scientific reports

Check for updates

OPEN

# Prediction of all-cause mortality in older adults in Colombia: real-world evidence from a middle-income country

Oscar Espinosa[1,2]✉, Valeria Bejarano[2], Felipe Rojas[3], Sandeep Pagali[4], Jordan Weiss[5] & Giancarlo Buitrago[6,7]

**Accurate mortality prediction in older adults is a critical component of precision public health and risk stratification in healthcare systems worldwide. In middle-income countries, however, real-world applications remain scarce due to limitations in data quality and methodological scalability. This study aims to develop and benchmark state-of-the-art predictive models for all-cause mortality among adults aged 80 years and older in Colombia's five major cities, leveraging a uniquely rich and comprehensive administrative health dataset. We constructed two predictive cohorts from high-dimensional, real-world data covering over 300,000 and 50,000 individuals for one- and two-year models, respectively. The dataset includes detailed records on thousands of prescribed medications, medical procedures, ICD-10 diagnoses, service provider and insurer identifiers, care modalities (outpatient, inpatient, emergency, home-based), and financial payment mechanisms, reflecting routine clinical practice across hundreds of healthcare providers and insurers. We implemented and compared five advanced modeling strategies: artificial neural networks, autoencoders, extreme gradient boosting (XGBoost), Bayesian regression, and TabTransformer (a self-attention architecture tailored for structured data). Model performance was assessed using area under the curve (AUC), F1-score, accuracy, and Brier score. Our research demonstrates that diagnostic codes, pharmacological profiles, and hospitalization patterns are the most predictive domains of late-life mortality, whereas operational variables (e.g., provider or insurer) add little incremental value. The highest discriminative performance was achieved with XGBoost (AUC = 0.92 for two-year, 0.89 for one-year predictions), followed by autoencoders (AUC = 0.85 and 0.87) and TabTransformer (AUC = 0.82 and 0.84). Neural networks showed moderate performance, and Bayesian regression faced convergence issues due to computational demands. Notably, some models achieved AUC values above 0.88, indicating strong discriminatory performance for mortality classification in older adults using real-world data, although F1 scores reflected moderate classification balance. This study demonstrates the feasibility and potential of deploying advanced machine learning techniques on richly granular administrative health data in a middle-income country. Our findings offer a robust methodological foundation for the early identification of high-risk older adults and provide actionable insights for health insurers, care providers, and policymakers within Colombia's General System of Social Security in Health. The integration of such models could inform targeted preventive strategies and resource allocation to improve outcomes in aging populations.**

**Keywords** Predictive medicine, Analytics, Real-world evidence, Old age, Machine learning

[1]Population Policy Research Center, Seoul National University, Seoul, South Korea. [2]Economic Models and Quantitative Methods Research Group, Centro de Investigaciones para el Desarrollo, Universidad Nacional de Colombia, Bogotá, D.C, Colombia. [3]Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy. [4]Division of Hospital Medicine & Section of Geriatrics, Department of Medicine, Mayo Clinic, Rochester, USA. [5]Optimal Aging Institute & Division of Precision Medicine, NYU Grossman School of Medicine, New York City, NY, USA. [6]Instituto de Investigaciones Clínicas, Universidad Nacional de Colombia, Bogotá, D.C., Colombia. [7]Subdirección de Investigación, Fundación Cardioinfantil - Instituto de Cardiología, Bogotá, D.C., Colombia. ✉email: oaespinosaa@unal.edu.co

Since the beginning of the 21st century, predictive medicine and data-driven health risk stratification have been increasingly discussed. This field encompasses analytical approaches aimed at identifying patterns associated with health outcomes and supporting decision-making for patients, physicians, health ministries, insurers, healthcare providers, and surveillance and regulatory entities within health systems[1–3].

The potential of predictive medicine lies in its added value for clinical practice. However, it is vital that certain theoretical-methodological criteria are met to facilitate solid, positive, and useful decision-making in patient care[4–6]. Although it was initially thought that clinical prediction would be primarily determined by genetics, it has been shown that other factors, such as demographic, environmental, sociocultural, and economic factors, among others, also play an essential role in shaping health outcomes within a population[1,7,8].

The continuous improvement in the collection, storage, and processing of millions of records has enabled advancements in predictive techniques in the clinical field. Real-world evidence, when subjected to rigorous quality processes such as data cleaning, parameterization, and standardization, can provide added value for managing health outcomes, from a generalizable perspective, as valuable data is typically collected from tens of thousands, if not millions, of patients over a reasonable time frame[9,10].

In recent years, international literature has seen a considerable increase in scientific publications on clinical prediction models regarding future hospital admissions[11–13], using artificial intelligence techniques (e.g. machine learning[1]) and classical statistics methods (e.g. multivariate logistic regression analysis and Cox proportional hazards models), utilizing administrative records, pre-hospital data, claims data, patient surveys, and other sources[14–30]. The most frequently included variables in these studies have been specific medical diagnoses, sex, medications, and age. As performance measures, a good percentage of works use the C-Statistic/AUC and the squared R, while the review of the classification results typically involves examining specificity and sensitivity[14–30].

The prediction of mortality in individuals aged 80 and older has become a key objective in public health, intersecting with broader discussions of healthspan vs. lifespan, equity in late-life care, and system-level preparedness. As people reach older ages, the risks associated with multiple comorbidities, functional decline, and chronic diseases rise[31,32]. Understanding and predicting mortality in this age group not only enables better allocation of healthcare resources but also facilitates the implementation of personalized prevention policies and palliative care strategies by stakeholders. Moreover, accurate assessment of the factors influencing life expectancy in old age can contribute to the design of effective interventions to improve the quality of life at this crucial stage.

The objective of this research was to develop mortality prediction models among individuals aged 80 years and older enrolled in the contributory regime of the Colombian health system for the five main cities, using real-world evidence. The strength of this type of information lies in the large number of records per patient that can be accessed, which accurately reflects normal medical practice in relation to the prescription and effective delivery of the various health technologies[33–37].

## Methods

### Claims data and study population

The database used, called *Suficiencia*, contains detailed records of the different health technologies applied to individuals enrolled within the contributory regime of the health system in Colombia (that is, those with formal jobs and their beneficiaries). These services are financed through the collective protection mechanism known as the Capitation Payment Unit (CPU)[2], for the years 2018 and 2019. Each record (row) includes information on the prescribed medicine or medical procedure, patient birthdate, date of healthcare use, medical diagnosis according to the International Classification of Diseases 10th Revision (ICD-10), place of care of the health service, identification code of the insurer and the health service provider, type of care settings and contractual payment mechanism between insurer and provider.

*Suficiencia* is reported by health insurers (called *Entidades Promotoras de Salud*, EPS, for its acronym in Spanish) to the regulatory body, which applies validation meshes to ensure data quality[3] resulting in complete datasets with no missing values (therefore, no imputation was necessary). Table 1 presents descriptive statistics of the analysis database (additional descriptive analysis can be found in Figures S1-S11 in Supplement 1), covering over 90% of contributory regime enrollees of years 80 and older for the five main cities of Colombia, in the years of study. Additionally, deaths are reported by authorized health facilities to the *Unique Register of Affiliates*. This database has been positively evaluated by international institutions, capturing 99% of births (2012–2016) and 91% of deaths (2016) in Colombia[38]. By linking both databases using anonymized identifiers, mortality outcomes were reliably captured. This research was approved by the Ethics Committee of the Faculty of Medicine at the Universidad Nacional de Colombia (Approval No. 012–121/2020), in accordance with institutional and national ethical guidelines.

Initially, we identified 440,827 individuals eligible for one-year analysis and 182,548 for two-year analysis. Due to computational constraints, we randomly undersampled the non-event group (i.e., those who were censored at the end of the study period) in the one-year cohort. For the two-year cohort, random undersampling of the

[1] Random forests, gradient boosted decision trees, kernelized and sparse support vector machines, neural network, among others.

[2] The CPU can be understood as an annual premium that the government gives to each health insurer, according to sex, age and region of the enrollee, so that they manage the health risk of their respective insured, based on a set of specific health technologies.

[3] Quality checks include verifying consistency between procedures and patients' diagnoses using ICD-10 codes. Inpatient and outpatient visits are reviewed for duplicates within time frames, and procedures are cross-checked historically to ensure uniqueness over a lifetime.

| | | 2018 | 2019 |
|---|---|---|---|
| **Female age** | | **Mean: 85.4 (sd: 4.55) Median: 84 (IQR: 6)** | **Mean: 85.6 (sd: 4.59) Median: 85 (IQR: 7)** |
| Male age | | Mean: 84.9 (sd: 4.25) Median: 84 (IQR: 7) | Mean: 85.1 (sd: 4.30) Median: 84 (IQR: 6) |
| % of patients by city | Bogotá | 49.2% | 50.0% |
| | Medellín | 21.4% | 21.1% |
| | Cali | 17.0% | 16.9% |
| | Barranquilla | 8.4% | 8.3% |
| | Cartagena | 4.1% | 4.0% |
| Number of centenarians | | 1,082 | 1,315 |
| Average number of different ICD-10 registered per user (sd) | | 8.39 (5.74) | 9.19 (6.45) |
| Number of medicines per capita (sd) | | 26.4 (16.5) | 19.3 (11.9) |
| Per capita cost for medicines (sd) | | USD 192 (536) | USD 231 (538) |
| Number of health procedures per capita (sd) | | 29.6 (18.0) | 31.8 (17.7) |
| Per capita cost for health procedures (sd) | | USD 857 (1,949) | USD 901 (1,853) |
| % of hospitalized inpatients | | 26.6% | 24.3% |
| % of patients with ambulatory care | | 98.8% | 99.2% |
| % of patients with domiciliary care | | 17.6% | 18.0% |
| % of patients with emergency care | | 35.5% | 35.7% |
| Average number of health service providers who treated the same patient (sd) | | 6.62 (3.46) | 8.24 (4.14) |
| Total number of health service providers in the database | | 3,749 | 4,252 |
| Average number of health insurers who treated the same patient (sd) | | 1.01 (0.12) | 1.03 (0.16) |
| Total number of health insurers in the database | | 9 | 10 |

**Table 1**. Descriptive statistics by year, 2018 and 2019. Note: sd - standard deviation, IQR - interquartile range.

majority class was implemented to address the extreme class imbalance (mortality prevalence below 0.045%) and improve computational feasibility[39]. As a result, the final analytic samples included 320,827 and 57,548 individuals for the one- and two-year cohorts with 22% and 1.41% of deaths, respectively. In both cases, we confirmed that the excluded individuals did not differ systematically from those retained in the analyses in terms of key baseline characteristics, thereby minimizing potential selection bias.

## Analytical models

Our investigation follows the updated checklists of *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis* (TRIPOD)[40] and *Prediction model Risk Of Bias ASsessment Tool* (PROBAST)[41]. The following analytical approaches were developed to ensure a robust classification framework, leveraging both traditional and advanced machine learning techniques. Neural networks were applied to capture complex, non-linear relationships in the data[42], an algorithm that has demonstrated strong performance in various clinical applications[43,44]. Extreme Gradient Boosting (XGBoost), a highly efficient and scalable tree-based ensemble algorithm, was employed for its superior predictive performance and ability to model intricate feature interactions as it is broadly used with good performance[45–47]. Autoencoders, despite their common use in unsupervised learning, were adapted for supervised classification by leveraging their encoding layers for feature extraction, thereby improving downstream predictive accuracy[48,49].

Bayesian regression with Horseshoe prior, was used to integrate statistical modeling with machine learning techniques, providing a flexible and principled approach to variable selection and regularization, an advantage often highlighted in the literature for improving predictor inclusion[50–53]. Finally, we considered the TabTransformer architecture, a novel deep learning approach specifically designed to improve predictive performance on tabular data. By applying Transformer layers to the embeddings of categorical variables, the model generates rich contextual representations that capture complex relationships across features, this mechanism enhances the robustness of the learned representations and has been shown to achieve superior prediction accuracy, often outperforming conventional deep learning and tree-based methods[54].

The dependent variable was dichotomous, indicating whether a person in the prediction year died. Mortality was determined based on the vital records database for the corresponding prediction year. The independent variables included in the models are presented in Table 2, which summarizes their corresponding categories and data types.

| N° | Variable group | Category | Data type | Description / Coding |
|---|---|---|---|---|
| 1 | Year of birth | Sociodemographic | Numeric (count) | Year of birth of the individual |
| 2 | Sex | Sociodemographic | Binary indicator | Female/Male |
| 3 | ICD-10 diagnoses | Clinical | Binary indicator | Medical diagnoses classified according to ICD-10 at the three-digit level |
| 4 | Medications | Pharmaceutical / Economic | Monetary value (continuous) | Medications classified according to the Anatomical Therapeutic Chemical (ATC) system at the five-digit level |
| 5 | Health procedures | Healthcare utilization / Economic | Monetary value (continuous) | Health procedures classified according to the Colombian CUPS system at the three-digit level |
| 6 | Total hospitalization days | Healthcare utilization | Numeric (count) | Total number of hospitalization days |
| 7 | Healthcare setting of procedures | Healthcare utilization | Frequency/count variable | Frequency of healthcare use across outpatient, inpatient, emergency, and domiciliary settings |
| 8 | Health service provider | Administrative | Binary indicator | Health service provider identifier at the ten-digit level |
| 9 | Health insurer (EPS) | Administrative | Binary indicator | Health insurer identifier |
| 10 | Payment and reimbursement modality | Administrative / Economic | Frequency/count variable | Frequency of use of different payment mechanisms: capitation payment, payment per event, payment per case/comprehensive care set, diagnosis-related group package, authorized payment, and direct payment |

**Table 2**. Summary of independent variable groups included in the mortality classification models.

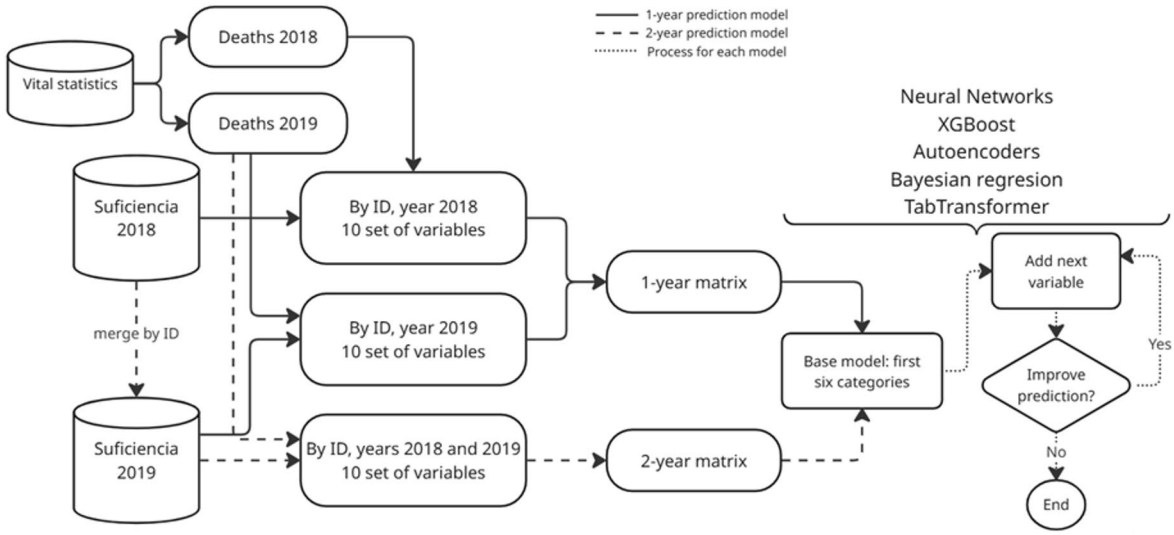


**Fig. 1**. Model prediction methodology. Source: own elaboration.

Based on data from year $t$, the model predicted mortality within the same year $t$ (for example, using information collected from 1 January to 15 September, knowing that the person passed away on 15 September). Additionally, by incorporating data from years $t$ and $t-1$, the model predicted mortality in $t$ (e.g., using 2018 and 2019 data to predict 2019 outcomes). Predictor variables were constructed using aggregated healthcare utilization data available during the observation period, including information accumulated up to the time of death for deceased individuals. Therefore, the modelling strategy was designed as a retrospective mortality classification approach based on patterns of healthcare utilization, rather than as a strictly prospective prediction model anchored at a fixed baseline timepoint. The first six categories of variables (Table 2) were fixed in each model based on authors' expert knowledge of the domain, while the remaining four categories were selected through an iterative forward algorithm. Figure 1 illustrates the methodological workflow (no data transformation, rescaling, or standardization procedures were performed[4]).

The dataset was randomly divided into three groups, a training set (75%), a validation set (12.5%), and a test set (12.5%), so their internal characterization is similar across subsets without data leakage. The performance of predictive models was evaluated using the metrics of accuracy, precision and recall, summarized in the F1-score (F1) and the area under the curve (AUC) based on the Receiver Operating Characteristic (ROC) curve. In addition to traditional metrics, the Brier score was calculated to assess the calibration of the probabilistic predictions, providing an additional measure of accuracy in the model's estimates.

[4] Given the high dimensionality of the dataset, we applied principal component analysis to reduce the number of variables; however, the results did not outperform those obtained using the original disaggregated data.

Given that the target variable was highly imbalanced, with deaths representing the minority class, additional techniques were employed to enhance the models' sensitivity to this critical outcome. The Synthetic Minority Over-Sampling Technique (SMOTE), which generates synthetic samples of the minority class by interpolating between existing examples, was applied with its extension, Borderline-SMOTE to improve the model's ability to detect rare mortality events[55]. After splitting the dataset into training, validation, and test sets, SMOTE procedures were applied exclusively to the training set. Validation and test sets remained unmodified to prevent information leakage and optimistic bias in model performance estimation. Borderline-SMOTE was evaluated to concentrate synthetic sample generation near the decision boundary, where classification is most challenging[56], without requiring model-specific adjustments like loss weighting or cost functions. This method concentrates synthetic sample generation, helping mitigate bias toward the majority class and enhancing the model's capacity to generalize to rare but clinically important outcomes such as mortality.

We developed the programming code in R 4.4.0 and Python 3.11.7 software. The modeling was run on a computer with 512 GB RAM, with an Intel Xeon Silver 4410Y processor with no graphics processing unit, belonging to the HPC Unit of the Clinical Research Institute at the Universidad Nacional de Colombia. Supplement 2 contains detailed information on the hyperparameters considered among the models developed.

## Results

For the calibration process, we initially performed experiments on a subset of the data, testing variations of neural network models combined with the balancing algorithms and threshold adjustments. The selection of optimal configurations was guided by average performance across accuracy, F1, and AUC metrics, using a grid search of hyperparameters. Based on these initial results, we selected the Borderline-SMOTE balancing algorithm, which was well suited to the imbalance in the dataset (low number of deaths), and applied a decision threshold of 0.75 to penalize false negatives (missed deaths) -as this value yielded the best performance-. With the balancing algorithm in place, a set of hyperparameters was explored during model tuning.

After this initial setup, we applied the models to a one-year dataset comprising individuals with claims in 2018 and 2019, where death occurred in 2018 and 2019 respectively (see Table 3). The models tested included neural networks, XGBoost, autoencoders, Bayesian regression and TabTransformer, initially using the first six categories of variables. The average performance metrics achieved were 76.8% (accuracy: 91.0%; F1: 57.6%; AUC: 81.8%) for neural networks, 80.5% (accuracy: 96.4%; F1: 59.1%; AUC: 86.1%) for XGBoost, 79.6% (accuracy: 96.5%; F1: 59.7%; AUC: 82.7%) for autoencoders and 80.3% (accuracy: 96.3%; F1: 60.8%; AUC: 83.8%) for TabTransformer. Convergence was not achieved for the Bayesian model, even after 350 h of computation.

Using a forward selection strategy, we sequentially added variable categories related to healthcare setting, service provider (HSP), insurer, and payment method to evaluate potential performance improvements. The neural network and XGBoost models identified the healthcare setting as an additional important predictor, showing improved average performance over the baseline of 77.2% (accuracy: 91.9%; F1: 58.3%; AUC: 81.4%) and 82.6% (accuracy: 96.4%; F1: 62.5%; AUC: 89.0%) respectively. No further variables contributed to enhanced prediction in these models. For autoencoders, including the health service provider, increased performance to 81.5%; adding the form of recognition and payment further improved it to 81.7%, with the healthcare setting ultimately yielding the highest performance at 82.3% (accuracy: 97.1%; F1: 62.5%; AUC: 87.3%).

We then applied the same modeling strategy using a dataset based on claims from two years, 2018 and 2019, where death occurred in 2019 (see Table 4). The models were tested initially using the first six categories of variables. The average performance metrics achieved were 72.4% (accuracy: 94.3%; F1: 55.4%; AUC: 67.6%) for neural networks, 82.8% (accuracy: 96.5%; F1: 60.0%; AUC: 91.8%) for XGBoost, 79.2% (accuracy: 98.2%; F1: 54.9%; AUC: 84.4%) for autoencoders, 72.9% (accuracy: 95.8%; F1: 53.8%; AUC: 69.2%) for the Bayesian approach, and 78.8% (accuracy: 97.6%; F1: 58.0%; AUC: 80.8%) for TabTransformer.

Subsequently, each variable category was individually reassessed across the previous models to determine whether its inclusion improved predictive accuracy. Only the autoencoder and Bayesian models identified the payment method as an additional important predictor, showing improved average performance over the baseline of 80.4% (accuracy: 98.1%; F1: 58.1%; AUC: 84.9%) and 73% (accuracy: 95.9%; F1: 53.8%; AUC: 69.4%) respectively. Neural networks highlighted the health insurance company as a relevant additional feature with performance of 73.5% (accuracy: 93.0%; F1: 55.9%; AUC: 71.7%) while TabTransformer improved its performance to 79.9% (accuracy: 97.6%; F1: 59.7%; AUC: 82.4%) by adding the health service provider. For

| Model | Neural network | | XGBoost | | Autoencoders | | | | TabTransformer |
|---|---|---|---|---|---|---|---|---|---|
| **Metric** | Base | Base + Setting | Base | Base + Setting | Base | Base + HSP | Base + HSP + Payment | Base + HSP + Payment + Setting | Base |
| **Accuracy*** | 91.0 | 91.9 | 96.4 | 96.4 | 96.5 | 96.9 | 96.8 | 97.1 | 96.3 |
| **F1** | 57.6 | 58.3 | 59.1 | 62.5 | 59.7 | 61.9 | 62.3 | 62.5 | 60.8 |
| **AUC** | 81.8 | 81.4 | 86.1 | 89.0 | 82.7 | 85.6 | 85.9 | 87.3 | 83.8 |
| **Mean** | 76.8 | ***77.2*** | 80.5 | ***82.6*** | 79.6 | 81.5 | 81.7 | ***82.3*** | ***80.3*** |
| **Computation time (hours)** | 19.7 | 22.7 | 20.3 | 22.5 | 86 | 76 | 96.5 | 86.7 | 45.7 |

**Table 3**. Performance metrics and computational time of developed models for 1-year approach. Note: Bold and italicized numbers indicate the best mean performance metric obtained by the model. * The high accuracy could be partly due to imbalance and should not be overinterpreted.

| Model | Neural network | | XGBoost | Autoencoders | | Bayesian | | TabTransformer | |
|---|---|---|---|---|---|---|---|---|---|
| **Metric** | Base | Base + EPS | Base | Base | Base + Payment | Base | Base + Payment | Base | Base + HSP |
| **Accuracy*** | 94.3 | 93.0 | 96.5 | 98.2 | 98.1 | 95.8 | 95.9 | 97.6 | 97.6 |
| **F1** | 55.4 | 55.9 | 60.0 | 54.9 | 58.1 | 53.8 | 53.8 | 58.0 | 59.7 |
| **AUC** | 67.6 | 71.7 | 91.8 | 84.4 | 84.9 | 69.2 | 69.4 | 80.8 | 82.4 |
| **Mean** | 72.4 | ***73.5*** | ***82.8*** | 79.2 | ***80.4*** | 72.9 | ***73.0*** | 78.8 | ***79.9*** |
| **Computation time (hours)** | 5 | 7 | 10 | 43 | 53 | 57 | 50 | 7.2 | 26.3 |

**Table 4**. Performance metrics and computational time of developed models for 2-years approach. Note: Bold and italicized numbers indicate the best mean performance metric obtained by the model. * The high accuracy could be partly due to imbalance and should not be overinterpreted.

| Model | One-year | Two-year |
|---|---|---|
| XGBoost | 0.0303 | 0.0300 |
| Autoencoders | 0.0316 | 0.0189 |
| TabTransformer | 0.0360 | 0.0219 |
| Neural Networks | 0.0649 | 0.0702 |
| Bayesian regression | - | 0.1827 |

**Table 5**. Brier scores for best models.

XGBoost, however, the inclusion of these additional variables did not enhance performance beyond the initial model.

Compared to international literature[32], our models achieved AUC values among the highest reported for one- and two-year mortality prediction in older adults (with several models exceeding 0.84), indicating excellent overall discriminative ability. However, F1 scores were moderate (ranging from approximately 0.54 to 0.63), which is expected given the low prevalence of the positive class (mortality) in our study population. This imbalance affects the trade-off between precision and sensitivity when applying a fixed classification threshold.

We also computed the Brier scores to evaluate the accuracy of probabilistic predictions for binary outcomes; a metric commonly applied in medical prediction models and clinical risk assessment. The values reported in Table 5 show approximations close to zero (below 0.1 except for Bayesian regression), indicating models with good performance in terms of both discrimination and calibration.

While AUC remains our primary metric of interest due to its threshold-independent nature and robustness in evaluating model discrimination, we also report F1 scores and accuracy to provide complementary insights into classification performance. Accuracy, although high (>94% in most models), should be interpreted with caution in imbalanced datasets, as it may overestimate performance by favoring the majority class. In contrast, F1 offers additional context on the balance between false positives and false negatives in a single measure. Thus, these results demonstrate that real-world evidence, specifically patient administrative data, can be used with relative success to identify individuals at risk for 1- and 2-year mortality.

## Discussion and conclusions

Our research demonstrates that diagnostic codes, pharmacological profiles, and hospitalization patterns are the most predictive domains of late-life mortality, whereas operational variables (e.g., provider or insurer) add little incremental value. These findings provide actionable early-warning signals for the design and implementation of targeted preventive interventions aimed at mitigating mortality risk in vulnerable older adult groups[57,58].

Interestingly, some of our models showed improved predictive performance with the inclusion of healthcare setting variables (e.g., outpatient, inpatient, emergency, domiciliary). However, this contrasts with findings from recent qualitative study in Colombia that observed no association between home-based health care and health outcomes among centenarians[59]. In the international literature, evidence regarding the impact of care settings on health outcomes in older adults remains heterogeneous and inconclusive, highlighting the need for further investigation in this area[60,61]. It is important to emphasize that our study does not aim to establish causal inference, but rather to identify predictive patterns based on large-scale administrative data. The purpose of these models is not to explain why mortality occurs, but to help anticipate who might be at greater risk, supporting proactive, data-driven decision-making in public health and clinical settings.

Among the different analytical models tested, XGBoost consistently demonstrated the highest predictive performance. This result underscores the relevance and power of advanced machine learning algorithms, particularly ensemble-based methods, for processing complex, high-dimensional healthcare datasets. XGBoost's strength lies in its ability to capture non-linear relationships and intricate interactions among diverse features, making it a superior tool in clinical risk prediction compared to traditional statistical models[62–65]. Additionally, this approach is favored for its speed and efficiency compared to the other tested algorithms. In contrast, the Bayesian model exhibited computational limitations, as it failed to converge for the one-year prediction horizon. This highlights an area for improvement in such an approach, which often demands substantial computational

resources. Enhancing the scalability and efficiency of Bayesian methods could expand their applicability in real-world healthcare settings, especially when dealing with large and complex datasets[66].

This study is particularly significant as it focuses on a population segment that exceeds the national average life expectancy in Colombia, an age group (80 years and older) that is often underrepresented in predictive analytics. Given the rapid aging of the population and the growing proportion of individuals reaching advanced ages, understanding the mortality predictors in this demographic is vital for informing public health priorities, resource planning, and care delivery strategies.

Recently, Ho et al.[32] published a systematic review evaluating 36 studies and 64 mortality prediction models in older adults, focused on prognostic horizons of up to three years. The authors noted that while many models incorporated clinically relevant predictors (such as comorbidities, age, and sex) the overall predictive performance was generally limited, with most models reporting discrimination (AUC) values below 0.80 for one-year mortality prediction. In contrast, our study reports multiple models achieving AUCs exceeding 0.80 for both one- and two-year prediction horizons. Furthermore, unlike several previous studies that developed only a single model, our study implemented and evaluated five different analytical approaches.

Our research provides novel evidence from a Latin American population-based perspective - a context that remains underrepresented in this domain, where most investigations are conducted in high-income countries. To our knowledge, this is one of the few international studies on late-life mortality prediction using a nationwide cohort of over 300,000 adults aged 80 years and older, providing unprecedented scale and representativeness. By leveraging high-quality administrative records and a robust cohort, we systematically assessed the performance of several mortality prediction models in older adults. In line with the recommendations by Ho et al.[32], our study adheres to rigorous methodological standards as outlined in the TRIPOD and PROBAST reporting guidelines.

This study also illustrates how advanced analytics, and real-world data can be harnessed to support a shift from reactive to proactive healthcare. By identifying patients at elevated risk of mortality through administrative claims data, institutions such as hospitals, insurers, and the Colombian Ministry of Health and Social Protection can implement evidence-based strategies for early intervention. These may include more frequent medical evaluations, home visits, prioritization of palliative care consultations, or tailored public health campaigns, thus optimizing resource allocation and improving health outcomes in aging populations.

Nevertheless, our study has several limitations. First, the analysis was restricted to individuals enrolled in the contributory regime of the Colombian health system, so caution is warranted when generalizing the findings to the subsidized population, which may have different socioeconomic and health profiles. Second, the models were built using administrative claims data, which, while abundant and structured, lack clinical nuance, lifestyle factors, socioeconomic details, and genetic information that could further enhance predictive accuracy. Third, although our models reliably estimate mortality risk, they do not yet provide specific guidance regarding the most effective preventive interventions per individual case.

Future research should aim to integrate multisource data, including electronic health records, socioeconomic indicators and social determinants of health to build more robust and generalizable predictive frameworks. Applying external validation across diverse low- and middle-income countries contexts will be essential to ensure transportability. Beyond prediction, coupling these models with prescriptive analytics, which recommend concrete courses of action, and incorporating causal inference frameworks could significantly enhance their utility for guiding interventions in clinical and public health settings.

In conclusion, this study offers a scalable and replicable approach for the development of predictive models using real-world evidence in middle-income countries. It demonstrates how the intelligent application of machine learning techniques (particularly ensemble models like XGBoost) can enhance the precision and effectiveness of preventive medicine, moving health systems closer to data-driven, person-centered care.

## Data availability

The data used for this study cannot be shared publicly due to legal and confidentiality restrictions. This study uses fully anonymized secondary data obtained from Ministry of Health and Social Protection of Colombia. The data were collected and provided in accordance with applicable data protection regulations. Code available on request to the corresponding author.

Received: 5 November 2025; Accepted: 22 June 2026
Published online: 25 June 2026

## References

1. Jen, M., Shahrokhi, M. & Varacallo, M. Predictive medicine. StatPearls [Internet]. Treasure Island: StatPearls Publishing; 2021 [cited 2022 Sep 26]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK441941/
2. Elton, J. & Ural, A. Predictive medicine depends on analytics [Internet]. Harvard Business Review. 2014 [cited 2022 Oct 10]. Available from: https://hbr.org/2014/10/predictive-medicine-depends-on-analytics
3. Bellazzi, R. & Zupan, B. Predictive data mining in clinical medicine: current issues and guidelines. *Int. J. Med. Inf.* **77** (2), 81–97 (2008).
4. Cowley, L., Farewell, D., Maguire, S. & Kemp, A. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagn. Progn Res.* **3** (1), 16 (2019).
5. Bellazzi, R., Ferrazzi, F. & Sacchi, L. Predictive data mining in clinical medicine: a focus on selected methods and applications. *Wiley Interdiscip Rev. Data Min. Knowl. Discov.* **1** (5), 416–430 (2011).
6. Tuena, C., Semonella, M., Fernández-Álvarez, J., Colombo, D. & Cipresso, P. Predictive precision medicine: towards the computational challenge. In: Pravettoni, G., Triberti, S. (eds): P5 eHealth: an agenda for the health technologies of the future. Cham: Springer International Publishing; 71–86 (2020).
7. Bouwmeester, W. et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med.* **9** (5), e1001221 (2012).

8. Brigham, K. & Johns, M. *Predictive health: how we can reinvent medicine to extend our best years.* 1st edn. New York: Basic Books (2012).
9. Liu, M., Qi, Y., Wang, W. & Sun, X. Toward a better understanding about real-world evidence. *Eur. J. Hosp. Pharm.* **29** (1), 8–11 (2022).
10. Corrigan-Curay, J., Sacks, L. & Woodcock, J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA* **320** (9), 867 (2018).
11. Monahan, A. & Feldman, S. Models predicting hospital admission of adult patients utilizing prehospital data: systematic review using PROBAST and CHARMS. *JMIR Med. Inf.* **9** (9), e30022 (2021).
12. de Ruijter, U. et al. Prediction models for future high-need high-cost healthcare use: a systematic review. *J. Gen. Intern. Med.* **37** (7), 1763–1770 (2022).
13. Klunder, J. et al. Prediction models for the prediction of unplanned hospital admissions in community-dwelling older adults: a systematic review. *PLoS One*. **17** (9), e0275116 (2022).
14. Webb, B. et al. Simple scoring tool to estimate risk of hospitalization and mortality in ambulatory and emergency department patients with COVID-19. *PLoS One*. **17** (3), e0261508 (2022).
15. Lee, J-T., Hsieh, C-C., Lin, C-H., Lin, Y-J. & Kao, C-Y. Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department. *Sci. Rep.* **11** (1), 19472 (2021).
16. Wollenstein-Betech, S., Cassandras, C. & Paschalidis, I. Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: hospitalizations, mortality, and the need for an ICU or ventilator. *Int. J. Med. Inf.* **142**, 104258 (2020).
17. Bradley, J. et al. Predicting hospitalisation for heart failure and death in patients with, or at risk of, heart failure before first hospitalisation: a retrospective model development and external validation study. *Lancet Digit. Heal.* **4** (6), e445–e454 (2022).
18. Brisimi, T., Xu, T., Wang, T., Dai, W. & Paschalidis, I. Predicting diabetes-related hospitalizations based on electronic health records. *Stat. Methods Med. Res.* **28** (12), 3667–3682 (2019).
19. Brisimi, T. S. et al. Predicting chronic disease hospitalizations from electronic health records: an interpretable classification approach. *Proc. IEEE*. **106** (4), 690–707 (2018).
20. Marcusson, J., Nord, M., Dong, H-J. & Lyth, J. Clinically useful prediction of hospital admissions in an older population. *BMC Geriatr.* **20** (1), 95 (2020).
21. King, Z. et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. *npj Digit. Med.* **5** (1), 104 (2022).
22. Yi, S. et al. Predicting hospitalisations related to ambulatory care sensitive conditions with machine learning for population health planning: derivation and validation cohort study. *BMJ Open.* **12** (4), e051403 (2022).
23. Lemke, K., Weiner, J. & Clark, J. Development and validation of a model for predicting inpatient hospitalization. *Med. Care*. **50** (2), 131–139 (2012).
24. Louis, D. et al. Predicting risk of hospitalisation or death: a retrospective population-based analysis. *BMJ Open.* **4** (9), e005223–e005223 (2014).
25. Billings, J., Georghiou, T., Blunt, I. & Bardsley, M. Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. BMJ Open **3** (8), e003352 (2013).
26. McAna, J. et al. A predictive model of hospitalization risk among disabled medicaid enrollees. *Am. J. Manag Care.* **19** (5), e166–e174 (2013).
27. Stylianou, N., Young, J., Peden, C. & Vasilakis, C. Developing and validating a predictive model for future emergency hospital admissions. *Health Inf. J.* **28** (2), 146045822211015 (2022).
28. Kleinjung, F. et al. Real-world data-driven risk prediction of hospitalization for heart failure in non-diabetic CKD. *Eur. Hear. J. - Digit. Heal.* **2** (4), ztab104.3057 (2021).
29. Morawski, K., Dvorkis, Y. & Monsen, C. Predicting hospitalizations from electronic health record data. *Am. J. Manag Care*. **26** (1), e7–13 (2020).
30. Verhoeff, M., de Groot, J., Burgers, J. & van Munster, B. Development and internal validation of prediction models for future hospital care utilization by patients with multimorbidity using electronic health record data. *PLoS One*. **17** (3), e0260829 (2022).
31. Hanlon, P. et al. Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants. *Lancet Public. Heal.* **3** (7), e323–e332 (2018).
32. Ho, L. et al. Performance of models for predicting 1-year to 3-year mortality in older adults: a systematic review of externally validated models. *Lancet Heal Longev.* **5** (3), e227–e235 (2024).
33. Sherman, R. et al. Real-world evidence — What is it and what can it tell us? *N Engl. J. Med.* **375** (23), 2293–2297 (2016).
34. Crown, W. Real-world evidence, causal inference, and machine learning. *Value Heal.* **22** (5), 587–592 (2019).
35. Galbraith, K., Ward, A. & Heneghan, C. A real-world approach to evidence-based medicine in general practice: a competency framework derived from a systematic review and Delphi process. *BMC Med. Educ.* **17** (1), 78 (2017).
36. Concato, J. & Corrigan-Curay, J. Real-world evidence — Where are we now? *N Engl. J. Med.* **386** (18), 1680–1682 (2022).
37. Makady, A. et al. Policies for use of real-world data in health technology assessment (HTA): a comparative study of six HTA Agencies. *Value Heal.* **20** (4), 520–532 (2017).
38. Toro, J., Iunes, R. & Mills, S. Achieving health outcomes in Colombia: civil registration and vital statistics system, unique personal identification number, and unified beneficiary registry system for births and deaths [Internet]. Washington, D.C.; (Health, Nutrition and Population). Report No.: 142548. (2019). Available from: http://documents.worldbank.org/curated/en/709921570776529158
39. Riley, R. D. et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* **368,** m441. (2020).
40. Collins, G. S. et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **162** (1), e078378 (2024).
41. Moons, K. G. M. et al. PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. ***BMJ*** 388, e082505. (2025).
42. Dreiseitl, S. & Ohno-Machado, L. Logistic regression and artificial neural network classification models: a methodology review. *J. Biomed. Inf.* **35** (5–6), 352–359 (2002).
43. Kim, J. S. et al. Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine*. **43** (12), 853–860 (2018).
44. Almansour, N. A. et al. Neural network and support vector machine for the prediction of chronic kidney disease: a comparative study. *Comput. Biol. Med.* **109**, 101–111 (2019).
45. Olang, O. et al. Artificial intelligence-based models for prediction of mortality in ICU patients: a scoping review. *J. Intensive Care Med.* **40** (12), 1240–1246 (2024).
46. Li, B. et al. Using machine learning (XGBoost) to predict outcomes after infrainguinal bypass for peripheral artery disease. *Ann. Surg.* **279** (4), 705–713 (2024).
47. Hernesniemi, J. A. et al. Extensive phenotype data and machine learning in prediction of mortality in acute coronary syndrome – the MADDEC study. *Ann. Med.* **51** (2), 156–163 (2019).
48. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Sci.* **313** (5786), 504–507 (2006).
49. Miotto, R., Li, L., Kidd, B. A. & Dudley, J. T. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* **6** (1), 26094 (2016).
50. Carvalho, C., Polson, N. & Scott, J. Handling Sparsity via the Horseshoe. *Proc. Mach. Learn. Res.* **5**, 73–80 (2009).

51. Fanconi, C., de Hond, A., Peterson, D., Capodici, A. & Hernandez-Boussard, T. A Bayesian approach to predictive uncertainty in chemotherapy patients at risk of acute care utilization. *eBioMedicine* **92**, 104632 (2023).
52. Sullivan, B. et al. Comparing conventional and Bayesian workflows for clinical outcome prediction modelling with an exemplar cohort study of severe COVID-19 infection incorporating clinical biomarker test results. *BMC Med. Inf. Decis. Mak.* **25** (1), 123 (2025).
53. Truong, V. T. T. et al. Bayesian regularization to predict neuropsychiatric adverse events in smoking cessation with pharmacotherapy. *BMC Med. Res. Methodol.* **23** (1), 107 (2023).
54. Huang, X., Khetan, A., Cvitkovic, M. & Karnin, Z. TabTransformer: tabular data modeling using contextual embeddings (2020). Available from: http://arxiv.org/abs/2012.06678
55. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
56. Han, H., Wang, W-Y. & Mao, B-H. In. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D., Zhang, X., Huang, G. (eds) Advances in intelligent computing. ICIC 2005. Lecture Notes in Computer Science, 3644. Berlin: Springer; 878–887 (2005).
57. Abdi, S., Spann, A., Borilovic, J., de Witte, L. & Hawley, M. Understanding the care and support needs of older people: a scoping review and categorisation using the WHO international classification of functioning, disability and health framework (ICF). *BMC Geriatr.* **19** (1), 195 (2019).
58. Li, Y. et al. Healthy lifestyle and the likelihood of becoming a centenarian. *JAMA Netw. Open.* **7** (6), e2417931 (2024).
59. Lozada-Martinez, I., Correa-Diaz, P., Correa-Rosales, J. & Anaya, J. Association between home care and health phenotype of centenarians: Is it a necessary new criterion in home health care? *Australas J. Ageing* **44** (1), e70004 (2025).
60. Low, L-F., Yap, M. & Brodaty, H. A systematic review of different models of home and community care services for older persons. *BMC Health Serv. Res.* **11** (1), 93 (2011).
61. Boland, L. et al. Impact of home care versus alternative locations of care on elder health outcomes: an overview of systematic reviews. *BMC Geriatr.* **17** (1), 20 (2017).
62. Liu, J. et al. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. *PLoS One.* **16** (2), e0246306 (2021).
63. Hu, C-A. et al. Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. *BMJ Open.* **10** (2), e033898 (2020).
64. Park, Y., Kim, Y., Kang, D., Lee, S. & Yoon, S. Prediction of all-cause mortality in Parkinson's disease with explainable artificial intelligence using administrative healthcare data. *npj Park Dis.* **11** (1), 144 (2025).
65. Kawano, K. et al. Prediction of mortality risk of health checkup participants using machine learning-based models: the J-SHC study. *Sci. Rep.* **12** (1), 14154 (2022).
66. Gelman, A. et al. *Bayesian data analysis* 3rd edn. New York: Chapman and Hall/CRC (2013).


## Acknowledgements

Oscar Espinosa acknowledges the HPC Unit of the Instituto de Investigaciones Clínicas at Universidad Nacional de Colombia for providing HPC resources. Giancarlo Buitrago acknowledges the NIHR (NIHR150067), using UK international development funding from the UK Government to support global health research, for strengthening the HPC Unit.

## Author contributions

Conceptualization: OE; data curation: OE, VB, FR; formal analysis: OE, VB, FR, SP, JW, GB; investigation: OE; methodology: OE, VB, FR; project administration: OE, GB; resources: OE, GB; software: OE, VB, FR; supervision: OE; validation: OE, SP, JW, GB; visualization: OE, VB, FR; writing – original draft: OE; and writing – review & editing: OE, VB, FR, SP, JW, GB. OE, VB and FR directly accessed and verified the underlying data reported in this manuscript.

## Declarations

## Competing interests

The authors declare no competing interests.

## Ethics approval

Ethical approval was obtained from the Ethics Committee of the Faculty of Medicine, Universidad Nacional de Colombia (approval letter number 012–121/2020), in accordance with institutional and national ethical guidelines. All data were anonymized by the Ministry of Health and Social Protection prior to access and analysis. The requirement for informed consent was waived by the Ethics Committee due to the use of fully anonymized secondary data. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-59543-2.

**Correspondence** and requests for materials should be addressed to O.E.