

# BMJ Open Validation and adaptation of the Charlson Comorbidity Index using administrative data from the Colombian health system: retrospective cohort study

Henry Oliveros,<sup>1,2</sup> Giancarlo Buitrago<sup>3</sup>

**To cite:** Oliveros H, Buitrago G. Validation and adaptation of the Charlson Comorbidity Index using administrative data from the Colombian health system: retrospective cohort study. *BMJ Open* 2022;**12**:e054058. doi:10.1136/bmjopen-2021-054058

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-054058>).

Received 03 June 2021

Accepted 23 February 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>PhD Program in Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics. (PhD cand.), Pontificia Universidad Javeriana, Bogotá, Colombia

<sup>2</sup>Facultad de medicina Departamento de Epidemiología, Universidad de La Sabana, Chia, Colombia

<sup>3</sup>PhD Program in Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics, Pontificia Universidad Javeriana, Bogotá, Colombia

## Correspondence to

Dr Henry Oliveros;  
[henryor@unisabana.edu.co](mailto:henryor@unisabana.edu.co)

## ABSTRACT

**Objective** To validate the Charlson Comorbidity Index (CCI) for the Colombian population using administrative databases from the health insurance system.

**Design** Retrospective cohort study.

**Setting** Database reports of events related to services that insurers provided (Health Promoter Enterprises, EPS in Spanish) in the Colombian health system, which covered 22.19 million residents in 2016.

**Participants** The study cohort comprised 3 849 849 patients aged 18 years and up admitted to the Colombian hospitals between 1 January and 31 December 2016.

**Primary outcome** The study aimed to gauge the CCI's predictive value for mortality by comparing the calibration and discrimination of three different versions of the index, with mortality information obtained from death certificates, including date of death and diagnoses associated with cause of death. Follow-up was conducted for at least 1 year.

**Results** Altogether, 46 429 patients died within 1 year (1.21%). Discriminatory power in predicting 1-year mortality was calculated for three versions of the ICC. In the original CCI model, the area under the Receiver operating characteristic (ROC) curve was 0.906 (95% CI (0.906 to 0.907),  $p < 0.001$ ). In the version for Colombia, it was 0.908 (95% CI (0.908 to 0.909),  $p < 0.001$ ) and for the new model it was 0.909 (95% CI (0.908 to 0.910),  $p < 0.001$ ).

**Conclusions** Adapting the CCI based on the 14 predictive variables of the new model resulted in an adequate predictive value for 1-year mortality in patients who were hospitalised for all causes. These findings support the use of the modified CCI in the Colombian population.

## INTRODUCTION

The past two decades have seen a progressive increase in epidemiological studies' use of administrative databases<sup>1,2</sup> due to growing availability of data from multiple sources. While conducting investigations using administrative databases has its advantages, it is also subject to problems. Administrative databases can be used to study population groups

## Strengths and limitations of this study

- This large population study used administrative data from health insurers and the national death registry to provide complete longitudinal health records.
- This is the first study to report comorbidities' prevalence in an extended cohort of patients hospitalised in the Colombian territory.
- The study describes a new Charlson Comorbidity Index model with good predictive value for mortality at 1 year from 14 variables.
- The study could not validate Hospital Episodes Statistics coding directly due to its anonymisation before research use.

that typically are excluded from controlled clinical experiments—such as children, older adults, pregnant women and patients with comorbidities. Nevertheless, these studies are conducted without a randomised process, which contributes to selection bias and confounders, eliciting the same limitations that affect observational studies. Thus, different instruments have been developed to adjust for the biases associated with observational studies, one of which notably is the Charlson Comorbidity Index (CCI),<sup>3,4</sup> which is designed to adjust for confounding variables. This index is one of the most widely used gauges, with 9850 citations, and has been adapted to the largest number of countries. It was developed in New York City in 1984 using a cohort of 559 patients, and it included 19 comorbidity conditions that, individually or in combination, predicted 1-year mortality risk.

While the index's original design was based on a review of medical records, later adaptations used administrative databases. As precursors of this modality, Deyo *et al* and Romano *et al*<sup>5,6</sup> used the index with administrative

data from the USA, including diagnostic information coded according to the International Classification of Diseases, which was current at the time (ICD-9). In both adaptations, the 19 original comorbidity categories were reduced to 17, in which 'lymphoma', 'leukaemia' and 'any malignant tumour' were grouped into one category, 'any tumour'. Altogether, 11 adaptations of the ICD-10 have been performed to date.<sup>7-16</sup>

With the exception of Brazil,<sup>15</sup> Latin American countries do not have a CCI adapted based on administrative databases. The present study's objectives were to adapt and validate the index to predict 1-year mortality for hospitalised adult patients in Colombia based on administrative databases from the Colombian health insurance system.

## METHODS

### Type of study and information sources

A retrospective cohort study was performed using administrative data and the per capita payment unit (UPC in Spanish) sufficiency database from the Colombian Ministry of Health's Integrated Social Protection Information System (SISPRO in Spanish). This database contains reports of the events related to services that the insurers provided (Health Promoter Enterprises; EPS in Spanish) in the Colombian health system, which covered 22.19 million Colombians in 2016 (48% of the population). This database is highly standardised and contains service provider codes (CUPS in Spanish), prescribed medications, dates when the service was provided, age, gender, insurer, municipality, ICD-10 codes and medical care costs. Mortality information was obtained from the death certificates contained in the corresponding database (RUAF in Spanish) and included date of death and diagnoses associated with cause of death. Based on these sources, a cohort of hospitalised patients was constructed, and corresponding data were compiled into a database that contained records of demographic variables, comorbidities and date of death for patients who died during the follow-up year. The entire database was used to calculate each comorbidity's prevalence and to produce a descriptive analysis of the demographic variables.

The total population was divided randomly into two subpopulations (population A and population B) with a 50:50 ratio. Population A was used as the model's test population (derivation group), and population B was used for internal validation (validation group).

### Population

The study population included patients over age 18 years who were admitted to hospitals in Colombia between 1 January and 31 December 2016. For patients who had been hospitalised several times, the first hospitalisation was used. All patients were followed for a minimum period of 1 year from the date of admission or until a mortality event. [Figure 1](#) presents the process for including subjects.

### Study variables

The present study included two types of variables: demographic and clinical. The demographic variables included gender, age, hospital stay and date of death. The clinical variables included the 17 comorbidities found in the Deyo *et al* and Romano *et al* adaptations of the Charlson Index.

The clinical variables were established based on three sources: the ICD-10 diagnosis assigned to the hospitalisation event; procedures specifically associated with the comorbidity and medications specifically associated with the comorbidity.

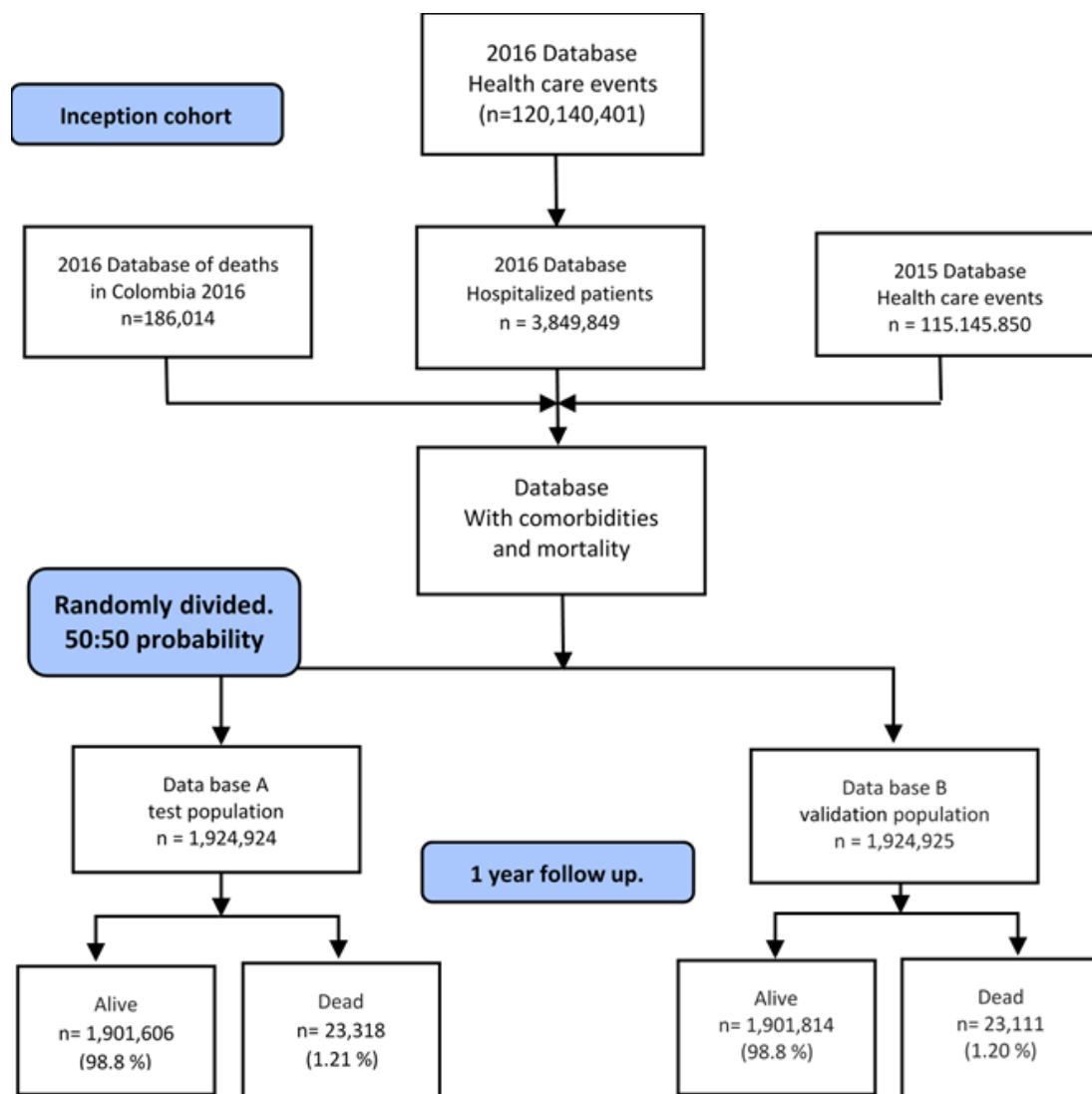
In all cases, the primary source used to determine comorbidities was the Quan *et al*<sup>13</sup> adaptation of the ICD-10 diagnosis, which is well recognised for its high discriminative ability to predict mortality. Nevertheless, in some cases, various sources helped determine comorbidity. Comorbidities such as peripheral vascular disease, kidney disease and ulcer disease were identified using specific procedures associated with these illnesses based on how they were recorded in the service provider codes (CUPS in Spanish) (online supplemental annex 1). As a supplemental source for determining comorbidities such as diabetes, chronic pulmonary disease (CPD), dementia and HIV, measurements specifically associated with these diseases were used. Nevertheless, identifying a code in any of the three sources mentioned generally was viewed as sufficient to establish the corresponding comorbidity's presence. Finally, the search for codes covered up to 12 months before the date of inclusion in the study, that is, 12 months before the date of the first registered hospitalisation in 2016.

Standardised mean differences (SMDs) were calculated to determine the degree of association between the variables and mortality, making it possible to establish percentage differences between the averages of variables pertaining to the comparison groups—namely, survivors and deaths at 1 year. SMDs present these differences in absolute values, with which comparisons can be established that are independent of not only the units in which the variable is expressed, but also the sample size. An SMD value over 10% is associated with statistically significant differences.

### Assignment of scores to calculate the CCI

For each patient, a CCI score was obtained based on the sum of weights corresponding to comorbidities that the patient presented. All the comorbidities were viewed as accumulative, with the exception of solid metastatic tumours, lymphoma, leukaemia, diabetes and liver disease, in which only one category was used. Comorbidity severity was omitted in diabetes and liver disease cases.

Calculating the CCI score also involved integrating the age category by adding one (1) point for each decade of life after 50 years.<sup>17</sup> The present study used the CCI modification proposed by Quan *et al*<sup>9</sup> as a guide, which groups comorbidities such as diabetes and liver disease into a single category, omitting degree of severity.



**Figure 1** Flow chart of subjects included based on administrative health data.

Weights were assigned for each comorbidity according to the values described in the original CCI model and based on HR ranges. Scores in the range  $\geq 1.2 < 1.5$  were assigned a weight of 1, those  $\geq 1.5 < 2.5$  were assigned a weight of 2, those  $\geq 2.5 < 3.5$  were assigned a weight of 3, those  $\geq 3.5 < 4.5$  were assigned a weight of 4 (no comorbidities had an HR of  $\geq 4.5$ ).

The present study evaluated the performance of three versions of the CCI. The first was obtained from the sum of the weights that the index originally assigned to each of the comorbidities.<sup>3</sup> The second was obtained by recalibrating the weights assigned to each comorbidity in the first version based on the HR for 1-year survival, which was obtained for each of the comorbidities using the Cox proportional-hazards model, adjusted by age and gender. The third was obtained by recalibrating the weights of the variables retained based on their statistical significance ( $p < 0.05$ ) and the magnitude of the HR resulting from the Cox proportional-hazards model, in accordance with the original methodology.<sup>3</sup> The score obtained in this third version was divided into four categories to distinguish

1-year mortality ranges, resulting in scores equal to 0, scores of 1 and 2, scores of 3 and 4, and scores equal to or greater than 5.

#### Duration of follow-up and outcome

This study's date of inclusion was the date of the first hospitalisation in 2016. Follow-up with patients was conducted for 12 months after the date of hospitalisation or until the date of death. The outcome was death from any cause during or after hospitalisation. Given the database's characteristics, the possibility of loss to follow-up could be dismissed.

#### Patient and public involvement

No patients were involved.

#### Statistical analysis

First, for purposes of verifying the balance of the variables between population A and population B, the frequencies and averages of the demographic and clinical variables mentioned previously were compared. Kaplan-Meier

curves were used to evaluate 1-year survival, and both in-hospital and 1-year mortality were calculated. In-hospital mortality was determined based on the percentage of patients who died during the hospitalisation period, and 1-year mortality was determined by the percentage of patients who died during the follow-up period.

Population A was used to recalibrate the weights of the comorbidities in the CCI, as well as develop the new Cox model. This recalibration was performed by constructing a Cox proportional-hazards model that included the 17 comorbidities recognised in the Quan *et al*<sup>13</sup> validations for the prediction of 1-year mortality for the Colombian population, from which crude and adjusted HRs were obtained by age and gender. This new Cox model was developed with only the variables retained due to their statistical significance ( $p \leq 0.05$ ) and HR magnitude, from which crude and adjusted HRs again were obtained by age and gender. Furthermore, Schoenfeld residuals were calculated to evaluate each variable's proportionality assumption.

Population B was used to validate the three versions of the CCI included in this study, namely the original CCI model, the version derived from recalibrating the weights based on the HR obtained with the model applied to the Colombian population and the version obtained with a new model for the Colombian population that included only the variables retained because of their statistical significance and HR magnitude.

These three versions' scores were validated by determining the ability to discriminate and the adjustment of each one. The ability to discriminate was determined by constructing the respective ROC curves for each score, and goodness of fit was evaluated using Harrell's C-index.<sup>18 19</sup>

The Cox model was used to evaluate the adjustment of each score, which included the score as an independent variable in the prediction of 1-year survival.

All statistical analyses and graphs were generated with the R statistical package (V.4.0.3) using the 'survival' and 'ggplot2' libraries. The code used can be found in online supplemental annex 2.

## RESULTS

### Characterisation of the population

As shown in [table 1](#) and [figure 1](#), 3 849 849 adult patients were admitted to hospitals in Colombia's contributory health system in 2016. During the first year of follow-up, there were 46 429 (1.21%) deaths, 19 169 (0.5%) of which occurred during hospitalisation. For the hospitalised population, the average age was 42.9 ( $\pm 17.9$  years), 41% of the patients were male and the median overall survival time was 213.9 days. The results regarding the total population's 1-year survival function are provided in [figure 2](#).

The three most frequent comorbidities were diabetes without complications (6.51%), CPD (6.21%) and ulcer disease (4.31%). The three least frequent comorbidities

were hemiplegia (0.06%), moderate or severe liver disease (0.09%) and congestive heart failure (0.56%). The differences between the prevalence of comorbidities in patients who survived and those who died are demonstrated in standardised differences, with greater differences found in age and presentation of congestive heart failure, CPD and cerebrovascular disease, while the presentation of HIV did not indicate differences in both groups. Furthermore, for the hospitalised population, 3 064 966 patients had no comorbidities (79.61%), 716 350 had one or two (18.61%), 63 712 had three or four (1.65%) and 4821 had five or more (0.13%). The average comorbidity per subject was 0.29 ( $\pm 0.66$ ).

As shown in [table 2](#), crude mortality gradually rose as patients' age increased. While crude mortality was 0.14% for patients under age 49 years, it reached 35.46% for patients over 100 years old.

### Validation of the original CCI score (first version of the index) Recalibration of the CCI for the Colombian population

A second version of the CCI was obtained by recalibrating the weights of the 17 comorbidities contained in the first version of the index ([table 3](#)). Regarding the first version of the CCI, this study determined that metastatic tumours and HIV were the two morbidities with the greatest differences, originally having a weight of 6 - and assigned new weights of 4 and 0, (1) respectively. Meanwhile, no differences were found in the weights assigned to myocardial infarction, congestive heart failure, peripheral vascular disease and hemiplegia (see online supplemental figure S1).

### New version of the CCI

Furthermore, the results obtained made it possible to generate a third version of the CCI ([table 3](#)), in which 14 of the variables proposed in the first version were retained because of each comorbidity's statistical significance and HR magnitude. Retention of variables in the index's third version is explained below.

First, one comorbidity was excluded for having a new weight of 0, specifically connective tissue disease. Second, considering that it was difficult to distinguish among the degrees of severity with liver disease based on the diagnostic codes available in the administrative databases—as previously reported by Armitage and van der Meulen<sup>16</sup>—mild, moderate and severe liver disease categories all were unified into one 'liver disease' category. Third, also due to the difficulty of distinguishing among degrees of severity, the categories 'chronic diabetes without complications' and 'diabetes with end-organ damage' were combined into a single 'diabetes' category.

### Survival according to CCI score categories

The score obtained in the third version of the CCI was transformed into a categorical variable to estimate the population's survival during the follow-up year, indicating significant differences among the

**Table 1** General characteristics of patients hospitalised in 2016 in Colombia

Baseline characteristics	Full sample	Alive	Mortality	SMD*
Total population, n (%)	3 849 849 (100)	3 802 982 (98.79)	46 867 (1.21)	
Age, mean (SD)	42.88 (17.9)	42.49 (17.5)	73.9 (15.4)	189.5
Male gender, n (%)	1 578 853 (41.0)	1 556 067 (40.91)	22 786 (49.08)	16.5
Charlson Index comorbidities, n (%)				
Myocardial infarct	67 796 (1.76)	63 276 (1.66)	4520 (9.74)	35.4
Congestive heart failure	21 374 (0.56)	19 380 (0.51)	1994 (0.55)	109.3
Peripheral vascular disease	23 251 (0.60)	21 581 (0.57)	1670 (3.60)	21.3
Cerebrovascular disease	53 271 (1.38)	47 103 (1.24)	6168 (13.28)	47.7
Dementia	39 317 (1.02)	33 735 (0.89)	5542 (11.94)	46.3
Chronic pulmonary disease	239 178 (6.21)	224 365 (5.90)	14 813 (31.90)	70.4
Connective tissue disease	74 124 (1.93)	72 121 (1.90)	2003 (4.31)	14
Ulcer disease	166 075 (4.31)	158 916 (4.18)	7159 (15.42)	38.5
Diabetes without chronic complications	254 476 (6.51)	244 687 (6.43)	9789 (21.08)	43.5
Diabetes with end-organ damage	68 706 (1.78)	64 862 (1.71)	3844 (8.28)	30.5
Mild liver disease	4503 (0.12)	4009 (0.11)	494 (1.06)	12.6
Moderate or severe liver disease	3355 (0.09)	2878 (0.08)	477 (1.03)	12.9
Moderate or severe renal disease	19 566 (0.51)	16 576 (0.44)	2990 (6.44)	33.4
Hemiplegia	1958 (0.05)	1831 (0.05)	127 (0.27)	21.4
Any tumour (including leukaemia and lymphoma)	10 219 (0.27)	8626 (0.23)	1593 (3.43)	24.1
Metastatic solid tumour	31 924 (0.83)	28 000 (0.74)	3924 (8.45)	37.5
HIV/AIDS	26 274 (0.68)	25 624 (0.67)	650 (1.40)	7.2
Number of Charlson comorbidities				
0	3 064 966 (79.61)	3 052 031 (80.24)	12 935 (27.86)	123
1–2	716 350 (18.61)	692 393 (18.20)	23 957 (51.60)	74.8
3–4	63.712 (1.65)	55 161 (1.45)	8551 (18.42)	59.2
≥5	4821 (0.13)	3.835 (0.10)	986 (2.12)	19.4
Comorbidities mean (SD)	0.29 (0.66)	0.27 (0.63)	1.46 (1.30)	115
Mortality, n (%)				
In-hospital mortality	19 169 (0.5)			
30-day mortality	18 651 (0.48)			
1-year mortality	46 429 (1.21)			
Survival time days, mean (SD)	213.9 (105.8)			

\*SMD between alive and mortality cohort.  
SMD, standardised mean difference.

categories. While mortality was 0.07% for the category corresponding to a CCI score of 0, it was 10.46% for the category corresponding to CCI scores equal to or greater than 5. [Figure 3](#) and online supplemental figure S2 present survival curves by category.

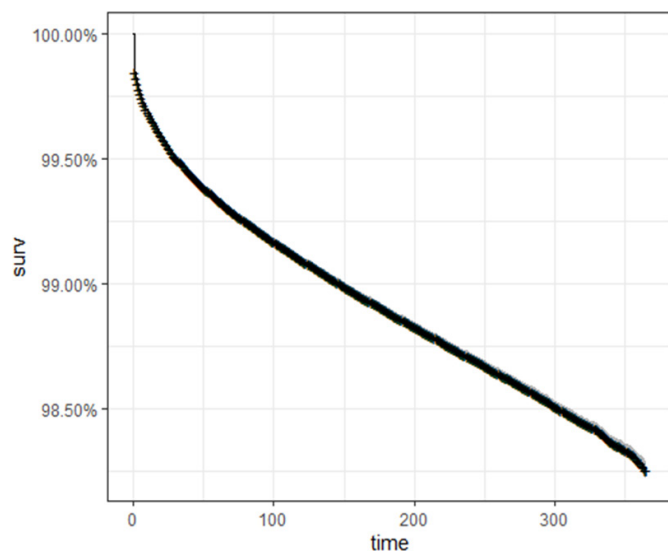
### The models' discriminative ability

As shown in [figure 4](#), population B was used to determine the discriminative ability of each of the three versions of the CCI to calculate 1-year survival. In the index's first version, the area under the ROC curve was 0.906 (95% CI (0.906 to 0.907),  $p < 0.001$ ). In the second version, it was 0.908 (95% CI (0.908 to 0.909),

$p < 0.001$ ). In the third version, it was 0.909 (95% CI (0.908 to 0.910),  $p < 0.001$ ).

### Adjustment of the models

The three models' goodness of fit was evaluated using Harrell's index of concordance. For the original model, a concordance of 0.901 (SE=0.001) was found, while for the score obtained by recalibrating the weights based on the risk ratios in the Colombian population, the Harrell's index was 0.903 (SE=0.001). Finally, the score obtained with the new model, which retained 14 comorbidity variables, presented a concordance of 0.904 (SE=0.001).



**Figure 2** Kaplan-Meier survival curve for the general population of hospitalised patients in 2016.

## DISCUSSION

The present study's findings can be classified in two ways: (1) those related to the characterisation of the Colombian population examined, that is, age, comorbidities' prevalence and 1-year mortality, and (2) those related to the CCI score's performance in predicting 1-year survival. Regarding the findings related to the characterisation of the Colombian population examined in this study, the comorbidities with the greatest prevalence were diabetes (8.09%) and CPD (6.21%). Altogether, 82.2% did not have any comorbidities: The average of comorbidities per individual was 0.24 ( $\pm 0.6$ ), and 1-year mortality was 1.21%. Regarding the CCI score's performance, the three tested scores resulted in very similar values in terms of discriminative ability and adjustment.

Furthermore, regarding the population's characterisation, significant age differences and prevalence of comorbidities were found between the Colombian population and other populations. A comparison with previous studies found that the percentage of hospitalised patients in Colombia who were aged 65 years and up (14.2%) was notably less than that of other countries, such as Japan (56.9%), New Zealand (37.2%) and Korea (26%).<sup>20</sup> Prevalence of comorbidities in the Colombian population also differed from other countries. While the percentage

of hospitalised patients with at least one comorbidity was 17.8% in Colombia, it was 29.6% in France and 74.43% in Korea.<sup>13 21 22</sup> The differences observed between these populations suggest an association between prevalence of comorbidities and hospitalised patients' ages. In-hospital mortality data also suggest an association between age and number of comorbidities, in that the Colombian population was found to have an in-hospital mortality rate of 0.5%, which is lower than the range found in reports of other populations<sup>13</sup>—from 1.4% (New Zealand) to 4.4% (Canada).

Regarding the recalibration of the weights assigned to CCI comorbidities, it is notable that its application to the Colombian population resulted in significant variations. Variations in scores assigned to a comorbidity greater than or equal to two points were deemed significant, considering that a variation of one point was viewed as small in terms of the risk of 1-year mortality and could be attributable to the particular comorbidity's interaction with other comorbidities, rather than to variations in the individual risk of the comorbidity itself.

Three comorbidities exhibited a descending variation of two or more scores with respect to the original CCI score: moderate or severe liver disease; solid metastatic tumour and HIV. These variations could be explained by the notable increase in the availability and effectiveness of treatments for these comorbidities, which were developed since the publication of the original CCI scores. In the case of moderate or severe liver disease, the decrease of two points may be associated with the implementation of liver transplants, as well as advances in pharmacological management. In the case of solid metastatic tumours, the reduction of two points could be related to advances in treatments such as chemotherapy, radiation therapy and surgery. As for HIV, the reduction of five points may be attributable to the implementation of more effective antiretroviral therapies.<sup>23</sup>

Meanwhile, chronic kidney disease was the only comorbidity with an ascending variation of two or more points, which may be associated with a stronger evolution towards a chronic condition.

Regarding the generation of the model's third version, variables were retained based on their levels of statistical significance. Two procedures were used to reduce the number of variables in this third version, from 17 to 14: regrouping (diabetes and liver disease) and exclusion (connective tissue disease, which had a score of 0). The regroupings in the third version were consistent with proposals from other authors<sup>13 14 20 22</sup> who noted the difficulty of establishing a clear cut-off point between the two comorbidities' degrees of severity. Regarding the exclusion of connective tissue disease, this was based on the lack of a statistically significant association between this comorbidity and mortality. Furthermore, the decrease in the score assigned to this comorbidity may be related to an increase in the effectiveness of treatments for this condition.

**Table 2** Crude mortality by age groups

Age	Alive cohort, n (%)	Mortality cohort, n (%)
18–49	2 564 595 (99.86)	3709 (0.14)
50–59	529 339 (99.21)	4230 (0.79)
60–69	358 887 (97.92)	7615 (2.08)
70–79	217 876 (95.11)	11 201 (4.89)
80–89	112 483 (89.05)	13 836 (10.95)
90–99	19 825 (77.94)	5610 (22.06)
100–109	415 (64.54)	228 (35.46)

**Table 3** Comparison of the weights of each comorbidity for 1-year mortality with the original Charlson weights

Comorbidity	Original weight	Original model HR (95% CI)	New weight	New model HR (95% CI)	New model weight
Myocardial infarct	1	1.23 (1.17 to 1.28)	1	1.23 (1.17 to 1.23)	1
Congestive heart failure	1	1.21 (1.13 to 1.29)	1	1.21 (1.14 to 1.30)	1
Peripheral vascular disease	1	1.13 (1.06 to 1.22)	1	1.14 (1.06 to 1.22)	1
Cerebrovascular disease	1	1.79 (1.72 to 1.87)	2	1.79 (1.72 to 1.87)	2
Dementia	1	1.52 (1.45 to 1.59)	2	1.52 (1.45 to 1.59)	2
Chronic pulmonary disease	1	1.64 (1.59 to 1.69)	2	1.64 (1.59 to 1.69)	2
Connective tissue disease	1	0.99 (0.94 to 1.06)	0	—	
Ulcer disease	1	1.49 (1.44 to 1.55)	2	1.50 (1.44 to 1.55)	2
Moderate or severe renal disease	2	3.59 (3.41 to 3.79)	4	3.61 (3.42 to 3.81)	4
Hemiplegia	2	1.74 (1.36 to 2.21)	2	1.74 (1.37 to 2.22)	2
Diabetes without chronic complications*	1	1.11 (1.07 to 1.15)	1	—	
Diabetes with end-organ damage*	2	1.15 (1.09 to 1.20)	1	—	
Any tumour (including leukaemia and lymphoma)	2	2.60 (2.41 to 2.80)	3	2.60 (2.42 to 2.80)	3
Metastatic solid tumour	6	4.03 (3.84 to 4.23)	4	4.03 (3.84 to 4.23)	4
HIV/AIDS	6	1.27 (1.14 to 1.42)	1	1.28 (1.14 to 1.43)	1
Mild liver disease	1	2.43 (2.14 to 2.75)	2	—	
Moderate or severe liver disease*	3	2.83 (2.48 to 3.22)	3	—	
Single diabetes category**		—		1.12 (1.09 to 1.15)	1
Single liver disease category**		—		2.61 (2.38 to 2.86)	3

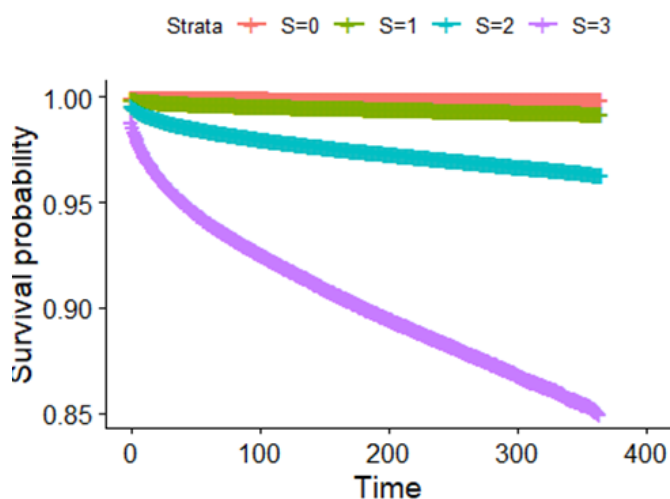
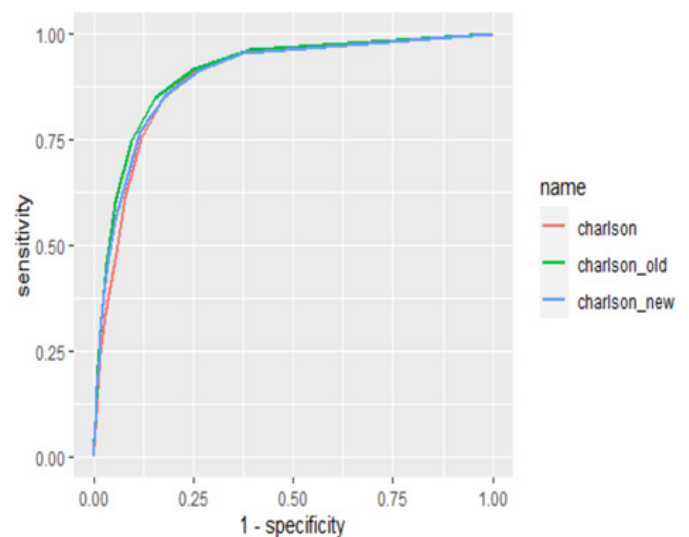
Weight of 1 between  $\geq 1.0$  and  $< 1.5$ ; weight of 2 between  $\geq 1.5$  and  $< 2.5$ ; weight of 3 between  $\geq 2.5$  and  $< 3.5$ ; weight of 4 between  $\geq 3.5$  and  $< 4.5$ .

\*\*Grouped categories - \*\*A single categories; adjusted by sex and age.

The Quan *et al*<sup>9</sup> model notably excludes peripheral vascular disease and myocardial infarction due to these comorbidities' lack of statistical significance. Nevertheless, the present study retained these variables in the model because it found a statistically significant association between these conditions and mortality.

Furthermore, researchers such as Quan *et al*<sup>13</sup> and Armitage and van der Meulen<sup>16</sup> have proposed excluding

ulcer disease because of the possibility of reversing the chronic condition with pharmacological treatments. Nevertheless, the recalibration herein of the CCI scores for the Colombian population resulted in an increase in the weight assigned to this comorbidity (from one to two


**Figure 3** Charlson Comorbidity Index score categories and 1-year survival.

**Figure 4** The three Charlson Comorbidity Index models' comparative curves.

points) due to the increase in the magnitude of the risk of death.

Regarding the models' ability to discriminate, the area below the ROC curve indicates that the three versions of the CCI examined herein exhibited values that not only were very similar (0.906 for the first version, 0.908 for the second version and 0.909 for the third version), but also surpassed those of the previous models generated with other populations, as indicated by the C-statistical values reported in the study by Quan for six countries (2010).

The Harrell's C-index values<sup>24</sup> were very similar for the three versions of the CCI studied herein (0.901, SE=0.001 for the first version; 0.903, SE=0.001 for the second version; and 0.904, SE=0.001 for the third version) and also exceeded the adjustment threshold of 0.7, which was deemed acceptable for predicting survival.

With regard to evaluating adjustment of the model by comparing predicted survival with estimated observed survival, as described by Harrell *et al.*,<sup>18</sup> calibration was poor when the likelihood of survival was under 0.7 and adequate with values over 0.7, as other authors also reported.<sup>21</sup> This lack of adjustment in the model is due to the few events in the lowest-risk categories.

The study herein presented limitations in studies that used administrative databases, such as a lack of detailed comorbidity records (laboratory results, interpretations of diagnostic tests), a lack of data that enable establishing each comorbidity's degree of severity and under-recording diagnoses. The latter warrants special attention because it can result in underestimating the prevalence of comorbidities that are more difficult to diagnose, while not necessarily involving bias. Nevertheless, it is notable that generally, this study should correct the risk of under-recording comorbidities by taking into account ICD-10 codes, CUPS codes and records of medications administered to patients for up to 1 year before hospitalisation.

Finally, it is important to highlight the present study's strengths. First, it is notable that few studies have been conducted with a heterogeneous cohort of patients, a similar sample size and the inclusion of all hospitalisation causes. This increases both the generalisability of this study's findings and its accuracy in weighting included variables. Second, the predictive value of mortality attained by the model herein is notable, which, after reducing the number of retained comorbidities to 14, reached the same level of discrimination as that of previous models.

**Contributors** Both authors contributed to the study's design. HO and GB were responsible for data integrity and the accuracy of data analysis. Both wrote, edited/revised and approved the final manuscript. HO as guarantor, accepts all responsibility for the work and/or conduct of the study, had access to the data and controlled the decision to publish.

**Funding** Pontificia Universidad Javeriana funded this study (grant number not applicable).

**Competing interests** None declared.

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

**Patient consent for publication** Not required.

**Ethics approval** In the present study, there was no patient participation, as the data comprised anonymised records. The Ethics Review Board of Pontificia Universidad Javeriana in Bogota, Colombia, approved this study (record no. 14/2020).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplemental information. Henry Oliveros (c).ORCID ID: 0000-0002-7632-906X; Program in Clinical Epidemiology, Department of Clinical Epidemiology and Biostatistics, Pontificia Universidad Javeriana, Bogotá, Colombia. Email :h.oliverosr@javeriana.edu.co.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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

## REFERENCES

- Roder D, Buckley E. Administrative data provide vital research evidence for maximizing health-system performance and outcomes. *Asia Pac J Clin Oncol* 2017;13:111–4.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
- 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.
- Sharabiani MTA, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Med Care* 2012;50:1109–18.
- Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45:613–9.
- Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *J Clin Epidemiol* 1993;46:1075–9.
- Halfon P, Egger Y, van Melle G, *et al.* Measuring potentially avoidable hospital readmissions. *J Clin Epidemiol* 2002;55:573–87.
- Sundararajan V, Henderson T, Perry C, *et al.* New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;57:1288–94.
- Quan H, Sundararajan V, Halfon P, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.
- Nuttall M, van der Meulen J, Emberton M. Charlson scores based on ICD-10 administrative data were valid in assessing comorbidity in patients undergoing urological cancer surgery. *J Clin Epidemiol* 2006;59:265–73.
- Christensen S, Johansen MB, Christiansen CF, *et al.* Comparison of Charlson comorbidity index with SAPs and APACHE scores for prediction of mortality following intensive care. *Clin Epidemiol* 2011;3:203–11.
- Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. *J Clin Epidemiol* 2011;64:1426–33.
- Quan H, Li B, Couris CM, *et al.* Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol* 2011;173:676–82.
- Thygesen SK, Christiansen CF, Christensen S, *et al.* The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol* 2011;11:83.
- Ramirina RA, Ramirina BL, Almeida RMVR, *et al.* Comorbidity adjustment index for the international classification of diseases, 10th revision. *Rev Saude Publica* 2008;42:590–7.



- 16 Armitage JN, van der Meulen JH, Royal College of Surgeons Co-morbidity Consensus Group. Identifying co-morbidity in surgical patients using administrative data with the Royal College of surgeons Charlson score. *Br J Surg* 2010;97:772–81.
- 17 Charlson M, Szatrowski TP, Peterson J, *et al*. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994;47:1245–51.
- 18 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;15:361–87.
- 19 Moons KGM, Kengne AP, Grobbee DE, *et al*. Risk prediction models: II. external validation, model updating, and impact assessment. *Heart* 2012;98:691–8.
- 20 Choi JS, Kim M-H, Kim YC, *et al*. Recalibration and validation of the Charlson comorbidity index in an Asian population: the National health insurance Service-National sample cohort study. *Sci Rep* 2020;10:13715.
- 21 Bannay A, Chaignot C, Blotière P-O, *et al*. The best use of the Charlson comorbidity index with electronic health care database to predict mortality. *Med Care* 2016;54:188–94.
- 22 Kil S-R, Lee S-I, Khang Y-H, *et al*. Development and validation of comorbidity index in South Korea. *Int J Qual Health Care* 2012;24:391–402.
- 23 Simon V, Ho DD, Abdool Karim Q. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. *Lancet* 2006;368:489–504.
- 24 Harrell FE, Lee KL, Califf RM, *et al*. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;3:143–52.