

Identifying subgroups of adult high-cost health care users: a retrospective analysis

James Wick MSc, David J.T. Campbell MD PhD, Finlay A. McAlister MD MSc, Braden J. Manns MD MSc, Marcello Tonelli MD SM, Reed F. Beall PhD, Brenda R. Hemmelgarn MD PhD, Andrew Stewart MSc, Paul E. Ronksley PhD

Abstract

Background: Few studies have categorized high-cost patients (defined by accumulated health care spending above a predetermined percentile) into distinctive groups for which potentially actionable interventions may improve outcomes and reduce costs. We sought to identify homogeneous groups within the persistently high-cost population to develop a taxonomy of subgroups that may be targetable with specific interventions.

Methods: We conducted a retrospective analysis in which we identified adults (≥ 18 yr) who lived in Alberta between April 2014 and March 2019. We defined “persistently high-cost users” as those in the top 1% of health care spending across 4 data sources (the Discharge Abstract Database for inpatient encounters; Practitioner Claims for outpatient primary care and specialist encounters; the Ambulatory Care Classification System for emergency department encounters; and the Pharmaceutical Information Network for medication use) in at least 2 consecutive fiscal years. We used latent class analysis and expert clinical opinion in tandem to separate the persistently high-cost population into subgroups that may be targeted by specific interventions based on their distinctive clinical profiles and the drivers of their health system use and costs.

Results: Of the 3919388 adults who lived in Alberta for at least 2 consecutive fiscal years during the study period, 21115 (0.5%) were persistently high-cost users. We identified 9 subgroups in this population: people with cardiovascular disease ($n = 4537$; 21.5%); people receiving rehabilitation after surgery or recovering from complications of surgery ($n = 3380$; 16.0%); people with severe mental health conditions ($n = 3060$; 14.5%); people with advanced chronic kidney disease ($n = 2689$; 12.7%); people receiving biologic therapies for autoimmune conditions ($n = 2538$; 12.0%); people with dementia and awaiting community placement ($n = 2520$; 11.9%); people with chronic obstructive pulmonary disease or other respiratory conditions ($n = 984$; 4.7%); people receiving treatment for cancer ($n = 832$; 3.9%); and people with unstable housing situations or substance use disorders ($n = 575$; 2.7%).

Interpretation: Using latent class analysis supplemented with expert clinical review, we identified 9 policy-relevant subgroups among persistently high-cost health care users. This taxonomy may be used to inform policy, including identifying interventions that are most likely to improve care and reduce cost for each subgroup.

A small proportion of patients account for a disproportionate amount of health care spending in developed countries with advanced health care systems. Specifically, patients in the top 1% of health care spending in North America account for 25% to 40% of total health care expenditures in a given year.^{1,2} Increasing age, receipt of end-of-life care, multimorbidity, mental health conditions and social disadvantage are each associated with high-cost health system use.² However, the heterogeneous nature of this patient population makes it difficult to develop and implement strategies to improve care and curb health care spending for all high-needs, high-cost patients.^{3,4} Further complicating this goal, two-thirds of the high-cost population in a given fiscal year are not high-cost in the following year, making it difficult to identify the appropriate target population for intervention.¹

There is growing interest in stratifying high-cost populations into subgroups to facilitate targeted interventions.⁵⁻⁹ A recent Canadian study of high-cost patients described the clinical and demographic profiles of the top 10% of individuals by cumulative cost accrued during a 2-year window.¹⁰ They identified 7 high-cost subgroups: “long-term care,” “palliative care,” “major acute illness,” “major chronic diseases,” “major cancer,” “major newborn” and “mental health.”

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: Paul Ronksley, peronksl@ucalgary.ca

CMAJ Open 2022 April 19. DOI:10.9778/cmajo.20210265

These subgroups were mutually exclusive segments of the high-cost population but were clinically heterogeneous and did not support the design of strategies to improve care and reduce costs.

A recent study of a US Kaiser Permanente cohort in northern California used latent class analysis to identify 7 subgroups of medically complex patients.¹¹ However, the authors developed their classification scheme in a subset of patients with high comorbidity and high likelihood of acute care use, rather than in the general population.

The studies of high-cost users in Canada¹⁰ and of medically complex patients in northern California¹¹ both defined their populations for a single time period (i.e., episodic health care use) rather than for those who were continually high-cost users in multiple subsequent years, and who arguably had the greatest potential for targeted interventions and cost savings.¹²

Much of the previous work in this area lacks granularity and neglects the high turnover in the high-cost population over time. Statistical techniques such as latent class analysis, applied over several consecutive years at a population level, may allow for more precise characterization of subgroups in the heterogeneous population of those with persistently high health care costs. Such an approach would facilitate the design of strategies to meet the distinctive needs of each subgroup.

To this end, we sought to identify homogeneous groups of persistently high-cost patients from longitudinal, multiyear, population-based data in Alberta, Canada, using latent class analysis and expert clinical opinion to develop a taxonomy of subgroups that may be targetable with specific interventions.

Methods

Study design and setting

We used a retrospective cohort design with population-level data from Alberta. Individuals were eligible for inclusion in the cohort based on registration in the Alberta Health Care Insurance Plan (AHCIP) for at least 2 consecutive fiscal years in the study window. The AHCIP is a publicly funded health care plan that provides coverage for medically necessary physician services, along with some dental and oral surgical services; about 99% of the Alberta population is eligible based on permanent residence in Alberta.¹³ We linked this population to administrative health care data in the province.¹⁴

We followed the guidelines of the Reporting of Studies Conducted Using Observational Routinely-collected Data (RECORD) statement.¹⁵

Study population

We identified adults (≥ 18 yr) who lived in Alberta between April 2014 and March 2019. We then calculated each individual's cumulative cost for all health care encounters in each fiscal year (April to March). To create a focused population of only the most consistently expensive patients, we defined those in the top 1% of cumulative spending as "high-cost." An individual was defined as "persistently high-cost" if their annual cumulative spending was in the top 1% of costs for at least 2 consecutive fiscal years¹ between 2014 and 2019.

Data sources

We used the previously described Alberta Kidney Disease Network administrative health data repository¹⁴ to define 4 types of health system encounters: inpatient encounters from the Discharge Abstract Database; emergency department and hospital-based ambulatory encounters from the Ambulatory Care Classification System; outpatient primary care and specialist encounters from Practitioner Claims; and medication dispensation records from the Pharmaceutical Information Network.¹⁴ Hospital and emergency department information was based on the *International Statistical Classification of Diseases and Health Related Problems, Tenth Revision* (ICD-10), and outpatient physician visits were based on ICD-9-CM coding. The accuracy and completeness of the data sources has been validated in Alberta.¹⁶ We obtained demographic information from the Alberta Health Registry, linked to neighbourhood-level 2016 Canadian census data.¹⁷

Variable definitions and classifying characteristics

We estimated encounter costs for the 4 types of health system encounters (defined above) and used these to identify persistently high-cost patients. For hospital and emergency department encounters, and for hospital-based ambulatory care encounters, we multiplied the resource intensity weight of the encounter (an estimate of the relative cost associated with elements of a given encounter, including demographic characteristics, diagnoses and procedures¹⁸) by the mean provincial cost of a typical encounter. Physician visits were costed using the Alberta Health *Schedule of Medical Benefits*, which determines the reimbursement rate for physicians for each encounter, including interpretation fees. Medication costs were estimated by multiplying the quantity of medication dispensed by the provincial formulary's list price for that medication (including relevant dispensing fees and commercial markups). We estimated all costs using information for the fiscal year in which they took place.

We used Andersen's Behavioral Model of Health Care Utilization to identify a set of relevant variables to inform our classification scheme.¹⁹ These included characteristics that predisposed an individual to health system contact (age, sex, urban or rural status, history of homelessness); enabling characteristics that influenced the likelihood of (further) health care utilization (primary care provider attachment, previous specialist visits, long-term care residence); and characteristics related to health care need (e.g., pre-existing chronic conditions).¹⁹

Demographic characteristics that predisposed individuals to health system contact were available from the Alberta Health Registry and the Canadian census, except for history of homelessness, which we defined based on the presence of the ICD-10 code Z59.0 in hospital or emergency department records in the previous year, or 45 days subsequent to the index date.²⁰

Enabling characteristics that influenced the likelihood of (further) health care utilization were as follows. We defined primary care provider attachment as the proportion of all primary care visits made to the most commonly visited provider

in a 2-year period, categorized as less than 50%, 50% to 74%, and 75% or more of all visits.²¹ We defined previous specialist visits based on at least 2 visits to providers of the same type within the previous year (defined for 30 classified specialist types and accounting for 98% of all encounters that were not primary care). We defined a history of long-term care residence as a discharge disposition from hospital to a continuing care facility; a “delivery centre” type of “long-term care” for an outpatient physician visit record; or residence in a postal code assigned to a continuing care facility.

For characteristics related to health care need, we ascertained the presence of 13 chronic comorbidities using validated algorithms in the administrative health data and ICD-9 or ICD-10 coding. We included alcohol misuse, depression, schizophrenia, chronic obstructive pulmonary disease, dementia, diabetes, coronary artery disease, heart failure, hypertension, chronic kidney disease, autoimmune conditions, stroke and cancers.²² The data sources for all variables are included in Appendix 1, Table S1, available at www.cmajopen.ca/content/10/2/E390/suppl/DC1.

Data analysis

Latent class analysis

Among persistently high-cost individuals, we used latent class analysis to identify and group patients who shared similar characteristics. The value of latent class analysis lies in its ability to classify individuals in a heterogeneous population into (latent) subgroups where no further context beyond the content of the data are available.²³ In statistical terms, latent class analysis is a maximum likelihood method in which observed variables are used to categorize individuals from a heterogeneous sample into otherwise unobserved homogeneous “classes” through patterns of conditional probabilities for each individual.²³ A latent class model estimates the proportion of the overall cohort that fits best in each class and the probability of each classifying characteristic in a particular latent class.

Using the classifying characteristics described above, we began by specifying a model with 2 classes (representing subgroups) and added classes iteratively until the Bayesian information criterion was minimized, indicating the appropriate number of classes to maximize model fit and minimize the potential for overfitting.²⁴ Subgroups were mutually exclusive, in that each individual in the cohort was assigned to a single subgroup that best represented their characteristics.

To aid in the identification of important characteristics for each subgroup, we tabulated the prevalence of classifying characteristics, including demographics, measures of system use, relational continuity of care and specialty involvement, and comorbidities; we also calculated the ratio of the prevalence of the characteristic in the subgroup compared to its overall prevalence across all subgroups. We presented the differences across subgroups for select characteristics graphically by plotting the standardized differences between each subgroup and the overall cohort. We determined the proportion

of cost attributable to each type of encounter (inpatient, emergency department, outpatient, medication), as well as the most costly and most frequent diagnoses and medications within each subgroup. All analyses were conducted using Stata 16 (StataCorp).

Expert opinion

Used in isolation, latent class analysis is open to issues of reification, where the set of identified subgroups may be erroneously interpreted to form a complete and literal representation of all types (i.e., classes) of individuals in a population.²⁵ Such interpretation is at odds with clinical context, because patients with disparate classifying characteristics may require similar clinical care. Combining the results from latent class analysis of administrative data with clinical judgment ensures that the identified subgroups are clinically relevant and potentially actionable. It also ensures that experts start with an objective view of the population that is unbiased by clinical gestalt and pre-existing notions. To this end, we developed a method of combining the previously described latent class analysis with expert opinion.

We assembled a panel of 6 content experts with expertise in epidemiology (B.H., P.R.), health economics (B.M.), health policy (B.M., R.B.) and health services research (B.H., D.C., B.M., M.T., P.R.); 4 of the panel members were also specialists in internal medicine (endocrinology [D.C.], general internal medicine [F.M.] and nephrology [B.H., B.M., M.T.]). Panel members were invited to participate based on their clinical and research expertise, and they were part of a larger interdisciplinary research network, the Interdisciplinary Chronic Disease Collaboration.

The panel reviewed the compiled analyses to determine key similarities and differences between subgroups, whether each subgroup represented a clinically meaningful population, and whether multiple subgroups identified in the latent class analysis could be combined based on similar characteristics, clinical need and potential care pathways. We repeated this process until we achieved consensus, at which point we described the demographic and clinical characteristics, details of health system use and costs for each of the finalized subgroups, using proportions or means with standard deviations.

Ethics approval

The University of Calgary Conjoint Health Research Ethics Board approved this study.

Results

For each fiscal year from April 2014 to March 2019, we identified 3.4 to 3.7 million adults (≥ 18 yr) who lived in Alberta. We identified 136 554 (3.5%) adults who were high-cost in at least 1 fiscal year, and 21 115 (0.5%) who were persistently high-cost, based on high-cost status in at least 2 consecutive fiscal years (99th percentile of cost in a given year 2014/15: \$46 399; 2015/16: \$47 367; 2016/17: \$48 145; 2017/18: \$48 908; 2018/19: \$49 783; Figure 1; Appendix 1).

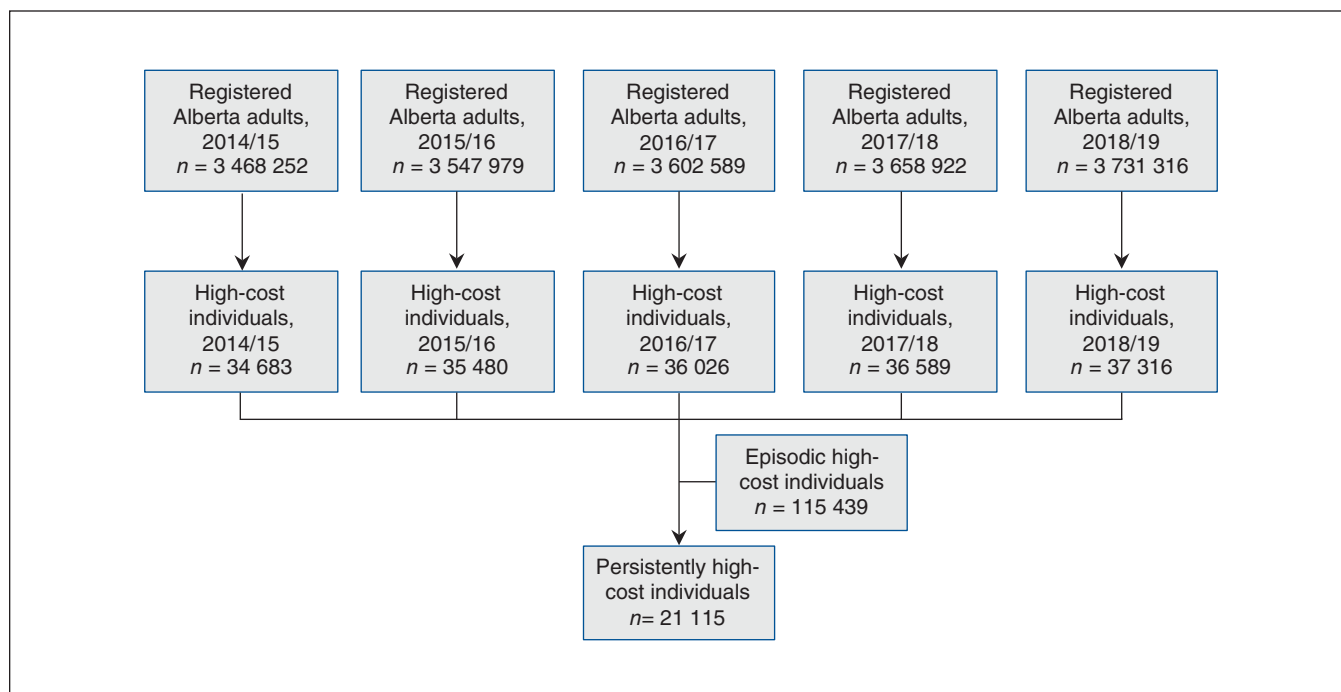


Figure 1: Identification of persistently high-cost individuals across at least 2 consecutive fiscal years. Persistently high-cost users were those in the top 1% of health care spending in at least 2 consecutive fiscal years. Episodic high-cost users were those who were in the top 1% of health care spending for 1 year only, or for nonconsecutive years.

Derivation of persistently high-cost subgroups

Using latent class analysis, we identified 21 potential subgroups of persistently high-cost individuals based on 20 classifying variables (Appendix 1, Tables S2 and S3). Consultation with clinical experts led us to decrease the number of clinically meaningful persistent high-cost subgroups from 21 to 9, based on similarities in patient characteristics and clinical needs (Figure 2).

The taxonomy of 9 subgroups was as follows: patients with cardiovascular disease (e.g., coronary artery disease or heart failure; 21.5%, $n = 4537$); patients receiving rehabilitation after surgery or recovering from complications of surgery (16.0%, $n = 3380$); patients with severe mental health conditions (e.g., schizophrenia or other psychosis; 14.5%, $n = 3060$); patients with advanced chronic kidney disease (12.7%, $n = 2689$); patients receiving biologic therapies for autoimmune conditions (e.g., inflammatory bowel disease, rheumatoid arthritis or multiple sclerosis; 12.0%, $n = 2538$); patients with dementia and awaiting community placement (11.9%, $n = 2520$); patients with chronic obstructive pulmonary disease or other respiratory conditions (4.7%, $n = 984$); patients receiving treatment for cancers (3.9%, $n = 832$); and patients with unstable housing situations, substance use disorders or both (2.7%, $n = 575$).

Sources of health care costs

The persistent high-cost population ($n = 21\,115$) was responsible for \$2.8 billion (adjusted to 2018 dollars) in health care spending annually, and subgroup costs ranged from \$112 000 per person per year for patients receiving

biologic therapies for autoimmune conditions to \$179 000 per person per year for patients with severe mental health conditions (Table 1).

Across all subgroups hospitalizations accounted for 48.3% of costs, medication use for 26.5%, emergency department encounters for 15.0% and outpatient physician encounters for 10.2%. However, each subgroup had a distinctive distribution of health care spending. For example, among patients with dementia and awaiting community placement, 69.1% of health care spending was attributable to hospital encounters; among patients with advanced chronic kidney disease, 33.7% of costs were attributable to emergency department encounters; and among patients with severe mental health conditions, 40.3% of costs were related to prescription medications (Figure 3).

Characteristics of persistently high-cost subgroups

Age varied across the 9 subgroups, from a mean age of 41.8 years (patients receiving biologic therapies for autoimmune conditions) to 81.1 years (patients with dementia and awaiting community placement), for an overall mean \pm standard deviation of 60.2 ± 18.4 years. Sex distribution also varied, from 34.6% male (patients with dementia and awaiting community placement; $n = 871$) to 62.1% male (patients receiving rehabilitation after surgery or recovering from complications of surgery; $n = 2098$), for an overall proportion of 53.9% male ($n = 11\,371$) for the entire cohort.

Measures of health system contact varied across subgroups; overall, 42.3% ($n = 8933$) of high-cost patients had no emergency department visits in the previous year (proportions varied

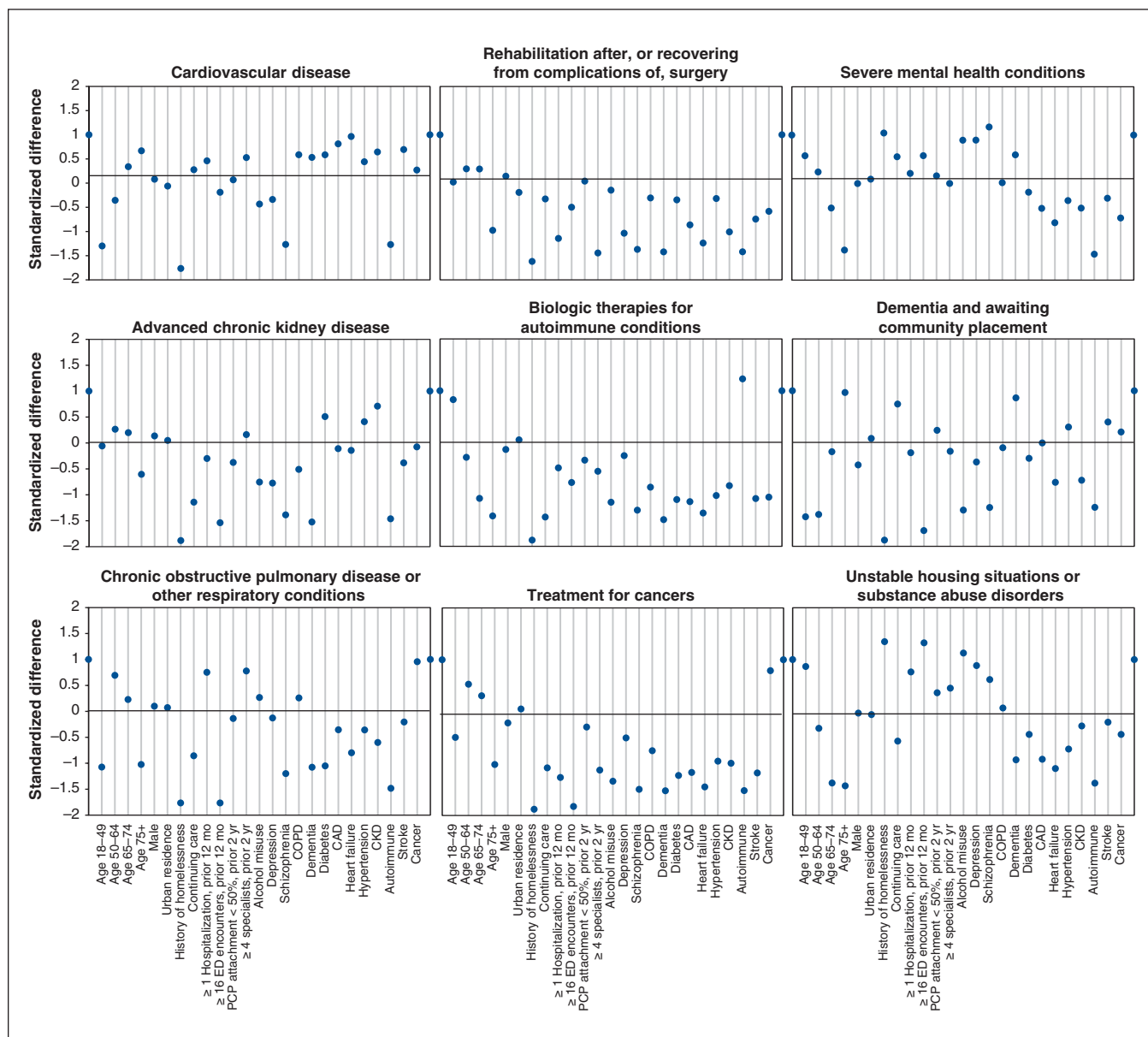


Figure 2: Standardized difference of classifying characteristics for each group relative to the overall population. Note: CAD = coronary artery disease, CKD = chronic kidney disease, COPD = chronic obstructive pulmonary disease, ED = emergency department, PCP = primary care provider.

from 0.0% to 76.7%). Approximately one-third of persistently high-cost patients had 1 or more hospitalizations in the previous 12 months (36.9%, $n = 7570$; range 4.0% to 85.7%). About two-thirds (66.5%, $n = 14037$) had poor relational continuity of care with a primary care provider (i.e., < 50% of all primary care visits with the most commonly seen provider; 45.3% in the most attached group and 96.3% in the least attached group). Almost one-fifth had not received care from a medical specialist (19.4%, $n = 4089$; range 0.0% to 72.5%; Table 1).

We tabulated the most prevalent and costly diagnoses for hospital, emergency department, and physician encounters, as well as prescribed medications for each subgroup (Appendix 1, Table S4).

Interpretation

We defined a population of persistently high-cost patients in Alberta based on high health care spending over consecutive years. Using latent class analysis supplemented by expert opinion, we defined a taxonomy of 9 clinically relevant subgroups based on demographic and clinical characteristics, relational continuity of care, and measures of system use. We examined how health care spending was distributed in each subgroup. This classification is a first step toward developing targeted interventions in the persistently high-cost population.

Previous literature exploring the persistently high-cost population has focused on specific subgroups (e.g., severe

Table 1 (part 1 of 2): Characteristics of persistently high-cost subgroups

Characteristic	Subgroup, %*									
	Overall, n (%) n = 21 115	CVD n = 4537	Rehab. or care after surgery n = 3380	Severe mental health conditions n = 3060	Advanced CKD n = 2689	Biologic therapies for autoimmune conditions n = 2538	Dementia and awaiting comm. placement n = 2520	COPD or other respiratory conditions n = 984	Treatment for cancers n = 832	Unstable housing or substance use n = 575
Age, yr										
Mean ± SD	60.2 ± 18.4	73.5 ± 11.6	56.5 ± 15.5	47.4 ± 15.7	59.0 ± 14.5	41.8 ± 13.4	81.1 ± 7.6	61.2 ± 9.7	59.1 ± 12.1	41.2 ± 11.5
18–49	5810 (27.5)	2.6	28.1	50.5	26.0	72.7	0.2	6.5	16.5	77.6
50–64	6182 (29.3)	20.5	39.4	37.1	38.2	22.1	1.1	63.2	51.1	21.2
65–74	4008 (19.0)	26.7	25.4	11.3	23.1	4.6	16.0	23.8	25.8	1.0
75+	5115 (24.2)	50.2	7.2	1.1	12.8	0.5	82.7	6.5	6.6	0.2
Male	11 371 (53.9)	58.4	62.1	53.8	61.7	47.3	34.6	59.5	43.4	52.3
Urban residence	18 130 (86.2†)	81.1	71.1	94.1	90.5	91.0	93.4	92.7	90.8	81.3
History of homelessness	486 (2.3)	0.2	0.4	7.9	0.0	0.0	0.0	0.2	0.0	38.3
Continuing care	5957 (28.2)	37.2	20.3	50.5	5.5	0.1	65.2	10.4	6.5	15.5
≥ 1 hospitalizations in the previous 12 mo	7570 (35.9)	57.8	6.8	44.2	26.6	21.7	29.6	84.3	4.0	85.7
Emergency department visits in the previous 12 months										
0	8933 (42.3)	21.9	69.6	31.8	51.7	50.6	46.2	14.2	76.7	0.0
1–4	9122 (43.2)	57.6	21.1	47.8	40.9	40.1	50.1	70.7	23.2	10.3
5–15	2575 (12.2)	18.5	7.8	16.6	6.8	8.0	3.4	14.8	0.0	60.5
16+	485 (2.3)	2.0	1.6	3.8	0.5	1.3	0.3	0.2	0.1	29.2
Primary care provider attachment in the previous 2 years										
< 50%	14 037 (66.5)	71.2	69.3	77.8	45.3	47.2	84.4	58.0	49.3	96.3
50%–74.9%	4908 (23.2)	21.8	22.5	16.7	31.4	31.9	14.4	31.3	35.8	3.7
> 75%	2084 (9.9)	6.7	7.8	5.4	22.5	19.3	1.1	10.7	14.9	0.0
Few visits	86 (0.4)	0.3	0.4	0.1*	0.8	1.5	0.0	0.0	0.0	0.0
Number of specialist types visited at least twice in the previous year										
0	4089 (19.4)	9.5	72.5	10.4	4.1	8.0	22.1	0.0	1.0	2.1
1	4470 (21.2)	13.9	19.6	24.0	20.2	36.2	20.6	0.0	49.8	8.2
2–3	7673 (36.3)	36.6	7.8	42.5	48.6	42.8	37.7	43.8	44.5	53.0
4+	4883 (23.1)	40.1	0.0	23.1	27.1	13.0	19.6	56.2	4.8	36.7

mental health²⁶) rather than on creating clear classifications of the broader population. Efforts have been made to classify nonpersistent (i.e., episodic) high-cost populations; however, because most high-cost users in any given year are not high-cost users in subsequent years, interventions targeting these groups are less feasible. For example, Anderson and colleagues¹⁰ used the population grouping methodology from

the Canadian Institute for Health Information to define subgroups of the high-cost population in Saskatchewan, Canada. They defined 16 mutually exclusive health profiles and considered 8 to be the most costly: “long-term care,” “palliative,” “major acute,” “major chronic,” “major cancer,” “major newborn,” “major mental health” and “moderate chronic,” based on measures of health system use, comorbidity and

Table 1 (part 2 of 2): Characteristics of persistently high-cost subgroups

Characteristic	Overall, n (%) [*] n = 21 115	Subgroup, % [*]								
		CVD n = 4537	Rehab. or care after surgery n = 3380	Severe mental health conditions n = 3060	Advanced CKD n = 2689	Biologic therapies for autoimmune conditions n = 2538	Dementia and awaiting comm. placement n = 2520	COPD or other respiratory conditions n = 984	Treatment for cancers n = 832	Unstable housing or substance use n = 575
Underlying medical diagnoses										
Alcohol misuse	3364 (15.9)	10.3	13.8	46.7	7.0	3.2	1.7	20.7	1.3	82.1
Depression	5001 (23.7)	16.9	6.2	70.4	10.0	18.5	16.3	20.9	14.1	68.9
Schizophrenia	1987 (9.4)	1.4	0.8	55.1	0.7	1.2	1.5	1.8	0.0	17.9
COPD	6702 (31.7)	59.3	23.3	32.2	18.7	11.6	28.8	41.1	13.7	33.9
Dementia	1510 (7.2)	12.3	0.5	13.2	0.0	0.2	19.5	2.0	0.0	2.6
Diabetes	7439 (35.2)	65.8	24.8	29.4	60.0	7.7	26.0	8.7	4.8	22.4
CAD	7289 (34.5)	88.8	12.5	20.2	31.0	6.7	34.3	24.1	6.0	11.3
Heart failure	4398 (20.8)	70.3	2.9	8.3	18.1	1.3	8.9	8.5	0.0	4.7
Hypertension	13264 (62.8)	99.3	45.5	43.8	95.8	16.4	85.2	43.6	19.0	28.0
CKD	6771 (32.1)	64.7	8.8	18.9	70.9	12.3	14.4	17.0	9.1	24.3
Autoimmune condition	1552 (7.4)	1.2	0.5	0.3	0.3	56.0	1.3	0.2	0.0	0.7
Stroke	4055 (19.2)	41.2	8.5	14.2	13.1	4.6	28.7	15.7	3.4	15.7
Cancer	3318 (15.7)	20.5	8.6	7.3	14.6	4.1	19.2	52.1	38.6	10.1
Mean per person cost, \$	136 735	121 454	133 880	178 946	143 291	112 200	119 263	150 302	178 920	119 407
Outpatient	13 542	13 953	14 028	18 227	14 149	6 808	11 905	14 119	11 937	17 909
Emergency department	19 912	27 582	15 018	12 448	46 423	8 447	9 297	16 225	14 041	15 815
Hospital	64 076	70 520	77 355	75 948	49 156	22 181	81 439	61 269	54 776	68 866
Medication	39 205	9 398	27 479	72 323	33 563	74 765	16 623	58 689	98 167	16 817
Median per person cost, \$	76 915	86 544	83 924	86 207	82 093	61 927	88 990	77 745	68 940	85 440
Outpatient	12 033	12 550	12 255	16 085	12 488	3 052	10 803	12 413	10 245	16 320
Emergency department	7 418	9 386	7 260	6 055	29 292	3 396	4 688	8 840	6 725	12 222
Hospital	52 593	60 229	61 299	58 846	34 643	2 789	70 641	49 738	42 351	52 624
Medication	4 871	4 379	3 111	5 221	5 670	52 689	2 858	6 754	9 620	4 274

Note: CKD = chronic kidney disease, comm. = community, COPD = chronic obstructive pulmonary disease, CVD = cardiovascular disease, ED = emergency department, rehab. = rehabilitation, SD = standard deviation.
 Light grey shading indicates that the ratio of the characteristic relative to the overall persistent high-cost population was less than 0.5. Dark grey shading indicates that the ratio of the characteristic relative to the overall persistent high-cost population was 2.0 or greater, or had a prevalence of 90% or higher.
^{*}Except where indicated otherwise.
[†]Data were available for n = 21 032 individuals.

socioeconomic status. Cancer and mental health were key characteristics in 2 subgroups of their population, and long-term care was a key determinant of another; each were important factors in our study as well. However, the final stratification of the high-cost population in that study lacked specificity. For example, although an individual's high-cost status may result from 1 or more chronic conditions, the strat-

egies needed to improve care or reduce cost might differ substantially depending on specific condition(s); more information would be needed to define targeted interventions.

Latent class analysis allows for the development of a more precise taxonomy. In a population of medically complex adult patients, Grant and colleagues¹¹ used latent class analysis in 1 year of data for the Kaiser Permanente Northern California

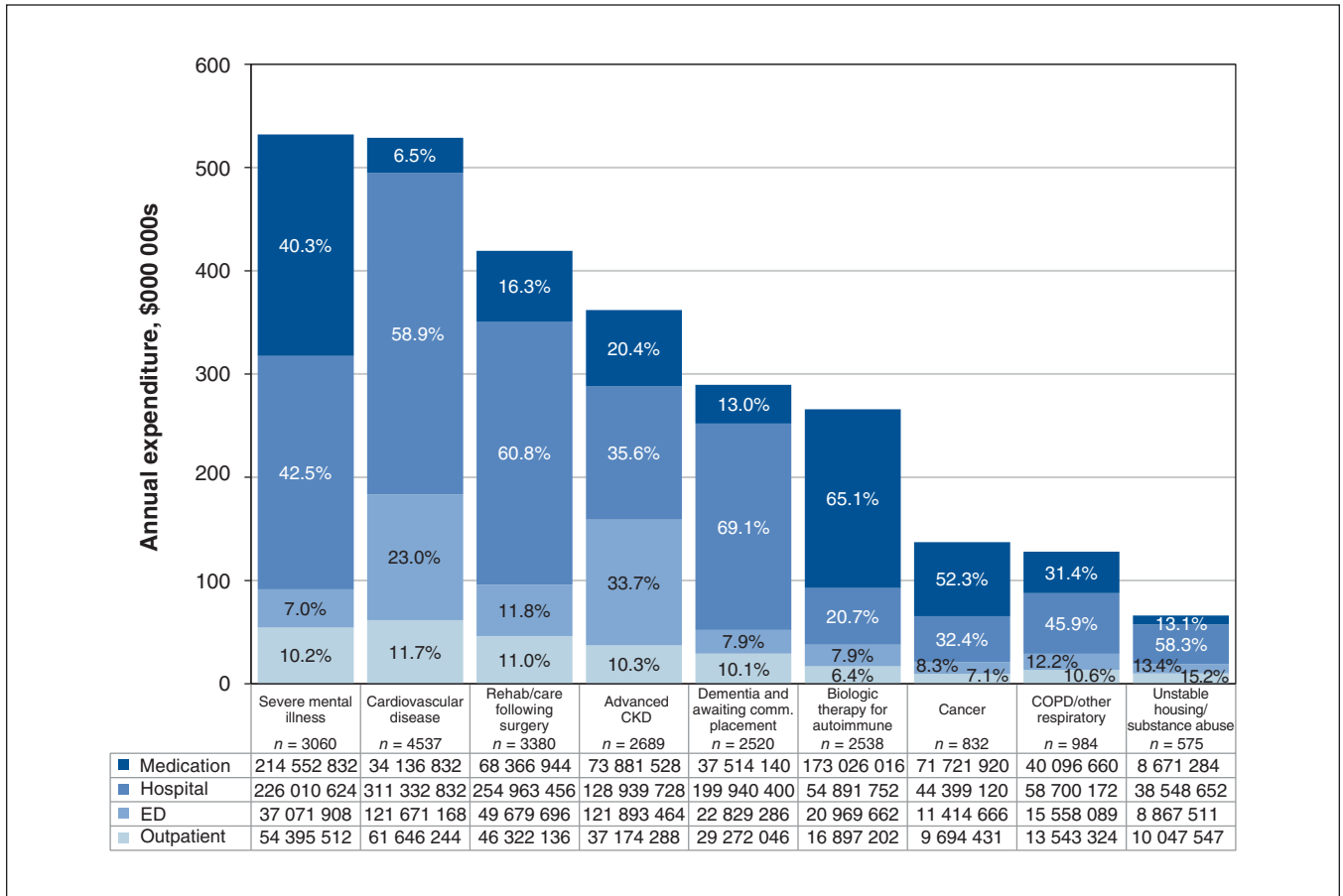


Figure 3: Breakdown of health care spending for each high-cost subgroup, by encounter type. Note: CKD = chronic kidney disease, comm. = community, COPD = chronic obstructive pulmonary disease, ED = emergency department.

population. They identified 7 classes of complex patient profiles: “less engaged,” “older patients with cardiovascular complications,” “frail elderly,” “pain management,” “psychiatric illness,” “high acuity” and “cancer treatment.” Three of these groups also appeared in our analysis: patients with cardiovascular disease, patients with severe mental health conditions and patients receiving treatment for cancers. However, rather than defining the high-cost population based on a single year, in the present study, we applied latent class methodology to a persistently high-cost population. This approach allowed us to develop a taxonomy that permitted a more precise approach to improving care for persistently high-cost patients by offering avenues for prioritization to match the right intervention to the right population at the right time.

Many interventions can be matched to each subgroup to improve patient outcomes and reduce health care costs (examples of proposed interventions for each subgroup are outlined in Appendix 2, available at www.cmajopen.ca/content/10/2/E390/suppl/DC1). These interventions relate broadly to pharmaceutical policy, decreased reliance on inpatient services and improved care in the community. For example, patients receiving biologic therapies for autoimmune conditions could benefit from a policy intervention that promotes the use of biosimilars over their more costly brand-name counterparts.

Among patients with inflammatory bowel disease, the brand-name form of infliximab costs twice as much as the equivalent biosimilar, and the switch from brand-name to generic pharmaceuticals is generally well tolerated by patients.^{27,28}

Because inpatient encounters tend to be more costly than community-based alternatives, interventions that decrease reliance on hospital services may be promising in all subgroups where viable (acknowledging that not every inpatient encounter is avoidable). Matching patients with cardiovascular disease to outpatient-based cardiac function clinics, and intensive case management programs such as hospital-at-home services may serve to improve patient outcomes and quality of life, as may improving outpatient support for self-management for individuals living with the condition.²⁹⁻³¹

For patients with advanced chronic kidney disease, a discussion of all available treatment options, including conservative care and kidney transplantation where appropriate, may reduce health care costs. Among those for whom in-centre hemodialysis is the only option, satellite hemodialysis units may reduce costs as a result of lower staffing levels.^{32,33} Such interventions, while discussed as promising in the literature, would naturally require feasibility assessment in local contexts, including consideration of barriers and facilitators to implementation.

Limitations

Our results should be interpreted in the light of the following limitations. We created a clinical classification using a population-based cohort and multiple cost sources, but other data sources — including long-term care costs and important measures of the social determinants of health (e.g., education level, food insecurity) — were not captured. Further, our definition of homelessness captured only those who received hospital-based care and may have missed a subset of this population with different health care needs and utilization patterns.

Although each individual was assigned to the group that fit them best in latent class analysis, they may have been eligible for another group. For example, a patient in the group whose health system use was driven by advanced chronic kidney disease might also have had encounters related to mental health conditions, and vice versa. Still, identifying the key characteristics of each subgroup highlighted areas for potential cost containment.

Finally, the decision to use a combination of latent class models and clinical expertise to develop our taxonomy may have introduced some subjectivity into the process. Another clinical panel may have interpreted the latent class information differently given their areas of expertise and clinical understanding of each high-cost subgroup. Further, from a statistical standpoint, this method did not provide a statistically representative Bayesian information criterion value that could be used to assess model fit. However, we believe that the face validity and clinical relevance of this final taxonomy minimizes these concerns and provides a pragmatic approach to the creation of high-cost patient subgroups that can be matched to potential interventions.

Conclusion

In this population-based study, we identified a taxonomy of 9 subgroups of persistently high-cost patients defined by differences in demographic and clinical characteristics, continuity and management of care, and measures of system use. Recognizing that no single intervention will reduce the health care needs of all persistent high-cost users, these subgroups can be used to match patients to a variety of subgroup-specific interventions that have been evaluated in the medical literature. Applying this methodology to high-cost populations has the potential to provide needed precision to care for this diverse group that drives health care spending.

References

1. Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. *CMAJ* 2016;188:182-8.
2. Wammes JJ, van der Wees PJ, Tanke MA, et al. Systematic review of high-cost patients' characteristics and health care utilisation. *BMJ Open* 2018; 8:e023113.
3. Stadhouder N, Kruse F, Tanke M, et al. Effective health care cost-containment policies: a systematic review. *Health Policy* 2019;123:71-9.
4. Blumenthal D, Chernof B, Fulmer T, et al. Caring for high-need, high-cost patients — an urgent priority. *N Engl J Med* 2016;375:909-11.
5. Davis AC, Shen E, Shah NR, et al. Segmentation of high-cost adults in an integrated health care system based on empirical clustering of acute and chronic conditions. *J Gen Intern Med* 2018;33:2171-9.

6. Powers BW, Yan J, Zhu J, et al. Subgroups of high-cost Medicare advantage patients: an observational study. *J Gen Intern Med* 2019;34:218-25.
7. Figueroa JF, Joynt Maddox KE, Beaulieu N, et al. Concentration of potentially preventable spending among high-cost Medicare subpopulations: an observational study. *Ann Intern Med* 2017;167:706-13.
8. Wong ES, Yoon J, Piegari RI, et al. Identifying latent subgroups of high-risk patients using risk score trajectories. *J Gen Intern Med* 2018;33:2120-6.
9. Keeney T, Belanger E, Jones RN, et al. High-need phenotypes in Medicare beneficiaries: drivers of variation in utilization and outcomes. *J Am Geriatr Soc* 2020;68:70-7.
10. Anderson M, Revie CW, Stryhn H, et al. Defining 'actionable' high-cost health care use: results using the Canadian Institute for Health Information population grouping methodology. *Int J Equity Health* 2019;18:171.
11. Grant RW, McCloskey J, Hatfield M, et al. Use of latent class analysis and k-means clustering to identify complex patient profiles. *JAMA Netw Open* 2020;3:e2029068.
12. Billings J, Mijanovich T. Improving the management of care for high-cost Medicaid patients. *Health Aff (Millwood)* 2007;26:1643-54.
13. *Alberta Health Care Insurance Plan (AHCIP) eligibility*. Edmonton: Government of Alberta; 2021. Available: <https://www.alberta.ca/ahcip-eligibility.aspx> (accessed 2021 Feb. 17).
14. Hemmelgarn BR, Clement F, Manns BJ, et al. Overview of the Alberta kidney disease network. *BMC Nephrol* 2009;10:30.
15. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12:e1001885.
16. Quan H, Li B, Duncan Saunders L, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008;43:1424-41.
17. *Census of population*. Ottawa: Statistics Canada; 2016. Available: <https://www12.statcan.gc.ca/census-recensement/2016/index-eng.cfm> (accessed 2021 Dec. 10).
18. Resource indicators: DAD resource intensity weights and expected length of stay. Ottawa: Canadian Institute for Health Information; 2022. Available: www.cihi.ca/en/resource-indicators-dad-resource-intensity-weights-and-expected-length-of-stay (accessed 2022 Mar. 21).
19. Andersen RM. Revisiting the behavioral model and access to medical care: Does it matter? *J Health Soc Behav* 1995;36:1-10.
20. Richard L, Hwang SW, Forchuk C, et al. Validation study of health administrative data algorithms to identify individuals experiencing homelessness and estimate population prevalence of homelessness in Ontario, Canada. *BMJ Open* 2019;9:e030221.
21. Breslau N, Reeb KG. Continuity of care in a university-based practice. *J Med Educ* 1975;50:965-9.
22. Tonelli M, Wiebe N, Fortin M, et al. Methods for identifying 30 chronic conditions: application to administrative data. *BMC Med Inform Decis Mak* 2015;15:31.
23. Porcu M, Giambona F. Introduction to latent class analysis with applications. *J Early Adolesc* 2017;37:129-58.
24. Yang CC, Yang CC. Separating latent classes by information criteria. *J Classif* 2007;24:183-203.
25. Lanza ST, Rhoades BL. Latent class analysis: an alternative perspective on subgroup analysis in prevention and treatment. *Prev Sci* 2013;14:157-68.
26. de Oliveira C, Mason J, Kurdyak P. Characteristics of patients with mental illness and persistent high-cost status: a population-based analysis. *CMAJ* 2020;192:E1793-801.
27. Moura CS, Choquette D, Boire G, et al. Inflectra and Remicade use and cost in Canada under provincial drug plans in 2016 [abstract]. *Arthritis Rheumatol* 2018;(Suppl 10):70. Available: <https://acrabstracts.org/abstract/inflectra-and-remicade-use-and-cost-in-canada-under-provincial-drug-plans-in-2016/> (accessed 2021 Feb. 17).
28. Gaither CA, Kirking DM, Ascione FJ, et al. Consumers' views on generic medications. *J Am Pharm Assoc* 2001;41:729-36.
29. Wood RL, Migliore LA, Nasshan SJ, et al. Confronting challenges in reducing heart failure 30-day readmissions: lessons learned with implications for evidence-based practice. *Worldviews Evid Based Nurs* 2019;16:43-50.
30. Shepperd S, Iliffe S, Doll HA, et al. Admission avoidance hospital at home. *Cochrane Database Syst Rev* 2008;(4):CD007491.
31. Wright SP, Walsh H, Ingleby KM, et al. Uptake of self-management strategies in a heart failure management programme. *Eur J Heart Fail* 2003;5:371-80.
32. Beaudry A, Ferguson TW, Rigatto C, et al. Cost of dialysis therapy by modality in Manitoba. *Clin J Am Soc Nephrol* 2018;13:1197-203.
33. Lee H, Manns B, Taub K, et al. Cost analysis of ongoing care of patients with end-stage renal disease: the impact of dialysis modality and dialysis access. *Am J Kidney Dis* 2002;40:611-22.

Affiliations: Department of Medicine (Wick, Campbell, Manns, Tonelli), Cumming School of Medicine; Department of Community Health Sciences (Campbell, Manns, Tonelli, Beall, Stewart, Ronksley), Cumming School of Medicine; Division of General Internal Medicine, Department of Medicine (McAlister); Department of Medicine, Faculty of Medicine & Dentistry (Hemmelgarn), University of Alberta, Edmonton, Alta.

Contributors: James Wick, David Campbell, Finlay McAlister, Braden Manns, Marcello Tonelli, Reed Beall, Brenda Hemmelgarn, Andrew Stewart and Paul Ronksley designed the study. Braden Manns, Marcello Tonelli, Brenda Hemmelgarn and Paul Ronksley acquired the data, which James Wick analyzed. James Wick and Paul Ronksley wrote the article, which David Campbell, Finlay McAlister, Braden Manns, Marcello Tonelli, Reed Beall, Brenda Hemmelgarn and Andrew Stewart reviewed. All of the authors revised the article critically for important intellectual content, gave final approval of the version to be published and agreed to be accountable for all aspects of the work.

Funding: David Campbell is supported by a Clinician Fellowship from Alberta Innovates. Finlay McAlister is funded by an Alberta Health Services Chair in Cardiovascular Outcomes Research. Braden Manns is supported by the Svare Chair in Health Economics. Marcello Tonelli is supported by the David Freeze Chair in Health Services Research. The funding sources had no role in study design, data collection, analysis, reporting, or the decision to submit for publication.

Disclaimer: The interpretation and conclusions contained herein are those of the researchers and do not necessarily represent the views of the Government of Alberta or Alberta Health Services. Neither the Government of Alberta, Alberta Health or Alberta Health Services express any opinion in relation to this study.

Content licence: This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY-NC-ND 4.0) licence, which permits use, distribution and reproduction in any medium, provided that the original publication is properly cited, the use is noncommercial (i.e., research or educational use), and no modifications or adaptations are made. See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Data sharing: This study is based in part on data provided by Alberta Health and Alberta Health Services. We are not able to make our data set available to other researchers due to our contractual arrangements with the provincial health ministry (Alberta Health), who is the data custodian.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/10/2/E390/suppl/DC1.