BMJ Open

Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy

Giovanni Corrao,[1,2] Federico Rea,[1,2] Mirko Di Martino,[3] Rossana De Palma,[4] Salvatore Scondotto,[1,5] Danilo Fusco,[3] Adele Lallo,[3] Laura Maria Beatrice Belotti,[4] Mauro Ferrante,[6] Sebastiano Pollina Addario,[1,5] Luca Merlino,[1,7] Giuseppe Mancia,[8] Flavia Carle[1,9]

To cite: Corrao G, Rea F, Di Martino M, *et al*. Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy. *BMJ Open* 2017;**7**:e019503. doi:10.1136/bmjopen-2017-019503

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-019503).

Received 6 September 2017
Revised 6 November 2017
Accepted 17 November 2017

For numbered affiliations see end of article.

Correspondence to
Professor Giovanni Corrao;
giovanni.corrao@unimib.it

ABSTRACT

Objective To develop and validate a novel comorbidity score (multisource comorbidity score (MCS)) predictive of mortality, hospital admissions and healthcare costs using multiple source information from the administrative Italian National Health System (NHS) databases.

Methods An index of 34 variables (measured from inpatient diagnoses and outpatient drug prescriptions within 2 years before baseline) independently predicting 1-year mortality in a sample of 500 000 individuals aged 50 years or older randomly selected from the NHS beneficiaries of the Italian region of Lombardy (training set) was developed. The corresponding weights were assigned from the regression coefficients of a Weibull survival model. MCS performance was evaluated by using an internal (ie, another sample of 500 000 NHS beneficiaries from Lombardy) and three external (each consisting of 500 000 NHS beneficiaries from Emilia-Romagna, Lazio and Sicily) validation sets. Discriminant power and net reclassification improvement were used to compare MCS performance with that of other comorbidity scores. MCS ability to predict secondary health outcomes (ie, hospital admissions and costs) was also investigated.

Results Primary and secondary outcomes progressively increased with increasing MCS value. MCS improved the net 1-year mortality reclassification from 27% (with respect to the Chronic Disease Score) to 69% (with respect to the Elixhauser Index). MCS discrimination performance was similar in the four regions of Italy we tested, the area under the receiver operating characteristic curves (95% CI) being 0.78 (0.77 to 0.79) in Lombardy, 0.78 (0.77 to 0.79) in Emilia-Romagna, 0.77 (0.76 to 0.78) in Lazio and 0.78 (0.77 to 0.79) in Sicily.

Conclusion MCS seems better than conventional scores for predicting health outcomes, at least in the general population from Italy. This may offer an improved tool for risk adjustment, policy planning and identifying patients in need of a focused treatment approach in the everyday medical practice.

Strengths and limitations of this study

- The multisource comorbidity score (MCS) combines data from administrative health sources currently available in all Italian regions into a tool able to measure comorbidity, and to predict 1-year mortality, and even other adverse outcomes.
- The study was based on a very large unselected population, which was made possible because in Italy a public-funded healthcare system involves virtually all citizens.
- MCS was both internally and externally validated, and tested on 2 million of individuals, a sample very large and representative of the whole Italian population.
- Data on outpatient services, education, functional status, caregiver availability and markers of social instability were not included in the prediction model.

INTRODUCTION

Comorbidity has been defined as the total burden of illnesses unrelated to the patient's principal diagnosis.[1] Ideally, in any given individual, assessment of comorbidity should be based on complete information on his/her clinical and demographic profile. However, this is so time consuming and costly that for large populations attention has been directed to measures that make use of data available via computerised information systems.[2] The Charlson Comorbidity Score[3] and the Chronic Disease Score (CDS),[4] that is, two popular indices respectively based on diagnostic coding system and prescribed medications, are extensively used comorbidity scores based on available computerised data.[5]

Most diagnosis-based comorbidity scores have been developed from hospital-based surveys reviewing inpatients' medical records, and only later they were adapted for use with population-based administrative data.[3–13] Conversely, few instruments have been developed from administrative data,[14] without

however providing a weighting system for scoring comorbidity indices.[5] As sick people are likely to receive pharmacotherapy and because the number of prescribed drugs has been shown to increase with the number of chronic disease conditions,[15] medication-based scores offer an alternative tool for measuring comorbidities.[16] However, convincing evidence suggesting the superiority of an approach (eg, of medication-based towards diagnosis-based comorbidity scoring) in predicting health outcomes is not currently available.[5 17–19]

Our population-based study was performed under the auspices of the Italian Health Ministry. We aimed to develop and validate a novel comorbidity score predictive of mortality, hospital admissions and healthcare costs using multiple source information from the administrative Italian National Health System (NHS) databases.

METHODS

All Italian citizens have equal access to healthcare services as part of the NHS. Computerised information systems of healthcare utilisation databases have been created within each of the 21 Italian regions to collect a variety of information, at least including: (1) demographic and administrative data of residents who receive NHS assistance; (2) hospital discharge records reporting information on primary diagnosis, up to five coexisting conditions and procedures coded according to the International Classification of Diseases, 9th Revision (ICD-9) Clinical Modification classification system (http://icd9.chrisendres.com/); (3) drug prescriptions reimbursed by the NHS coded according to the Anatomical Therapeutic Chemical (ATC) classification system (https://www.whocc.no/atc_ddd_index/). Record linkage between databases was performed within each region by means of the identification (ID) code assigned to each NHS beneficiary. In order to preserve the privacy of the beneficiaries, ID codes were deidentified, and the conversion table was deleted.

The Ethical Committee of the University of Milano-Bicocca evaluated the protocol and decided that the study (1) was exempt from informed consent, and (2) provided sufficient guarantees of individual records' deidentification.

The healthcare utilisation data were used for empirically developing a risk prediction model using the methods described by May *et al*,[20] Royston *et al*[21] and Riley *et al*.[22]

Candidate predictors

Starting from the lists included under the Charlson, Elixhauser and Chronic Disease Scores (respectively denoted CCI, EI and CDS), we developed a list of 46 diseases and conditions classified as infectious and parasitic diseases (2), neoplasms (4), endocrine, nutritional and metabolic diseases, and immunity disorders (6), diseases of the blood and blood-forming organs (2), mental disorders (7), diseases of the nervous (5), circulatory (9), respiratory (2), digestive (3), genitourinary (3) systems, diseases of the musculoskeletal system and connective tissue (1), and other conditions (2). Of the 46 included conditions, 18 were traced from inpatient diagnostic codes only, 6 from outpatients prescribed drugs only, and the remaining 22 from both, diagnostic and therapeutic codes, depending on availability of specific diagnostic codes and drug therapies supplied free of charge from the Italian NHS. Two of us (FR and GM) independently chose ICD-9 and ATC codes capturing individuals who experienced each of the 46 included conditions. Discrepancies were resolved in conference.

The entire list of candidate predictors, and the corresponding codes, are reported in the online supplementary table S1.

Score development

With the aim of selecting conditions independently able to predict 1-year mortality (ie, the main outcome of interest), we proceeded as follows. First, a training (derivation) set of 500 000 individuals was randomly selected from individuals who in 2008 were: (1) aged 50 years or older, (2) NHS beneficiaries, and (3) resident in Lombardy from at least 2 years. Data were retrieved from the databases of Lombardy, a region of Italy that accounts for about 16% of its population, being almost 4 million those aged 50 years or more. Second, the relationship between the selected covariates and the time to death was investigated by fitting parametric survival models based on the Weibull distribution. Covariates included into the model were gender, age (in 1 January 2008), and the 46 above reported diseases or conditions which were made available, respectively, by patient hospitalisations and outpatient prescriptions in the years 2006 and 2007. These data entered as dichotomous variables into the model, with value 0 or 1 according to whether the specific condition was not or was recorded at least once within 2 years prior to baseline (2006–2007). Third, the least absolute shrinkage and selection operator (LASSO) method was applied for selecting the diseases/conditions able to predict 1-year mortality.[23] LASSO selects variables correlated to the measured outcome by shrinking coefficient weights, down to zero for the ones not correlated to outcome. Finally, the coefficients estimated from the model were used for assigning a score at each selected covariate. In particular, the coefficients were converted into scores by multiplying them by 10 and rounding them to the nearest whole number,[24] which were sequentially summed to produce a total aggregate score. To simplify the system, that is, with the aim of accounting for excessive heterogeneity of the total aggregate score, the latter was categorised by assigning increasing values of 0, 1, 2, 3 and 4 to the categories of the aggregate score of 0–4, 5–9, 10–14, 15–19 and ≥20, respectively. The index so obtained was termed multisource comorbidity score (MCS).

Model validation

Internal and external validity of MCS was investigated by applying the score developed from individuals belonging to the training set, to several validation sets. These latter

were selected by applying the same inclusion/exclusion criteria of the training set.

The following two-stage validation procedure was applied. First, the MCS performance was explored with respect to other prognostic scores by applying the current multisource comorbidity, the CCI, the EI and the CDS, to an internal validation set of 500 000 NHS beneficiaries from Lombardy. Two approaches were used with this aim. One, the discriminatory power was assessed by constructing the receiver operating characteristic (ROC) curve and calculating the area under the ROC curves (AUC). Two, the net reclassification improvement (NRI) was calculated to assess the improvement of risk classification of MCS with respect to CCI, EI and CDS.[25] The NRI measures the net proportion of subjects correctly reclassified by MCS by evaluating the predicted probability among those who experienced and those who did not experience the outcome.

Second, three external validation sets, each consisting of 500 000 NHS beneficiaries, were selected from a Northern (Emilia-Romagna), Central (Lazio) or Southern (Sicily) Italian region and considered jointly with the internal validation set. The total population of these regions amounts to about 21.4 million NHS beneficiaries, that is, more than one-third of the Italian population (35.3%). Due to the heterogeneity of data availability, different periods had to be considered for different regions, that is, 2008 for Lombardy, 2010 for Emilia-Romagna and Lazio, and 2013 for Sicily. Between-regions consistence of MCS performance was tested by comparing AUC estimates and Kaplan-Meier 1-year survival probabilities stratified by MCS.

Sensitivity analysis and secondary outcomes

Because of the arbitrary nature of score categorisation (see the above reported description of the score development), in a secondary analysis we verified the MCS robustness in predicting 1-year mortality by comparing the probability of survival (Kaplan-Meier curves) of the internal validation sample as stratified according to MCS categories alternative to that used in the main analysis.

Further analyses were performed for evaluating whether MCS may predict other secondary outcomes including: (1) the 5-year all-cause mortality; (2) the 1-year and 5-year hospital admissions for all causes; and (3) the 2-year hospital costs measured from the perspective of the Italian NHS. Secondary outcomes were referred to 1000 person-years and calculated along the categories of MCS within the internal validation set.

RESULTS

MCS score

Factors which mostly contributed to the total aggregate score were metastatic cancer, alcohol abuse, cancer without metastasis and tuberculosis, while arrhythmia, obesity and hypothyroidism provided small, although significant, contributions (table 1).

Table 1 Assignment of weights in building the multisource comorbidity score (MCS) through a time-to-death multivariate Weibull model

Disease/condition	Regression coefficient (SE)	Weight
Metastatic cancer	1.63 (0.04)	18
Alcohol abuse	0.99 (0.16)	11
Cancer, without metastasis	0.91 (0.03)	10
Tuberculosis	0.88 (0.28)	10
Psychoses	0.77 (0.05)	8
Liver disease	0.72 (0.05)	8
Anxiety medication	0.52 (0.23)	6
Weight loss	0.51 (0.12)	6
Dementia	0.51 (0.06)	6
Malignancy medication	0.49 (0.05)	5
Parkinson's disease	0.49 (0.09)	5
Lymphoma	0.46 (0.08)	5
Hemiplegia and hemiparesis	0.46 (0.09)	5
Coagulation defects	0.43 (0.10)	5
Disorders of fluid, electrolyte and acid-base balance	0.40 (0.06)	4
Kidney diseases	0.39 (0.04)	4
Kidney dialysis	0.36 (0.17)	4
Heart failure	0.35 (0.02)	4
Other neurological diseases	0.32 (0.07)	3
Rheumatoid arthritis	0.27 (0.11)	3
Anaemias	0.26 (0.04)	3
Cerebrovascular diseases	0.25 (0.03)	3
Diabetes	0.20 (0.02)	2
Vascular diseases	0.20 (0.05)	2
Gout	0.18 (0.03)	2
Epilepsy	0.18 (0.03)	2
Chronic pulmonary diseases	0.16 (0.02)	2
Peptic ulcer	0.16 (0.02)	2
Acute myocardial infarction	0.11 (0.04)	1
Coronary and peripheral vascular disease	0.11 (0.02)	1
Valvular diseases	0.10 (0.06)	1
Arrhythmia	0.09 (0.02)	1
Obesity	0.08 (0.10)	1
Hypothyroidism	0.07 (0.09)	1

Overall, 86.4% and 1.2% of NHS beneficiaries respectively had the lowest (0) and the highest (4) MCS value. The less favourable prognosis of men and elderly people with respect to women and young people was caught by the novel prognostic score. The prevalence of NHS beneficiaries belonging to the lowest MCS category progressively decreased with the increasing categories of age from 94% to 64% in men and from 95% to 72% in women (figure 1).

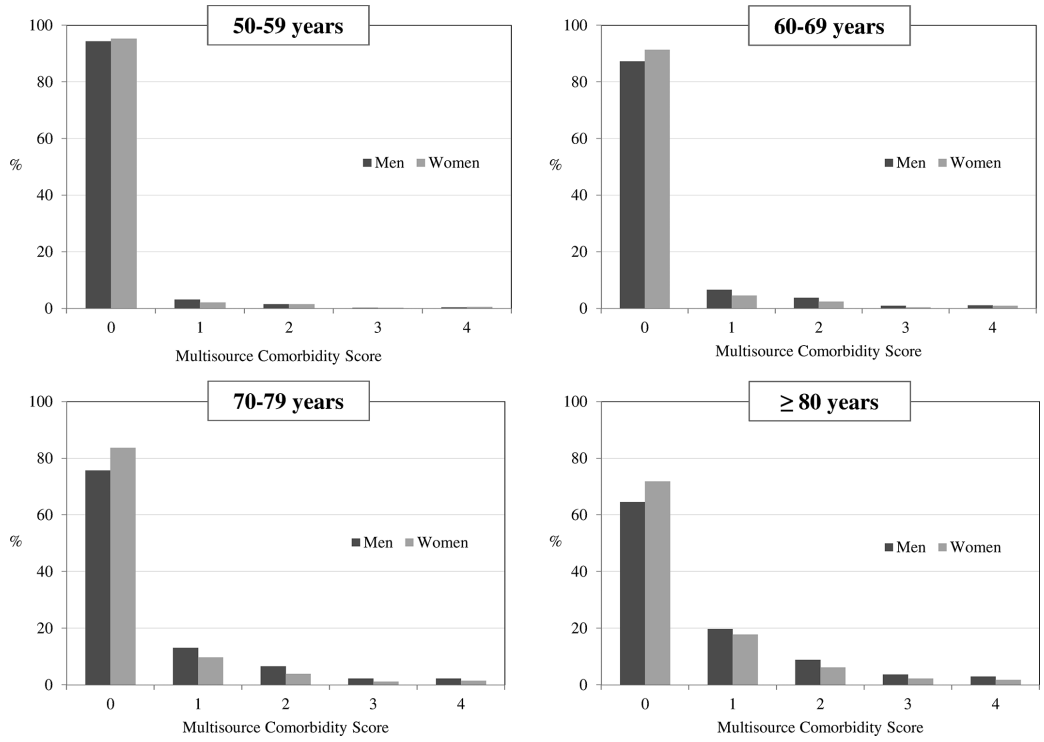

Figure 1 Multisource comorbidity score distribution among National Health System (NHS) beneficiaries (internal validation set) according to their gender and age category.

MCS compared with other comorbidity scores

The AUC values (95% CI) of MCS, CCI, EI and CDS were 0.78 (0.77 to 0.79), 0.69 (0.68 to 0.70), 0.65 (0.64 to 0.66) and 0.69 (0.68 to 0.70), respectively (figure 2).

Performance analyses using NRI showed that MCS significantly improved the net 1-year mortality reclassification by all other scores, the magnitude of the improvement being 38.8% (95% CI 36.9 to 40.7; P<0.0001) when compared with the CCI, 68.8% (95% CI 66.8 to 70.7; P<0.0001) when compared with the EI and 27.2% (95% CI 25.3 to 29.1; P<0.001) when compared with the CDS. With respect to the CDS (the medication-based score), MCS improved by 17% the sensitivity of the correct reclassification of individuals who experienced the outcome (the deceased ones), whereas with respect to CCI and EI (ie, the diagnosis-based scores), it improved the correct reclassification of individuals who did not experience the outcome (the survivors) by 37% and 67%, respectively.

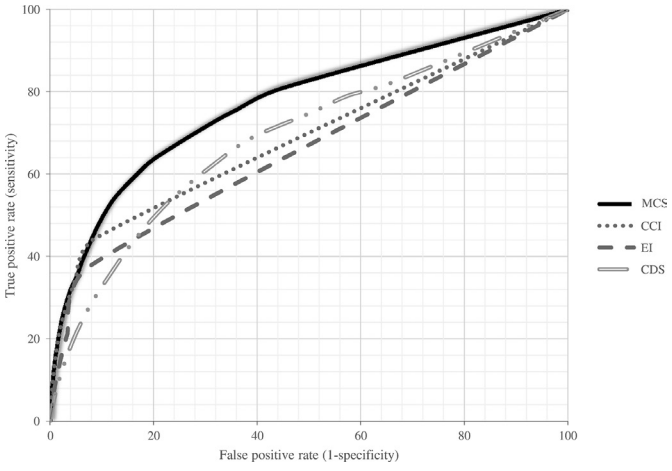

Figure 2 Receiver operating characteristic (ROC) curves comparing discriminant power of multisource comorbidity score (MCS), Charlson Comorbidity Index (CCI), Elixhauser Index (EI) and Chronic Disease Score (CDS) in predicting 1-year survival among National Health System (NHS) beneficiaries (internal validation set).

MCS model performance across Italian regions

The AUC values (95% CI) of MCS showed superimposable values in the four regions, that is, 0.78 (0.77 to 0.79), 0.78 (0.77 to 0.79), 0.77 (0.76 to 0.78) and 0.78 (0.77 to 0.79) in Lombardy, Emilia-Romagna, Lazio and Sicily, respectively (figure 3). In addition, in all four regions there was a progressively reduction of 1-year survival as MCS increased (figure 4).

Sensitivity analyses and other secondary outcomes

A reduced 1-year survival with increasing MCS values was observed also when alternative criteria for categorising the MCS were employed (online supplementary figure S1). This was the case also when secondary outcomes, rather the 1-year mortality, were considered (figure 5), the NHS beneficiaries with the highest MCS score (MCS=4) exhibiting 5-year mortality rates, 1-year and 5-year hospital admission rates and 2-year hospital costs, respectively, ninefold, eightfold, sixfold and eightfold higher than NHS beneficiaries with the lowest MCS score (MCS=0).

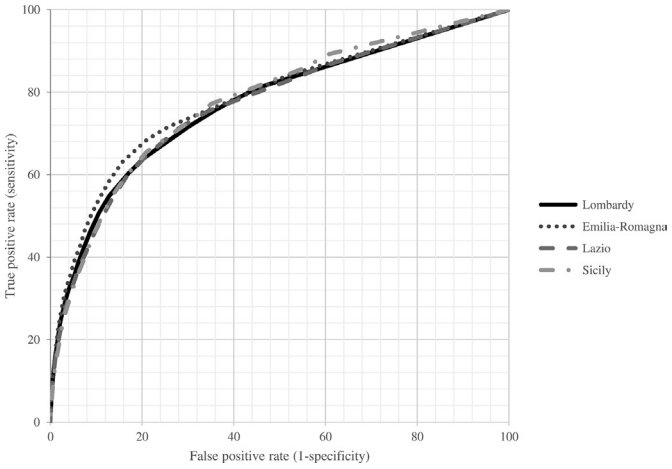

Figure 3 Receiver operating characteristic (ROC) curves comparing discriminant power of multisource comorbidity score (MCS) in predicting 1-year survival in four Italian regions (internal and external validation sets).

DISCUSSION

Our study shows that a simple score based on hospital diagnoses and drug prescriptions derived from current administrative data is able to stratify beneficiaries of Italian NHS according to their 1-year risk of death. It further shows that this score significantly improves the discriminatory power and net reclassification of commonly used prognostic scores, such as the CCI, the EI and the CDS. It finally shows that the score performance: (1) was comparable in northern, central and southern Italian general populations; and (2) was similarly valid for predicting long-term mortality, short-term and long-term number of hospital admissions, and 2-year cost of hospitalisations as calculated from the NHS perspective.

Although MCS was derived from the entire list of 46 diseases and conditions already used for developing CCI, EI and CDS, our score used more information than any of the previously validated comorbidity scores. In general, our MCS identified more individuals at higher risk of experiencing clinical outcomes than the CDS, another comorbidity score that integrates information about medications into its scoring. The MCS also was able to exclude more individuals at low risk of adverse outcomes than the other diagnosis-based comorbidity scores.

The present study has several strengths. First, although previous studies already identified predictors of mortality and other health outcomes,[5] to our knowledge MCS is the first combining inpatient diagnoses and outpatient drug prescriptions to stratify NHS beneficiaries according to comorbidities related to relevant clinical outcomes. Second, our study was based on a very large unselected population, which was made possible because in Italy a public-funded healthcare system involves virtually all citizens. Third, MCS was validated and tested on 2 million of NHS beneficiaries, a sample very large and representative of the entire Italian population. Fourth, because pharmacists are required to report drug prescriptions in detail in order to obtain reimbursement, and incorrect reports about the dispensed drugs have legal consequences the drug prescription database provided highly accurate data.[26] Finally, we avoided the selection of comorbidities based on opinion of experts[27 28] and prevalence data.[22 29 30] Moreover, with the aim of overcoming the limitation of conventional stepwise selection when several predictors must be analysed,[31 32] the LASSO model has been adopted. By shrinking variables with very unstable estimates towards zero, the LASSO model can effectively exclude some irrelevant variables generating sparse estimations.[33]

Several potential limitations must also be taken into account. First, predictors are restricted to those routinely

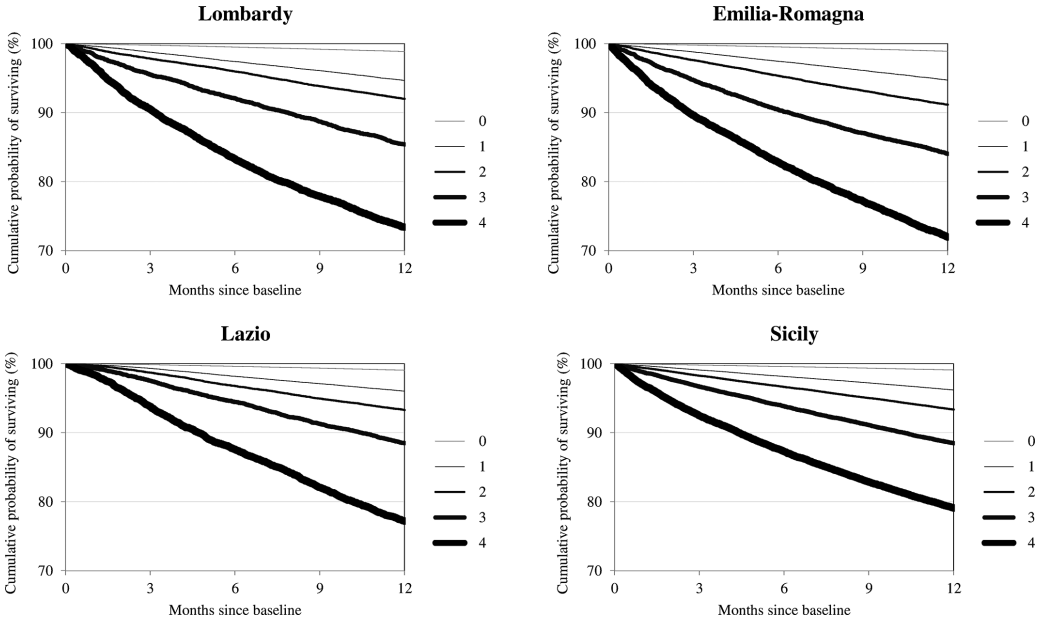

Figure 4 One-year Kaplan-Meier survival curves according to the value of the multisource comorbidity score (MCS) in four Italian regions (internal and external validation sets).

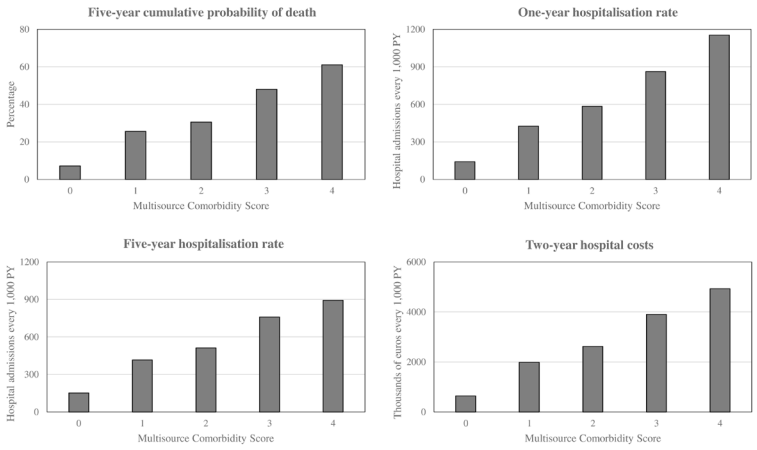

Figure 5 Five-year mortality, and hospital admissions and hospital cost annual rates according to the value of the multisource comorbidity score (MCS) of National Health System (NHS) beneficiaries (internal validation set). PY, person-years.

collected in all regions of Italy. This means that some data potentially relevant to clinical outcomes and healthcare costs such as outpatient services (including visits and diagnostic tests performed by specialised physicians and laboratories accredited by the NHS, payment exemptions, drugs directly delivered to inpatients and emergency room visits) were not considered because they were not ubiquitously available. Furthermore, the administrative databases did not contain information on the educational level, the functional patient's status, the caregiver availability and the markers of social instability, which have been shown to have a predictive value for the outcomes explored in our study.[34] This emphasises the interest of future research on additional predictors, and implies that there is potential for scores that predicts outcomes even more accurately than ours.

Second, our scoring system did not capture health services supplied from private providers. For example, the lack of evidence that depression predicts comorbidity-related outcomes might be due to our inability to capture patients who are not treated from public mental health services. However, given that the Italian NHS covers entirely essential healthcare needs, it is unlikely that diseases strongly affecting mortality escape its databases.

Third, misdiagnosis (due to poor accuracy in reporting diagnoses and comorbidities[35]) and upcoding (in pursuit of higher reimbursements[36]) of hospital records might have generated a conservative estimate of MCS performance. However, these diagnostic errors would affect similarly all diagnosis-based comorbidity scores, thereby failing to question our main result, that is, that MCS had a better performance than both the Charlson and Elixhauser scores.

Fourth, since outcomes are markedly influenced by the nature and quality of the healthcare system,[37] our scoring system might perform differently in countries other than Italy, which means that its applicability elsewhere in Europe will have to be tested. In this context, however, it is important to emphasise that the MCS performance showed an impressive stability throughout Italian regions where important differences in quality of, and accessibility to, healthcare services have been reported.[38] This suggests that its predictive value for mortality and other outcomes of medical relevance may persist under different settings.

Finally, we must be aware that the MCS may not apply to every relevant outcome and quantify the role of all conditions that may increase patients' risk of death. For example, our score cannot take into account of: (1) conditions that do not affect 1-year mortality; (2) NHS beneficiaries suffering a given condition who did not leave 'footprints' of routine medical care able to detect that condition (eg, untreated hypertensions); and (3) patients who did not survive at least 2 years after the onset of an acute condition (eg, fatal myocardial infarction).

CONCLUSIONS

In summary, we developed and validated a simple multisource prognostic score derived from data usually used for health system management, useful for predicting short-term and long-term risk of death, hospitalisation and high health costs of each individual NHS beneficiary.

MCS can represent a useful tool for risk adjustment in clinical and epidemiological studies, for assessing and health system performance and health policy planning, as well as an instrument for the identification of patients in need of a focused approach in everyday medical practice.

Author affiliations
[1]National Centre for Healthcare Research & Pharmacoepidemiology, at the University of Milano-Bicocca, Milan, Italy
[2]Laboratory of Healthcare Research & Pharmacoepidemiology, Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
[3]Department of Epidemiology, Lazio Regional Health Service, Rome, Italy
[4]Authority for Healthcare and Welfare, Emilia-Romagna Regional Health Service, Bologna, Italy
[5]Epidemiologic Observatory, Sicily Regional Health Service, Palermo, Italy
[6]Department of Culture and Society, University of Palermo, Palermo, Italy
[7]Epidemiologic Observatory, Lombardy Regional Health Service, Milan, Italy
[8]University of Milano-Bicocca, (Emeritus Professor), Milan, Italy
[9]Center of Epidemiology and Biostatistics, Polytechnic University of Marche, Ancona, Italy

Contributors GC and FC conceived the idea for this manuscript. GC designed the study and drafted the manuscript. FR, MDM, AL, LMBB, MF and SPA performed the data analysis. RDP, SS, DF and LM extracted the data and authorised their utilisation. GM assisted in interpreting the results under clinical prospective. All authors assisted the results interpretation and manuscript revision. All authors read and approved the final manuscript.

Funding This work was supported by the Italian Ministry of the Education, University and Research ('Fondo d'Ateneo per la Ricerca' portion, year 2015), grant number 2015-ATE-0524.

Disclaimer The Italian Ministry of the Education, University and Research had no role in the design of the study, the collection, analysis and interpretation of data, or the decision to approve publication of the finished manuscript.

Competing interests GC received research support from the European Community (EC), the Italian Agency of Drug (AIFA) and the Italian Ministry of Education, University and Research (MIUR). GC took part in a variety of projects that were funded by pharmaceutical companies (ie, Novartis, GSK, Roche, AMGEN, BMS). GC also received honoraria as member of Advisory Board from Roche.

Ethics approval The Ethical Committee of the University of Milano-Bicocca approved the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Leal JR, Laupland KB. Validity of ascertainment of co-morbid illness using administrative databases: a systematic review. *Clin Microbiol Infect* 2010;16:715–21.
2. Fishman PA, Goodman MJ, Hornbrook MC, *et al*. Risk adjustment using automated ambulatory pharmacy data: the RxRisk model. *Med Care* 2003;41:84–99.
3. 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.
4. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol* 1992;45:197–203.
5. 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.
6. 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.
7. 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.
8. Romano PS, Roost LL, Jollis JG. Further evidence concerning the use of a clinical comorbidity index with ICD-9-CM administrative data. *J Clin Epidemiol* 1993;46:1085–90.
9. Ghali WA, Hall RE, Rosen AK, *et al*. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol* 1996;49:273–8.
10. D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. *Methods Inf Med* 1993;32:382–7.
11. D'Hoore W, Bouckaert A, Tilquin C. Practical considerations on the use of the Charlson comorbidity index with administrative data bases. *J Clin Epidemiol* 1996;49:1429–33.
12. 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.
13. 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.
14. Elixhauser A, Steiner C, Harris DR, *et al*. Comorbidity measures for use with administrative data. *Med Care* 1998;36:8–27.
15. Veehof L, Stewart R, Haaijer-Ruskamp F, *et al*. The development of polypharmacy. A longitudinal study. *Fam Pract* 2000;17:261–7.
16. George J, Vuong T, Bailey MJ, *et al*. Development and validation of the medication-based disease burden index. *Ann Pharmacother* 2006;40:645–50.
17. Sharabiani MT, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. *Med Care* 2012;50:1109–18.
18. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. *Int J Epidemiol* 2000;29:891–8.
19. Needham DM, Scales DC, Laupacis A, *et al*. A systematic review of the Charlson comorbidity index using Canadian administrative databases: a perspective on risk adjustment in critical care research. *J Crit Care* 2005;20:12–19.
20. May M, Royston P, Egger M, *et al*. Development and validation of a prognostic model for survival time data: application to prognosis of HIV positive patients treated with antiretroviral therapy. *Stat Med* 2004;23:2375–98.
21. Royston P, Moons KG, Altman DG, *et al*. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604–7.
22. Riley RD, Ensor J, Snell KI, *et al*. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353:i3140.
23. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16:385–95.
24. Gagne JJ, Glynn RJ, Avorn J, *et al*. A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* 2011;64:749–59.
25. Pencina MJ, D'Agostino RB, D'Agostino RB, *et al*. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
26. Strom BL. Overview of automated databases in pharmacoepidemiology. In: Strom BL, ed. *Pharmacoepidemiology*. 4th edn. New York: Wiley, 2005.
27. Normand SL, Morris CN, Fung KS, *et al*. Development and validation of a claims based index for adjusting for risk of mortality: the case of acute myocardial infarction. *J Clin Epidemiol* 1995;48:229–43.
28. Desai MM, Bogardus ST, Williams CS, *et al*. Development and validation of a risk-adjustment index for older patients: the high-risk diagnoses for the elderly scale. *J Am Geriatr Soc* 2002;50:474–81.
29. Fleming ST, Pearce KA, McDavid K, *et al*. The development and validation of a comorbidity index for prostate cancer among Black men. *J Clin Epidemiol* 2003;56:1064–75.
30. Holman CD, Preen DB, Baynham NJ, *et al*. A multipurpose comorbidity scoring system performed better than the Charlson index. *J Clin Epidemiol* 2005;58:1006–14.
31. Innocenti F, Tozzi C, Donnini C, *et al*. SOFA score in septic patients: incremental prognostic value over age, comorbidities, and parameters of sepsis severity. *Intern Emerg Med* 2017.

32. Zampieri FG, Colombari F. The impact of performance status and comorbidities on the short-term prognosis of very elderly patients admitted to the ICU. *BMC Anesthesiol* 2014;14:59.
33. Guo P, Zeng F, Hu X, *et al*. Improved variable selection algorithm using a LASSO-Type penalty, with an application to assessing hepatitis b infection relevant factors in community residents. *PLoS One* 2015;10:e0134151.
34. Low LL, Liu N, Wang S, *et al*. Predicting 30-day readmissions in an asian population: building a predictive model by incorporating markers of hospitalization severity. *PLoS One* 2016;11:e0167413.
35. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
36. Schonberger RB, Dutton RP, Dai F. Is There evidence for systematic upcoding of ASA physical status coincident with payer incentives? A regression discontinuity analysis of the national anesthesia clinical outcomes registry. *Anesth Analg* 2016;122:243–50.
37. Tsai TC, Joynt KE, Orav EJ, *et al*. Variation in surgical-readmission rates and quality of hospital care. *N Engl J Med* 2013;369:1134–42.
38. Costa G, Marinacci C, Caiazzo A, *et al*. Individual and contextual determinants of inequalities in health: the Italian case. *Int J Health Serv* 2003;33:635–67.