Development of Deep Learning Models for Predicting in-Hospital Mortality using an Administrative Claims Database

Hiroki Matsui (ptmatsui-tky@umin.ac.jp)
The University of Tokyo https://orcid.org/0000-0003-0004-4743
Hayato Yamana
The University of Tokyo: Tokyo Daigaku
Kiyohide Fushimi
Tokyo Medical and Dental University: Tokyo Ika Shika Daigaku
Hideo Yasunaga
The University of Tokyo: Tokyo Daigaku

Research

Keywords: Prognostic model, Deep learning, Real world data

Posted Date: February 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-176518/v1

Abstract

Background: To develop and validate deep learning–based prediction models for in-hospital mortality of acute-care patients.

Methods: The main model was developed using only administrative claims data (age, sex, diagnoses, and procedures on the day of admission). We also constructed disease-specific models for acute myocardial infarction, heart failure, stroke, or pneumonia using common severity indices for these diseases. Using the Japanese Diagnosis Procedure Combination data from July 2010 to March 2017, we identified 46,665,933 inpatients and divided them into derivation and validation cohorts in a ratio of 95:5. The main model was developed using a 9-layer deep neural network with four hidden dense layers that had 1000 nodes and were fully connected to adjacent layers. We evaluated model discrimination ability by an area under the receiver operating characteristics curve and calibration ability by calibration plot.

Results: Among the eligible patients, 2,005,035 (4.3%) died. Discrimination and calibration of the models were satisfactory. The AUC of the main model in the validation cohort was 0.954 (95% confidential interval 0.9537–0.9547). The main model had higher discrimination ability than the disease-specific models.

Conclusions: Our deep learning-based model using diagnoses and procedures produced valid predictions of in-house mortality.

Introduction

Administrative claims databases have been used widely in clinical and epidemiological studies because they have large sample sizes and are easily available. However, administrative data generally lack clinical information [1, 2] and do not distinguish between comorbidities at admission and complications after admission [3]. Risk adjustment is not necessarily feasible in studies that use administrative databases because of the lack of data on disease severity, and inadequate risk adjustment can result in confounding by indications.

Various risk-adjustment models have been developed for administrative data, but their validity and usability remain controversial [1, 4–6]. For example, the Charlson comorbidity index is a risk adjustment tool for predicting mortality that was developed using patients' demographic characteristics and comorbid illnesses in administrative data [4].

Previous studies showed that additional clinical information improved the performance of risk-adjustment models using administrative databases. In a previous study, we developed a procedure-based prediction model using the Japanese Diagnosis Procedure Combination (DPC) database, a nationwide administrative claims database [7]. However, these previous studies used simple logistic regression models that included only limited numbers of predictors.

Recent advances in deep neural network (deep learning) methods have made it possible to handle large amounts of information and complex models [8, 9]. In this study, we developed and validated deep learning-based models for predicting in-hospital mortality using the DPC database, including a huge amount of diagnoses and procedure data. To test the performance of the main model, we also constructed disease-specific models for predicting in-hospital mortality of patients with acute myocardial infarction (AMI), heart failure (HF), stroke, or pneumonia, using common severity indices for these diseases. Then, we compared the prediction abilities between the main model and the disease-specific models.

Methods

Data source

The data from July 2010 to March 2017 were collected from the DPC database. All the patients in the database were included to maximize the generalizability of the results. During the study period, 1569 hospitals contributed to the database. The patients in the database represented about 50% of all the acute-care inpatients in Japan.

The following data are included in the DPC database: age, sex, admission date, discharge date, diagnoses, and procedures (drugs, examinations, and surgical and non-surgical treatments) for each patient. In the DPC database, comorbidities present at admission are clearly distinguished from complications arising after admission. All diagnoses were recorded using the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10) codes. Procedure records were coded with Japanese original codes.

The DPC database also includes several severity indices, namely, the Killip classification for AMI [10, 11], New York Heart Association classification for HF [12], Barthel index score for activity of daily living at admission [13], Japan Coma Scale of consciousness level at admission [14]; and A-DROP, the Japan Respiratory Society community-associated pneumonia severity index [15, 16]. The Japan Coma Scale is used widely in Japan to measure impaired consciousness: a score of 0 indicates alert consciousness; single-digit scores (1, 2, 3) indicate being awake without stimuli; double-digit scores (10, 20, 30) indicate patients can be aroused by some stimuli; and triple-digit scores (100, 200, 300) indicate coma. A-DROP is a system for scoring severity of pneumonia that includes age (men $\geq$ 70 years, women $\geq$ 75 years), dehydration (serum urea nitrogen $\geq$ 21 mg/dL), respiratory failure (oxygen saturation by pulse oximetry $\leq$ 90% or PaO2 $\leq$ 60 mm Hg), orientation disturbance (confusion), and low blood pressure (systolic blood pressure $\leq$ 90 mm Hg).

Our study was approved by the Ethics Committee of the University of Tokyo School of Medicine (approval number: 3501-(4)).

Patient selection

We extracted the data of inpatients who were discharged from a hospital between 01 July 2010 and 31 March 2017. The study population was divided randomly into a derivation cohort (95%) and a validation cohort (5%). Patients who were discharged or died on the day of hospitalization were excluded from the validation cohort.

Variables

The outcome variable was in-hospital death. For predictive variables, we used patients' demographic information (age, sex, and history of hospitalization in the 180 days before admission), all the ICD-10-based diagnoses at admission, and all the procedures performed on the day of admission. Age was handled as a continuous variable; the other variables were handled as dichotomous variables (0 or 1).

Development of the main model

We developed a deep neural network model as the main model for predicting in-hospital death for all the patients, using nine layers with four hidden dense layers [17, 18]. All the layers had 1000 nodes and were fully connected to adjacent layers. We used a softmax layer with two nodes as the output layer. Because the numbers of dead and alive patients were very different, we weighted the dead cases with the reciprocal of the proportion of dead cases (that is, 1/0.045 = 22.3). We used stochastic gradient descent to obtain neural network weights iteratively. To avoid overfitting, 20% drop-out layers were sandwiched within each of the dense layers and an early stopping procedure involving learning steps using 3% data in the derivation cohort was employed. Details of the weight optimization process are described in the **Supplementary Material**.

Development of the disease-specific models

We constructed disease-specific models for predicting in-hospital mortality in subgroups with AMI, HF, stoke, or pneumonia. The four models included patient backgrounds (age, sex, and history of hospitalization in the 180 days before admission) and diagnoses, and none of the models included procedures. For the AMI-specific model, we selected patients with AMI and included the Killip classification. For the HF-specific model, we selected patients with HF and included the New York Heart Association classification. For the stroke-specific model, we selected patients with stroke and included the Barthel index and the Japan Coma Scale at admission.

For the pneumonia-specific model, we selected patients with pneumonia and included the A-DROP scores.

Comparing prediction abilities between the main model and the disease-specific models

We applied the main model to the subgroups of patients with AMI, HF, stoke, and pneumonia and compared its prediction performance with the prediction performances of the disease-specific models for AMI, HF, stoke, and pneumonia.

We evaluated the performance of each model by calculating performance measures in the validation cohort. Performance measures included the area under the receiver operating characteristic curve (AUC) as model discrimination. We calculated the 95% confidence interval (CI) of the AUC using Delong's method [19] and plotted a calibration curve as a goodness of fit.

Results

We obtained the data for 46,665,942 patients from the DPC database during the study period and divided them randomly into two groups with a ratio of 95:5 as the derivation (n = 44,334,477) and validation (n = 2,331,465) cohorts. We excluded patients who died or were discharged within one day of admission from the validation cohort according to the exclusion criteria, which left 2,277,968 patients as the validation cohort (Fig. 1). The characteristics of the derivation and validation cohorts are shown in Table 1. The average lengths of stay were 14.2 days and 14.5 days and in-hospital mortality was 4.3% and 3.7 % in the derivation and validation cohorts, respectively. Patients in the validation cohort were slightly older and had more comorbidities than those in the derivation cohort.

Table 1
Characteristics of the patients in the derivation and validation cohorts.

	Derivation cohort (n = 44,334,477)	Validation cohort (n = 2,277,968)	p value
Death, n (%)	1,905,286 (4.3)	83,292 (3.7)	< 0.001
Length of hospital stay (days), mean (sd)	14.2 (24.1)	14.5 (24.2)	< 0.001
Age (years), mean (sd)	60.1 (24.4)	60.4 (24.2)	< 0.001
Sex (male), n (%)	23,480,628 (53.0)	1,207,886 (53.0)	0.066
History of hospitalization within 180 days, n (%)	12,282,386 (27.7)	632,362 (27.8)	0.066
Charlson comorbidity index, n (%)			< 0.001
0–1	28,734,890 (64.8)	1,465,779 (64.3)	
2–3	11,432,403 (25.8)	594,500 (26.1)	
≥ 4	4,165,579 (9.4)	217,605 (9.6)	

Table 2
A: Structure of main model.

Layer	Input	Output	Number of weights
1: Input	49,297	1,000	49,297,000
2: Drop-out			
3: Hidden 1	1,001	1,000	1,001,000
4: Drop-out			
5: Hidden 2	1,001	1,000	1,001,000
6: Drop-out			
7: Hidden 3	1,001	1,000	1,001,000
8: Drop-out			
9: Output	1,001	2	2,002
Sum of weights			52,302,002

Table 2
B. Summary of the main and disease-specific models.

Model	Input node	Total number of weights
Main model	49297	52,302,002
Acute myocardial infarction model	9	3,014,002
Stroke model	54	3,059,002
Heart failure model	9	3,014,002
Pneumonia model	9	3,014,002

The structure of the main model is shown in Table 2A. There were 49,297 predictor variables, including 3 demographic variables (age, sex, history of hospitalization in the 180 days before admission), 19,930 diagnoses at admission, and 29,364 procedures (drugs, examinations, surgical and non-surgical treatments). We inserted a dropout layer between the layers to avoid overfitting. Overall, 52,302,002 weights (= 49297 × 1000 + 1001 × 1000 + 1001 × 1000 + 1001 × 1000 + 1001 × 2) of links between the layers were optimized in the derivation. The script for the deep learning model including model weights is available on our website (https://researchmap.jp/ptmatsui).

An overview of the main and disease-specific models used in this study is given in Table 2B. Total number of weights = the number of input nodes × 1000 + 1001 × 1000 + 1001 × 1000 + 1001 × 1000 + 1001 × 2.

The AUC of the main model in the validation cohort was 0.954 (95% CI 0.9537–0.9547).

The calibration curves of the observed and estimated mortality in the validation cohort are shown in Fig. 2.Observed and estimated mortality were strongly correlated, but the estimated mortality was slightly lower than the observed mortality.

The AUCs of the main and disease-specific models are shown in Table 3. The AUCs of the main model for the AMI, HF, stroke, and pneumonia subgroups were 0.944, 0.832, 0.921, and 0.918, respectively. The AUCs of the disease-specific models for the AMI, HF, stroke, and pneumonia subgroups were 0.876, 0.745, 0.894, and 0.863, respectively. The main model showed significantly higher discriminant ability than the disease-specific models for all four subgroups.

Table 3
Performances of the main and disease-specific models.

Population	n	Main model AUC (95% CI)	Disease-specific model AUC (95% CI)
Acute myocardial infarction	14,213	0.944 (0.938–0.950)	0.876 (0.866–0.887)
Heart failure	43,792	0.831 (0.825–0.837)	0.745 (0.738–0.753)
Stroke	82,454	0.921 (0.918–0.925)	0.894 (0.890–0.898)
Pneumonia	87,775	0.918 (0.915–0.920)	0.863 (0.859–0.867)
AUC, area under the receiver operating characteristic curve; CI, confidence interval.			

The calibration curves for main and disease specific models for the subgroups are shown in Fig. 3. The correlations between the observed and estimated mortality were better with the main model than with the disease-specific models for AMI, HF, and stroke subgroups (Fig. 3A–C). For the pneumonia subgroup, the correlations were similar between the main and disease-specific models when the predicted mortality was ≤ 0.8. However, the disease specific model failed to estimate morality well when the predicted mortality was ≥ 0.8. (Fig. 3D).

Discussion

We constructed deep leaning-based prediction models for in-hospital mortality, using a large Japanese inpatient database. Patient backgrounds, diagnoses, and treatments within 1 day of admission were entered into the models. The overall discriminant abilities of the models were high in the subgroups of

patients with AMI, HF, stroke, and pneumonia. The main model had better discriminant abilities than the disease-specific models using common severity indices.

Risk scores derived from administrative claims databases have been developed previously. For example, the Charlson and Elixhauser models that use comorbidity information to predict long-term survival have been used for risk adjustment in clinical and epidemiological studies [20, 21]. In the present study, the new prediction model for in-hospital mortality, which was developed using administrative claims data, showed high discriminatory power (AUC = 0.945). We believe that our model also can be used for risk adjustment in clinical and epidemiological studies using administrative claims data including diagnoses and procedures.

In a previous study, we constructed a prediction model for in-hospital mortality by incorporating comorbidities and several selected procedures (blood tests, radiography, echocardiogram) on the day of admission into the model [7]. However, that model lacked generalizability; for example, it was not applicable for critically ill patients. The newly constructed model can be used for risk adjustment for patients with a wide range of disease severity.

In a previous study, the predictive abilities of models with administrative claims data alone were compared with those of models with electronic medical records combined with administrative claims data [22]. The predictive abilities of the models with electronic medical records were higher because the electronic medical records included sophisticated information on disease-specific severity. In the present study, the deep learning model that used only massive administrative data had higher predictive ability than the models that used disease-specific severity information. On the basis of our results, we consider that large-scale administrative data can be used to predict in-hospital mortality more accurately than the generally used severity indices. Therefore, we propose that patient outcome studies can be conducted using administrative data alone without the need for data on disease severity.

This study has several limitations. First, we did not conduct an external validation. Second, we did not use the various methods for machine learning (for example, random forest, Lasso regression, XGBoost, and their ensembles), so the prediction performance of our deep neural network model compared with those of these machine learning methods was not ascertained. Third, because the database used in this study is for acute hospitalization, we could not obtain data on long-term outcomes. Fourth, model accuracy is not always guaranteed for all diseases, so the applicability of the model to other populations needs to be considered.

In conclusion, we constructed a deep neural network model to predict in-hospital mortality using all the data on diagnoses and procedures performed on the day of admission in a Japanese administrative claims database. The model showed higher prediction ability than those using generally used severity indices. We propose that prognostic models using data on diagnoses and procedures available from administrative claims databases can predict in-hospital mortality and can be used for risk adjustment in clinical and epidemiological studies using only administrative claims

Declarations

Ethics approval and consent to participate: This study was approved by the Institutional Review Board of The University of Tokyo (Approval Number: 3501-(1)). Because all data were de-identified, the requirement for patient informed consent was waived.

Consent for publication: Not applicable.

Availability of data and material: Because individual privacy could be compromised, the datasets analysed during the current study are not publicly available. But they are available from the corresponding author on reasonable request.

Funding: This work was supported by grants from the Ministry of Health, Labour and Welfare, Japan (19AA2007 and 20AA2005) and the Ministry of Education, Culture, Sports, Science and Technology, Japan (20H03907 and 17H05077).

Competing interests: All authors have completed declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; and no other relationships or activities that could appear to have influenced the submitted work.

Authors' Contributions: HM and HY1 andHY2 contributed to the conception and design of the study. KF and HY2 contributed to the data collection. HM contributed to the data analysis. All authors contributed to the data interpretation and drafting of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

References

1. Sung SF, Hsieh CY, Kao Yang YH, Lin HJ, Chen CH, Chen YW, et al. Developing a stroke severity index based on administrative data was feasible using data mining techniques. J Clin Epidemiol. 2015;68:1292–300. https://doi.org/10.1016/j.jclinepi.2015.01.009.
2. Virnig BA, McBean M. Administrative data for public health surveillance and planning. Annu Rev Public Health. 2001;22:213–30. https://doi.org/10.1146/annurev.publhealth.22.1.213.
3. Yamana H, Matsui H, Sasabuchi Y, Fushimi K, Yasunaga H. Categorized diagnoses and procedure records in an administrative database improved mortality prediction. J Clin Epidemiol. 2015;68:1028–35. https://doi.org/10.1016/j.jclinepi.2014.12.004.
4. Sundararajan V, Quan H, Halfon P, Fushimi K, Luthi J-C, Burnand B, et al. Cross-national comparative performance of three versions of the ICD-10 Charlson index. Med Care. 2007;45:1210–5. https://doi.org/10.1097/MLR.0b013e3181484347.

5. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. Med Care. 1998;36:8–27. https://doi.org/10.1097/00005650-199801000-00004.
6. Pine M, Jordan HS, Elixhauser A, Fry DE, Hoaglin DC, Jones B, et al. Enhancement of claims data to improve risk adjustment of hospital mortality. J Am Med Assoc. 2007;297:71–6. https://doi.org/10.1001/jama.297.1.71.
7. Yamana H, Matsui H, Fushimi K, Yasunaga H. Procedure-based severity index for inpatients: Development and validation using administrative database. BMC Health Serv Res. 2015;15:261. https://doi.org/10.1186/s12913-015-0889-x.
8. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. J Biomed Inform. 2018;83:112–34. https://doi.org/10.1016/j.jbi.2018.04.007.
9. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. Npj Digital Medicine. 2018;1:18. https://doi.org/10.1038/s41746-018-0029-1.
10. Killip T, Kimball JT. Treatment of myocardial infarction in a coronary care unit. A two year experience with 250 patients. Am J Cardiol. 1967;20:457–64. https://doi.org/10.1016/0002-9149(67)90023-9.
11. Shiraishi J, Kohno Y, Nakamura T, Yanagiuchi T, Hashimoto S, Ito D, et al. Predictors of In-hospital Outcomes after Primary Percutaneous Coronary Intervention for Acute Myocardial Infarction in Patients with a High Killip Class. Intern Med. 2014;53:933–9. https://doi.org/10.2169/internalmedicine.53.1144.
12. Paul Dudley White MMM. The classification of cardiac diagnosis. J Am Med Assoc. 1921;77:1414–5. https://doi.org/10.1001/jama.1921.02630440034013.
13. Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability (Inter-rater Agreement) of the Barthel Index for Assessment of Stroke Survivors. Stroke. 2013;44:462–8. https://doi.org/10.1161/STROKEAHA.112.678615.
14. Shigematsu K, Nakano H, Watanabe Y. The eye response test alone is sufficient to predict stroke outcome-reintroduction of Japan Coma Scale: A cohort study. BMJ Open. 2013;3:e002736. https://doi.org/10.1136/bmjopen-2013-002736.
15. Miyashita N, Matsushima T, Oka M. The JRS Guidelines for the Management of Community-acquired Pneumonia in Adults:An Update and New Recommendations. Intern Med. 2006;45:419–28. https://doi.org/10.2169/internalmedicine.45.1691.
16. Ahn JH, Choi EY. Expanded A-DROP, Score. A New Scoring System for the Prediction of Mortality in Hospitalized Patients with Community-acquired Pneumonia. Sci Rep. 2018;8:14588. https://doi.org/10.1038/s41598-018-32750-2.
17. Chollet F others. Keras[Internet]. 2015 [cited 27 Dec 2020]. Available form: https://keras.io/.
18. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-scale machine learning on heterogeneous systems[Internet]. 2015[cited 27 Dec 2020]. Available form: https://www.tensorflow.org/.

19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–45.
20. Matsui H, Jo T, Fushimi K, Yasunaga H. Outcomes after early and delayed rehabilitation for exacerbation of chronic obstructive pulmonary disease: a nationwide retrospective cohort study in Japan. Respir Res. 2017;18:68. https://doi.org/10.1186/s12931-017-0552-7.
21. Matsui H, Koike S, Fushimi K, Wada T, Yasunaga H. Effect of neurologic specialist staffing on 30-day in-hospital mortality after cerebral infarction. Annals of Clinical Epidemiology. 2019;1:86–94. https://doi.org/10.37737/ace.1.3_86.
22. Zeltzer D, Balicer RD, Shir T, Flaks-Manov N, Einav L, Shadmi E. Prediction Accuracy With Electronic Medical Records Versus Administrative Claims. Med Care. 2019;57:551–9. https://doi.org/10.1097/MLR.0000000000001135.

Figures

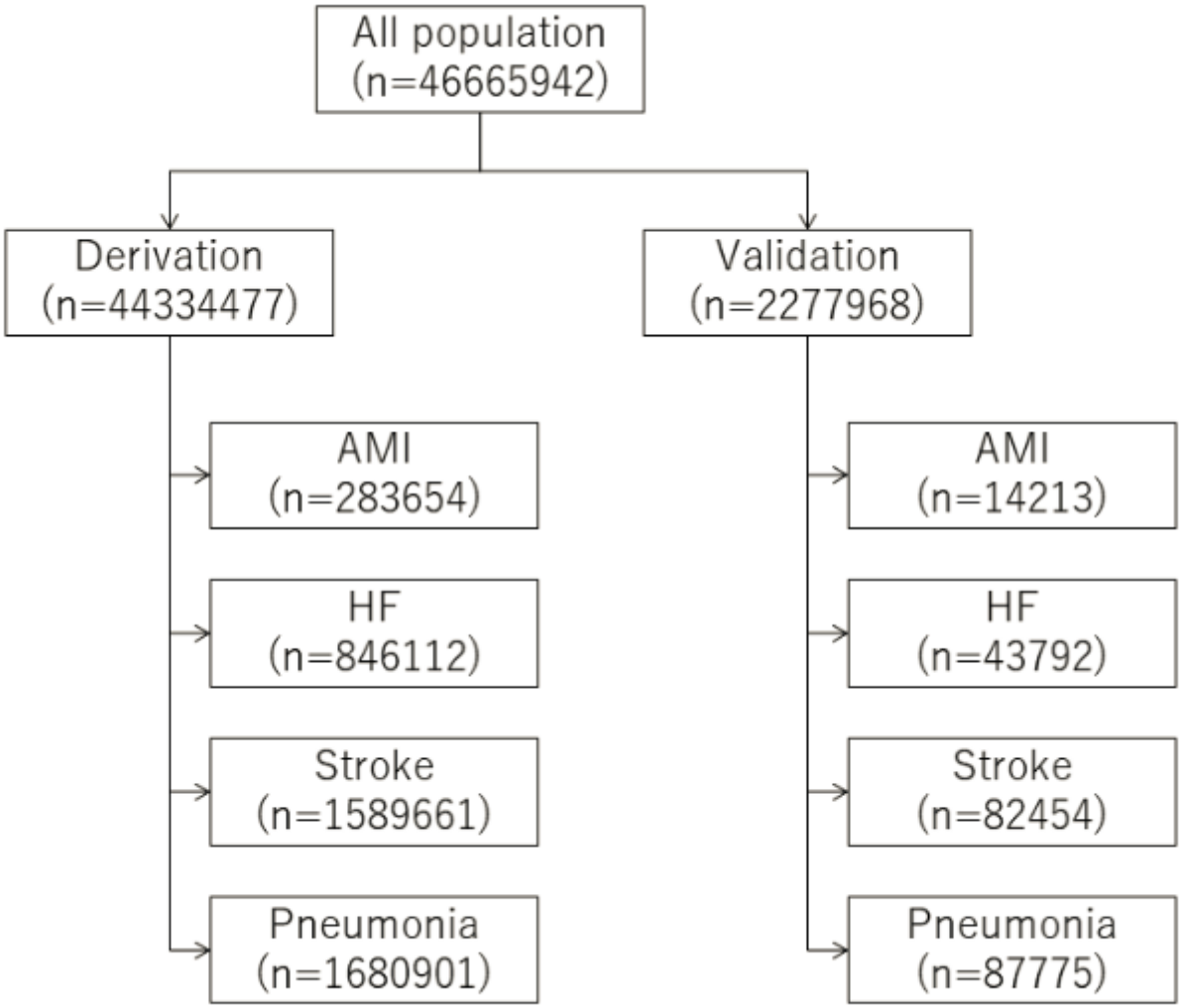

Figure 1

Numbers of patient in the derivation and validation cohorts and disease-specific subgroups. AMI: acute myocardial infarction, HF: heart failure

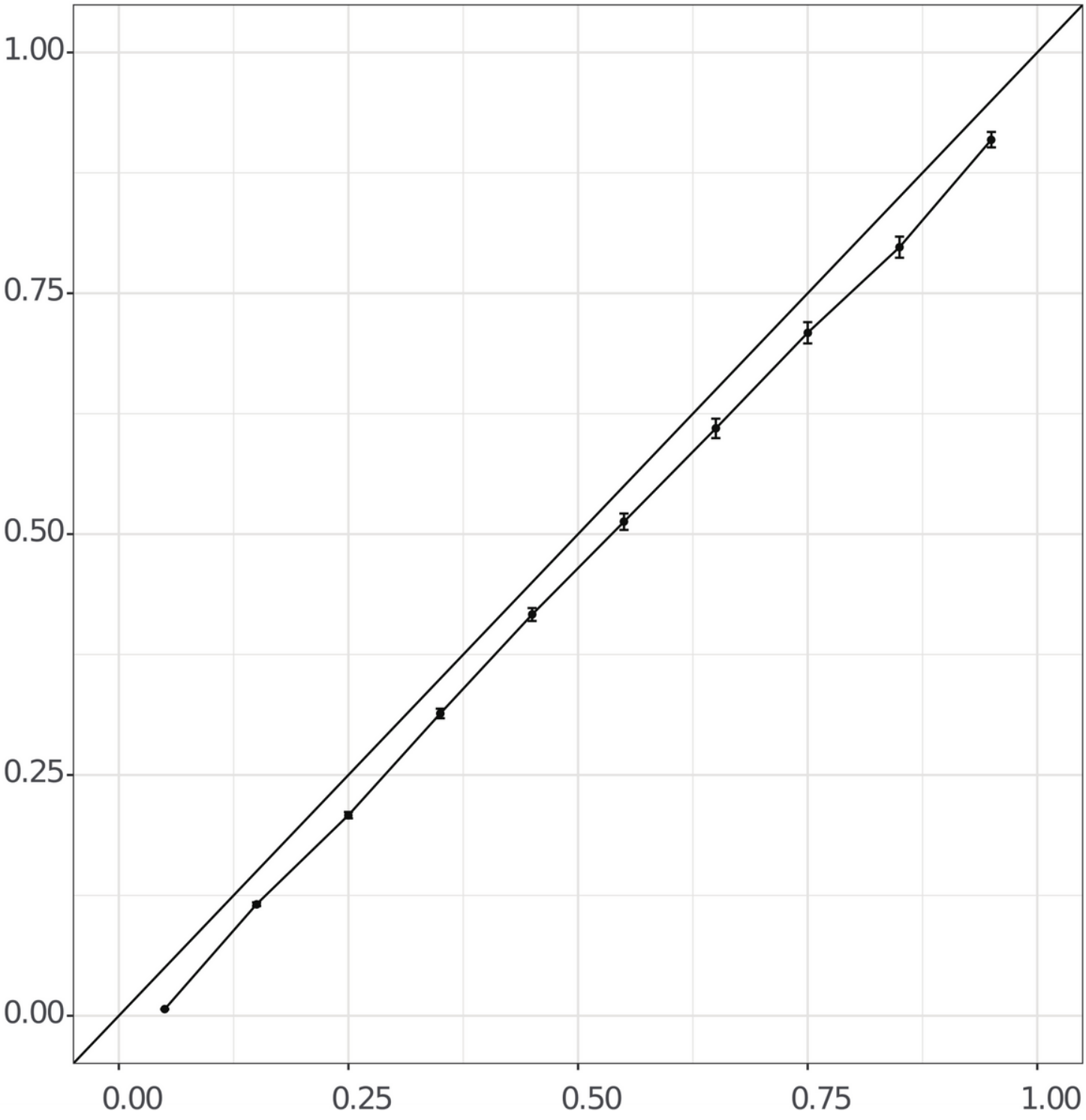

Figure 2

Calibration curves for the observed and estimated mortality in the validation cohort with the main model X-axis, predicted mortality; Y-axis, actual mortality. X-axis indicates predicted mortality, and Y-axis indicates actual mortality.

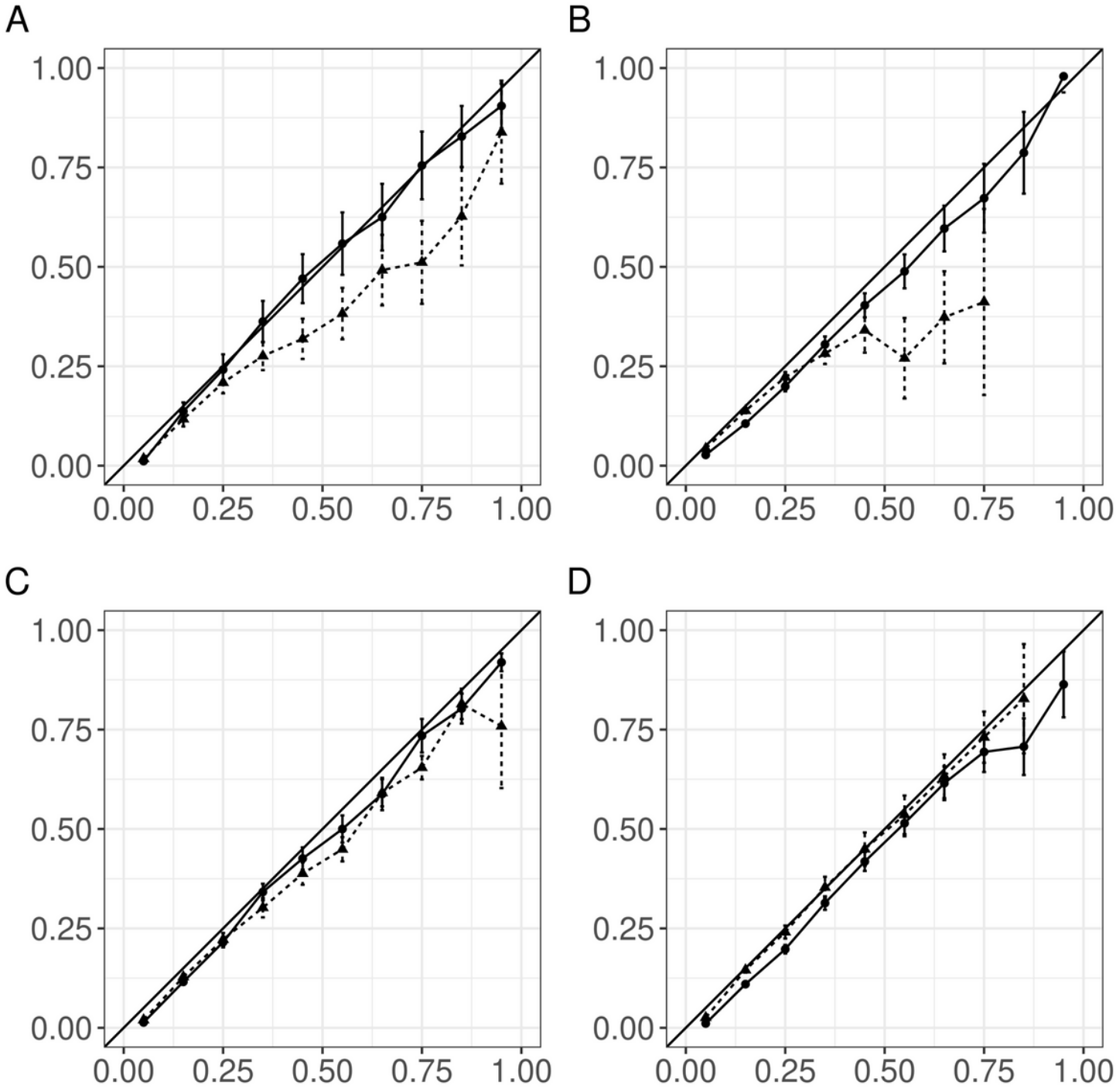

Figure 3

Calibration curves for the observed and estimated mortality in the validation cohort with the disease-specific models Models for (A) acute myocardial infarction, (B) heart failure, (C) stroke, and (D) pneumonia. X-axis, predicted mortality; Y-axis, actual mortality. Solid line, main model; dotted line, disease-specific models.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial.docx