ESC HEART FAILURE
ESC Heart Failure 2025; **12**: 859–868
Published online 19 November 2024 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.15056

ORIGINAL ARTICLE

Machine learning-based prediction of elevated N terminal pro brain natriuretic peptide among US general population

Yuichiro Mori[1], Shingo Fukuma[1], Kyohei Yamaji[2], Atsushi Mizuno[3], Naoki Kondo[4] and Kosuke Inoue[4,5]*

[1]*Department of Human Health Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan;* [2]*Department of Cardiovascular Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan;* [3]*Department of Cardiovascular Medicine, St. Luke's International Hospital, Tokyo, Japan;* [4]*Department of Social Epidemiology, Graduate School of Medicine, Kyoto University, Kyoto, Japan; and* [5]*Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan*

Abstract

Aims Natriuretic peptide-based pre-heart failure screening has been proposed in recent guidelines. However, an effective strategy to identify screening targets from the general population, more than half of which are at risk for heart failure or pre-heart failure, has not been well established. This study evaluated the performance of machine learning prediction models for predicting elevated N terminal pro brain natriuretic peptide (NT-proBNP) levels in the US general population.

Methods and results Individuals aged 20–79 years without cardiovascular disease from the nationally representative National Health and Nutrition Examination Survey 1999–2004 were included. Six prediction models (two conventional regression models and four machine learning models) were trained with the 1999–2002 cohort to predict elevated NT-proBNP levels (>125 pg/mL) using demographic, lifestyle, and commonly measured biochemical data. The model performance was tested using the 2003–2004 cohort. Of the 10 237 individuals, 1510 (14.8%) had NT-proBNP levels >125 pg/mL. The highest area under the receiver operating characteristic curve (AUC) was observed in SuperLearner (AUC [95% CI] = 0.862 [0.847–0.878], $P < 0.001$ compared with the logistic regression model). The logistic regression model with splines showed a comparable performance (AUC [95% CI] = 0.857 [0.841–0.874], $P = 0.08$). Age, albumin level, haemoglobin level, sex, estimated glomerular filtration rate, and systolic blood pressure were the most important predictors. We found a similar prediction performance even after excluding socio-economic information (marital status, family income, and education status) from the prediction models. When we used different thresholds for elevated NT-proBNP, the AUC (95% CI) in the SuperLearner models 0.846 (0.830–0.861) for NT-proBNP > 100 pg/mL and 0.866 (0.849–0.884) for NT-proBNP > 150 pg/mL.

Conclusions Using nationally representative data from the United States, both logistic regression and machine learning models well predicted elevated NT-proBNP. The predictive performance remained consistent even when the models incorporated only commonly available variables in daily clinical practice. Prediction models using regularly measured information would serve as a potentially useful tools for clinicians to effectively identify targets of natriuretic-peptide screening.

Keywords Machine learning; NHANES; NT-proBNP; Pre-heart failure; Screening

Received: 30 November 2023; Revised: 1 August 2024; Accepted: 21 August 2024
**Correspondence to: Kosuke Inoue, Department of Social Epidemiology, Graduate School of Medicine, Kyoto University, Yoshida-Konoecho, Sakyo-ku, Kyoto-shi, Kyoto 6068315, Japan. Email: inoue.kosuke.2j@kyoto-u.ac.jp*

Introduction

Natriuretic peptide-based pre-heart failure (pre-HF) screening is gaining increasing interest as a preventive measure against the progression of HF. In 2022, the American Heart Association/American College of Cardiology (AHA/ACC) guidelines for HF introduced the use of brain natriuretic peptide (BNP) or N terminal pro BNP (NT-proBNP) as one of the tools for pre-HF screening among individuals at risk for HF (class IIa recommendation).[1] Moreover, NT-proBNP has been demonstrated to be an independent risk factor for both all-cause and cardiovascular mortality in the general population.[2] However, adhering to the guideline recommendation for measuring natriuretic peptides could be re-

source-intensive, given that nearly 60% of the general population is reportedly at risk for HF or pre-HF.[3] Therefore, identifying specific populations with higher probabilities of positive screening is essential to enhance the feasibility of natriuretic peptide-based pre-HF screening.

Machine learning techniques have the potential to predict the risk of HF accurately by nonparametrically handling the complex interactions among variables.[4] Previous investigations have revealed that machine learning generally improves the performance of prediction models for cardiovascular disease (CVD), including HF, compared with conventional regressions, especially when social risks were incorporated into the models.[5] Although several HF risk predictors, including age, physiological sex, and diabetes, are also reported to be associated with elevated NT-proBNP,[6] no studies to date have evaluated the prediction model performance for elevated NT-proBNP levels in the general population.

To address this knowledge gap, using nationally representative data from the United States, we aimed to evaluate the performance of machine-learning-based prediction models for elevated NT-proBNP levels in the general population without a history of CVD, incorporating well-known HF risk predictors. Additionally, we assessed the predictive performance of models using only commonly available non-invasive data without detailed information on socioeconomic status, which is sometimes difficult to obtain (i.e., marital status, family income, and education status) because prediction models that are applicable in usual clinical settings are key to enhancing the potential of natriuretic peptide-based pre-HF screening as a promising strategy in preventive cardiology.

Methods

Study setting

The National Health and Nutrition Examination Survey (NHANES), conducted by the National Center for Health Statistics, is a multiphase, nationally representative survey of the non-institutionalized US population. The survey continuously collects structured interview data, physical examinations, and laboratory testing data, and releases the information in two-year cycles. The comprehensive design of the NHANES cohort and participants has been discussed in other sources.[7] All individuals provided written consent to participate in the NHANES study protocols, according to the National Center for Health Statistics research ethics review board.[8]

Study samples

This study utilized data from 10 237 US adults aged 20 to 79 years without a history of CVD who participated in three NHANES cycles (1999–2000, 2001–2002, and 2003–2004) and agreed to store their blood samples for future research. Using these stored specimens, NT-proBNP levels were measured for research purpose during 2018–2020 at the University of Maryland School of Medicine, Baltimore, Maryland.[9] The upper age limit (79 years old) was determined to align with the subjects of the AHA/ACC Atherosclerotic Cardiovascular Disease risk calculator[10] and the Pooled Cohort equations to Prevent HF (PCP-HF) score.[11]

Outcome

We defined elevated NT-proBNP levels as NT-proBNP greater than 125 pg/mL aligning with the NT-proBNP Selected Prevention of Cardiac Events in a Population of Diabetic Patients without a History of Cardiac Disease (PONTIAC) study, which investigated the NT-proBNP-based screening and treatment strategy for pre-HF with diabetes.[12] We also used different thresholds for sensitivity analyses (100 and 150 pg/mL). NT-proBNP levels were measured using Roche e601 autoanalyser (Roche Diagnostics, Basel, Switzerland).[2,9]

Predictors

In our prediction models, we included variables previously used in prediction models of HF[11,13] and CVD,[10,14] including age, sex, race/ethnicity (non-Hispanic White, non-Hispanic Black, Mexican American, or others), marital status (currently married or not), educational status (<9th grade, 9th–11th grade, high school/general education development, or >high school), family income (<$20 000, $20 000–$55 000, or >$55 000), smoking status (never, former, or current), body mass index, systolic blood pressure, history of hypertension, history of diabetes, statin use, and biochemical profile (listed below). Participants who smoked at least 100 cigarettes in their lifetime were considered current or former smokers. Self-reported demographic, lifestyle, and medical information were collected by trained interviewers using a computer-assisted interviewing tool.[15] Information on statin use was also collected by trained interviewers who visually confirmed the medication containers.[16]

The biochemical profile included albumin, haemoglobin, aspartate aminotransferase, alanine aminotransferase, total cholesterol, high-density lipoprotein cholesterol, sodium, potassium, glycated haemoglobin A1c, blood urea nitrogen, and the estimated glomerular filtration rate (eGFR). The eGFR was calculated with the 2021 Chronic Kidney Disease Epidemiology Collaboration equation.[17] We used total and high-density lipoprotein cholesterol levels as indicators of cholesterol profile because triglycerides were measured only in participants who were examined in the morning (less than half of all participants), and low-density lipoprotein

cholesterol levels were only available through calculation with triglyceride and other cholesterol levels.[18] Across those variables, missing data were imputed using random forest algorithm with 100 trees and a maximum of 10 times iteration without predictive mean matching.[19,20]

Statistical analysis

In accordance with the TRIPOD statement,[21] the data were divided into a training set (NHANES 1999–2002, *n* = 6800)) and a test set (NHANES 2003–2004, *n* = 3437). Two conventional prediction models and four machine-learning-based models were trained to predict elevated NT-proBNP levels. The conventional models were logistic regression models with and without restricted cubic splines for systolic blood pressure and body mass index. These splines were applied because nonlinear relationships between these variables and elevated NT-proBNP levels would exist.

Four machine learning-based prediction models were built using the previously mentioned predictors. Those models included lasso regression,[22,23] random forest,[20] gradient-boosting machines (GBM),[24] and SuperLearner.[25] Lasso regression is a standard regression model with machine-learning-based selection and weighting of important predictors. Both random forest and GBM use decision trees; random forest combines outputs from randomly generated decision trees, while GBM is an additive model of decision trees estimated through gradient boosting. Lastly, the SuperLearner is an ensemble of multiple machine-learning algorithms (logistic regression, lasso regression, random forest, and GBM) that optimally weights the results of included models. The lambda in lasso regression and hyperparameters in machine learning algorithms (*Table* S1) were tuned using 10-fold cross-validation.

In the test set, each model's predictive performance was assessed by computing the area under the receiver operating characteristic curve (AUC) and prospective prediction performance (i.e., sensitivity, specificity, positive predictive value, and negative predictive value). The AUCs were compared using DeLong's test (reference: the logistic regression model without splines). Considering the class imbalance in the outcome, the Youden index[26] was used to define the threshold probability of elevated NT-proBNP. Calibration was evaluated using calibration intercept and slope for each model.[27] To visualize the contributions of predictor variables to the outputs of machine-learning models, we utilized the Shapley additive explanations (SHAP) analysis for GBM algorithms.[24,28] All analyses were performed using R version 4.2.2.

We conducted the following seven sensitivity analyses: (i) we trained models without marital status, family income, and education status because it is sometimes difficult to obtain detailed information in usual clinical settings; (ii) to further improve the generalizability, we trained models using only commonly available, non-invasive information including age, sex, body mass index, systolic blood pressure, history of hypertension, history of diabetes, and statin use; (iii) to assess optimal prediction threshold, the threshold probability of elevated NT-proBNP was set to its observed prevalence, instead of the threshold determined with the Youden index, for all models, and set to various thresholds for the model that achieved the best AUC; (iv) we restricted participants to those who were free from shortness of breath on stairs or inclines; (v) we repeated the main analysis among participants with complete data; vi) we also repeated the main analysis with different definitions of elevated NT-proBNP: greater than 100 and 150 pg/mL; and vii) we repeated the main analysis with an oversampling technique, the Random Over-Sampling Examples (ROSE), to produce dataset with a 1:1 ratio of those with or without elevated NT-proBNP.[29]

Results

Among 10 237 participants, the mean age was 45.3 years (standard deviation [SD]: 16.6 years), and 53.0% (*n* = 5425) were female. Of those 10 237 participants, 1510 (14.8%) had elevated NT-proBNP. Participants with elevated NT-proBNP were more likely to be older, female, non-Hispanic White, have lower education levels, have lower family income levels, have histories of hypertension and diabetes, take statins, and have higher blood pressure (*Table* 1). The comparison between before and after imputing missing data is presented in Table S2, and the comparison between training and test sets is presented in Table S3. The prevalence of missing data was lower than 0.5% in almost all variables except for family income (5.8%), systolic blood pressure (4.0%), marital status (3.3%), and body mass index (1.9%).

Across the six models, the highest predictive performance was shown in SuperLearner (AUC [95% confidence interval: CI] = 0.862 [0.846–0.878], $P < 001$, sensitivity = 0.781, specificity = 0.780, positive predictive value = 0.395, and negative predictive value = 0.951). The conventional logistic regression models showed comparable predictive performance (without splines [reference]: AUC = 0.854 [0.837–0.871], sensitivity = 0.811, specificity = 0.734, positive predictive value = 0.360, and specificity = 0.955; and with splines: AUC = 0.857 [0.841–0.874] *P* = 0.08, sensitivity = 0.766, specificity = 0.772, positive predictive value = 0.382, and specificity = 0.947). Three machine learning models showed performances almost equal to or slightly lower than the logistic regression model without splines: lasso regression (AUC = 0.848 [0.830–0.865], *P* = 0.55); random forest (AUC = 0.848 [0.831–0.865], P = 0.32); and GBM (AUC = 0.854 [0.837–0.870], P = 0.92; *Table* 2 and *Figure* 1). All six models were well calibrated with intercepts between

Table 1 Baseline characteristics of study participants by NT-proBNP level

	NT-proBNP ≤ 125 pg/mL n = 8727	NT-proBNP > 125 pg/mL n = 1510
Age, year	42.56 (15.32)	60.95 (14.78)
Female	4402 (50.4)	1023 (67.7)
Race/ethnicity		
Non-Hispanic White	4085 (46.8)	884 (58.5)
Non-Hispanic Black	1675 (19.2)	227 (15.0)
Mexican American	2179 (25.0)	304 (20.1)
Others	788 (9.0)	95 (6.3)
Married	5080 (58.2)	884 (58.5)
Education status		
<9th grade	1093 (12.5)	316 (20.9)
9th–11th grade	1407 (16.1)	257 (17.0)
High school or GED	2087 (23.9)	360 (23.8)
>High school	4140 (47.4)	577 (38.2)
Family income		
<$20 000	2592 (29.7)	584 (38.7)
$20 000–55 000	3559 (40.8)	644 (42.6)
>$55 000	2576 (29.5)	282 (18.7)
Smoking status		
Never	4619 (52.9)	733 (48.5)
Former	1983 (22.7)	490 (32.5)
Current	2125 (24.3)	287 (19.0)
Hypertension	1919 (22.0)	716 (47.4)
Diabetes	572 (6.6)	209 (13.8)
Statin use	475 (5.4)	200 (13.2)
Body mass index, kg/m^2	28.40 (6.17)	28.48 (6.48)
Systolic blood pressure, mmHg	120.78 (16.53)	135.86 (24.99)
eGFR, mL/min/1.73 m^2	105.25 (19.61)	86.32 (23.83)
HbA1c, %	5.49 (0.97)	5.69 (1.11)
Aspartate aminotransferase, IU/L	25.16 (19.48)	25.48 (44.05)
Alanine aminotransferase, IU/L	27.05 (31.83)	22.87 (52.54)
Blood urea nitrogen, mg/dL	12.48 (4.30)	14.99 (7.02)
Haemoglobin, g/dL	14.40 (1.54)	13.79 (1.46)
Albumin, g/dL	4.30 (0.38)	4.14 (0.36)
Total cholesterol, mg/dL	203.32 (42.35)	205.84 (42.34)
High-density lipoprotein cholesterol, mg/dL	52.38 (15.51)	56.11 (17.07)
Sodium, mEq/L	138.88 (2.37)	139.12 (2.80)
Potassium, mEq/L	4.02 (0.32)	4.06 (0.38)
NT-proBNP, pg/mL	42.40 (30.15)	471.66 (1939.88)
NHANES cycle		
1999–2000	2643 (30.3)	455 (30.1)
2000–2001	3181 (36.5)	521 (34.5)
2002–2003	2903 (33.3)	534 (35.4)

Values are n (%) in counts and mean (SD) in continuous value.
eGFR, estimated glomerular filtration rate; GED, General Educational Development; HbA1c, glycated haemoglobin A1c; NHANES, National Health and Nutrition Examination Survey; NT-proBNP, N-terminal prohormone of brain natriuretic peptide.

−0.10 and 0.10, and slopes between 0.90 and 1.10 (*Figure* S1). In the SHAP analysis, age was the most important predictor, followed by albumin, haemoglobin, sex systolic blood pressure, and eGFR (*Figure* 2 and Figure S2). The increase in age and the decrease in albumin and eGFR showed monotonic positive contributions to the predicted probability of elevated NT-proBNP. In contrast, the contributions of haemoglobin and systolic blood pressure appeared to have J-curve characteristics. These predictor-contribution relationships did not obviously differ across age groups.

The predictive performance did not qualitatively change (i) when we excluded marital status, family income, and education status from the models (*Table* 2 and Figure S3), (ii) when we trained models only with commonly available, non-invasive information (Table S4), (iii) when we set the threshold probability for elevated NT-proBNP to its observed prevalence (Table S5), (iv) when we analysed the data among participants without shortness of breath on stairs or inclines (Table S6 and Figure S4), and (v) when we conducted complete-case analysis without imputation (Table S7).

When we used different thresholds for elevated NT-proBNP (i.e., >100 or 150 pg/mL), higher cut-off values of NT-proBNP resulted in higher predictive performance (Figure S5 and Table S8). For example, in SuperLearner models, AUCs (95% CI) were 0.846 (0.830–0.861) for NT-proBNP > 100 pg/mL and 0.866 (0.849–0.884) for NT-proBNP > 150 pg/mL (*Figure* 3). The variable importance was also almost consistent in the higher order (Figure S6). The AUC, sensitivity, and specificity were also consistent when the models were trained and tested using a dataset

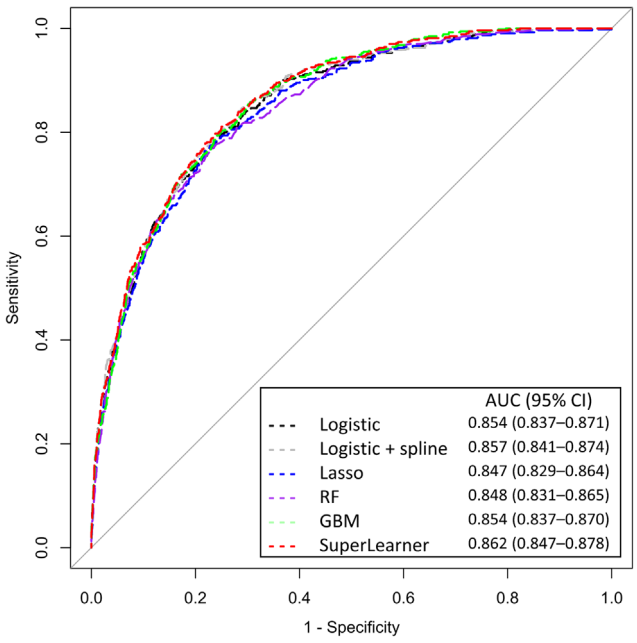
Table 2 Predictive ability of the logistic regression model, tree-based algorithms, and SuperLearner for elevated NT-proBNP

	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	*P* value
Main analysis						
Logistic regression	0.854 (0.837–0.871)	0.811	0.734	0.360	0.955	Reference
Logistic regression + spline models	0.857 (0.841–0.874)	0.803	0.755	0.376	0.954	0.08
Lasso regression	0.847 (0.829–0.864)	0.788	0.752	0.369	0.951	0.55
Random forest	0.848 (0.831–0.865)	0.753	0.788	0.395	0.945	0.32
Gradient boosting	0.854 (0.837–0.870)	0.846	0.713	0.352	0.962	0.92
SuperLearner	0.862 (0.847–0.878)	0.781	0.780	0.395	0.951	0.001
Sensitivity analysis without marital status, family income, and education status						
Logistic regression	0.855 (0.838–0.871)	0.833	0.726	0.359	0.959	Reference
Logistic regression + spline models	0.859 (0.842–0.875)	0.850	0.709	0.349	0.963	0.047
Lasso regression	0.848 (0.830–0.865)	0.837	0.704	0.343	0.959	0.56
Random forest	0.848 (0.831–0.865)	0.742	0.794	0.398	0.943	0.25
Gradient boosting	0.858 (0.842–0.874)	0.803	0.756	0.377	0.954	0.50
SuperLearner	0.863 (0.847–0.879)	0.766	0.772	0.382	0.947	0.002

Results of the main analysis (upper) and a sensitivity analysis without marital status, family income, and education status (lower) are displayed. In the main analysis, models were trained using age, sex, race/ethnicity, marital status, education status, family income, smoking status, body mass index, systolic blood pressure, history of hypertension, history of diabetes, statin use, and biochemical profile. The AUCs were compared using DeLong's test. The threshold probability of elevated NT-proBNP in the main analysis, determined based on the Youden's index, was 0.133 in logistic regression, 0.146 in logistic regression + splines, 0.146 in lasso regression, 0.196 in random forest, 0.109 in the gradient boosting, and 0.157 in SuperLearner.

AUC, area under the receiver operating characteristic curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

Figure 1 Receiver operating characteristic curve of the logistic regression model, tree-based algorithms, and SuperLearner to predict elevated NT-proBNP levels. Models were trained using age, sex, race/ethnicity, marital status, education status, family income, smoking status, body mass index, systolic blood pressure, history of hypertension, history of diabetes, statin use, and biochemical profile. AUC, area under the receiver operating characteristics curve; CI, confidence interval; GBM, gradient-boosting machines; Lasso, logistic regression with lasso regularization; Logistic, logistic regression model; RF, random forest.



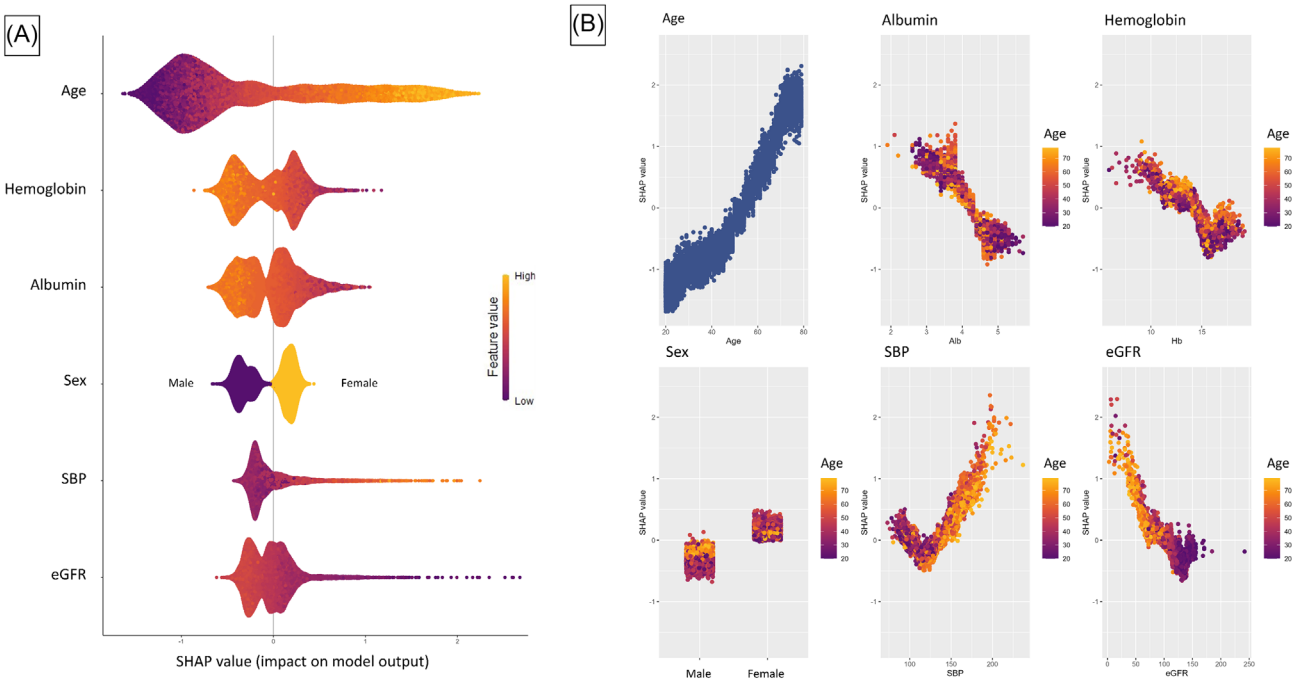
generated by an oversampling technique to achieve a 1:1 ratio of those with and without elevated NT-proBNP (Table S9). When various prediction thresholds were applied to SuperLearner model, threshold probability around 0.10 seemed to achieve high sensitivity for screening purpose with maintaining specificity (Figure S7).

Discussions

In this study, including 10 237 participants from nationally representative data in the United States, both conventional logistic regression models and machine learning-based models predicted elevated NT-proBNP levels with AUCs of approximately 0.860. The SuperLearner model showed the best performance among the six models investigated, but the conventional logistic regression models also showed comparable performance. Some key variables for HF or CVD largely contributed to the prediction performance of our models, including age, albumin, haemoglobin, sex, systolic blood pressure, and eGFR. Model performance remained consistent when the models incorporated only commonly available information in daily clinical practice.

This study provides new insights into the natriuretic peptide-based pre-HF screening. To the best of our knowledge, this is the first study that evaluated prediction model performance for elevated NT-proBNP in the general population. The AHA/ACC guidelines provide class IIa recommendations for screenings in individuals at risk for HF based on two randomized controlled trials.[12,30] However, defining a standardized screening strategy has been challenging because the risk of HF is a continuous concept without a clear boundary, and nearly 60% of the general population is reportedly at risk for HF or pre-HF,[3] resulting in borderline cost-effectiveness of the pre-HF screening.[31] By utilizing

Figure 2 Title: Shapley additive explanations (SHAP) analysis for variables with high importance in the GBM model. (A) The SHA*P* value (x-axis) indicates the contribution of predictor variables (*y*-axis) to the model output (the logit of the probability of NT-proBNP > 125 pg/mL). Each dot in a row represents a single participant. Colour indicates the value of predictor variables. (B) Each dot in each scatter plot represents a single participant as well as the panel (A). Those plots represent the relationships between the predictor values and the SHAP values, which could differ across participants. Colour represents the age of participants to visualize whether predictor-contribution relationships differ across age groups. eGFR, estimated glomerular filtration rate; SBP, systolic blood pressure; SHAP, Shapley additive explanations.



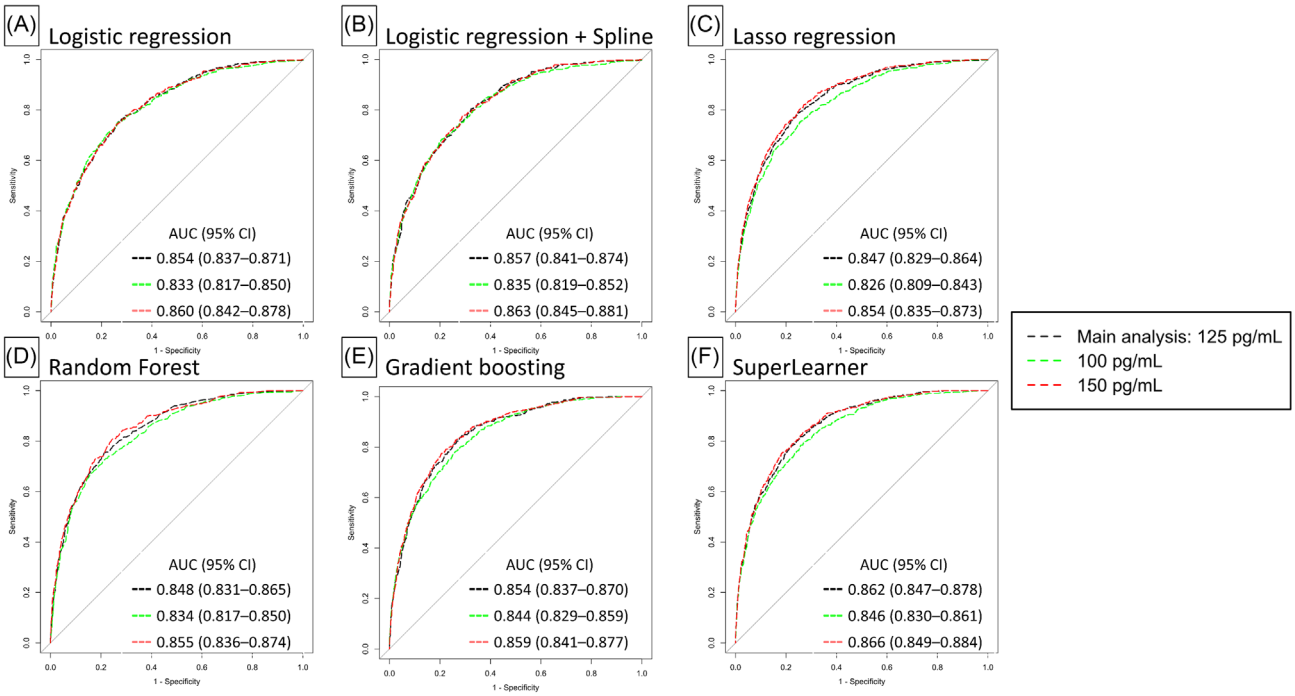
prediction models for elevated NT-proBNP with commonly collected information, natriuretic peptide-based pre-HF screening may achieve better cost-effectiveness and feasibility, which helps healthcare providers adhere to guideline recommendations.

Recent advancements in drug therapy for pre-HF also enhance the value of the screening. For example, the Personalized Prospective Comparison of ARNI With ARB in Patients With Natriuretic Peptide Elevation (PARABLE) study[32] suggested that sacubitril/valsartan may reduce cardiovascular risk in pre-HF with preserved ejection fraction. Concurrently, guideline-recommended drug therapy already exists for pre-HF with reduced ejection fraction.[1] Considering the advancements in the field of pre-HF, efficient identification of elevated NT-proBNP and pre-HF cases among the general population is vital in addressing the HF pandemic—an emerging public health concern worldwide.[33]

In our analysis, conventional logistic regression models showed similar performance with other machine learning models in predicting elevated NT-proBNP. To date, some studies have reported improvements in the prediction performance for incident HF using machine learning compared with logistic regression models.[5] This discrepancy between our findings and previous reports might be due to the fact that NT-proBNP levels predominantly represent a specific aspect of physiology—elevated cardiac pressure[34] although future HF progression involves a multifactorial phenomenon (e.g., physiological, behavioural, and socio-economic). Moreover, predictor variables used in our study have been extensively validated for their association with the risk of HF or CVD,[10,11,13,14] that might lead to the high prediction performance regardless of the models (i.e., either machine learning or conventional logistic regression models). In this context, conventional logistic regression model—a less computationally intense approach than machine learning—would be sufficient to predict pre-HF as long as we measure key risk factors for HF.

Of interest, in addition to major risk factors for future onset of HF such as hypertension, diabetes, or dyslipidemia, we found a large contribution of albumin and haemoglobin to the prediction of elevated NT-proBNP levels. Physiologically, low albumin and haemoglobin levels could modify the haemodynamic status and place stress on the myocardium, potentially resulting in elevated NT-proBNP levels. Indeed, the Atherosclerosis Risk in Communities study included haemoglobin in its prediction tools for HF with preserved ejection fraction.[35] Although the causal relationships between these variables and pre-HF should be further elucidated, our findings indicate that including such variables that affect haemodynamic status in risk prediction models could

Figure 3 Receiver operating characteristic curves of different models with different thresholds of defining elevated NT-proBNP. Models were trained using age, sex, race/ethnicity, marital status, education status, family income, smoking status, body mass index, systolic blood pressure, history of hypertension, history of diabetes, statin use, and biochemical profile. The ROCs and AUCs for model with different definitions of elevated NT-proBNP were displayed for each model. Panel (A): logistic regression. Panel (B): logistic regression with splines. Panel (C): lasso regression. Panel (D): random forest. Panel (E): gradient boosting. Panel (F): SuperLearner. AUC, area under the receiver operating characteristics curve; CI, confidence interval.



be helpful in identifying patients at elevated risk of HF progression. It is also important to note that variable importance is a population-level indicator. The individual-level of importance of risk factors with low prevalence in the study cohort, such as diabetes, might be not fully reflected in the variable importance.

Recently, the role of socio-economic status in HF management and prevention has received substantial attention.[36,37] The cumulative psychological stress due to social risks (e.g., low income levels, low education levels, etc.) can induce a status called 'allostatic load', which is considered to have a significant impact on CVD risk equivalent to traditional CVD risk factors[38] as well as the new onset of HF.[39] Such stress influences the function of hypothalamic–pituitary–adrenal axis, increasing cortisol release[40] which induces cardiac and vascular remodelling and regulates renal sodium and water homeostasis. Indeed, a previous study of Multi-Ethnic Study of Atherosclerosis showed the relationship of elevated urinary cortisol levels with increased risk of incident hypertension and cardiovascular events among general population.[41] Meanwhile, detailed information on such socio-economic status is not always readily available in daily clinical care. Given our results in which the prediction performance for elevated NT-proBNP did not significantly differ when the models only included downstream physiological information of such socio-economic status, we still can identify pre-HF screening targets effectively even under the situation where we cannot measure such key social information.

Patients with pre-HF are typically asymptomatic, making it crucial to maximize opportunities of pre-HF screenings when heart failure risk is identified—sometimes, by chance. The prediction of NT-proBNP elevation prediction offers a valuable tool for primary care physicians to efficiently identify patients who may benefit from such screenings. Recent studies have demonstrated that implementing prediction models embedded within electronic health record systems significantly increased the initiation of preventive intervention.[42] Given the large number of potential candidates for NT-proBNP screening in the current guideline, there would also be synergy between the prediction model of elevated NT-proBNP and such decision support systems using administrative data (Figure S8).

There are several limitations in this study. First, as our analysis depends on data collected during 1999–2004 because NT-proBNP were not measured in blood samples of later NHANES cycles, predictors for elevated NT-proBNP in the current clinical practice could differ given the recent advancement in HF risk management (e.g., development of novel drugs including sodium-glucose cotransporter-2 inhibitors,

proprotein convertase subtilisin/kexin type 9 inhibitors, and non-steroidal mineralocorticoid receptor antagonists). Nevertheless, the cross-sectional relationships between measured variables and NT-proBNP, which primarily reflect cardiac stress, are likely to remain consistent compared with the association between NT-proBNP and longitudinal outcomes. Second, although we followed the TRIPOD guideline and evaluated the prediction performance in the test data, our model was not externally validated using data outside of NHANES. Third, our findings might suffer from information bias as NT-proBNP was measured only once for each participant. Moreover, because sociodemographic information was self-reported, we cannot rule out the possibility of misclassification of these predictors. Fourth, the threshold of NT-proBNP for pre-HF screening has not yet established.[43–45] Therefore, we confirmed model performance with different thresholds in sensitivity analyses. Fifth, the cut-offs we reported were chosen to balance sensitivity and specificity for comparing different models. The optimal cut-off for screening would vary depending on clinical and policy contexts. Therefore, we conducted the sensitivity analyses with different prediction threshold. Sixth, the models were not designed to handle missing data. Therefore, we conducted sensitivity analyses with readily available data that are typically complete in most clinical settings. Lastly, we did not utilize the NHANES survey weight, and thus may limit the generalizability. Further investigation is needed to validate our findings and assess the applicability of our prediction models in the current clinical practice. Future work should focus on the external validity of our prediction models and assess the utility of these models as a pre-HF screening strategy.

In conclusion, both conventional logistic regression models and machine learning-based models effectively predicted elevated NT-proBNP levels in the US general population. Results were consistent when the models included only commonly available variables in daily clinical practice. Utilizing such prediction models may enhance the effectiveness and feasibility of the natriuretic peptide-based pre-HF screening among the general population.

Clinical perspectives

Clinical competencies

Both conventional logistic regression models and machine learning-based models effectively predicted elevated NT-proBNP levels among the US general population. We found a similar predictive performance even when we did not include detailed socio-economic status in our models, increasing the applicability of these models in usual clinical settings as a promising strategy in preventive cardiology.

Translational outlook

The successful application of machine learning in predicting elevated NT-proBNP levels paves the way for further discussion on integrating such models into standardized screening strategies for pre-HF. Future work should focus on the external validity of those prediction models and the cost-effectiveness of natriuretic peptide-based pre-HF screening strategies combined with such prediction models.

Conflict of interest

None declared.

Funding

KI was supported by grants 22K17392 and 23KK0240 from the Japan Society for the Promotion of Science, the Japan Agency for Medical Research and Development (AMED; JP22rea522107), the Japan Science and Technology (JST PRESTO; JPMJPR23R2), and the Program for the Development of Next-generation Leading Scientists with Global Insight (L-INSIGHT) sponsored by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. Study sponsors were not involved in study design, data interpretation, writing, or the decision to submit the article for publication.

Data availability statement

Statistical code: https://github.com/YuichiroMori56/NT-proBNP-prediction/ (CC-BY-NC-SA). Original data generated and analysed during this study are publicized by the Center for Disease Control and Prevention, the United States: https://wwwn.cdc.gov/nchs/nhanes/Default.aspx (accessed July 23, 2024).

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: Hyperparameters automatically tuned in machine learning algorithms.
Table S2: Comparison of before and after imputing missing data.
Table S3: Baseline characteristics of training and test data.
Table S4: Results of sensitivity analysis with commonly available, non-invasive information.

Table S5: Results of sensitivity analysis when probability thresholds were set to observed prevalence of elevated NT-pro BNP.
Table S6: Results of sensitivity analysis in individuals free from shortness of breath on stairs/inclines.
Table S7: Results of sensitivity analysis among participants with complete data.
Table S8: Results of sensitivity analysis when elevated NT-pro BNP were defined as NT-pro BNP $>$ 100 or 150 pg/dL.
Table S9: Results of sensitivity analysis among participants with oversampling method.
Figure S1: Calibration plots in the main analysis.
Figure S2: The beeswarm plot of the SHAP analysis in the GBM model.
Figure S3: Receiver operating characteristic curves in a sensitivity analysis without marital status, family income, and education status.
Figure S4: Receiver operating characteristic in a sensitivity analysis among those free from shortness of breath on stairs/inclines.
Figure S5: Sensitivity analysis with different definitions of elevated NT-pro BNP.
Figure S6: Variable importance in SHAP analysis with main analysis and sensitivity analyses.
Figure S7: Results in the SuperLerner model with various prediction thresholds.
Figure S8: Current practice in pre-HF screening and potential of pre-HF screening with prediction models.

References

1. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, *et al*. 2022 AHA/ACC/HFSA guideline for the Management of Heart Failure: a report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines. *Circulation American Heart Association* 2022;**145**:e895-e1032. doi:10.1161/CIR.0000000000001063
2. Echouffo-Tcheugui JB, Zhang S, Daya N, McEvoy JW, Tang O, Juraschek SP, *et al*. NT-proBNP and all-cause and cardiovascular mortality in US adults: a prospective cohort study. *J Am Heart Assoc Ovid Technologies (Wolters Kluwer Health)* 2023;**12**: doi:10.1161/JAHA.122.029110
3. Xanthakis V, Enserro DM, Larson MG, Wollert KC, Januzzi JL, Levy D, *et al*. Prevalence, Neurohormonal correlates, and prognosis of heart failure stages in the community. *JACC Heart Fail* 2016;**4**:808-815. doi:10.1016/j.jchf.2016.05.001
4. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med* 2017;**376**:2507-2509. doi:10.1056/NEJMp1702071
5. Zhao Y, Wood EP, Mirin N, Cook SH, Chunara R. Social determinants in machine learning cardiovascular disease prediction models: a systematic review. *Am J Prev Med* 2021;**61**:596-605. doi:10.1016/j.amepre.2021.04.016
6. Welsh P, Campbell RT, Mooney L, Kimenai DM, Hayward C, Campbell A, *et al*. Reference ranges for NT-proBNP (N-terminal pro-B-type natriuretic peptide) and risk factors for higher NT-proBNP concentrations in a large general population cohort. *Circ Heart Fail* 2022;**15**:e009427. doi:10.1161/CIRCHEARTFAILURE.121.009427
7. National Health and Nutrition Examination Survey. National Center for Health Statistics. 2023. https://www.cdc.gov/nchs/nhanes/index.htm. Accessed 19 April 2023
8. NHANES - NCHS Research Ethics Review Board Approval. National Center for Health Statistics. 2022. https://www.cdc.gov/nchs/nhanes/irba98.htm. Accessed 19 April 2023
9. National Health and Nutrition Examination Survey 1999–2004 Data Documentation, Codebook, and Frequencies N-terminal Pro-BNP (Surplus) (SSBNP_A). National Center for Health Statistics. 2022. https://wwwn.cdc.gov/Nchs/Nhanes/1999-2000/SSBNP_A.htm. Accessed 17 April 2023
10. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, *et al*. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014;**129**:S49-S73. doi:10.1161/01.cir.0000437741.48606.98
11. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, *et al*. 10-year risk equations for incident heart failure in the general population. *J Am Coll Cardiol* Elsevier BV 2019;**73**:2388-2397. doi:10.1016/j.jacc.2019.02.057
12. Huelsmann M, Neuhold S, Resl M, Strunk G, Brath H, Francesconi C, *et al*. PONTIAC (NT-proBNP selected prevention of cardiac events in a population of diabetic patients without a history of cardiac disease): a prospective randomized controlled trial. *J Am Coll Cardiol* 2013;**62**:1365-1372. doi:10.1016/j.jacc.2013.05.069
13. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, *et al*. Prediction of incident heart failure in general practice: the atherosclerosis risk in communities (ARIC) study. *Circ Heart Fail* 2012;**5**:422-429. doi:10.1161/CIRCHEARTFAILURE.111.964841
14. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham heart study. *Stat Med* 1990;**9**:1501-1515. doi:10.1002/sim.4780091214
15. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National health and nutrition examination survey: plan and operations, 1999-2010. *Vital Health Stat* 2013;**56**:1-37.
16. National Health and Nutrition Examination Survey 1999–2000 Data Documentation, Codebook, and Frequencies Prescription Medications (RXQ_RX). National Center for Health Statistics. 2009. https://wwwn.cdc.gov/nchs/nhanes/1999-2000/RXQ_RX.htm. Accessed 18 April 2023
17. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, *et al*. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med* 2021;**385**:1737-1749. doi:10.1056/NEJMoa2102953
18. NHANES 1999-2000: Cholesterol - LDL & Triglycerides Data Documentation, Codebook, and Frequencies. 2023. https://wwwn.cdc.gov/nchs/nhanes/1999-2000/LAB13AM.htm. Accessed 18 April 2023
19. Tang F, Ishwaran H. Random forest missing data algorithms. *Stat Anal Data Min* 2017;**10**:363-377. doi:10.1002/sam.11348
20. Wright MN, Ziegler A. Ranger: a fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw Foundation for Open Access Statistic* 2017;**77**:1-17. doi:10.18637/jss.v077.i01

21. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;**162**:W1-W73. doi:10.7326/M14-0698
22. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw Foundation for Open Access Statistic* 2010;**33**:1-22.
23. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. *J Stat Softw Foundation for Open Access Statistic* 2023;**106**:1-31. doi:10.18637/jss.v106.i01
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. arXiv [cs. LG]. 2016.
25. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;**6**:25. doi:10.1039/d4cy00484a
26. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden index to discriminate individuals using pooled blood samples. *Epidemiology* 2005;**16**:73-81. doi:10.1097/01.ede.0000147512.81966.ba
27. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic group 'evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;**17**:230. doi:10.1186/s12916-019-1466-7
28. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., eds. *Neural Inf Process Systems*. Vol.**30**. Curran Associates, Inc.; 2017:4765-4774. doi:10.3389/fsoc.2024.1367517
29. Lunardon N, Menardi G, Torelli N. ROSE: a package for binary imbalanced learning. *R J The R Foundation* 2014;**6**:79.
30. Ledwidge M, Gallagher J, Conlon C, Tallon E, O'Connell E, Dawkins I, *et al*. Natriuretic peptide-based screening and collaborative care for heart failure: the STOP-HF randomized trial. *JAMA American Medical Association (AMA)* 2013;**310**:66-74. doi:10.1001/jama.2013.7588
31. Ledwidge MT, O'Connell E, Gallagher J, Tilson L, James S, Voon V, *et al*. Cost-effectiveness of natriuretic peptide-based screening and collaborative care: a report from the STOP-HF (St Vincent's screening to prevent heart failure) study. *Eur J Heart Fail* 2015;**17**:672-679. doi:10.1002/ejhf.286
32. Ledwidge M, Dodd JD, Ryan F, Sweeney C, McDonald K, Fox R, *et al*. Effect of sacubitril/valsartan vs valsartan on left atrial volume in patients with pre-heart failure with preserved ejection fraction: the PARABLE randomized clinical trial. *JAMA Cardiol* 2023;**8**:366-375. doi:10.1001/jamacardio.2023.0065
33. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, *et al*. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol* 2014;**63**:1123-1133. doi:10.1016/j.jacc.2013.11.053
34. Gremmler B, Kunert M, Schleiting H, Kisters K, Ulbricht LJ. Relation between N-terminal pro-brain natriuretic peptide values and invasively measured left ventricular hemodynamic indices. *Exp Clin Cardiol* 2003;**8**:91-94.
35. Thorvaldsen T, Claggett BL, Shah A, Cheng S, Agarwal SK, Wruck LM, *et al*. Predicting risk in patients hospitalized for acute decompensated heart failure and preserved ejection fraction: the atherosclerosis risk in communities study heart failure community surveillance. *Circ Heart Fail* 2017;**10**: doi:10.1161/CIRCHEARTFAILURE.117.003992
36. Schultz WM, Kelli HM, Lisko JC, Varghese T, Shen J, Sandesara P, *et al*. Socioeconomic status and cardiovascular outcomes: challenges and interventions. *Circulation* 2018;**137**:2166-2178. doi:10.1161/CIRCULATIONAHA.117.029652
37. Khariton Y, Nassif ME, Thomas L, Fonarow GC, Mi X, DeVore AD, *et al*. Health status disparities by sex, race/ethnicity, and socioeconomic status in outpatients with heart failure. *JACC Heart Fail* 2018;**6**:465-473. doi:10.1016/j.jchf.2018.02.002
38. Hamad R, Penko J, Kazi DS, Coxson P, Guzman D, Wei PC, *et al*. Association of low socioeconomic status with premature coronary heart disease in US adults. *JAMA Cardiol* 2020;**5**:899-908. doi:10.1001/jamacardio.2020.1458
39. Potter EL, Hopper I, Sen J, Salim A, Marwick TH. Impact of socioeconomic status on incident heart failure and left ventricular dysfunction: systematic review and meta-analysis. *Eur Heart J Qual Care Clin Outcomes* 2019;**5**:169-179. doi:10.1093/ehjqcco/qcy047
40. Kivimäki M, Steptoe A. Effects of stress on the development and progression of cardiovascular disease. *Nat Rev Cardiol* 2018;**15**:215-229. doi:10.1038/nrcardio.2017.189
41. Inoue K, Horwich T, Bhatnagar R, Bhatt K, Goldwater D, Seeman T, *et al*. Urinary stress hormones, hypertension, and cardiovascular events: the multi-ethnic study of atherosclerosis. *Hypertension* 2021;**78**:1640-1647. doi:10.1161/HYPERTENSIONAHA.121.17618
42. Adusumalli S, Kanter GP, Small DS, Asch DA, Volpp KG, Park S-H, *et al*. Effect of nudges to clinicians, patients, or both to increase statin prescribing: a cluster randomized clinical trial. *JAMA Cardiol* 2023;**8**:23-30. doi:10.1001/jamacardio.2022.4373
43. Bozkurt B, Coats A, Tsutsui H. Universal definition and classification of heart failure. *J Card Fail* 2021;**27**:387-413. doi:10.1016/j.cardfail.2021.01.022
44. Bayes-Genis A, Doherty KF, Petrie MC, Januzzi JL, Mueller C, Andreson L, *et al*. Practical algorithms for early diagnosis of heart failure and heart stress using NT-proBNP: a clinical consensus statement from the heart failure association of the ESC. *Eur J Heart Fail* 2023;**25**:1891-1898. doi:10.1002/ejhf.3036
45. Jehn S, Mahabadi AA, Pfohl C, Vogel L, Al-Rashid F, Luedike P, *et al*. BNP and NT-proBNP thresholds for the assessment of prognosis in patients without heart failure. *JACC: Advances* 2023;**2**:100688. doi:10.1016/j.jacadv.2023.100688