use. ⁴⁵	T1FF 2011–2014, NHS 2011 and Census 2006
After-tax low-income status of census family	Categorical	It refers to the after-tax income situation of the census family compared with the low-income measure in a reference year. A family could be classified as low-income or not low-income.	Multiple studies in Canada showed that lower-income is associated with a higher risk of being HCUs. On the contrary, studies in the US showed that higher income is associated with higher costs. ^{7 11 34 35}	T1FF 2011–2014, NHS 2011 and Census 2006
Income adequacy deciles among Canadian residents	Continuous	Values are assigned according to where each person's family after-tax gap ratio falls in the deciles ranging from 1 to 10. The higher the decile, the larger the gap ratio is. The after-tax gap ratio refers to the amount that the person's adjusted family income falls short of the low-income measure.		T1FF 2011–2014
Highest level of education - respondent	Categorical	It refers to the highest level of education that a person has successfully completed.	Studies showed that lower education degree is associated with higher costs. ^{7 15}	NHS 2011 and Census 2006

DAD, Discharge Abstract Database; ED, emergency department; HCUs, high-cost users; HSU, high system user; HSUS, High System Users database; NHS, National Household Survey; T1FF, T1 Family File.

Dependent variable (outcome measure)

The dependent variable for the primary analysis in our study is a dichotomous variable indicating whether a patient is an HCU or non-HCU of acute care. A patient is an HCU if he/she is among the top 10% of his/her province's highest cumulative acute care cost adult patients in a specific fiscal year. Non-HCUs are patients who were randomly selected from the remaining 90% of that year's adult acute care cohort. Given the fact that the characteristics of HSUs can

be different when different metrics are used to define this population,⁹ dependent variables including the status of being HSU or non-HSU defined by the total length of stay, the number of hospitalisations and the number of ED visits will be used in sensitivity analyses to examine the robustness of primary analysis results and to explore different HSU characteristics across different definitions of HSUs (table 3). All the dependent variables will be obtained from the HSUS database.

Table 3 Proposed methods for primary analysis and sensitivity analysis

Objectives	Outcome variable	Predictor variables	Method of analysis
Primary analysis			
To identify the independent factors associated with being acute care HCU and potential effect modifiers	The classification of being HCUs or non-HCUs (ie, being HSUs or non-HSUs defined by acute care cost)	<p>Clinical factors: Admission category, the Elixhauser comorbidity score.</p> <p>Sociodemographic factors: Patient's age, sex, rurality of residence, marital status, immigrant status and visible minority.</p> <p>Socioeconomic factors: Work activity, occupation classification, the after-tax low-income status of a family, income adequacy deciles among Canadians, and the highest level of education.</p> <p>Interaction terms: Comorbidity scores and age, comorbidity scores and sex, comorbidity scores and income level</p>	Mixed effects logistic regression
Sensitivity analyses			
To analyse the robustness of results when HSUs are defined using other metrics	The classification of being HSUs or non-HSUs defined by the total length of stay, frequency of hospitalisations and frequency of ED visits	<p>Clinical factors: Admission category, the Elixhauser comorbidity score.</p> <p>Sociodemographic factors: Patient's age, sex, rurality of residence, marital status, immigrant status and visible minority.</p>	Mixed effects logistic regression
To analyse the robustness of results when missing data is handled using multiple imputation	The classification of being acute care HCUs or non-HCUs (ie, being HSUs or non-HSUs defined by acute care cost)	<p>Socioeconomic factors: Work activity, occupation classification, the after-tax low-income status of a family, income adequacy deciles among Canadians, and the highest level of education.</p> <p>Interaction terms: Comorbidity scores and age, comorbidity scores and sex, comorbidity scores and income level</p>	

ED, emergency department; HCU, high-cost user; HSU, high system user.

Clinical factors

Patients' admission type and diagnosis codes will be obtained using the DAD.²⁶ The admission type reflects the circumstance under which a patient is admitted and indicates the priority and urgency of his/her admission. It could range from urgent to elective. The diagnosis codes are the International Classification of Diseases, 10th Revision, with Canadian Enhancements (ICD-10-CA) codes assigned to patients. They will be used to classify a patient's major condition to one of the 31 categories of the Elixhauser comorbidity index and to derive the Elixhauser comorbidity score with the van Walraven algorithm.^{29–31}

Demographic and socioeconomic factors

The variables that will be included in our study are listed in [table 2](#). Demographic factors include age, sex, rurality of residence, marital status, immigrant status and visible minority. Socioeconomic factors include work activity during the reference year, occupation classification, the after-tax low-income status of a family, income adequacy deciles among Canadian residents and the highest level of education. The rurality of a patient's residence will be categorised into rural or urban areas that accommodate the categorisations in Census 2006 and NHS 2011.^{24 25} As for patients' income, the after-tax income and the income adequacy deciles among Canadian residents will be used.²³ Using the low-income measure after tax (LIM-AT) as the cut-off value, a fixed 50% of median census family income with adjustment for family needs, the after-tax family income will be classified into two levels: low income and non-low-income.²³ The income adequacy deciles will be used to assess the extent to which a person's income gap ratio is compared with the LIM-AT value among the Canadian population.²³ The higher the decile is, the larger the after-tax income gap ratio is, and the less adequate the person's family income is for his/her family needs. The other variables will be classified according to the options in the dataset.

Sample size

As a general rule, the number of events per variable (EPV) should be at least 10 to prevent major problems in logistic regression (eg, overestimation or underestimation of regression coefficients).^{32 33} To be more conservative, we will use an EPV of 20 in our study. Considering that there are 12 variables with 21 degrees of freedom in our study, the minimum number of events (ie, being HCUs) is 420. Since the sampling ratio between the HCU group and the non-HCU group is 4:1 in our study, the minimum sample size will be 2100. Using the data published on the website of the linked dataset,²² the province with the smallest number of HCUs has a total number of HCUs and non-HCUs of around 7000 over the 4-year period, which is larger than the minimum sample size and sufficient to do the analysis.

Data analyses

The data analysis will involve two steps. The first step will be to compare the characteristics of HCUs and non-HCUs ([table 4](#)). Categorical variables will be summarised using count and percentage, and continuous variables will be summarised using mean and SD for normally distributed data and median and IQR for non-normally distributed data. The second step will be multilevel mixed-effect logistic modelling to identify risk factors associated with the high acute care cost. The plans for primary analysis and sensitivity analyses in this step are listed in [table 3](#). For the included clinical, demographic and socioeconomic factors, we hypothesise that older age, being male, increased rurality, being low-income, being immigrant, being visible minority, a higher level of comorbidity, certain types of conditions and occupations are associated with higher risk of being HCUs.^{7 11 34 35} Being married, higher work activity and having a certificate, diploma or degree are associated with decreased risk of being HCUs.⁷ The significance level of 0.05 will be used to identify significant factors. For each independent variable, the unadjusted OR and 95% CI will be estimated to determine if it is statistically significant ([table 5](#)).

Given that the response variable is dichotomous and the data is hierarchical within which some individuals are nested within the same province, we will conduct the mixed-effects logistic regression to address the dependence between observations and to explore the provincial variations of risk factors associated with being HCUs or non-HCUs.³⁶ The technique of multilevel model (MLM) estimating subject-specific effect rather than the generalised estimation equations (GEE) approach estimating marginal or population-averaged effect will be used in our study.^{37 38} Although both methods are commonly used for the analysis of binary outcome data violating the independence assumption of traditional regression models, MLM treats dependence between observations as interest, and it is the more efficient way to account for the dependence. In contrast, GEE treats it as a nuisance.^{37 38} With the capacity of partitioning the covariance structure of the outcomes within and between provinces, MLM is more appropriate to address our research question.

Because the management and delivery of healthcare services in different provinces in Canada are highly decentralised,³⁹ we treat province as a random effect rather than a fixed effect in our primary analysis. The analysis will model within-province and between-province variations simultaneously. Patient-level risk factors of being HCUs are listed in [table 2](#). Since province will be included in the model as a random effect, province-level predictions such as the percentage of patients older than 65 years old and gross domestic product per capita will not be included as the independent factors in our analysis. Interactions between predictors will be explored through consulting experts in health economics. Sensitivity analyses will also be performed to investigate variations when high users are defined using different metrics including the length of stay, frequency of hospitalisations

Table 4 Characteristics of HCUs of acute care, 2011/2012–2014/2015, HSUS-T1FF-CENSUS-NHS

	Non-HCUs	HCUs	Total
Age (years), mean (SD)			
Sex, %			
Male			
Female			
Rurality, %			
Urban area			
Rural area			
Marital status, %			
Married or common-law			
Other			
Province, %			
AB			
BC			
MB			
NB			
NL			
NS			
ON			
PE			
SK			
Northern Canada			
Immigrant status, %			
Canadian-born			
Non-Canadian born			
Visible minority, %			
Visible minority			
Non-visible minority			
Work activity, %			
Did not work			
Work part time			
Work full time			
Occupation category, %			
Management occupations			
Business, finance and administration occupations			
Natural and applied sciences and related occupations			
Health occupations			
Occupations in education, law and social, community and government services			
Occupations in art, culture, recreation and sport			
Sales and service occupations			
Trades, transport and equipment operators and related occupations			
Natural resources, agriculture and related production occupations			
Occupations in manufacturing and utilities			
Income status (after tax), %			
Low income			
Non-low income			

Continued

Table 4 Continued

	Non-HCUs	HCUs	Total
Income adequacy deciles among Canadian residents, median (IQR)			
Highest education, %			
No certificate, diploma or degree			
With a certificate, diploma or degree			
Elixhauser Comorbidity Index, %			
Congestive heart failure			
Cardiac arrhythmias			
Valvular disease			
Pulmonary circulation disorders			
Peripheral vascular disorders			
Hypertension, uncomplicated			
Hypertension, complicated			
Paralysis			
Other neurological disorders			
Chronic pulmonary disease			
Diabetes, uncomplicated			
Diabetes, complicated			
Hypothyroidism			
Renal failure			
Liver disease			
Peptic ulcer disease, excluding bleeding			
AIDS/HIV			
Lymphoma			
Metastatic cancer			
Solid tumour without metastasis			
Rheumatoid arthritis/collagen vascular diseases			
Coagulopathy			
Obesity			
Weight loss			
Fluid and electrolyte disorders			
Blood loss anaemia			
Deficiency anaemia			
Alcohol abuse			
Drug abuse			
Psychoses			
Depression			
Elixhauser Comorbidity Score, median (IQR)			

AB, Alberta; BC, British Columbia; HCU, high-cost user; HSU, high system user; HSUs-T1FF-CENSUS-NHS, the high system users linked to T1 family file - census of the population long-form - national household survey; IQR, interquartile range; MB, Manitoba; NB, New Brunswick; NL, Newfoundland and Labrador; NS, Nova Scotia; ON, Ontario; PE, Prince Edward Island; SD, standard deviation; SK, Saskatchewan.

and frequency of ED visits and to examine the robustness of findings when missing data are handled using different methods (complete case analysis and multiple imputation by fully conditional specification algorithm).^{23 37}

There are assumptions in logistic regression models: linearity between logit and independent variables, absence of multicollinearity and binomial distribution of errors.⁴⁰ A violation of any assumptions can result in

Table 5 Results of primary analysis and sensitivity analysis, HSUS-T1FF-CENSUS-NHS

Variables	Categories	Primary analysis OR (95% CI)	Sensitivity analyses OR (95% CI)			
			Different HSU metrics			
			Length of stay	Frequency of hospitalisations	Frequency of ED visits	Missing values
Age, years	5 year unit increase					
Sex	Male versus female					
Rurality	Rural area versus urban area					
Marital status	Married/common-law versus other					
Immigrant status	Non-Canadian born versus Canadian-born					
Visible minority	Visible minority versus non-visible minority					
Work activity	Work part-time versus did not work					
	Work full-time versus did not work					
Occupation category	Other categories versus management occupations					
Income status (after tax)	Low-income versus non-low-income					
Income adequacy deciles	1-decile unit increase					
Education	With a certificate, degree or diploma versus no certificate, degree or diploma					
Elixhauser Comorbidity Score	1-unit increase					

CI, confidence interval; ED, emergency department; HCU, high-cost user; HSU, high-system user; HSUS-T1FF-CENSUS-NHS, the high system users linked to T1 family file - census of the population long-form - National Household Survey; OR, odds ratio.

a biased or invalid effect estimate. Thus, these assumptions will be tested in our study. A smoothed scatter plot will be used to check the linearity of the logit graphically. The fractional polynomial method will also be used to test the assumption of linearity in the model.⁴⁰ If the test result is not significant, then there is linearity in the logit. For collinearity, the tolerance statistics will be used.⁴⁰ A tolerance of less than 0.20 indicates a concern of collinearity among the independent variables, and highly correlated risk factors will be removed from the model.⁴¹

The overall fit of the established model will be assessed using C-statistic.⁴⁰ A value over 0.7 indicates that the developed model is good. All the data analyses will be performed using R statistical software, V.4.0.1.

Patient and public involvement

There will be no patient or public participation in the design, conduct, reporting and dissemination of this study.

Implications of the study results

It is known that heterogeneity exists among high-need, high-cost patients.^{9 42} Our study will provide insights into the understanding of the heterogeneity and social complexity of acute care HCUs and inform HCU predictions and policy-makings. The national perspective adopted in our study will provide a full picture regarding demographic, socioeconomic and clinical characteristics of acute care HCUs for Canada. By examining potential provincial variations of HCU characteristics in our study, the findings could be potentially useful to inform decision making at the provincial level. It is possible that HCUs in Northern Canada could be younger with more acute disorders compared with HCUs in other provinces, which may require more strategies targeting the prevention of acute disorders rather than the management of chronic disorders. By exploring the variations of high user characteristics across different definitions of HSUs, the results can be used to develop management strategies with specific target metrics such as hospital admission or ED visits.

Ethics and dissemination

The researchers in this study will follow the Code of Conduct and the Values and Ethics Code of Statistics Canada and the security and confidentiality requirements of the Research Data Centre (RDC) at McMaster University. Only researchers who are listed on the Microdata Research Contract and have completed the personnel security clearance can examine the data using authorised computers in the RDC. High priority will be given to the confidentiality of respondents' personal information in the database. All researchers will adhere to the principles of physical protection, computer protection, confidentiality vetting and 'deemed employee' responsibilities to maintain the culture of confidentiality. The manuscript will be reviewed by all researchers and submitted by the principal investigator on behalf of the research team. The study findings will be communicated in peer-reviewed journals and academic conferences to inform further research and policymaking about HCUs.

Contributors MZ and LT conceptualised the study. FX and JM were involved in the design of the study. MZ drafted the manuscript, and all the other authors contributed to the revision of it. All the authors have read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Mengmeng Zhang <http://orcid.org/0000-0001-9431-9909>

REFERENCES

- Alberta Health Services. Health care glossary, 2019. Available: <https://www.albertahealthservices.ca/about/Page12677.aspx> [Accessed 16 May 2019].
- Hirshon JM, Risko N, Calvillo EJB, *et al.* Health systems and services: the role of acute care. *Bull World Health Organ* 2013;91:386–8.
- Lee JY, Muratov S, Tarride J-E, *et al.* Managing high-cost healthcare users: the International search for effective evidence-supported strategies. *J Am Geriatr Soc* 2018;66:1002–8.
- Rais S, Nazerian A, Ardal S, *et al.* High-cost users of Ontario's healthcare services. *Heal Policy* 2013;9:44–51.
- Canadian Institute for Health Information. *National health expenditure trends, 1975 to 2018*. Ottawa, ON: CIHI, 2018. https://secure.cihi.ca/free_products/NHEX-trends-narrative-report-2018-en-web.pdf
- Canadian Institute for Health Information. *National health expenditure trends, 1975 to 2019*. Ottawa, ON: CIHI, 2019. <https://www.cihi.ca/sites/default/files/document/nhex-trends-narrative-report-2019-en-web.pdf>
- Wammes JGG, van der Wees PJ, Tanke MAC, *et al.* Systematic review of high-cost patients' characteristics and healthcare utilisation. *BMJ Open* 2018;8:e023113.
- Reid R, Evans R, Barer M, *et al.* Conspicuous consumption: characterizing high users of physician services in one Canadian Province. *J Health Serv Res Policy* 2003;8:215–24.
- Wick J, Hemmelgarn B, Manns B, *et al.* Comparison of methods to define high use of inpatient services using population-based data. *J Hosp Med* 2017;12:596–602.
- Urbanoski K, Cheng J, Rehm J, *et al.* Frequent use of emergency departments for mental and substance use disorders. *Emerg Med J* 2018;35:220–5.
- Chechulin Y, Nazerian A, Rais S, *et al.* Predicting Patients with High Risk of Becoming High-Cost Healthcare Users in Ontario(Canada). *Hcpol* 2014;9:68–79.
- Joynt KE, Gawande AA, Orav EJ, *et al.* Contribution of preventable acute care spending to total spending for high-cost Medicare patients. *JAMA* 2013;309:2572–8.
- Ronskley PE, Kobewka DM, McKay JA, *et al.* Clinical characteristics and preventable acute care spending among a high cost inpatient population. *BMC Health Serv Res* 2016;16:165.
- Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. *CMAJ* 2016;188:182–8.
- Fitzpatrick T, Rosella LC, Calzavara A, *et al.* Looking beyond income and education: socioeconomic status gradients among future high-cost users of health care. *Am J Prev Med* 2015;49:161–71.
- Rashidi B, Kobewka DM, Campbell DJT, *et al.* Clinical factors contributing to high cost hospitalizations in a Canadian tertiary care centre. *BMC Health Serv Res* 2017;17:777.
- Singer A, Kosowan L, Katz A, *et al.* Characterizing patients with high use of the primary and tertiary care systems: a retrospective cohort study. *Health Policy* 2020;124:291–7.
- Government of Canada. Canada's health care system, 2019. Available: <https://www.canada.ca/en/health-canada/services/health-care-system/reports-publications/health-care-system/canada.html> [Accessed 19 Jun 2020].
- Statistics Canada. Annual demographic estimates: Canada, provinces and territories: section 2: population by age and sex. Available: <https://www150-statcan-gc.ca.libaccess.lib.mcmaster.ca/n1/pub/91-215-x/2012000/part-partie2-eng.htm> [Accessed 19 Jun 2020].
- Health Council of Canada. Where you live matters: Canadian views on health care quality. results from the 2013 Commonwealth fund international health policy survey of the general public. Canadian health care matters, Bulletin 8. Toronto, on: health Council of Canada, 2014. Available: <https://healthcouncilcanada.ca/806/> [Accessed 20 Jun 2020].
- Canadian Research Data Centre Network. HSUS-T1FF-CENSUS-NHS (high system users linked to T1 family file - census of the population long-form - National Household Survey), 2019. Available: <https://crdcn.org/taxonomy/term/4530> [Accessed 16 Jul 2019].
- Canadian Institutes of Health Research. Dynamic cohort of complex, high system users – 2011-2015, 2019. Available: http://www.cihr-irsc.gc.ca/e/50129.html#section_3_3 [Accessed 16 Jul 2019].
- Statistics Canada. Annual income estimates for census families and individuals (T1 family file), 2019. Available: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvVariableList&Id=795956> [Accessed 16 Jul 2019].
- Statistics Canada. Census of population, 2019. Available: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=assembleDESURV&DECID=303996&RepClass=586&Id=152274&DFID=244177> [Accessed 16 Jul 2019].
- Statistics Canada. National household survey (NHS), 2019. Available: <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5178> [Accessed 16 Jul 2019].
- Canadian Institute for Health Information. Discharge Abstract database metadata (DAD), 2019. Available: <https://www.cihi.ca/en/discharge-abstract-database-metadata> [Accessed 16 Jul 2019].
- Song JW, Chung KC. Observational studies: cohort and case-control studies. *Plast Reconstr Surg* 2010;126:2234–42.
- Woodward M. *Epidemiology: study design and data analysis*. 2nd edn. Boca Raton: Chapman & Hall/CRC, 2005: 297–300.
- Elixhauser A, Steiner C, Harris DR, *et al.* Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
- van Walraven C, Austin PC, Jennings A, *et al.* A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care* 2009;47:626–33.
- Garland A, Fransoo R, Olafson K, *et al.* *The epidemiology and outcomes of critical illness in Manitoba*, 2012.
- Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.



- 33 Steyerberg EW, Eijkemans MJ, Harrell FE, *et al.* Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000;19:1059–79.
- 34 Rosella LC, Fitzpatrick T, Wodchis WP, *et al.* High-cost health care users in Ontario, Canada: demographic, socio-economic, and health status characteristics. *BMC Health Serv Res* 2014;14:532.
- 35 Rosella LC, Kornas K, Yao Z, *et al.* Predicting high health care resource utilization in a single-payer public health care system: development and validation of the high resource user population risk tool. *Med Care* 2018;56:e61–9.
- 36 Twisk JWR. *Applied multilevel analysis: a practical guide for medical researchers*. New York, NY: Cambridge University Press, 2006: 38–47.
- 37 Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? *Stat Med* 2009;28:221–39.
- 38 Hubbard AE, Ahern J, Fleischer NL, *et al.* To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* 2010;21:467–74.
- 39 Marchildon G. *Health systems in transition: Canada*. 2nd edn. Canada: University of Toronto Press, 2013: 19–39.
- 40 Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. New York, NY: John Wiley & Sons, Incorporated, 2013: 186–202.
- 41 Menard S. *Logistic regression: from introductory to advanced concepts and applications*. Thousand Oaks, CA, 2019: 125–44.
- 42 Khullar D, Kaushal R. “Precision health” for high-need, high-cost patients. *Am J Manag Care* 2018;24:396–8.
- 43 Nguyen OK, Tang N, Hillman JM, *et al.* What’s cost got to do with it? Association between hospital costs and frequency of admissions among “high users” of hospital care. *J Hosp Med* 2013;8:665–71.
- 44 Ronksley PE, McKay JA, Kobewka DM, *et al.* Patterns of health care use in a high-cost inpatient population in Ottawa, Ontario: a retrospective observational study. *CMAJ Open* 2015;3:E111–8.
- 45 Ravesteijn B, van Kippersluis H, van Doorslaer E. The contribution of occupation to health inequality. *Res Econ Inequal* 2013;21:311–32.