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

Laura C. Rosella, PhD, MHSc,*†‡ Kathy Kornas, MSc,* Zhan Yao, MSc,†
Douglas G. Manuel, MD, MSc,†§ Catherine Bornbaum, PhD, MSc,*‖ Randall Fransoo, PhD, MSc,¶#
and Therese Stukel, PhD*†

Background: A large proportion of health care spending is incurred by a small proportion of the population. Population-based health planning tools that consider both the clinical and upstream determinants of high resource users (HRU) of the health system are lacking.

Objective: To develop and validate the High Resource User Population Risk Tool (HRUPoRT), a predictive model of adults that will become the top 5% of health care users over a 5-year period, based on self-reported clinical, sociodemographic, and health behavioral predictors in population survey data.

Research Design: The HRUPoRT model was developed in a prospective cohort design using the combined 2005 and 2007/2008

Canadian Community Health Surveys (CCHS) (N = 58,617), and validated using the external 2009/2010 CCHS cohort (N = 28,721). Health care utilization for each of the 5 years following CCHS interview date were determined by applying a person-centered costing algorithm to the linked health administrative databases. Discrimination and calibration of the model were assessed using c-statistic and Hosmer-Lemeshow (HL) χ^2 statistic.

Results: The best prediction model for 5-year transition to HRU status included 12 predictors and had good discrimination (c-statistic = 0.8213) and calibration (HL χ^2 = 18.71) in the development cohort. The model performed similarly in the validation cohort (c-statistic = 0.8171; HL χ^2 = 19.95). The strongest predictors in the HRUPoRT model were age, perceived general health, and body mass index.

Conclusions: HRUPoRT can accurately project the proportion of individuals in the population that will become a HRU over 5 years. HRUPoRT can be applied to inform health resource planning and prevention strategies at the community level.

Key Words: health care utilization, high resource users, prediction

BACKGROUND

The majority of health care spending is concentrated among a small proportion of the population, irrespective of the type of health care system in which the costs are incurred.[1] Within the single-payer universal health care system in Ontario, Canada, the top 5% of health care users account for almost 50% of total health care spending.[2,3] Similar patterns in the distribution of health care expenditures have been observed in jurisdictions around the world, including the United States,[4] Australia,[5] Japan,[6] and the European Union.[1,7,8] Within the context of fiscally constrained health care systems, analytic tools that can predict who will become a high resource user (HRU) in the future at the population-level and across population groups, are needed to better inform how health care resources should be coordinated and identify targets for prevention strategies.

Common clinical and sociodemographic determinants of high users of the health system have been established in the

existing literature, and serve as a useful starting point in developing a HRU prediction tool. In particular, HRU transitions are more common in old age and with the presence of multiple chronic conditions, which typically require coordinated and continuous care.[9–12] HRU transitions are also more common for those who experience economic disadvantage, such as having low income and living in materially deprived neighborhoods,[12–14] and among those that engage in modifiable risky behaviors, such as smoking and physical inactivity.[12] The persistence of HRU status has been shown to be relatively stable.[2,15] For instance, in Ontario, almost half of those that transitioned to the top 5% of health care users remained above the 90th percentile for 2 years, and one third remained a HRU for 3 years.[2]

Existing predictive models for high health care utilization have largely been developed for applications at the individual patient-level,[16–18] for specific health sectors (eg, emergency department),[19] require information derived from sources that are not widely accessible to decision-makers (eg, electronic medical records),[17,19] or do not consider the role of socioeconomic and health behavioral determinants.[17–21] We aimed to respond to the need for tools that can be used by health analysts, planners, and decision-makers to understand how future health resource utilization will be distributed across the population, identify targets for prevention strategies, and inform health resource planning.[22]

The purpose of this study was to develop and validate a population-based risk prediction tool for transition to the top 5% of health resource users over a 5-year period, based on publically available population health survey data, including socioeconomic and health behavioral information. Moreover we aimed to develop a predictive model that could capture HRUs across the major sectors of health care spending, including inpatient hospitalizations, physician visits, complex continuing care, long-term care, home care services, and assistive devices.

METHODS

This prospective cohort study utilized linked population health surveys and health administrative data held and accessed at the Institute for Clinical Evaluative Sciences (ICES) to develop and validate a population-based model for 5-year transition to HRU status given sociodemographic, health status, and health behavioral risk factors. All persons living in Ontario have a health card number, which is encoded to enable linkages with datasets held at ICES. This study was approved by the Research Ethics Boards at the University of Toronto and Sunnybrook Health Sciences Centre.

Development and Validation Cohorts

The development cohort consisted of 58,617 Ontario respondents of the combined 2005 and 2007/2008 Canadian Community Health Surveys (CCHS), and a validation cohort consisted of 28,721 Ontario respondents to the 2009/2010 CCHS. The CCHS is a cross-sectional survey administered by Statistics Canada that collects self-reported health-related data and is representative of 98% of the Canadian population aged 12 and older living in private dwellings; the detailed survey methodology is reported elsewhere.[23] Before 2007, the CCHS operated on a 2-year collection cycle (ie, 2001, 2003, and 2005) therefore no such cycle existed in 2006.

All Ontario residents are covered by a single-payer health insurance system that is funded from general taxation and paid for by the Ontario Ministry of Health and Long-Term Care (MOHLTC). Utilization of these services from across health sectors are tracked in the health administrative data. Annual health care utilization for each of the 5 years following CCHS interview date were determined for each individual in the cohort by applying a person-centered costing approach to the linked health administrative databases.[24] The administrative databases included the Ontario Health Insurance Plan claims database, Canadian Institute for Health Information Discharge Abstract Database, National Ambulatory Care Reporting System, Continuing Care Reporting System, National Rehabilitation Reporting System, Ontario Mental Health Reporting System, Home Care Database, and Ontario Drug Benefit claims database, which captures claims for prescription drugs for individuals 65 years and older. The costing algorithm estimates costs accrued by each person according to each individual health care encounter that are covered by the single-payer government insurer, MOHLTC, including inpatient hospital stay, emergency department visits, same day surgery, stays in complex continuing care hospitals, inpatient rehabilitation, long-term care, home care, inpatient psychiatric admissions, physician services, and prescriptions for individuals eligible for the Ontario Drug Benefit program; the costing methodology is described elsewhere.[24] Individuals were then ranked in each year according to their total annual per-person health care expenditures, with HRUs defined as the top 5% of users in any given year.

Individuals were excluded from the cohorts if they were below 18 years of age (development, n = 6143; validation, n = 3206), or could not be linked to the health administrative data (development, n = 13,246; validation, n = 6694). To apply a long-term perspective in predicting individuals who are on a high spending trajectory and at risk for transitioning to HRU status we also excluded those who were determined to be a HRU in the baseline year, that is, within 1-year following their CCHS interview date (development, n = 3084; validation, n = 1511). For individuals that appeared in multiple CCHS cycles, only data collected from their first CCHS interview was used.

Risk Factors for HRUs

We examined sociodemographic, health status, and health behavioral risk factor variables from the CCHS for consideration in the development of the HRU model; variables were selected apriori based on previous epidemiology literature that characterized HRUs,[12] and categories for certain variables were combined before analysis based on similarity. We screened potential candidate variables in a univariable analysis by computing the unadjusted odds ratios (OR) for all candidate predictor variables on HRU transition in the development cohort. To determine which variables were included in the final logistic model, we examined improvements to predictive accuracy and discrimination.[25,26] Sociodemographic and health status risk factors included in the final HRU model were: sex, age category (below 30, 30–39, 40–49, 50–59, 60–69, 70–79, 80 and above), ethnicity (white, nonwhite), immigrant status

(Canadian born, immigrant <10 y, immigrant ≥ 10 years), household income quintile, household food security (moderately/severely food insecure, food secure), chronic condition (self-reported having any of the following: asthma, arthritis, back problems, migraines, chronic obstructive pulmonary disease, diabetes, hypertension, heart disease, cancer, intestinal ulcers, stroke, urinary incontinence, bowel disease, Alzheimer, mood disorder, or anxiety) and self-reported general health (excellent/very good/good, fair, poor). Health behavioral risk factors in the HRU model are detailed in Table 1. Additional risk factors examined in the univariable analysis, but not included in the final logistic model were educational attainment, marital status, diet (based on fruit and vegetable intake), perceived mental health, perceived life stress, life satisfaction, consulted a mental health professional in last 12 months, and has a regular doctor. The list of CCHS questions used to define each risk factor variable can be found in the supplementary content (Table, Supplemental Digital Content 1, http://links.lww.com/MLR/B499).

Development of the HRU Model

We used logistic regression modeling to analyze transition to HRU status up to 5 years following CCHS interview.

TABLE 1. Definitions of Health Behavioral Risk Factors in the HRUPoRT Algorithm

Risk Factor	Definition
Body mass index (kg/m^2)*	
Underweight	< 18.5
Normal weight	18.5–24.9
Overweight	25.0–29.9
Moderately obese	30.0–34.9
Very obese	35.0–39.9
Severely obese	≥ 40.0
Smoking status	
Heavy smoker	Current smoker [≥ 1 pack (25 cigarettes)/d]
Light smoker	Current smoker [< 1 pack (25 cigarettes)/d]
Former heavy smoker	Former smoker [≥ 1 pack (25 cigarettes)/d]
Former light smoker	Former smoker [< 1 pack (25 cigarettes)/d]
Nonsmoker	Never smoker or former occasional smoker with <100 lifetime cigarettes
Physical activity quartile†	
Q1	Bottom 25% physically active
Q2	Bottom 26%–50% physically active
Q3	Bottom 51%–75% physically active
Q4	Top 25% physically active
Alcohol consumption (drinks/wk, past 12 mo)	
Heavy drinker	≥ 21 (men) or ≥ 14 (women) drinks in the previous week, or binging behavior on a weekly basis (≥ 5 drinks on any occasion)
Moderate drinker	4–21 (men) or 3–14 (women) drinks in the previous week
Light drinker	1–3 (men) or 1–2 (women) drinks in the previous week
Nondrinker	No alcohol consumption in the last 12 mo

*Calculated using self-reported height and weight.
†On the basis of average daily metabolic equivalent of task expenditure (kcal/kg/d) for participation in an aggregate list of leisure time physical activities in the past 3 months.
HRUPoRT indicates High Resource User Population Risk Tool.

To develop the HRU algorithm we selected variables using a stepwise approach based on the size of effect, clinical importance, and conservative *P*-values (<0.25). We began with chronic condition, sex, age group, and income quintile in the model. We then examined whether adding sociodemographic (education, ethnicity, food security, immigrant status), health status (general health, life stress), and health behavioral variables (smoking, body mass index, alcohol consumption, physical activity) improved the model's discrimination and calibration.[25,26] Missing values for risk factors were maintained as separate categories, except for chronic condition in which respondents with missing information were imputed as missing (<1%).

Assessment of Model Performance and Validation

The overall performance of the models were compared using the likelihood ratio (R^2) which measures the variation explained by a model, and the Brier score which measures the accuracy of probabilistic predictions by calculating the squared difference between the outcome and predictions.[26] Discrimination refers to the models ability to distinguish between those that will transition and those that will not transition to HRU status, and was measured using the *c*-statistic, which is identical to the area under the receiver operating characteristic curve.[26] Calibration refers to agreement between observed HRU transitions and predicted transitions from the model, and was measured using the Hosmer-Lemeshow (HL) χ^2 statistic.[26] Calibration plots were displayed by grouping observations into deciles of predicted risk and comparing agreement between observed and predicted HRU transitions across deciles of predicted risk. We then selected a final model, applied the risk factor coefficients to the validation cohort, and assessed model performance in this external cohort using the metrics described above.

General Statistical Analyses

All statistical analyses were conducted using SAS, version 9.4 (SAS institute, Cary, North Carolina). All estimates incorporated bootstrap replicate survey weights provided by Statistics Canada to accurately reflect the Ontario population and account for the complex survey design of the CCHS. Cycles of the CCHS in the development cohort were combined using the pooled approach.[27]

RESULTS

The characteristics of participants in the development and validation cohort are described in Table 2. At the end of the 5-year follow-up period, 6.0% (n = 3502) of respondents in the development cohort and 5.6% (N = 1611) in the validation cohort transitioned to HRU status. Overall, the cohorts differed in respect to sociodemographic and health behavioral characteristics, such as age, income, and smoking, but resembled each other relative to self-reported health status, such as history of chronic conditions, as shown in Table 2.

The best model for predicting 5-year transition to HRU status included 12 risk factors. Table 3 shows all ORs derived from the univariable analysis and final logistic model, including risk factor coefficients and referent categories. In the final model, HRU transition was most strongly associated

TABLE 2. Weighted Distribution of Baseline Characteristics Across HRU Transition Status in the Development and Validation Cohorts

Risk Factor	Development Cohort			Validation Cohort		
	Overall (N = 58,617)	HRU Top 5% (N = 3502)	Non-HRU Bottom 95% (N = 55,115)	Overall (N = 28,721)	HRU Top 5% (N = 1611)	Non-HRU Bottom 95% (N = 27,110)
Sex (male)	48.9	50.4	48.8	48.9	50.1	48.7
Age group (y)						
< 30	21.6	3.7	22.7	21.6	4.0	22.7
30–39	18.6	3.7	19.5	16.7	3.9	17.6
40–49	22.1	10.9	22.8	20.9	8.4	21.6
50–59	16.9	17.8	16.9	18.3	16.8	18.4
60–69	11.3	24.2	10.4	12.3	24.0	11.6
70–79	6.7	22.4	5.7	6.8	23.5	5.8
≥ 80	2.9	17.3	2.0	3.3	19.4	2.3
Ethnicity						
White	76.2	84.7	75.7	73.9	86.3	73.1
Nonwhite	21.2	12.8	21.7	23.7	11.3	24.5
Immigrant status						
Canadian born	67.2	64.5	67.4	65.7	66.8	65.6
Immigrant (< 10 y)	8.4	2.9	8.7	8.2	2.2	8.6
Immigrant (≥ 10 y)	24.2	32.3	23.7	25.8	30.8	25.5
Household income						
Q1 (lowest)	15.6	24.5	15.1	16.4	24.6	15.9
Q2	15.9	19.6	15.7	16.5	19.9	16.3
Q3	16.9	14.5	17.1	16.3	14.3	16.4
Q4	18.9	14.0	19.2	17.1	11.1	17.4
Q5 (highest)	20.0	11.2	20.5	18.6	13.4	18.9
Food security						
Food secure	93.1	92.6	93.2	91.9	90.6	92.0
Food insecure	5.3	6.4	5.3	7.3	8.5	7.3
Chronic conditions						
Yes	56.3	82.3	54.7	56.1	85.0	54.3
No	43.4	17.4	45.0	43.7	14.9	45.4
General health						
Excellent/very good/good	89.1	70.0	90.3	88.4	68.6	89.6
Fair	8.1	18.4	7.4	8.9	22.3	8.1
Poor	2.8	11.5	2.3	2.5	9.1	2.2
Body mass index						
Underweight	2.6	1.8	2.6	2.4	2.7	2.4
Normal weight	44.4	35.5	44.9	43.5	33.3	44.1
Overweight	33.2	36.4	33.0	32.7	32.8	32.7
Moderately obese	11.3	13.9	11.1	12.3	15.6	12.1
Very obese	3.0	4.6	2.9	3.4	4.4	3.3
Severely obese	1.4	2.4	1.3	1.5	2.4	1.4
Smoking status						
Heavy smoker	3.5	5.0	3.4	2.6	3.1	2.5
Light smoker	17.8	14.1	18.0	16.8	17.4	16.7
Former heavy smoker	6.2	12.6	5.8	5.1	10.1	4.8
Former light smoker	16.1	21.5	15.8	15.5	20.0	15.3
Nonsmoker	52.9	42.4	53.6	56.1	44.7	56.8
Physical activity quartile						
Q1 (lowest)	25.1	32.5	24.7	25.0	34.0	24.4
Q2	26.2	24.6	26.3	25.4	26.8	25.4
Q3	24.9	22.5	25.0	24.7	19.8	25.0
Q4 (highest)	21.6	14.5	22.1	23.4	13.0	24.0
Alcohol consumption						
Heavy drinker	8.4	5.5	8.6	7.5	3.8	7.8
Moderate drinker	21.6	20.4	21.7	21.1	18.8	21.2
Light drinker	13.8	12.2	13.9	13.8	12.2	13.9
Nondrinker	54.8	60.0	54.4	56.4	62.8	56.0

Numbers are weighted percentages using bootstrap weights by Statistics Canada. Column percentages do not total 100% where missing values are not reported.
HRU indicates high resource user.

TABLE 3. Weighted Unadjusted ORs and Estimated Parameters of the Logistic Model Predicting 5-Year HRU Transition in the Development Cohort (n = 58,617)

Risk Factor	Univariable Analysis		Logistic Model		
	OR	P	OR (95% CI)	β	P
Constant	—	—	—	−5.074	< 0.001
Sex (male vs. female)	1.06	0.064	1.27 (1.18–1.38)	0.2425	< 0.001
Age group (y)					
< 30	1.00 (ref)	—	1.00 (ref)	—	—
30–39	1.15	0.269	1.13 (0.88–1.44)	0.1202	0.341
40–49	2.91	< 0.001	2.51 (2.05–3.08)	0.9213	< 0.001
50–59	6.41	< 0.001	5.11 (4.19–6.23)	1.6306	< 0.001
60–69	14.12	< 0.001	10.55 (8.66–12.85)	2.3558	< 0.001
70–79	24.16	< 0.001	16.58 (13.54–20.31)	2.8083	< 0.001
≥ 80	53.70	< 0.001	37.29 (30.08–46.24)	3.6188	< 0.001
Ethnicity					
White	1.00 (ref)	—	1.00 (ref)	—	—
Nonwhite	0.53	< 0.001	0.78 (0.69–0.89)	−0.2426	< 0.001
Missing	0.86	0.172	1.01 (0.79–1.28)	0.0098	0.936
Immigrant status					
Canadian born	1.00 (ref)	—	1.00 (ref)	—	—
Immigrant (< 10 y)	0.35	< 0.001	0.79 (0.63–1.00)	−0.2358	0.046
Immigrant (≥ 10 y)	1.42	< 0.001	0.96 (0.88–1.05)	−0.0375	0.416
Missing	1.54	0.156	1.19 (0.60–2.40)	0.1781	0.616
Household income					
Q1 (lowest)	2.96	< 0.001	1.69 (1.47–1.95)	0.5256	< 0.001
Q2	2.28	< 0.001	1.50 (1.30–1.72)	0.4035	< 0.001
Q3	1.55	< 0.001	1.22 (1.06–1.41)	0.2000	0.006
Q4	1.34	< 0.001	1.25 (1.09–1.44)	0.2267	0.002
Q5 (highest)	1.00 (ref)	—	1.00 (ref)	—	—
Missing	2.36	< 0.001	1.47 (1.27–1.70)	0.3841	< 0.001
Food security					
Food secure	0.82	0.005	0.77 (0.65–0.90)	−0.2661	0.002
Food insecure	1.00 (ref)	—	1.00 (ref)	—	—
Missing	0.52	< 0.001	0.45 (0.30–0.68)	−0.7921	< 0.001
Chronic conditions					
Yes vs. no	3.89	< 0.001	1.44 (1.30–1.59)	0.3617	< 0.001
General health					
Excellent/very good/good	1.00 (ref)	—	1.00 (ref)	—	—
Fair	3.20	< 0.001	1.50 (1.36–1.67)	0.4080	< 0.001
Poor	6.49	< 0.001	2.89 (2.52–3.32)	1.0628	< 0.001
Missing	2.92	0.077	1.16 (0.33–4.05)	0.1470	0.818
Body mass index					
Underweight	0.89	0.392	1.02 (0.77–1.35)	0.0174	0.903
Normal weight	1.00 (ref)	—	1.00 (ref)	—	—
Overweight	1.40	< 0.001	1.08 (0.99–1.18)	0.0753	0.092
Moderately obese	1.59	< 0.001	1.12 (0.99–1.26)	0.1131	0.064
Very obese	1.99	< 0.001	1.44 (1.19–1.74)	0.3624	< 0.001
Severely obese	2.32	< 0.001	1.89 (1.46–2.45)	0.6384	< 0.001
Missing	1.65	< 0.001	1.10 (0.89–1.36)	0.0959	0.365
Smoking status					
Heavy smoker	1.86	< 0.001	1.58 (1.32–1.90)	0.4579	< 0.001
Light smoker	0.99	0.798	1.29 (1.15–1.45)	0.2569	< 0.001
Former heavy smoker	2.75	< 0.001	1.34 (1.18–1.52)	0.2926	< 0.001
Former light smoker	1.72	< 0.001	1.14 (1.03–1.26)	0.1296	0.012
Nonsmoker	1.00 (ref)	—	1.00 (ref)	—	—
Missing	1.60	< 0.001	0.94 (0.77–1.13)	−0.0661	0.497
Physical activity quartile					
Q1 (lowest)	2.01	< 0.001	1.19 (1.06–1.34)	0.1766	0.004
Q2	1.42	< 0.001	0.99 (0.88–1.11)	−0.0110	0.857
Q3	1.37	< 0.001	1.02 (0.90–1.15)	0.0172	0.780
Q4 (highest)	1.00 (ref)	—	1.00 (ref)	—	—
Missing	4.8	< 0.001	1.42 (1.13–1.78)	0.3490	0.003
Alcohol consumption					
Heavy drinker	0.73	< 0.001	0.99 (0.82–1.19)	−0.0105	0.913
Moderate drinker	1.07	0.269	0.93 (0.81–1.06)	−0.0757	0.259
Light drinker	1.00 (ref)	—	1.00 (ref)	—	—

(Continued)

TABLE 3. Weighted Unadjusted ORs and Estimated Parameters of the Logistic Model Predicting 5-Year HRU Transition in the Development Cohort (n = 58,617) (*continued*)

	Univariable Analysis		Logistic Model		
Risk Factor	OR	P	OR (95% CI)	β	P
Nondrinker	1.26	<0.001	1.04 (0.93–1.17)	0.0420	0.482
Missing	1.63	<0.001	1.06 (0.80–1.41)	0.0619	0.669
Education					
Less than secondary	2.99	<0.001	—	—	—
Secondary graduate	1.19	<0.001	—	—	—
More than secondary	1.00 (ref)	—	—	—	—
Marital status					
Married/common law vs. other	0.98	0.505	—	—	—
Perceived mental health					
Good vs. poor	0.60	<0.001	—	—	—
Consulted mental health professional					
Yes vs. no	1.23	<0.001	—	—	—
Has a regular doctor					
Yes vs. no	1.64	<0.001	—	—	—
Life satisfaction					
Very satisfied	1.00 (ref)	—	—	—	—
Satisfied	1.09	0.03	—	—	—
Neither	1.20	0.02	—	—	—
Dissatisfied	2.6	<0.001	—	—	—
Very dissatisfied	3.12	<0.001	—	—	—
Perceived life stress					
High vs. low	0.84	<0.001	—	—	—
Diet score					
<1.0 (lowest)	0.89	0.14	—	—	—
1.0–1.9	0.77	0.005	—	—	—
2.0–2.9	1.05	0.47	—	—	—
≥3 (highest)	1.0 (ref)	—	—	—	—

CI indicates confidence interval; HRU, high resource user; OR, odds ratio; ref, reference.

with older age, in that those aged 80+ years had a 37.29-fold increased risk compared with those aged below 30 [OR, 37.29; confidence interval (CI), 30.08–46.24]. Subsequent leading predictors for the odds of HRU transition were perceived general health, body mass index, and household income, specifically, 2.89 times higher for those with poor versus good general health (OR, 2.89; 95% CI, 2.52–3.32), 1.89 times higher for severely obese versus normal weight (OR, 1.89; 95% CI, 1.46–2.45), and 1.69 times higher for those in the lowest income quintile versus highest (OR, 1.69; 95% CI, 1.47–1.95). In the univariable analysis, education, perceived mental health, mental health professional consultation, regular doctor, life satisfaction, life stress, and diet were associated with HRU transition; however, these variables were excluded from the final model because they did not improve its predictive accuracy.

In the final model, the overall number of predicted (and observed) HRUs in the 5-year period was closely approximated with 560,055 (560,054) in the development cohort, and 617,148 (550,438) in the validation cohort. The performance of the HRU algorithm in the development and validation cohort is presented in Table 4. The model displayed good discriminative power (*c*-statistic = 0.8213) and good calibration (HL χ^2 = 18.71, P = 0.016) in the development cohort. The model performed similarly when it was applied to the validation cohort, with a *c*-statistic of 0.8171 and good calibration (HL χ^2 = 19.95, P = 0.011). On the basis of the final model, we compared the predicted versus

observed number of HRUs by decile risk groups (Fig. 1). In the development cohort, predicted and observed HRU estimates differed by <10% across all risk decile groups, except for a 19% underprediction in decile 4, and a 31% overprediction in decile 2. Similarly, in the validation cohort, differences between observed and predicted estimates were <10% across all decile groups, except for an overprediction by 11% in decile 10, 63% in decile 1 and 68% in decile 6. The discrepancies observed within specific decile groups relative to observed and predicted HRU estimates suggests that the importance of certain risk factors may vary across deciles of risk and that there may be additional risk factors that do not improve the predictive accuracy of HRU

TABLE 4. Performance of the Prediction Algorithm for 5-Year HRU Transition in the Development (n = 58,617) and Validation Cohort (n = 28,721)

Performance Measure	Development Cohort	Validation Cohort
Overall		
R^2 (likelihood ratio)	0.0859	0.0801
Brier score	0.0491	0.0469
Discrimination		
c-statistic	0.8213	0.8171
Calibration		
HL χ^2 test	χ^2 = 18.71	χ^2 = 19.95
	P = 0.016	P = 0.011

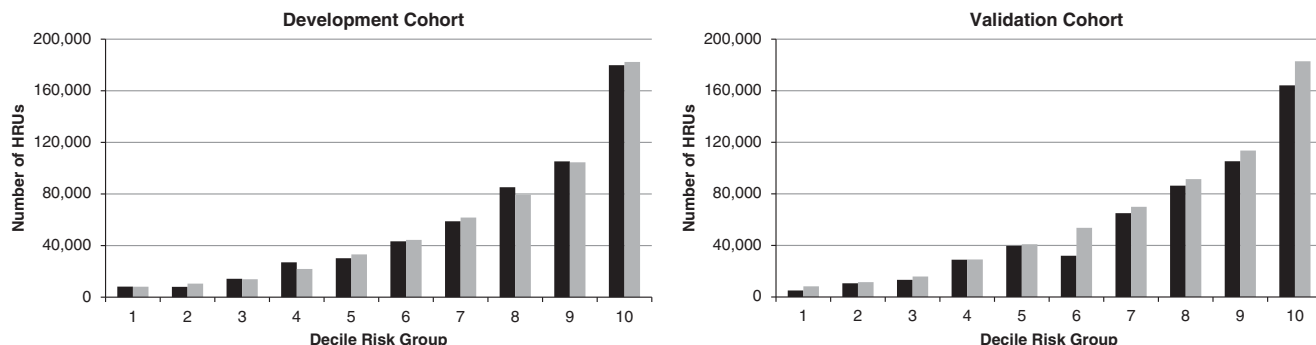HL indicates Hosmer-Lemeshow; HRU, high resource user.

FIGURE 1. Five-year observed and predicted number of HRU transitions in the development and validation cohort by HRU risk decile. HRU indicates high resource user.

estimates overall, but which may be more discriminating for predicting HRUs among groups who are in lower risk deciles.

High Resource User Population Risk Tool Algorithm

The HRUPoRT algorithm uses the β coefficients derived from the validated logistic model predicting 5-year risk of transition to HRU status. The predicted probability for each person is calculated by multiplying their risk factor values by the corresponding β coefficients and summing the products. The application of HRUPoRT to population survey data is described in the Appendix (Supplemental Digital Content 2, http://links.lww.com/MLR/B500).

DISCUSSION

This study presents a population-based decision support tool for projecting high health care resource utilization across the major sectors of health care spending. The HRUPoRT algorithm was validated in an external cohort, and shown to accurately predict the number of individuals in the population that will become a HRU over a 5-year time period. Overall, there was a close approximation between observed and predicted transitions to HRU status, especially for those classified in higher deciles of risk. HRUPoRT was developed for use on publically available population survey data, which facilitates its adoption by health system planners and decision-makers who need evidence on where future HRU transitions and resulting health care expenditures are likely to be concentrated to better inform how health care resources should be allocated. The self-reported nature of the risk factors which the algorithm is based on, enables its application at the individual level to determine patients who are at risk of becoming a HRU without the need for additional data. The HRUPoRT algorithm also includes socioeconomic and modifiable health behavioral risk factors, of which the baseline distributions can be manipulated, offering opportunities to test policy and intervention scenarios that can guide how prevention efforts should be focused to effectively respond to the HRU burden.

Efforts to address high-cost users of the health system have largely been focused on the elderly, managing multimorbidity, and the coordination and delivery of care.[28,29] In the HRUPoRT algorithm, age and presence of self-reported chronic conditions were found to be strong clinical predictors of HRU transitions. The prevalence of multimorbidity has been shown

to increase with age,[30] and it is well documented that persons with co-occurring conditions utilize greater health care services, partly because their conditions have complex care needs.[31,32] The challenge of coordinating health care services for persons with multiple chronic conditions is exacerbated in health systems that have traditionally been designed to treat individual diseases.[33–35] From the clinical perspective, predictive tools, such as HRUPoRT, that can project high health care use at the population-level and among priority subgroups are useful for health system planners and decision-makers that need guidance on how to better integrate health care services to appropriately manage multimorbidity and realize health system cost savings.

Importantly, the HRUPoRT algorithm recognizes important socioeconomic and health behavioral predictors of HRU transitions, which offers a means to incorporate a broader socioeconomic perspective in understanding and responding to HRU. In particular, the HRUPoRT model showed that perceived general health was a stronger predictor of HRU transition than self-reported chronic conditions, despite that the later, multimorbidity, is a more common target in high-cost user interventions.[29] Moreover, household income was the strongest socioeconomic driver, and obesity and smoking the strongest modifiable behaviors associated with HRU transition; findings that are consistent with other research.[12,13] A strength of HRUPoRT is that it can be applied to quantify the impact of changes to disadvantaged socioeconomic living conditions and modifiable risk factors, which offers opportunities to explore which investments in population-level and targeted interventions are suited to preventing high health care use in community settings.[22] Thus, HRUPoRT adds to existing predictive models that largely focus on clinical predictors and are more frequently applied to informing hot spotting and care management at the individual level.[16–18,36] Tools that can quantify the role of upstream social determinants on HRU are important for informing strategies for preventing HRU and improving health system sustainability.

A strength of this study is that HRUPoRT was developed using linked health care expenditures data incurred in a single-payer health system, capturing HRUs across the main sectors of health care spending. However, HRUPoRT does not capture HRUs in domains of health care spending that are not covered by Ontario's Universal Health Insurance Plan, including dental care, eye care, physiotherapy and other allied health services, such as prescription claims for those below 65 years old.

The relevance of HRUPoRT's application outside of Ontario is strengthened given that the algorithm was validated for use on self-reported risk factor information that are widely available in population health surveys, and that HRUs have been shown to share common sociodemographic and behavioral characteristics across jurisdictions. Nonetheless, it is recommended that the model be validated and calibrated before application in different settings to ensure its accuracy. Existing administratively linked population health surveys that exist in other jurisdictions and which may be suitable for international validation include the National Household Interview Survey (NHIS),[37] National Health Survey for England,[38] and the Scottish Health Survey.[39]

Despite these strengths, this study had a couple limitations. The HRUPoRT algorithm was not validated to capture HRU transitions among institutionalized persons (eg, those living in long-term care or complex continuing care facilities), full-time members of the Canadian Forces, and persons living on-reserve and other Aboriginal settlements, as the CCHS sampling frame excludes these populations. Similarly, HRUPoRT does not capture HRU transitions among children below 18 years old, who have been shown to contribute to a large proportion of HRUs.[2] In addition, HRUPoRT does not capture multiple HRU transitions experienced by an individual or new HRU transitions that occur within the first year. As a result, HRUPoRT projections for the general population may be an underestimate of the true HRU burden. Predictive models specific for HRU transitions among children warrant further research given that the determinants of health care use among children are unique, including factors related to child mental health and family functioning.[40] Finally, self-reported information in the CCHS has potential for measurement error due to recall or social desirability bias. For example, self-reported physical activity has been shown to be overreported in the CCHS compared with measures by accelerometer.[41] Despite these limitations, the use of self-reported risk factor measures in our algorithm were found to be discriminating for HRU transition.

CONCLUSIONS

Across jurisdictions, socioeconomic and health behavioral characteristics are important predictors of health care utilization; however, health planning tools that consider the upstream determinants of high-cost users of the health system are lacking. HRUPoRT was validated to accurately project the future number of high health care resource users across the main sectors of health care spending, based solely on publically available clinical, sociodemographic, and health behavioral risk factor information. HRUPoRT is intended to be used by health decision-makers and planners as an aid to integrate clinical and socioeconomic perspectives into effective strategies for population-based planning and prevention of HRU at the community level.

REFERENCES

1. French E, Kelly E. Medical spending around the developed world. *Fiscal Studies*. 2016;37:327–344.
2. Wodchis W, Austin P, Henry D. A 3-year study of high-cost users of health care. *CMAJ*. 2016;188:182–188.
3. Rais S, Nazerian A, Ardal S, et al. High-cost users of Ontario's healthcare services. *Healthc Policy*. 2013;9:44–51.
4. Cohen S. The concentration of health care expenditures in the US and predictions of future spending. *J Econ Soc Meas*. 2016;41:167–189.
5. Calver J, Brameld K, Preen D, et al. High-cost users of hospital beds in Western Australia: a population-based record linkage study. *Med J Aust*. 2006;184:393–397.
6. Ibuka Y, Chen S, Ohtsu Y, et al. Medical spending in Japan: an anlysis using administrative data from a citizen's health insurance plan. *Fiscal Studies*. 2016;37:561–592.
7. Bakx P, O'Donnell O, van Doorslaer E. Spending on health care in the Netherlands: not going so Dutch. *Fiscal Studies*. 2016;37:593–625.
8. Christensen B, Gortz M, Kallestrup-Lamb M. Medical spending in Denmark. *Fiscal Studies*. 2016;37:461–497.
9. Lubitz J, Cai L, Kramarow E, et al. Health, life expectancy, and health care spending among the elderly. *N Engl J Med*. 2003;349:1048–1055.
10. McPhail S. Multimorbidity in chronic disease: impact on health care resources and costs. *Risk Manag Healthcare Pol*. 2016;9:143–156.
11. Lehnert T, Heider D, Leicht H, et al. Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Med Care Res Rev*. 2011;68:387–420.
12. Rosella L, Fitzpatrick T, Wodchis W, et al. High-cost health care users in Ontario, Canada: demographic, socio-economic, and health status characteristics. *BMC Health Serv Res*. 2014;14:1–13.
13. Fitzpatrick T, Rosella L, 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;49:161–171.
14. Lemstra M, Mackenbach J, Neudorf C, et al. High health care utilization and costs associated with lower socio-economic status: results from a linked dataset. *Can J Public Health*. 2009;100:180–183.
15. Hayes S, Salzberg C, McCarthy D, et al. *High-need, High-Cost Patients: Who Are They and How Do They Use Health Care—A Population-Based Comparison of Demographics, Health Care Use, And Expenditures*. New York, NY: The Commonwealth Fund; 2016.
16. Chechulin Y, Nazerian A, Rais S, et al. Predicting patients with high risk of becomming high-cost healthcare users in Ontario (Canada). *Healthc Policy*. 2014;9:68–79.
17. Chang H, Boyd C, Leff B, et al. Identifying consistent high-cost users in a health plan: comparison of alternative prediction models. *Med Care*. 2016;54:852–859.
18. Billings J, Mijanovich T. Improving the management of care for high-cost medicaid patients. *Health Aff*. 2007;26:1643–1654.
19. Frost DW, Vembu S, Wang J, et al. Using electronic medical record to identify patients at high risk for frequent emergency department visits and high system costs. *Am J Med*. 2017;130:601.e17–601.e22.
20. Louffenburger J, Franklin J, Krumme A, et al. Logitudinal patterns of spending enhance the ability to predict costly patients: a novel approach to identify patients for cost containment. *Med Care*. 2017;55:64–73.
21. Hu Z, Hao S, Jin B, et al. Online prediction of health care utilization in the next six months based on electronic health record information: a cohort and validation study. *J Med Internet Res*. 2015;17:e219.
22. Manuel D, Rosella L, Hennessy D, et al. Predictive risk algorithms in a population setting: an overview. *J Epidemiol Commun Health*. 2012;66:859–865.
23. Beland Y. Canadian Community Health Survey—methodological overview. *Health Rep*. 2002;13:9–14.
24. Wodchis W, Bushmeneva K, Nikitovic M, et al. *Guidelines on Person-level Costing Using Administrative Databases in Ontario*. Toronto: Health System Performance Research Network; 2013.
25. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
26. Steyerberg E, Vickers A, Cook N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21:128–138.
27. Thomas S, Wannell B. Combining cycles of the Canadian Community Health Survey. *Health Rep*. 2009;20:55–60.
28. Ali-Faisal S, Colella T, Medina-Jaudes N, et al. The effectiveness of patient navigation to improve healthcare utilization outcomes: a

meta-analysis of ransomized controlled trials. *Patient Educ Couns.* 2017;100:436–448.

29. Bleich S, Sherrod C, Chiang A, et al. Systematic review of programs treating high-need and high-cost people with multiple chronic diseases or diabilities in the United States, 2008-2014. *Prev Chronic Dis.* 2015;12 (E197):1–16.

30. Marengoni A, Angleman S, Melis R, et al. Aging with multimorbidity: a systematic review of the literature. *Aging Res Rev.* 2011;10:430–439.

31. Palladino R, Tayu Lee J, Ashworth M, et al. Associants between multimorbidity, healthcare utilisation and health status: evidence from 16 European countries. *Age Aging.* 2016;45:431–435.

32. Zulman D, Pal Chee C, Wagner T, et al. Multimorbidity and healthcare utilisation among high-cost patients in the US Veterans Affairs Health Care System. *BMJ Open.* 2015;5:e007771.

33. Kuluski K, Peckham A, Williams P, et al. What gets in the way of person-centered care for people with multimorbidity? Lessons from Ontario, Canada. *Health Care Quart.* 2016;19:17–23.

34. Albreht T, Dyakova M, Schellevis F, et al. Many diseases, one model of care? *J Comorbidity.* 2016;6:12–20.

35. Barnett K, Mercer S, Norbury M, et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet.* 2012;380:37–43.

36. Jianying H, Wang F, Sun J, et al. A healthcare utilization analysis framework for hot spotting and contectual anomaly detection. *AMIA Annu Symp Proc.* 2012;2012:360–369.

37. Centers for Disease Control and Prevention. National Health Interview Survey. 2017. Available at: www.cdc.gov/nchs/nhis/index.htm. Accessed September 22, 2017.

38. Mindell J, Biddulph JP, Hirani V, et al. Cohort profile: the health survey for England. *Int J Epidemiol.* 2012;41:1585–1593.

39. Scottish Government Statistics. Scottish Health Survey publications. 2017. Available at: www.gov.scot/Topics/Statistics/Browse/Health/scottish-health-survey/Publications. Accessed Septemeber 22, 2017.

40. Janicke D, Finney J. Determinant of children's primary health care use. *J Clin Psychol Med Settings.* 2000;7:29–39.

41. Garriguet D, Colley R. A comparison of self-reported lesure-time physical activity and measured moderate-to-vigorous physical activity in adolescents and adults. *Health Rep.* 2014;25:3–11.