BMJ Open

Development and internal validation of a multimorbidity index that predicts healthcare utilisation using the Canadian Longitudinal Study on Aging

Zhuoyu Wang,[1] Laurence Boulanger,[1] David Berger,[1] Pierrette Gaudreau,[1,2] Ruth Ann Marrie,[3] Brian Potter,[1] Andrew Wister,[4] Christina Wolfson,[5,6] Genevieve Lefebvre,[7] Marie-Pierre Sylvestre ![ORCID] ,[1,8] M Keezer ![ORCID] [1,8,9]

Correspondence to
Dr M Keezer;
mark.keezer@umontreal.ca

ABSTRACT

Objectives We aimed to develop and internally validate a measure of multimorbidity burden using data from the Canadian Longitudinal Study on Aging (CLSA).

Design Data from 40 264 CLSA participants (52% men) aged 45–85 years (a mean of 63 years) were analysed. We used logistic regression models to predict overnight hospitalisation in the last 12 months in the development dataset (random two-thirds of the total) and used these to construct 10 multimorbidity indices (5 models, each treated with and without an age interaction term). Thirty-five chronic conditions were considered for inclusion in these models, in addition to age and sex. We assessed predictive and convergent validity for these 10 different multimorbidity indices in the validation dataset (remaining one-third of the total).

Results The absolute count of chronic conditions plus an interaction with age, displayed strong calibration properties, outperforming other candidate indices. Discrimination was modest for all of the indices that we internally validated, with C-statistics ranging from 0.66 to 0.68. The indices showed weak correlations (ie, convergent validity) with satisfaction with life, functional disability and mental health (absolute Pearson's correlation coefficients ranging from 0.11 to 0.30) but generally moderate correlations with self-rated general health (0.32–0.45).

Conclusions We investigated alternative methods to measure the multimorbidity burden of individuals, tailored to the CLSA. Our findings show that an absolute count of conditions, along with an age interaction term, has the strongest calibration for overnight hospitalisation in the last 12 months. The utility of an age interaction term in measuring multimorbidity burden may be applicable to the study of chronic disease in cohorts other than the CLSA.

Strengths and limitations of this study

► The Canadian Longitudinal Study Aging is a large, population-based longitudinal cohort and will undoubtedly remain an important platform to study chronic disease in the general population for the next several decades as participant follow-up assessments accrue.

► We carried out a robust and comprehensive assessment of 10 multimorbidity indices, comparing multiple measures of internal validity.

► We were limited to a dependent variable in all regression models that was likely coarse, unable to discriminate between nuanced differences in outcome.

► The overall performance of the multimorbidity indices we developed was improved, in particular calibration, through the inclusion of an age interaction term.

INTRODUCTION

Multimorbidity is defined as the co-occurrence of multiple chronic medical conditions in the same individual and affects at least 50% of individuals in the general population over age 65 years.[1] According to a study conducted in Ontario, Canada, the prevalence of multimorbidity increased by an average of 40% across all age groups (from 17.4% to 24.3%) between 2003 and 2009.[2] This prevalence will likely continue to increase as the elderly are projected to remain the fastest growing age demographic in high-income countries.[3] An increase in the number of chronic medical conditions is associated with increased resource utilisation and mortality, as well as decreased functional status and quality of life.[4]

An accurate understanding of the prevalence and severity of multimorbidity in a population is fundamental to inform healthcare planning and policy-making, from frontline to tertiary social and medical services. As an important determinant of numerous health and well-being outcomes, it is also important to consider the effect of multimorbidity to allow for meaningful comparisons across populations. When evaluating healthcare

utilisation, the efficacy of new healthcare policies, or comparing performance between different healthcare institutions or administrative units, controlling for differences in multimorbidity burden is essential.[5] Considering the impact of coexisting chronic diseases on outcomes for colorectal surgery, for instance, is commonly done when comparing performance between hospitals.[6]

The Canadian Longitudinal Study on Aging (CLSA) provides a unique multidisciplinary resource to study multimorbidity in the ageing Canadian population. Launched in 2010, the CLSA is a national, prospective, population-based cohort study that includes over 50 000 participants aged 45–85 years at the time of recruitment, all of whom will be followed with periodic reassessments at 3-year intervals for up to 20 years.[7] A validated measure of the overall multimorbidity burden among individuals, however, is not included in the CLSA.[8] While the Charlson[9] and Elixhauser[10] indices are the most frequently used weighted scores in research settings, neither can be computed/derived using data from the CLSA (42%–47% of conditions required for their calculation are missing in the CLSA). Several recent publications using CLSA data have used an unweighted count of chronic diseases to quantify multimorbidity burden.[11–13] Prior studies, however, have reported that an unweighted count of chronic diseases is a less valid measure of multimorbidity as compared with weighted indices.[14–16] An absolute count potentially lacks face validity as it makes no distinction between the unequal effects of different conditions on outcome (eg, migraine is presumably associated with a smaller risk of hospitalisation than metastatic cancer). Therefore, a study of the performance of different approaches to measuring multimorbidity, applicable to the CLSA and potentially other cohorts, is required.

The goal of the present study was to develop a set of new indices to measure the multimorbidity burden of individuals in the CLSA cohort and to compare their internal validity with reference to their ability to predict hospitalisation.

METHODS

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement was followed in reporting this study.[17]

Cohort description: the CLSA

The current study used the baseline CLSA data, collected on all participants during their first assessment, carried out between 2010 and 2015. All CLSA participants were asked to provide core information on demographics, lifestyle and behaviour, social, physical, clinical, psychological, economic and health status.[7] The CLSA cohort included 51 338 participants at baseline. Excluded from the CLSA at baseline were residents in the three Canadian territories, persons living on federal First Nations reserves, full-time members of the Canadian Armed Forces, individuals living in institutions, those unable to respond in English or French, and people who were cognitively impaired.

The CLSA is divided into two subsamples: The tracking and the comprehensive cohorts. Together these are referred to as the full cohort. Participants in the tracking cohort (n=21 241) completed their data collection through a 60 min computer-assisted telephone interview. Participants in the comprehensive cohort (n=30 097) underwent an at-home 90 min face-to-face interview followed by a 2–3 hour assessment at one of 11 data collection sites across Canada. All assessments were carried out in either English or French, depending on the preference of the participant. There were three CLSA sampling frames: participants in the Canadian Community Health Survey 4.2, provincial health registration data and random digit dialling. Participants in the tracking cohort were recruited from all 10 Canadian provinces while participants in the comprehensive cohort were required to live within 25 km (50 km in low-density population areas) of a data collection site (located in seven provinces; that is, excluding Saskatchewan, New Brunswick and Prince Edward Island). The full cohort was used in the present analysis.

Patient and public involvement statement

Participants are a central part of the CLSA. Participant newsletters and CLSA-sponsored webinars update participants on the progress of the study cohort. CLSA governance includes the CLSA advisory council, including 10–12 members, with representatives from health charities and the private sector, as well as lay members. CLSA participants were not involved in the design and conduct of this current analysis of CLSA data; the identification of CLSA participants and our directly contacting them are strictly prohibited by our data sharing agreement with the CLSA.

Variables

A lifetime history of 33 chronic conditions is queried in the CLSA full cohort. These are captured using the self-report question stem: 'Has a doctor ever told you that you have…' Some of the condition variables are a combination of multiple conditions which were combined if there was sufficient pathophysiological similarity between them (all variables are described in online supplementary table 1). Tobacco smoking and regular alcohol consumption were treated as additional chronic conditions, bringing the total to 35. All condition variables are dichotomous, denoting either the presence or absence of the condition in question.

Additional variables measured at the baseline assessment were life satisfaction, functional disability, as well as self-rated general health and mental health. Life satisfaction is the cognitive component of well-being (in contrast to the affective component that is reflected in outward manifestations). It is conceptualised as an overall assessment of one's global quality of life.[18] Life satisfaction is

measured in the CLSA using the Satisfaction with Life Scale (SWLS), a 5-item self-report questionnaire scored on a 35-point scale. The SWLS correlates well with depression, perceived stress, self-esteem and negative effect.[18 19] Functional disability is measured in the CLSA using the Older Americans Resources and Services measure of activities of daily living (ADL) scale.[20] This is a 14-item self-report questionnaire, divided between physical/basic and instrumental ADLs, scored on a 28-point scale. The English-language and French-language versions of this questionnaire are highly correlated with healthcare professional assessments of functional autonomy.[21] Self-rated general health status is assessed in the CLSA using the questionnaire item: 'In general, would you say your health is excellent, very good, good, fair, or poor'. A meta-analysis of 22 studies demonstrated that such a question is a strong predictor of mortality.[22] An almost identical item queries an individual's 'mental health'.

After the initial data collection interview, all CLSA participants were recontacted (planned for 18 months after the initial interview) as part of the 'maintaining contact interview'. This was a 30 min computer-assisted telephone interview during which additional questionnaires were administered, including a new set of items to measure healthcare utilisation. We used the self-reported occurrence of any overnight hospitalisation (not including emergency department visits) in the preceding 12 months as our proxy for multimorbidity severity (data on the frequency of medical visits are not available). Healthcare utilisation is often used to construct and validate measures of multimorbidity.[23] We excluded CLSA participants whose healthcare utilisation questionnaire was administered less than 12 months after the baseline interview, to ensure that the reported hospitalisation occurred after the report of chronic conditions. The administration of the chronic condition questionnaires (during the initial interview) and the healthcare utilisation questionnaire (during the maintaining contact initiative) was very unlikely done by the same CLSA research assistant, given the size and complexity of the CLSA data collection apparatus. Research assistants during the maintaining contact interview were generally not provided the opportunity to review the results of the previously collected chronic condition questionnaires.

Statistical analyses

The full cohort inflation analytical weights were used for the descriptive statistics and the regression analyses. These weights adjust for the age-stratified sampling strategy of the CLSA and the probability of inclusion in the CLSA (influenced by unequal sampling probabilities across sampling units and response rates) for each individual.[24]

To allow for internal validation, the full cohort was randomly divided into two datasets: a development dataset and a validation dataset (2:1 ratio). The regression model coefficients were estimated in the development dataset. We used occurrence of an overnight hospitalisation in the last 12 months as the dependent variable. We used 10-fold multiple imputation to handle missing data.[25]

We considered five predictive models, estimated using multivariable logistic regression. All variables were measured at baseline with the exception of hospitalisation which was queried at approximately 18 months later (minimum 12 months). All models were adjusted for age (as a continuous variable) and sex. Model 1 further included: the absolute count (ie, unweighted sum) of the total number of chronic conditions; model 2: the 35 chronic conditions, each treated as an individual dichotomous variable; models 3–5: all three began with the same predictor variables as in model 2, but we carried out variable selection using LASSO,[26] stepwise selection based on the Akaike information criterion (AIC),[27] or Bayesian model averaging (median probability model),[28] respectively.

From models 2–5, multimorbidity indices (correspondingly indices 2–5) were built by assigning a weight to each of the chronic conditions. Each weight was obtained by dividing the corresponding untransformed regression coefficient (ie, log-odds) by the smallest absolute value among all coefficients and then rounding to the nearest single decimal place. The multimorbidity index 'score' for a participant is the sum of the weights for those conditions by which they are affected. For index 1, the index score is the unweighted sum of chronic conditions.

We evaluated the predictive criterion validity of each multimorbidity index by assessing model goodness of fit, discrimination and calibration in the validation dataset. These aspects of predictive performance were measured using Nagelkerke's R^2 of the model, Harrell's C-statistic, the Pearson correlation between the predicted outcome probability and the observed outcome, and calibration plots with loess smoothers.[29] A C-statistic of 1.00 denotes a measure that perfectly discriminates between individuals with different outcomes while a C-statistic of 0.50 denotes a measure that performs no better than chance. A calibration plot graphs the smoothed relationship between the observed outcomes and predicted probabilities. The diagonal line (slope equal to 1) represents perfect calibration.[30 31] We hypothesised that the effect of multimorbidity on healthcare utilisation interacts with the effect of the age of a person and so we included an interaction term between each index and age in the regression models evaluating predictive validity. We tested the linearity of the relationship between continuous predictors and healthcare utilisation by introducing b-spline functions and observing whether this had any impact on model fit of the validation regression models (measured using AIC).

To evaluate convergent construct validity, we examined the correlation between the multimorbidity indices and constructs we expected a priori to correlate with the multimorbidity burden of an individual. These constructs were life satisfaction, functional disability, as well as self-rated general health and mental health. We assessed the correlation between the index and these additional

constructs using Pearson's correlation coefficient, stratified on sex. Correlations (absolute values) were characterised as very weak (0–0.19), weak (0.2–0.39), moderate (0.4–0.59), strong (0.6–0.79) or very strong (0.8–1.0).[32]

We used R V.3.4.2, along with the R packages mice,[33] mitools,[34] glmnet,[35] Bayesian model averaging (BMA),[36] pROC[37] and regression modeling strategies[38] to carry out the statistical analyses.

RESULTS

Of the 51 338 CLSA participants, 88 were removed from the analysis because they were not aged between 45 and 85 at the time of enrolment. Of the remaining participants, 10 986 were removed because the healthcare utilisation questionnaire items were administered less than 12 months after the baseline data collection. The remaining 40 264 were split randomly according to a 2:1 ratio into the development dataset and the internal validation dataset (online supplementary figure 1). A description of the chronic condition variables used in this study is presented in the online supplementary table 1. There were no clinically relevant differences observed between the distribution of values between the development and validation datasets. We found 1.8% of participants with 0 chronic conditions, 10.6% with 1, 16.5% with 2, 17.1%

with 3, 14.4% with 4, 11.1% with 5 and 20.0% with 6 or more conditions (examining only individuals with complete data).

The distribution of missing data is described in online supplementary table 1. The mean analytical weight, as well as the sex and age distributions between participants with versus without missing data did not reveal any clinically relevant differences (online supplementary table 2).

Online supplementary table 3 presents the estimated regression coefficients for the models from the development dataset. These five predictive models, as compared with a reference model including only age and sex, demonstrate small but consistent improvements in model fit, discrimination and correlation between predicted probabilities and observed outcomes. We did not find evidence of non-linearity between continuous predictors and healthcare utilisation using b-spline functions and comparing AIC.

Index 1, but with an interaction term between the index and age, showed the strongest concordance between the predicted and the observed probability of hospitalisation (figure 1). The calibration plots without an interaction term are reported in online supplementary figure 2. For the remaining multimorbidity indices, calibration was poor, especially when above a predicted probability
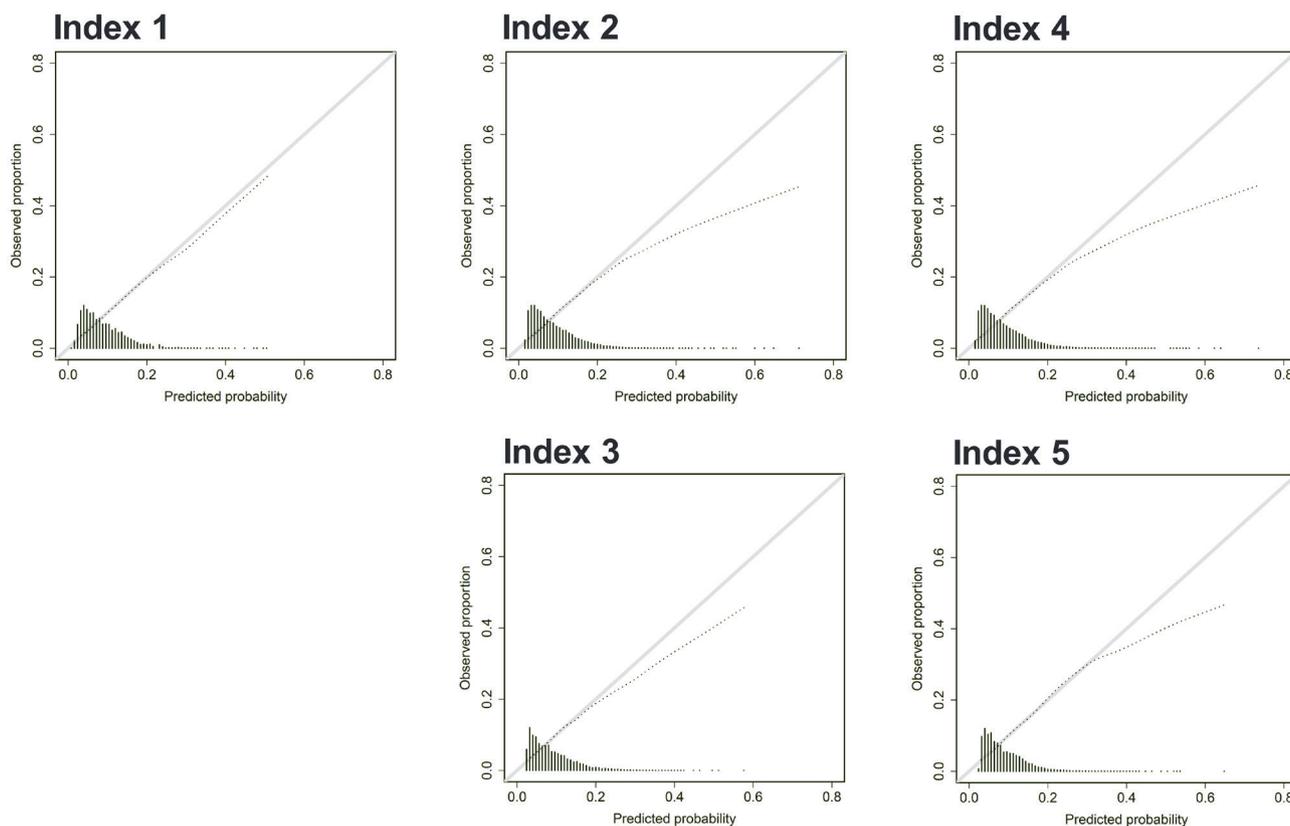

Figure 1 Calibration plots for the five multimorbidity indices, each including an interaction term with age. A calibration plot graphs the smooth relationship between the observed outcomes and predicted probabilities. The diagonal line indicates perfect calibration.[30 31] The histogram along the x-axis represents the relative count of individuals with the plotted predicted probability. The distribution of predicted probabilities varies between indices.

Table 1 The five multimorbidity indices

	Multimorbidity index weights*				
	Index 1—absolute count	Index 2—individual unweighted conditions	Index 3—weighted LASSO	Index 4—weighted AIC	Index 5—weighted BMA
Mood disorder	1	9	1.2	2.8	1.8
Anxiety	1	7.8	–	2.4	–
Osteoarthritis	1	15.3	5.5	4.7	–
Rheumatoid arthritis	1	3	–	–	2.8
Other type of arthritis	1	3.2	–	1.1	–
Asthma	1	2.5	–	–	–
COPD	1	13.8	3.8	4.5	4.1
Hypertension	1	10.6	3.6	3.3	–
Heart disease	1	14.6	7.4	4.6	–
Angina	1	4.7	1	1.5	4
Myocardial infarction	1	9.6	2.2	3.0	3.9
Peripheral vascular disease	1	10.3	2.1	3.3	–
Stroke or TIA	1	4	–	–	4
Gastrointestinal ulcer	1	11.9	2.2	3.7	2.4
Bowel disorder	1	5.8	–	1.9	1.9
Kidney disease or failure	1	13.6	1.7	4.2	4.5
Hypothyroidism	1	4.1	–	1.3	1
Hyperthyroidism	1	6.7	–	2.2	–
Diabetes	1	7.8	1.6	2.4	–
Osteoporosis	1	1.4	–	–	2
Cataracts	1	2.7	–	–	–
Glaucoma	1	–3.2	–	–	–
Macular degeneration	1	1.5	–	–	–
Epilepsy	1	19.1	–	5.8	–
Parkinson's disease	1	33.2	–	10.2	7.9
Multiple sclerosis	1	23.3	–	7.1	–
Migraine	1	1	–		–
Allergies	1	2.7	–	1	–
Skin cancer: melanoma	1	–6.7	–	–	–
Skin cancer: non-melanoma	1	–2.9	–	–	–
Solid cancer	1	6.4	–	2	2.9
Haematological and soft cancer	1	30.6	–	9.5	–
Ill-defined cancer	1	–5.1	–	–	–
Current smoker	1	4.4	–	1.3	–
Regular drinker	1	–10.5	–1.9	–3.3	–
Internal predictive validity of the regression models (in the validation dataset)†‡					
Nagelkerke's R^2§	0.06	0.07	0.06	0.07	0.06
C-statistic (95% CI)§	0.67 (0.65 to 0.68)	0.68 (0.67 to 0.70)	0.67 (0.65 to 0.68)	0.68 (0.66 to 0.69)	0.66 (0.65 to 0.68)
Pearson correlation (95% CI)§¶	0.88 (0.84 to 0.92)	0.89 (0.85 to 0.92)	0.77 (0.69 to 0.83)	0.89 (0.85 to 0.92)	0.86 (0.81 to 0.90)

Continued

Table 1 Continued

	Multimorbidity index weights*				
	Index 1—absolute count	Index 2—individual unweighted conditions	Index 3—weighted LASSO	Index 4—weighted AIC	Index 5—weighted BMA

*Index 1 is based on a regression model including the absolute count of chronic conditions as a single independent variable. Index 2 is based on a regression model where each chronic condition was treated as a dichotomous independent variable. Indices 3–5 are based on regression models that began with the same independent variables as those used for index 2, but then involved variable selection using LASSO, stepwise selection based on the Akaike information criterion (AIC), or Bayesian model averaging (BMA), respectively.
†There are many estimates for each measure due to multiple imputation. We report the median estimate.
‡For the model reference model including only age and sex, R^2=0.03, C-statistic=0.62 (95% CI: 0.60, 0.64), and correlation coefficient=0.83 (95% CI: 0.74, 0.89).
§These are based on the model including the index, sex, age and an interaction between age and the index.
¶Correlation between the predicted and the observed probability of the outcome, using 130 quantiles.
COPD, Chronic obstructive pulmonary disease; LASSO, least absolute shrinkage and selection operator; TIA, transient ischaemic attack.

of 0.20, where the curves diverged from a slope of 1.0 (figure 1). The correlation coefficients between the predicted and the observed probability of hospitalisation were generally strong for all indices (greater than 0.80) (table 1). Model fit, measured using R^2, was relatively weak for all indices. Discrimination was modest for all models (C-statistic=0.66–0.69) (table 1), although consistently higher than a model including only sex and age, with the exception of index 5 (table 1, online supplementary table 3).

All of the multimorbidity indices showed no more than weak correlations with satisfaction with life, functional disability and self-rated mental health (correlation coefficients=0.15–0.29) but generally moderate correlations with self-rated general health (correlation coefficients=0.35–0.45) (table 2). Stratification did not reveal subjectively large differences in correlation between men and women.

Box 1 describes the derivation and method to calculate an absolute sum of chronic conditions that considers the interaction between index 1 and age. This allows for the calculation of an 'age-adjusted' sum of chronic diseases, as opposed to a simple absolute count.

DISCUSSION

We developed five multimorbidity indices using the baseline CLSA data and assessed their internal validity. An unweighted sum of 35 chronic conditions, combined with an interaction term between age and this sum, resulted in a highly calibrated index (ie, able to accurately predict the probability of hospitalisation in a population). The discrimination of the five indices (ie, their ability to determine whether a specific person is more likely to be hospitalised as compared with another person) was modest. There were no more than weak correlations

Table 2 Convergent validity between the multimorbidity indices and relevant constructs*

	Sex	SWLS†	OARS†	Self-rated general health‡	Self-rated mental health‡
Index 1—absolute count§	Women	−0.20 (−0.21 to −0.19)	−0.26 (−0.27 to −0.25)	0.39 (0.38 to 0.40)	0.22 (0.21 to 0.23)
	Men	−0.15 (−0.17 to −0.14)	−0.18 (−0.19 to −0.16)	0.36 (0.35 to 0.37)	0.18 (0.17 to 0.20)
Index 2—individual unweighted conditions	Women	−0.24 (−0.25 to −0.23)	−0.30 (−0.31 to −0.29)	0.45 (0.43 to 0.46)	0.24 (0.23 to 0.25)
	Men	−0.18 (−0.20 to −0.17)	−0.21 (−0.23 to −0.20)	0.40 (0.39 to 0.41)	0.20 (0.19 to 0.22)
Index 3—weighted LASSO	Women	−0.16 (−0.18 to −0.15)	−0.24 (−0.26 to −0.23)	0.37 (0.36 to 0.39)	0.16 (0.15 to 0.17)
	Men	−0.11 (−0.13 to −0.10)	−0.15 (−0.17 to −0.14)	0.35 (0.34 to 0.36)	0.14 (0.12 to 0.15)
Index 4—weighted AIC	Women	−0.24 (−0.25 to −0.23)	−0.30 (−0.31 to −0.28)	0.44 (0.43 to 0.46)	0.24 (0.23 to 0.26)
	Men	−0.18 (−0.20 to −0.17)	−0.21 (−0.22 to −0.19)	0.40 (0.39 to 0.41)	0.20 (0.19 to 0.22)
Index 5—weighted BMA	Women	−0.19 (−0.20 to −0.18)	−0.24 (−0.26 to −0.23)	0.34 (0.33 to 0.35)	0.20 (0.19 to 0.22)
	Men	−0.14 (−0.15 to −0.13)	−0.19 (−0.20 to −0.18)	0.32 (0.3 to 0.33)	0.16 (0.15 to 0.18)

*There are many estimates for each measure due to multiple imputation. We report the median estimate, and in parenthesis, the minimum and maximum estimates.
†Lower Satisfaction with Life Scale (SWLS) or Older Americans Resources and Services measure of activities of daily living (OARS) scores represent worse outcomes.
‡Higher self-rated general and mental health represent worse outcomes.
§Index 1 is based on a regression model including the absolute count of chronic conditions as a single independent variable. Index 2 is based on a regression model where each chronic condition was treated as a dichotomous independent variable. Indices 3–5 are based on regression models that began with the same independent variables as those used for index 2, but then involved variable selection using LASSO, stepwise selection based on the Akaike information criterion (AIC), or Bayesian model averaging (BMA), respectively.

> **Box 1 Calculating a sum of chronic disease that considers the interaction with age**
>
> Recall that the regression model for index 1 in the validation dataset is:
>
> $$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\,\mathrm{SEX} + \beta_2\,\mathrm{AGE} + \beta_4\,\mathrm{AGE}*\mathrm{I}$$
>
> where I is index 1 (the absolute sum of chronic conditions). Note that we can also write the above as:
>
> $$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1\,\mathrm{SEX} + \beta_2\,\mathrm{AGE} + \beta_3\mathrm{I}\left(1 + \left(\frac{\beta_4}{\beta_3}\right)\mathrm{AGE}\right)$$
>
> This suggests that a new index taking the effect modification due to age into account is obtained by multiplying the usual index by the 'adjustment factor' λ, where:
>
> $$\lambda = 1 + \left(\frac{\beta_4}{\beta_3}\right)\mathrm{AGE}$$
>
> For the regression model including index 1, β_3=0.416 and β_4=−0.004, therefore
>
> $$\lambda = 1 - 0.00962\,\mathrm{AGE}$$
>
> **Example**
>
> The 'adjustment factor' for a person who is aged 45 years is λ=1−0.00962*45=0.567. For a person who is aged 85 years, the 'adjustment factor' is 0.183.
>
> Therefore, when considering the interaction between index 1 and age, for a person aged 45 years with 4 chronic conditions, the 'age-adjusted sum of chronic diseases' is 4*0.567=2.269 while for a person who is aged 85 years it is 0.731.

with satisfaction with life, functional disability and self-rated mental health, while correlations with self-rated general health were larger but remained no better than moderate. We also present a method of calculating an absolute sum of chronic conditions that adjusts for the interaction with age, as opposed to a simple count, which could be used in future research. The precise age adjustment factors presented are specific to the CLSA database and may not be transportable to other databases although the mathematical approach is applicable to any database.

The modest discrimination that we found is consistent with multiple prior studies, including those studying the performance of the Charlson and Elixhauser indices. One systematic review summarised the predictive validity of three variants of the Charlson index, reporting C-statistics between 0.62 and 0.76 in numerous populations (predicting in hospital, 30 days and 1-year mortality).[39] One study of almost 3500 community-dwelling adults found that the Charlson and Elixhauser indices resulted in C-statistics no higher than 0.695 for any hospitalisation or death in the year following the baseline assessment.[14] Previous investigators have commented that multimorbidity models show a modest improvement in predictive performance as compared with regression models that consider only age and sex.[14 40] This is likely related to the inherent challenge of reducing a complex construct into a single measure such as an index. Disease severity and duration, as well as the social determinants of health, are unaccounted for in most multimorbidity indices.[41]

In addition, the accuracy of chronic disease ascertainment in large databases is often wanting, relying on, for example, International Classification of Disease coding or self-reported questionnaires. Such challenges likely explain, at least in part, why some comparative studies have reported that an unweighted count of chronic diseases performs almost as well as an index based on a more complex regression model. Multimorbidity indices that are based on prescribed medications rather than diagnosed chronic conditions may perform better at predicting hospitalisation.[23] These data are not yet available in the CLSA.

In spite of their limitations, multimorbidity indices, including an unweighted sum of chronic conditions, nevertheless remain recommended, especially in exploratory studies and analyses.[23 39 42] This is due to their ease of use, that they standardise the operationalisation of multimorbidity, and facilitate comparisons between populations and studies. It is generally assumed that, although the ability of an index to measure the magnitude of a multimorbidity association is limited, they provide insight into the presence of such an association as well as its direction.[39 42]

Our study has several strengths. The CLSA is a large, population-based longitudinal cohort which includes data on 35 chronic conditions. The CLSA will undoubtedly remain an important platform to study chronic disease in the general population for the next several decades as participant follow-up assessments accrue. We studied three important aspects of predictive performance, as recommended: model fit, discrimination and calibration. We used three variable selection methods (LASSO, AIC-based stepwise selection and BMA), and examined the impact of adding interaction terms. Finally, in addition to studying predictive performance, we examined convergent validity, assessing the correlation of each index with measures of physical and mental health, as well as life satisfaction.

Our study has limitations. It is possible that healthy people are oversampled in the CLSA, although the sampling weights are meant to at least in part correct for this. Each chronic disease was treated as present or absent, with no measure of severity. As a result, we likely underestimated the impact of more severe cases given their relative rarity in this cohort. We used overnight hospitalisation within the last 12 months as the outcome as a proxy for multimorbidity severity on which we built our regression models and validated our indices. This outcome cannot distinguish between an individual with multiple hospital stays during this time period, a single but prolonged hospital stay or an individual who only spent a single night in hospital. The overall performance of the multimorbidity indices was likely also limited by the relative paucity of participants at high risk of poor outcome. A longer follow-up period would facilitate our ability to construct indices with higher predictive performance. Chronic diseases in the CLSA are self-reported by participants; therefore, there may be inaccuracies in case

ascertainment. CLSA participants cannot be contacted to confirm their self-reported diagnoses. Prior validation studies have shown that self-report questions can accurately identify certain conditions such as multiple sclerosis.[43]

We developed and comprehensively examined the internal validity of five indices to measure the multimorbidity burden of individuals in the general population, particularly tailored for use in the CLSA. Prior research has shown that using predictive weights that are study (ie, database) specific generally improves the predictive performance of these indices.[23][39] We show that an absolute count of chronic conditions, with an interaction term for age, has the greatest calibration. Our work lays the foundation for future multimorbidity research in the CLSA, a unique platform for the study of chronic conditions in the general Canadian population for the next several decades. The utility of an age interaction term in measuring multimorbidity burden may be applicable to the study of chronic disease in cohorts other than the CLSA.

Author affiliations
[1]Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CRCHUM), Montreal, Québec, Canada
[2]Department of Medicine, Université de Montréal, Montreal, Québec, Canada
[3]Department of Internal Medicine (Neurology) & Department of Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
[4]Department of Gerontology, Simon Fraser University, Burnaby, British Columbia, Canada
[5]Neuroepidemiology Research Unit, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada
[6]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[7]Department of Mathematics, Université du Québec à Montréal (UQAM), Montreal, Quebec, Canada
[8]Department of Social and Preventive Medicine, Université de Montréal, Montreal, Quebec, Canada
[9]Department of Neurosciences, Université de Montréal, Montreal, Québec, Canada

ORCID iDs
Marie-Pierre Sylvestre http://orcid.org/0000-0001-5803-4615
M Keezer http://orcid.org/0000-0002-2807-7347

REFERENCES

1 Salive ME. Multimorbidity in older adults. *Epidemiol Rev* 2013;35:75–83.
2 Pefoyo AJK, Bronskill SE, Gruneir A, *et al*. The increasing burden and complexity of multimorbidity. *BMC Public Health* 2015;15:415.
3 Vespa J, Armstrong D, Medina L. *Demographic turning points for the United States: population projections for 2020 to 2060, current population reports, 25-1144, US census bureau*. Washington, DC, 2018.
4 Gijsen R, Hoeymans N, Schellevis FG, *et al*. Causes and consequences of comorbidity: a review. *J Clin Epidemiol* 2001;54:661–74.
5 Elixhauser A, Steiner C, Harris DR, *et al*. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
6 Dobbins TA, Badgery-Parker T, Currow DC, *et al*. Assessing measures of comorbidity and functional status for risk adjustment to compare Hospital performance for colorectal cancer surgery: a retrospective data-linkage study. *BMC Med Inform Decis Mak* 2015;15:55.
7 Raina PS, Wolfson C, Kirkland SA, *et al*. The Canadian longitudinal study on aging (CLSA). *Can J Aging* 2009;28:221–9.
8 Canadian longitudinal study on aging, 2012. Available: www.clsa-elcv.ca [Accessed 28 Dec 2017].
9 Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
10 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.
11 Wolfson C, Fereshtehnejad S-M, Pasquet R, *et al*. High burden of neurological disease in the older general population: results from the Canadian longitudinal study on aging. *Eur J Neurol* 2019;26:356–62.
12 Wister A, Lear S, Schuurman N, *et al*. Development and validation of a multi-domain multimorbidity resilience index for an older population: results from the baseline Canadian longitudinal study on aging. *BMC Geriatr* 2018;18:170.
13 Harasemiw O, Newall N, Shooshtari S, *et al*. From social integration to social isolation: the relationship between social network types and perceived availability of social support in a national sample of older Canadians. *Res Aging* 2018;40:715–39.
14 Perkins AJ, Kroenke K, Unützer J, *et al*. Common comorbidity scales were similar in their ability to predict health care costs and mortality. *J Clin Epidemiol* 2004;57:1040–8.
15 Di Bari M, Virgillo A, Matteuzzi D, *et al*. Predictive validity of measures of comorbidity in older community dwellers: the Insufficienza cardiaca negli Anziani Residenti a Dicomano study. *J Am Geriatr Soc* 2006;54:210–6.
16 Huntley AL, Johnson R, Purdy S, *et al*. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. *Ann Fam Med* 2012;10:134–41.
17 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015;13:1.
18 Pavot W, Diener E. The satisfaction with life scale and the emerging construct of life satisfaction. *J Posit Psychol* 2008;3:137–52.

19 Vassar M. A note on the score reliability for the satisfaction with life scale: an Rg study. *Soc Indic Res* 2007;86:47–57.

20 Fillenbaum GG. *Multidimensional functional assessment of older adults: the Duke older Americans resources and services procedures*. Psychol Press, 2013.

21 McCusker J, Bellavance F, Cardin S, *et al*. Validity of an activities of daily living questionnaire among older patients in the emergency department. *J Clin Epidemiol* 1999;52:1023–30.

22 DeSalvo KB, Bloser N, Reynolds K, *et al*. Mortality prediction with a single General self-rated health question. A meta-analysis. *J Gen Intern Med* 2006;21:267–75.

23 Yurkovich M, Avina-Zubieta JA, Thomas J, *et al*. A systematic review identifies valid comorbidity indices derived from administrative health data. *J Clin Epidemiol* 2015;68:3–14.

24 Canadian longitudinal study on aging: sampling and computation of response rates and sample weights for the tracking (telephone interview) participants and comprehensive participants. Available: https://www.clsa-elcv.ca/doc/1041 [Accessed 16 Jul 2017].

25 Sv B, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. *J Stat Softw* 2011;45:1–68.

26 Tibshirani R. Regression shrinkage and selection via the LASSO. *J Royal Stat Soc B* 1996;58:267–88.

27 Bozdogan H. Model selection and Akaike's information criterion (AIc): the general theory and its analytical extensions. *Psychometrika* 1987;52:345–70.

28 Wasserman L. Bayesian model selection and model averaging. *J Math Psychol* 2000;44:92–107.

29 Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using Loess smoothers. *Stat Med* 2014;33:517–35.

30 Steyerberg EW, Vickers AJ, Cook NR, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128.

31 Altman DG, Vergouwe Y, Royston P, *et al*. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605.

32 The bmj: Correlation and regression, 2019. Available: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression

33 Buuren Svan, Groothuis-Oudshoorn K. Mice : Multivariate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:67.

34 Lumley T. Mitools: tools for multiple imputation of missing data, 2006. Available: http://CRAN R-project org

35 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33:22.

36 Raftery AE, Painter IS, Volinsky CT. Bma: an R package for Bayesian model averaging. *The Newsletter of the R Project Volume* 2005;5:2.

37 Robin X, Turck N, Hainard A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.

38 Harrell Jr F. *Rms: regression modeling strategies. R package version 5.1-0*. 2017, 2018.

39 Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000;29:891–8.

40 Wister AV, Levasseur M, Griffith LE, *et al*. Estimating multiple morbidity disease burden among older persons: a convergent construct validity study to discriminate among six chronic illness measures, CCHS 2008/09. *BMC Geriatr* 2015;15:12.

41 Griffith LE, Gruneir A, Fisher KA, *et al*. Key factors to consider when measuring multimorbidity: results from an expert panel and online survey. *J Comorbid*2018;8:2235042X18795306.

42 Schneeweiss S, Seeger JD, Maclure M, *et al*. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol* 2001;154:854–64.

43 Marrie RA, Cutter G, Tyry T, *et al*. Validation of the NARCOMS registry: diagnosis. *Mult Scler* 2007;13:770–5.