Appendix 1 (as supplied by the authors): Overview of CIHI's Population Grouping Methodology

CIHI's Population Grouping Methodology is a health risk predictive modelling platform developed using Canadian data. Patient-level administrative healthcare data are used to summarize clinical complexity of the population and explain or predict costs. Model development relied upon pooled data from three provinces (Ontario, Alberta and British Columbia) for fiscal years (FY) 2010 and 2011 (concurrent period; April 1, 2010-March 31, 2012) and FY 2012 (prospective period; April 1, 2012-March 31, 2013). Patient clinical profiles utilize diagnosis codes from all available encounter types and form the basis for modeling costs in the concurrent (baseline) and prospective (12 months post-baseline) periods.

CIHI's purpose-built diagnosis grouper utilizes approximately 10,000 ICD-9 and 18,000 ICD-10-CA diagnosis codes (recorded during inpatient, outpatient, emergency department and physician encounters) along with DSM-IV codes (from all available settings) and some elements from RAI-MDS 2.0© and RAI-MH© (recorded in the long term care setting). Each diagnosis code is classified into one of 226 clinically-meaningful and distinct Health Conditions (HCs) using algorithms developed in consultation with physician experts. HCs are groupings of clinically similar diseases with comparable resource-utilization patterns. HCs cover the spectrum of chronic and acute illnesses, disabilities, medical emergencies, signs and symptoms, and other health states, such as pregnancy. Patients may have multiple HCs. However, a set of "clinical override rules" is applied so that less serious HCs are recoded as null in the presence of a more serious HC that is considered to be part of the same disease process (e.g., if a patient has HCs for both seizure and epilepsy, only the epilepsy flag will remain after overrides are applied).

Healthcare costs were estimated by CIHI as the sum of payments recorded for inpatient stays (excluding mental health hospitalizations), day surgery, ED visits and physician care. In the case of shadow billed physician encounters, costs were imputed based on payments for the same fee codes billed under fee-for-service. For health system users with at least one health condition, costs were modelled as a function of 226 HCs and 460 HC interaction terms using ordinary least squares regression. For health system users with no conditions and health system non-users, costs were modelled separately based on age and gender only. The coefficients from the final additive models comprise the model weights that may be applied to new datasets in order to create risk scores, which are predictions of patient cost relative to others in the study sample or population.

CIHI has reported that their population grouper model explained nearly half of the variance in the dependent variable using age, gender, health conditions and condition interactions in the concurrent period (R^2=0.475) and nearly one-tenth in the prospective period (R^2=0.094) for their three-province model validation sample.[1] These findings were externally validated for Ontario.[2] Another external study found that the model performed somewhat better than the Johns Hopkins ACG® model that has been used in Ontario for many years.[3]

References

1. Canadian Institute for Health Information. CIHI's Population Grouping Methodology 1.1 (compiled code): Methodology Report. CIHI, April 2017.
2. Li Y, Weir S, Steffler M, Shaikh S, Wright J, Kantarevic, J. Using Diagnoses to Estimate Healthcare Cost Risk in Canada. Philadelphia, PA: Medical Care. 2019 Nov;57(11):875-881.

Appendix to: Weir S, Steffler M, Li Y, et al. Use of the Population Grouping Methodology of the Canadian Institute for Health Information to predict high-cost health system users in Ontario. *CMAJ* 2020. DOI:10.1503/cmaj.191297.

3. Cheng S, Austin P, Wodchis W, et al. Evaluation of Population Groupers. ICES Report, September 2016.