

Meraj Neyazi, Jan P. Bremer, Marius S. Knorr, Stefan Gross, Jan Bredericke, Nils Schweingruber, Dora Csengeri, Benedikt Schrage, Martin Bahls, Nele Friedrich, Tanja Zeller, Stephan Felix, Stefan Blankenberg, Marcus Dörr, Marcus Vollmer and Renate B. Schnabel*

Deep learning-based NT-proBNP prediction from the ECG for risk assessment in the community

<https://doi.org/10.1515/cclm-2023-0743>

Received July 24, 2023; accepted October 20, 2023;

published online November 20, 2023

Abstract

Objectives: The biomarker N-terminal pro B-type natriuretic peptide (NT-proBNP) has predictive value for identifying individuals at risk for cardiovascular disease (CVD). However, it is not widely used for screening in the general population, potentially due to financial and operational reasons. This study aims to develop a deep-learning model as an efficient means to reliably identify individuals at risk for CVD by predicting serum levels of NT-proBNP from the ECG.

Methods: A deep convolutional neural network was developed using the population-based cohort study Hamburg City Health Study (HCHS, n=8,253, 50.9 % women). External validation was performed in two independent population-based cohorts (SHIP-START, n=3,002, 52.1 % women, and SHIP-TREND, n=3,819, 51.2 % women). Assessment of model performance was conducted using Pearson correlation (R) and area under the receiver operating characteristics curve (AUROC).

Results: NT-proBNP was predictable from the ECG (R, 0.566 [HCHS], 0.642 [SHIP-START-0], 0.655 [SHIP-TREND-0]). Across

cohorts, predicted NT-proBNP (pNT-proBNP) showed good discriminatory ability for prevalent and incident heart failure (HF) (baseline: AUROC 0.795 [HCHS], 0.816 [SHIP-START-0], 0.783 [SHIP-TREND-0]; first follow-up: 0.669 [SHIP-START-1, 5 years], 0.689 [SHIP-TREND-1, 7.3 years]), comparable to the discriminatory value of measured NT-proBNP. pNT-proBNP also demonstrated comparable results for other incident CVD, including atrial fibrillation, stroke, myocardial infarction, and cardiovascular death.

Conclusions: Deep learning ECG algorithms can predict NT-proBNP concentrations with high diagnostic and predictive value for HF and other major CVD and may be used in the community to identify individuals at risk. Long-standing experience with NT-proBNP can increase acceptance of such deep learning models in clinical practice.

Keywords: artificial intelligence; biomarker; cardiovascular risk; general population; natriuretic peptide

Introduction

The timely identification of individuals in the general population at risk for cardiovascular diseases (CVD) allows for an early medical intervention and thus a potentially better

Meraj Neyazi and Jan P. Bremer contributed equally to this work and share first authorship.

***Corresponding author: Renate B. Schnabel**, Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, 20251 Hamburg, Germany; and German Centre for Cardiovascular Research (DZHK), Partner Site Hamburg/Kiel/Luebeck, Hamburg, Germany, Phone: +49 (0)40 7410 53979, Fax: +49 40 7410 53622, E-mail: r.schnabel@uke.de

Meraj Neyazi, Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; German Centre for Cardiovascular Research (DZHK), Partner Site Hamburg/Kiel/Luebeck, Hamburg, Germany; and Department of Genetics, Harvard Medical School, Boston, MA, USA

Jan P. Bremer, Marius S. Knorr, Dora Csengeri, Benedikt Schrage, Tanja Zeller and Stefan Blankenberg, Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; and German Centre for

Cardiovascular Research (DZHK), Partner Site Hamburg/Kiel/Luebeck, Hamburg, Germany

Stefan Gross, Martin Bahls, Stephan Felix and Marcus Dörr, Department of Internal Medicine B, University Medicine Greifswald, Greifswald, Germany; and German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, Germany

Jan Bredericke, Department of Cardiology, University Heart & Vascular Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Nils Schweingruber, Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Nele Friedrich, Institute for Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany; and German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, Germany

Marcus Vollmer, Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany; and German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, Germany

outcome [1, 2]. Yet, this can be difficult in outpatient settings, potentially due to less severe and unspecific symptoms at early stages [3].

The strong cardiovascular biomarkers Brain natriuretic peptide (BNP) and its N-terminal prohormone (NT-proBNP) alone can help identify individuals with heart failure in primary care [4] and represent one of the pillars for heart failure diagnosis and management [5]. This hormone is predominantly secreted by cardiomyocytes upon stimuli including mechanical stretch. Its biologic functions include regulation of blood pressure, blood volume and electrolyte homeostasis. Elevated serum levels of BNP and NT-proBNP are strongly correlated with multiple prevalent and incident cardiovascular diseases such as heart failure (HF), atrial fibrillation (AF), stroke, myocardial infarction (MI), and death [6–10]. Importantly, these associations have been validated extensively in patient cohorts and in the general population [11, 12]. However, only a fraction of patients with heart failure receives a BNP or NT-proBNP measurement before their diagnosis [13], despite this evidence and the low barrier for determining natriuretic peptide levels through point-of-care measurements [14]. Development and application of additional low-barrier methods for risk assessment may improve timely CVD diagnosis. One possible way of achieving this may lie in the application of deep learning (DL) algorithms on the widely available electrocardiogram (ECG). The ECG, a cheap and non-invasive routine diagnostic procedure, can be automatically and quickly analysed by DL models to infer medically relevant information. Previous studies have shown that ECG-based DL models can detect pathological states in populations at risk such as valve defects [15], AF in patients with sinus rhythm [16], or cardiac dysfunction [17], among others. However, their applicability in the general population is less effective [18].

We hypothesized that CV risk assessment in the general population with the benefits of ECG-DL models can be achieved by training a DL model to predict NT-proBNP serum levels from the ECG. By doing so, we can utilize a larger number of datapoints for model training and additionally build on existing evidence for this strong cardiovascular biomarker.

To test our hypothesis, we trained a DL model, AI-NT-proBNP, to predict NT-proBNP serum levels from the ECG using data from an urban population in Germany, and consistently tested generalizability in two external independent German population-based cohorts. We evaluated the correlation of predicted NT-proBNP serum levels (pNT-proBNP) to lab-measured NT-proBNP levels (mNT-proBNP) from two manufacturers and compared the predictive ability of pNT-proBNP to mNT-proBNP and a clinical

risk factor-based model for prevalent and incident diseases, including heart failure, atrial fibrillation, stroke, and cardiovascular death. To our knowledge, this study is the first to harness blood biomarker information for an ECG-DL model, and it shows that a model like AI-NT-proBNP can provide diagnostic value for cardiovascular risk assessment in the population comparable to NT-proBNP.

Materials and methods

Study cohorts

The Hamburg City Health Study (HCHS) is an ongoing population-based cohort study enrolling randomly selected individuals between 45 and 74 years from the Hamburg metropolitan area in Germany. For this study, the data release of the first 10,000 subjects enrolled between 2016 and 2019 was used. All participants received a baseline study examination following the published HCHS protocol [19]. External validation and prospective assessment were performed in the Study of Health in Pomerania (SHIP), a population-based study based in Western Pomerania, Germany, which includes multiple cohorts and follow-up examinations [20]. We utilized data from the two independent cohorts SHIP-START and SHIP-TREND. For SHIP-START, a total of 6,265 individuals aged from 20 to 79 years were recruited after a two-step stratified random sampling procedure between 1997 and 2001. Out of these, 4,308 received baseline examinations (SHIP-START-0). Follow-up examinations were conducted between 2002 and 2006 among 3,300 participants (SHIP-START-1), 2008 and 2012 among 2,333 participants (SHIP-START-2) and between 2014 and 2016 among 1,718 participants (SHIP-START-3). The median follow-up time from baseline was 5.0, 10.8 and 15.8 years for each cohort. The second SHIP cohort, SHIP-TREND, was initiated in 2008 and included a randomly selected independent group of individuals from Western Pomerania. Baseline examinations were conducted between 2008 and 2011 on 4,420 individuals (SHIP-TREND-0). A follow-up was performed between 2016 and 2019 on a sample of 2,507 individuals (SHIP-TREND-1) resulting in a median follow-up time of 7.3 years. All SHIP subjects received examinations according to the SHIP protocol [20].

Participants of all cohorts received 12-lead ECG recordings (HCHS: CARDIOVIT CS-200 Excellence, Schiller Medizintechnik GmbH, Germany; SHIP-START-0 to -1: Personal-120-LT, Esaote Biomedica, Genova, Italy; SHIP-START-2 to -3 and SHIP-TREND-0 to -1: Welch Allyn SE-PRO-600 Cardioperfect Pro Recorder/Hill-Rom Holdings, Batesville, Indiana, USA). Medical history and risk factor information were derived from questionnaires or computer-assisted interviews. Furthermore, all examined participants received blood draws and laboratory assessments for routine measures including the measurement of NT-proBNP concentrations. Samples from HCHS were measured with Alere NT-proBNP for ARCHITECT (Abbott Diagnostics, Wiesbaden, Germany). NT-proBNP concentrations from all SHIP cohorts were measured using Siemens VISTA (Siemens Healthcare Diagnostics, Eschborn, Germany) [21].

The studies were conducted in accordance with the Helsinki Declaration. The local Ethics Committee of the State of Hamburg Chamber of Medical Practitioners (PV5131) and the data protection commissioners of the University Medical Center Hamburg-Eppendorf

and of the Free and Hanseatic City of Hamburg approved the HCHS. The Ethics Committee at the University Medicine Greifswald approved the SHIP studies (approval number BB 39/08). All participants provided written, informed consent.

Outcomes

Prevalent and incident heart failure, stroke, MI, and cardiovascular death were assessed by echocardiography, questionnaires or computer-assisted interviews encompassing medical history and prior medication, laboratory analysis of blood samples, ECGs, and death certificates (see Supplementary Methods).

Model development and evaluation

For model development, we used a 1-dimensional ResNet convolutional neural network [22] to predict serum-level NT-proBNP (pNT-proBNP). The 12-channel ECGs, which were recorded with a sample rate of 250 Hz, were pre-processed utilizing a powerline and 5th-order Butterworth filter.

During model training, 3-min and 10-s ECGs were used. 10-s ECGs were used during inference on test datasets.

The model was trained on resting state ECGs from the HCHS to predict logarithmized NT-proBNP values. ECGs ($n=8,256$) were randomly split into 66 % development ($n=5,416$) and 34 % internal testing data ($n=2,840$). The development data was deployed in a 5-fold cross validation scheme, resulting in 5-final models. For the final prediction on test datasets the mean of these 5 models was calculated and subsequently used.

In order to build a more robust model, the ECG data was randomly augmented during training by dropout augmentation (setting one ECG channel to zero), baseline-shifts, voltage-scaling and generation of random starting points by randomly cropping approximately 8 s of every ECG, which translates to 2,048 consecutive samples within each recording [23].

Statistical analysis

To assess the performance of pNT-proBNP, we calculated Pearson's correlation coefficient (R) with NT-proBNP laboratory measurements (mNT-proBNP) on internal and external validation datasets. For comparison, two baseline linear regression models (Risk Factor Model) were fitted to predict the log-transformed NT-proBNP in the same 5-fold cross validation scheme: RF-ModelLin-1 was fitted using the characteristics age and sex only, while RF-ModelLin-2 was fitted with the addition of body mass index, hypertension, diabetes mellitus, smoking status, LDL-cholesterol concentrations, AF and CKD-epi estimated glomerular filtration rate (eGFR) [24]. The correlation coefficient was bootstrapped for 1,000 iterations in order to determine confidence intervals (CI) and to perform z-tests. For the evaluation of the diagnostic and prognostic value of mNT-proBNP and pNT-proBNP for adverse cardiovascular events, comprising incident heart failure, AF, MI, stroke, and cardiovascular death, five logistic regression models were computed: 1. pNT-proBNP alone as predictor, 2. mNT-proBNP alone as predictor, 3. pNT-proBNP as predictor adjusted for risk factors, 4. mNT-proBNP as predictor adjusted for risk factors, and 5. risk factors (age, sex, body mass index, hypertension, diabetes, smoking status, LDL-cholesterol, AF, and eGFR) without NT-proBNP as predictors (RF-ModelLog = reference model). AF as a risk factor was excluded when

adjusting models for the evaluation of diagnostic value for AF. Adjustment of pNT-proBNP and mNT-proBNP for risk factors was performed by generating regression models that included pNT-proBNP or mNT-proBNP and the above-mentioned risk factors. These logistic regression models were bootstrapped using random stratified sampling with replacement (drawing from positive and negative classes separately to retain the rate of incident events) and a stratified 20 % test split. The performance was assessed using the area under the receiver operating characteristic (AUROC). The accumulated AUROCs for different logistic regression models were tested for statistical significance by z-tests. Brier scores for each classification were determined. Additionally, net reclassification improvement (NRI) analysis was performed to compare pNT-proBNP to mNT-proBNP using a cut-off of 0.1 for risk categories. This analysis is in line with previously published statistical analyses on DL models [25] and the recommendations of the DZHK project group AI/ML [26]. Patients with cancer at baseline were excluded from the analysis. Also, patients with missing baseline variables were removed from further analysis. Model training was performed in Python 3 [27] using PyTorch [28] and torchvision [29]. All statistical analysis was performed using Python 3 and SciPy [30]. Plots were generated using Matplotlib [31].

All statistical analyses to reproduce the results reported in this paper will become open-source and available under MIT license at: <https://github.com/JanBrem/AI-NT-proBNP>.

Results

Study cohorts

The model AI-NT-proBNP was developed using resting ECGs from the HCHS. A total of 8,253 ECGs with corresponding NT-proBNP serum concentrations were available. The mean age of these individuals was 62.8 ± 8.6 years (50.9 % women). The median NT-proBNP serum concentration was 81.0 pg/mL. AI-NT-proBNP was validated on the SHIP-START and SHIP-TREND cohorts. For SHIP-START-0, we analysed a total of 3,002 subjects, with a mean age of 48.7 ± 16.4 years at baseline (52.1 % women). From SHIP-TREND-0, data of 3,819 individuals with a mean age 50.7 ± 15 years (51.2 % women) were analysed. Data from the follow-up studies were included to assess the diagnostic value of pNT-proBNP. Figure 1 provides a graphical overview of the analysis and of the study cohorts. Baseline characteristics for all participants included in this study are presented in Table 1.

Prediction of NT-proBNP serum concentrations from ECGs using AI-NT-proBNP

The performance of AI-NT-proBNP to predict pNT-proBNP from the ECG in HCHS is shown in Figure 2 and Table 2. On log-scale, pNT-proBNP values showed a decent correlation with mNT-proBNP concentrations ($R=0.566$, 95 % CI 0.563–

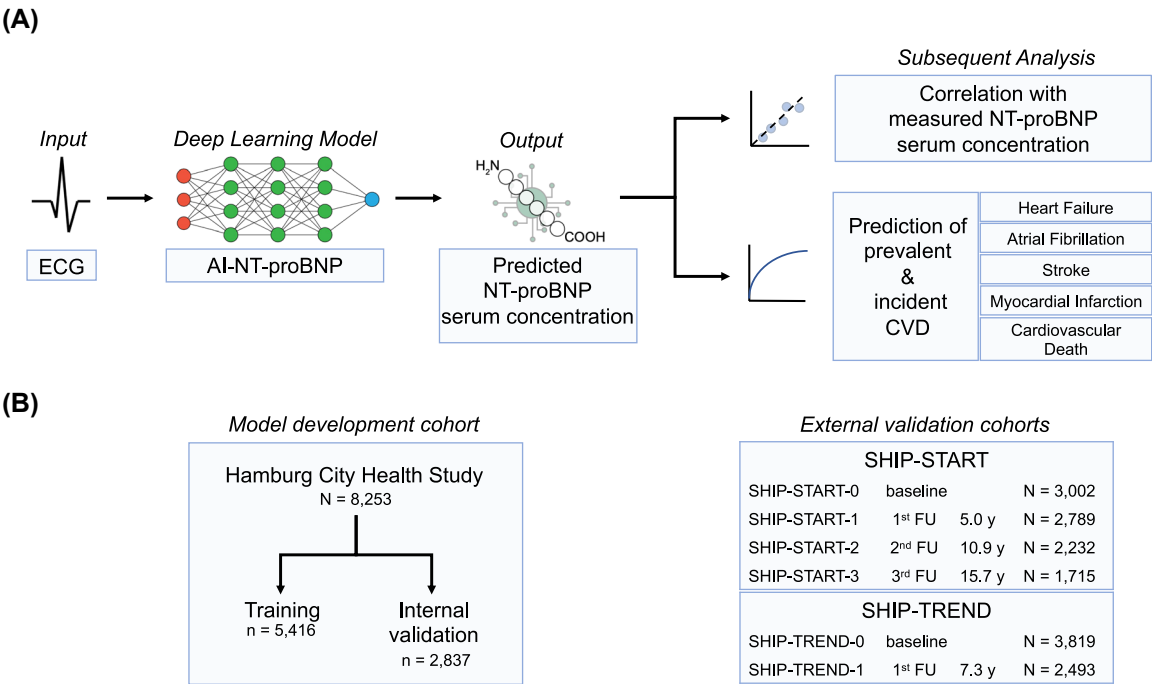


Figure 1: Study overview. (A) Analysis workflow. After training of the model, ECGs of study participants were analysed using AI-NT-proBNP, resulting in predicted NT-proBNP serum concentrations for each participant. Predicted NT-proBNP concentrations were correlated with lab-measured NT-proBNP values. Also, predicted NT-proBNP concentrations were analysed in regards to their diagnostic value for predicting incident and prevalent CVD, including heart failure, atrial fibrillation, stroke, myocardial infarction, and cardiovascular death. (B) Cohort overview. The population-based HCHS (n=8,253) served as model development cohort. The cohort was split into a training (n=5,416) and an internal validation set (n=2,837). External validation was performed on data from the SHIP-START and SHIP-TREND cohorts, which are also population-based. Both cohorts had FUs after baseline examinations. For each FU, the respective median FU time in years and the number of participants are presented. CVD, cardiovascular disease; HCHS, Hamburg City Health Study; FU, follow-up.

Table 1: Baseline characteristics by study cohort.

Variables	HCHS	SHIP-START-0	SHIP-START-1	SHIP-START-2	SHIP-START-3	SHIP-TREND-0	SHIP-TREND-1
n	8,253	3,002	2,789	2,232	1,715	3,819	2,493
Follow-up	Baseline	Baseline	1st	2nd	3rd	Baseline	1st
Follow-up time, years	–	–	5	10.9	15.7	–	7.3
Age, years	62.8 ± 8.6	48.7 ± 16.4	53.5 ± 15.1	57.5 ± 13.6	60.1 ± 12.8	50.7 ± 15.0	57.1 ± 13.9
Sex (women), %	50.9	52.1	52.1	53.3	53.7	51.2	51.7
NT-proBNP pg/mL	81 (43, 142)	54 (28, 107)	65 (33, 136)	72 (37, 143)	–	72 (29, 112)	–
Body mass index, kg/m ²	26.0 ± 4.7	27.2 ± 4.8	28.0 ± 4.9	28.3 ± 4.9	28.1 ± 5.0	28.0 ± 5.2	28.0 ± 4.9
Hypertension, %	66.1	48.8	57.3	64.2	67.5	55.1	56.9
Diabetes, %	8.2	9.2	12.1	10.3	11.9	11.3	9.7
Smoking status (never), %	19.9	35.9	41.7	49	57.3	51.5	56.2
eGFR, mL/min/1.73 m ²	79 ± 13	101 ± 18	96 ± 19	89 ± 18	88 ± 17	96 ± 16.9	88 ± 17
LDL-cholesterol, mmol/L	3.13 ± 0.96	3.32 ± 0.91	3.16 ± 0.88	3.34 ± 0.96	3.32 ± 0.93	3.34 ± 0.96	3.34 ± 0.96
Heart failure, %	4.4	2.9	3	3.3	3.9	2.6	3.4
Atrial fibrillation, %	5.6	1.1	1.9	2.6	2.4	1.2	1.9
History of myocardial infarction, %	2.8	3.1	4.3	4.1	3.9	2.3	3.3
History of stroke, %	3.4	1.9	1.1	1.5	3.2	1.9	3

Continuous variables are summarized with mean and standard deviation. For LDL cholesterol and NT-proBNP median and quartiles are reported. For follow-up times median is reported. eGFR, estimated glomerular filtration rate; HCHS, Hamburg City Health Study; LDL, low-density lipoprotein; NT-proBNP, N-terminal prohormone of brain natriuretic peptide; SHIP, Study of Health in Pomerania.

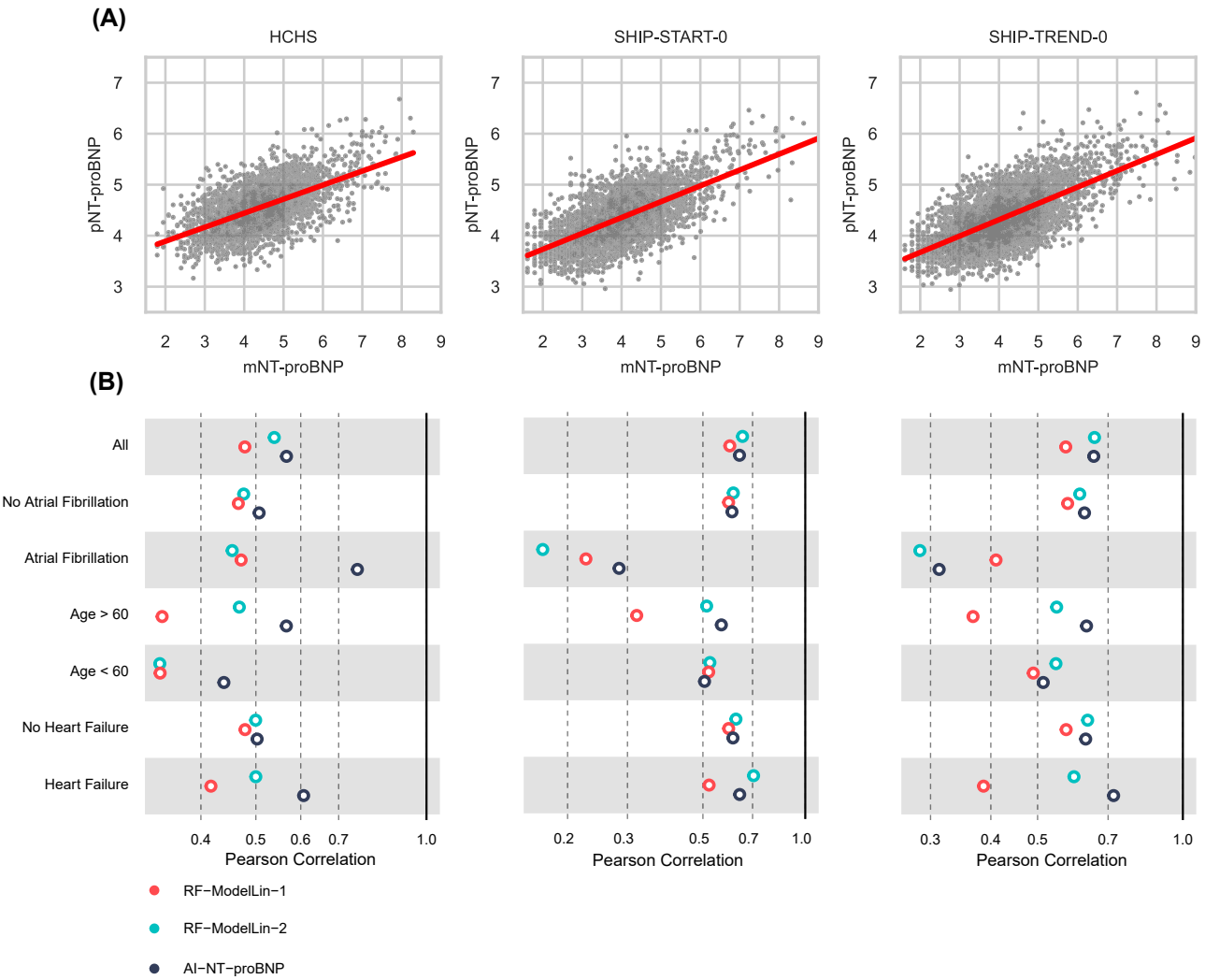


Figure 2: Correlation of different NT-proBNP prediction models with laboratory mNT-proBNP. (A) Scatter plots for correlations between pNT-proBNP and mNT-proBNP at baseline (HCHS, START, TREND). Concentrations are printed on logarithmic scale. (B) Pearson correlation coefficients of mNT-proBNP and pNT-proBNP for three different models (RF-ModelLin-1 (red), RF-ModelLin-2 (turquoise), AI-NT-proBNP (purple)) for all study participants and for different subcohorts. HCHS, Hamburg City Health Study; pNT-proBNP, predicted N-terminal prohormone of brain natriuretic peptide; mNT-proBNP, measured N-terminal prohormone of brain natriuretic peptide.

Table 2: Correlation of different NT-proBNP prediction models with laboratory mNT-proBNP.

Cohort	AI-NT-proBNP	RF-ModelLin-1 R (95 % CI)	RF-ModelLin-2
HCHS	0.566 ^{a,b} (0.563, 0.569)	0.478 (0.474, 0.481)	0.538 (0.535, 0.541)
SHIP-START-0	0.642 ^{a,b} (0.640, 0.644)	0.600 (0.598, 0.602)	0.654 (0.652, 0.656)
SHIP-START-1	0.681 ^{a,b} (0.679, 0.683)	0.582 (0.579, 0.584)	0.673 (0.671, 0.675)
SHIP-START-2	0.652 ^{a,b} (0.650, 0.655)	0.594 (0.592, 0.597)	0.688 (0.686, 0.670)
SHIP-TREND-0	0.655 ^{a,b} (0.653, 0.656)	0.574 (0.572, 0.575)	0.658 (0.656, 0.659)

Provided are Pearson's correlation coefficients (R) and 95 % bootstrapped confidence intervals. mNT-proBNP is correlated either with pNT-proBNP, predicted NT-proBNP levels from a linear regression model including age and sex (RF-ModelLin-1), or predicted NT-proBNP levels from a linear regression model including age, sex, BMI, hypertension, diabetes, smoking status, LDL-cholesterol, atrial fibrillation, and eGFR (RF-ModelLin-2). Models were fitted in the individual data sets. The bootstrapped R values of AI-NT-proBNP were tested for significance using z-tests in comparison to RF-ModelLin-1 and RF-ModelLin-2. ^ap<0.05 in comparison to RF-ModelLin-1. ^bp<0.05 in comparison to RF-ModelLin-2.

0.569). Multivariable linear regression models using the variables age and sex (RF-ModelLin-1) or the additional risk factors hypertension, diabetes mellitus, body mass index (BMI), LDL-cholesterol levels and smoking status (RF-ModelLin-2) yielded weaker correlations ($R=0.478$, 95 % CI 0.474–0.481 and $R=0.538$, 95 % CI 0.535–0.541) compared to AI-NT-proBNP.

The performance of AI-NT-proBNP generalized well in predicting NT-proBNP serum concentrations in the external cohorts SHIP-START-0 ($R=0.642$, 95 % CI 0.640–0.644) and SHIP-TREND-0 ($R=0.655$, 95 % CI 0.653–0.656). The generalizability remained good in the follow-up studies after five years in SHIP-START-1 ($R=0.681$, 95 % CI 0.679–

0.683) and ten years (median 10.8 years) in SHIP-START-2 ($R=0.652$, 95 % CI 0.650–0.655). In these cohorts, too, AI-NT-proBNP's performance was compared to the RF-ModelLin-1 and with RF-ModelLin-2. pNT-proBNP consistently yielded a stronger correlation with mNT-proBNP compared to RF-ModelLin-1, while RF-ModelLin-2's correlation strength came closer to pNT-proBNP, partially surpassing it. Given that NT-proBNP was measured using systems from different manufactures in HCHS (Abbott Diagnostics) and the SHIP (Siemens Healthcare Diagnostics) cohorts, this data additionally underlines that pNT-proBNP correlates well across these two systems.

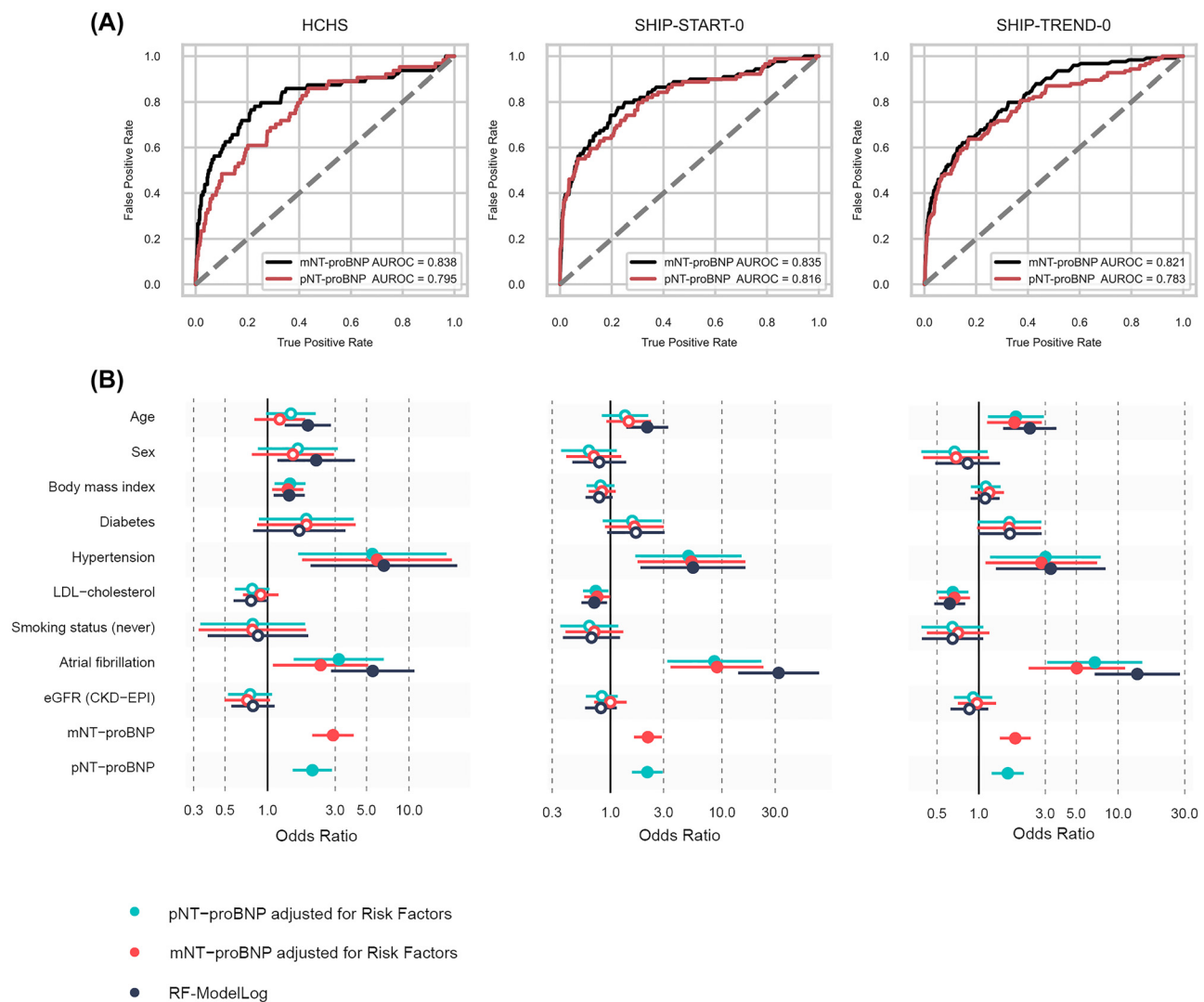


Figure 3: Classification of prevalent heart failure. (A) AUROC curves for the prediction of prevalent heart failure using unadjusted logistic regression models pNT-proBNP and mNT-proBNP in baseline cohorts (HCHS, START, TREND). (B) Odds ratios for three risk factor adjusted logistic regression models (pNT-proBNP adjusted for risk factors (turquoise), mNT-proBNP adjusted for risk factors (red), RF-ModelLog (purple)) predicting prevalent heart failure at baseline cohorts. HCHS, Hamburg City Health Study; pNT-proBNP, predicted N-terminal prohormone of brain natriuretic peptide; mNT-proBNP, measured N-terminal prohormone of brain natriuretic peptide; AUROC, area under receiver operating characteristic; LDL, low density lipoprotein; eGFR, estimated glomerular filtration rate; CKD-EPI, chronic kidney disease epidemiology collaboration.

Table 3: Area under the curve information for classification of prevalent heart failure.

Cohort	pNT-proBNP crude	mNT-proBNP crude	pNT-proBNP adjusted AUROC (95 % CI)	mNT-proBNP adjusted	RF-ModelLog	Number of participants with heart failure/total cohort
HCHS	0.795 (0.793, 0.796)	0.838 (0.836, 0.839)	0.843 ^{a,b} (0.838, 0.847)	0.851 ^b (0.846, 0.856)	0.823 (0.818, 0.828)	64/1,682
SHIP-START-0	0.816 (0.814, 0.818)	0.835 (0.833, 0.836)	0.866 ^b (0.862, 0.869)	0.865 ^b (0.861, 0.869)	0.844 (0.840, 0.848)	82/2,806
SHIP-START-1	0.821 (0.819, 0.8223)	0.842 (0.841, 0.843)	0.849 ^{a,b} (0.845, 0.853)	0.860 ^b (0.856, 0.864)	0.831 (0.826, 0.835)	77/2,524
SHIP-START-2	0.838 (0.837, 0.840)	0.856 (0.855, 0.858)	0.860 ^b (0.855, 0.865)	0.864 ^b (0.859, 0.870)	0.846 (0.841, 0.851)	57/2,013
SHIP-TREND-0	0.783 (0.781, 0.785)	0.821 (0.820, 0.823)	0.873 ^{a,b} (0.870, 0.876)	0.883 ^b (0.880, 0.886)	0.869 (0.866, 0.872)	86/3,722

Provided are AUROC values for the diagnosis of prevalent heart failure with 95 % bootstrapped confidence intervals. Baseline characteristics in the adjusted models are: age, sex, BMI, hypertension, diabetes, smoking status, LDL-cholesterol, atrial fibrillation, and eGFR. The bootstrapped AUROC values of pNT-proBNP adjusted were tested for significance using z-test in comparison to mNT-proBNP adjusted and RF-ModelLog. The bootstrapped AUROC values of mNT-proBNP adjusted were tested for significance using z-test in comparison to RF-ModelLog. ^ap<0.05 in comparison to mNT-proBNP adjusted. ^bp<0.05 in comparison to RF-ModelLog. AUROC, area under the receiver operating characteristics; BMI, body mass index; CI, confidence interval; eGFR, estimated glomerular filtration rate; LDL, low-density lipoprotein; mNT-proBNP, measured N-terminal prohormone of brain natriuretic peptide; pNT-proBNP, predicted N-terminal prohormone of brain natriuretic peptide.

pNT-proBNP and prevalent heart failure

To evaluate the predictive value of pNT-proBNP for identifying individuals with prevalent heart failure, we compared the bootstrapped AUROCs for heart failure classification with logistic regression models. These included crude models for pNT-proBNP or mNT-proBNP, respectively, risk factor adjusted models for pNT-proBNP or mNT-proBNP, and a model for the risk factors alone (RF-ModelLog). Each comparison was performed in each cohort with baseline and follow-up studies (Figure 3, Table 3). Brier scores and NRI were also calculated to investigate calibration and reclassification (Supplementary Table 1, Supplementary Table 3).

In these cohorts and follow up studies, mNT-proBNP alone consistently provided excellent discriminatory properties for prevalent heart failure, consistent with published data [6–8]. pNT-proBNP demonstrated similar properties in HCHS and in both external validation cohorts and their follow-up studies.

When adjusting for risk factors, the AUROCs of models with pNT-proBNP and mNT-proBNP consistently increased. Both models' performances were superior compared to the reference model RF-ModelLog in all studies. The AUROC of mNT-proBNP adjusted for risk factors was slightly higher compared to pNT-proBNP adjusted for risk factors in HCHS, SHIP-START-1 and SHIP-TREND-0. There was no statistically significant difference in the AUROCs between risk factor-adjusted models with pNT-proBNP and mNT-proBNP in

SHIP-START-0 and SHIP-START-2. Compared to RF-ModelLog, neither the addition of pNT-proBNP, nor mNT-proBNP provided a significant positive or negative effect on reclassification. Brier scores present comparable accuracy for adjusted models of mNT-proBNP and pNT-proBNP, as well as RF-ModelLog across cohorts.

pNT-proBNP and incident cardiovascular events

To assess the predictive value of pNT-proBNP for incident cardiovascular events, we followed a similar approach (Table 4, Supplementary Table 2, Supplementary Table 3).

The risk factor-adjusted model including pNT-proBNP showed comparable AUROCs for incident heart failure prediction compared to a risk factor-adjusted model with mNT-proBNP in all cohorts, being similar in SHIP-START-1 and SHIP-TREND-1, lower in SHIP-START-2 and higher in SHIP-START-3. It was superior to RF-ModelLog in SHIP-START-3, while the risk factor-adjusted model with mNT-proBNP was superior to RF-ModelLog in SHIP-START-2 (Figure 4). Unadjusted models with pNT-proBNP present predictive values comparable to unadjusted models with mNT-proBNP across all follow-ups and cohorts.

The prediction of incident stroke with a risk factor adjusted model including pNT-proBNP was superior compared to a model including mNT-proBNP in participants in SHIP-START-1 and SHIP-START-2, yet similar in

Table 4: Area under the curve information for classification of incident cardiovascular diseases.

Variable	Cohort	pNT-proBNP crude	mNT-proBNP crude	pNT-proBNP adjusted	mNT-proBNP adjusted	RF-ModelLog	Events/total number of participants
AUROC (95 % confidence interval)							
Heart failure	SHIP-START-1	0.669 (0.666, 0.672)	0.649 (0.646, 0.652)	0.741 (0.734, 0.748)	0.742 (0.735, 0.750)	0.732 (0.725, 0.739)	38/2,134
	SHIP-START-2	0.709 (0.707, 0.711)	0.686 (0.684, 0.688)	0.824 ^a (0.819, 0.829)	0.828 ^b (0.823, 0.833)	0.815 (0.810, 0.820)	68/1,565
	SHIP-START-3	0.708 (0.706, 0.710)	0.664 (0.662, 0.666)	0.835 ^{a,b} (0.831, 0.840)	0.830 (0.826, 0.834)	0.83 (0.826, 0.834)	98/1,123
	SHIP-TREND-1	0.689 (0.687, 0.691)	0.705 (0.703, 0.707)	0.848 (0.844, 0.853)	0.845 (0.84, 0.85)	0.846 (0.842, 0.851)	52/2,197
Stroke	SHIP-START-1	0.766 (0.763, 0.769)	0.642 (0.638, 0.646)	0.732 ^{a,b} (0.723, 0.740)	0.698 ^b (0.689, 0.707)	0.71 (0.702, 0.720)	21/2,255
	SHIP-START-2	0.74 (0.737, 0.742)	0.652 (0.649, 0.655)	0.800 ^{a,b} (0.793, 0.805)	0.788 (0.782, 0.795)	0.792 (0.785, 0.798)	36/1,650
	SHIP-START-3	0.715 (0.712, 0.717)	0.693 (0.691, 0.696)	0.806 (0.800, 0.812)	0.809 (0.803, 0.814)	0.807 (0.802, 0.813)	53/1,176
	SHIP-TREND-1	0.648 (0.646, 0.651)	0.677 (0.674, 0.679)	0.703 ^a (0.696, 0.710)	0.720 ^b (0.713, 0.727)	0.699 (0.691, 0.706)	42/2,221
Atrial fibrillation	SHIP-START-1	0.681 (0.677, 0.686)	0.782 (0.778, 0.785)	0.814 ^a (0.807, 0.821)	0.838 ^b (0.831, 0.845)	0.813 (0.807, 0.820)	23/2,078
	SHIP-START-2	0.75 (0.748, 0.753)	0.813 (0.811, 0.815)	0.854 ^{a,b} (0.848, 0.859)	0.869 ^b (0.864, 0.875)	0.849 (0.843, 0.854)	46/1,577
	SHIP-START-3	0.763 (0.760, 0.765)	0.82 (0.818, 0.821)	0.881 ^{a,b} (0.877, 0.885)	0.896 ^b (0.892, 0.900)	0.876 (0.871, 0.880)	58/1,192
	SHIP-TREND-1	0.659 (0.656, 0.661)	0.785 (0.782, 0.788)	0.858 ^a (0.853, 0.863)	0.870 ^b (0.864, 0.876)	0.859 (0.854, 0.864)	24/2,237
Myocardial infarction	SHIP-START-1	0.651 (0.648, 0.655)	0.58 (0.676, 0.683)	0.756 (0.748, 0.763)	0.745 (0.738, 0.753)	0.76 (0.753, 0.767)	24/2,134
	SHIP-START-2	0.614 (0.611, 0.617)	0.579 (0.577, 0.582)	0.798 (0.793, 0.803)	0.792 (0.787, 0.797)	0.799 (0.794, 0.804)	46/1,553
	SHIP-START-3	0.625 (0.623, 0.627)	0.577 (0.575, 0.579)	0.783 (0.778, 0.788)	0.781 (0.777, 0.786)	0.786 (0.781, 0.791)	67/1,175
	SHIP-TREND-1	0.665 (0.662, 0.667)	0.623 (0.620, 0.626)	0.788 ^{a,b} (0.781, 0.794)	0.776 (0.769, 0.782)	0.780 (0.774, 0.786)	36/2,212
Cardiovascular death	SHIP-START-1	0.829 (0.826, 0.832)	0.9 (0.898, 0.902)	0.937 (0.933, 0.940)	0.945 (0.942, 0.948)	0.939 (0.936, 0.942)	25/2,820
	SHIP-START-2	0.792 (0.791, 0.794)	0.837 (0.836, 0.838)	0.895 (0.892, 0.897)	0.9 (0.898, 0.903)	0.893 (0.891, 0.896)	82/2,820
	SHIP-START-3	0.742 (0.741, 0.743)	0.8 (0.800, 0.801)	0.881 (0.879, 0.883)	0.889 (0.887, 0.891)	0.88 (0.878, 0.882)	177/2,820
	SHIP-TREND-1	0.665 (0.663, 0.668)	0.778 (0.776, 0.800)	0.821 ^{a,b} (0.816, 0.826)	0.858 ^b (0.854, 0.863)	0.826 (0.821, 0.831)	54/3,788

Provided are AUROC values for the classification of incident cardiovascular diseases with 95 % bootstrapped confidence intervals. Baseline characteristics in the adjusted models are: age, sex, BMI, hypertension, diabetes, smoking status, LDL-cholesterol, atrial fibrillation, and eGFR. The bootstrapped AUROC values of pNT-proBNP adjusted were tested for significance using z-test in comparison to mNT-proBNP adjusted and RF-ModelLog. The bootstrapped AUROC values of mNT-proBNP adjusted were tested for significance using z-test in comparison to RF-ModelLog. ^ap<0.05 in comparison to mNT-proBNP adjusted. ^bp<0.05 in comparison to RF-ModelLog. AUROC, area under the receiver operating characteristics; BMI, body mass index; eGFR, estimated glomerular filtration rate; LDL, low-density lipoprotein; mNT-proBNP, measured N-terminal prohormone of brain natriuretic peptide; pNT-proBNP, predicted N-terminal prohormone of brain natriuretic peptide.

SHIP-START-3 and inferior in SHIP-TREND-1. A difference between a risk factor adjusted model including pNT-proBNP and RF-ModelLog was observed in SHIP-START-1 and SHIP-START-2.

For AF, a risk factor-adjusted model including pNT-proBNP showed inferior predictive value in all cohorts compared to a risk-factor adjusted model including mNT-proBNP, but superior to RF-ModelLog in SHIP-START-2 and SHIP-START-3. A

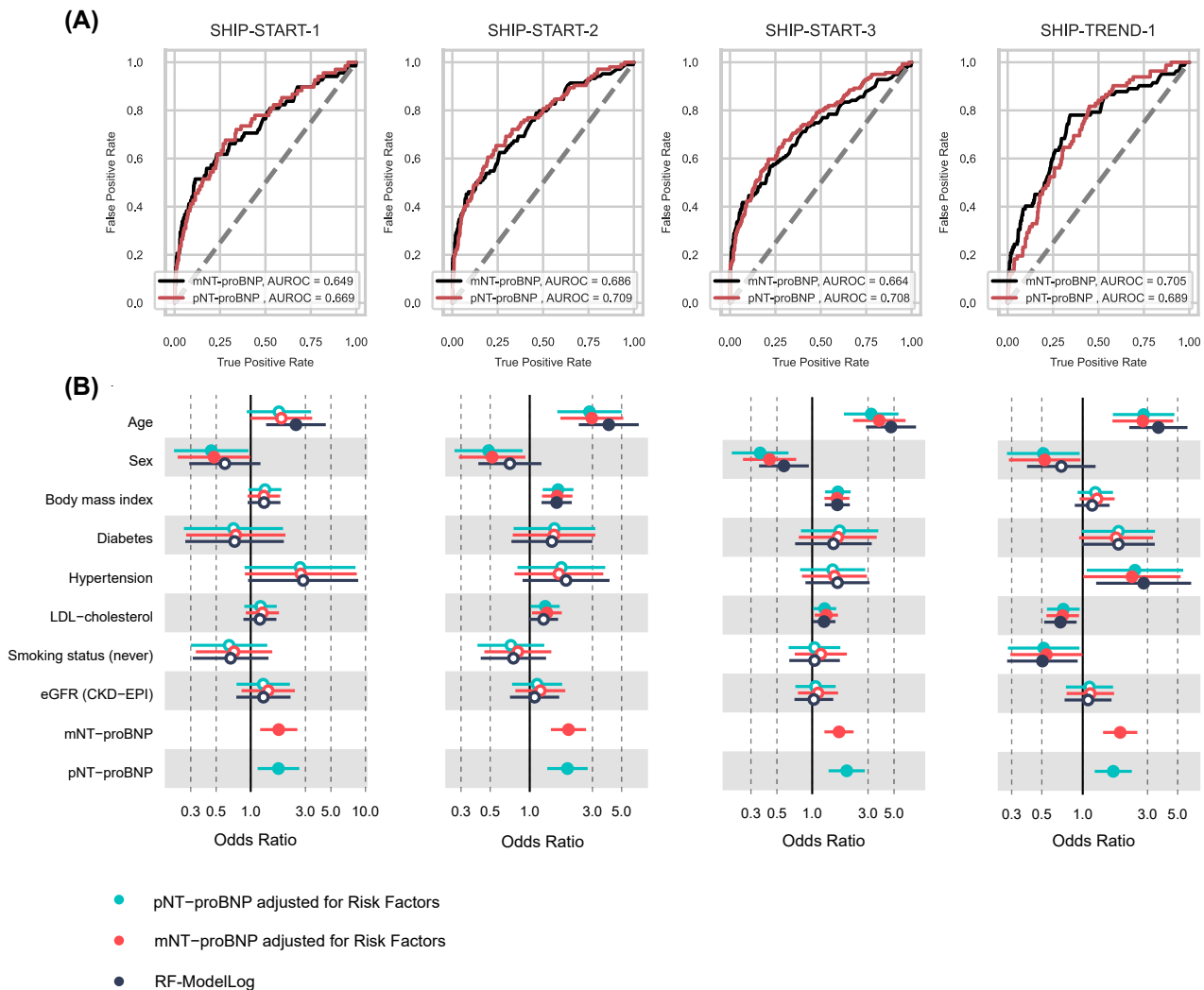


Figure 4: Classification of incident heart failure. (A) AUROC curves for the prediction of incident heart failure using unadjusted logistic regression models of pNT-proBNP or mNT-proBNP in follow-up cohorts. (B) Odds ratios for three risk factor adjusted logistic regression models (pNT-proBNP adjusted for risk factors (turquoise), mNT-proBNP adjusted for risk factors (red), RF-ModelLog (purple)) predicting incident heart failure at follow-up cohorts. AUROC, area under receiver operating characteristic; LDL, low density lipoprotein; eGFR, estimated glomerular filtration rate; CKD-EPI, chronic kidney disease epidemiology collaboration; pNT-proBNP, predicted N-terminal prohormone of brain natriuretic peptide; mNT-proBNP, measured N-terminal pro-hormone of brain natriuretic peptide.

risk factor adjusted model with mNT-proBNP had a greater AUROC compared to RF-ModelLog in all cohorts.

Neither risk factor-adjusted models including pNT-proBNP, nor including mNT-proBNP showed differing results compared to the reference RF-ModelLog in predicting incident MI or cardiovascular death in all SHIP-START follow-up studies. In SHIP-TREND-1, a risk factor-adjusted model with pNT-proBNP showed significantly better results in predicting MI compared to both a risk factor-adjusted model with mNT-proBNP or RF-ModelLog, while the AUROCs of the latter two were not significantly different. In the same study, a risk-factor adjusted model with pNT-proBNP showed a lower AUROC for cardiovascular death prediction compared to both an adjusted model

with mNT-proBNP or RF-ModelLog. NRI revealed that neither mNT-proBNP, nor pNT-proBNP yielded a significant effect on reclassification compared to RF-ModelLog for any incident event. Accuracy was comparable across cohorts and cardiovascular events.

Discussion

We developed a DL model (AI-NT-proBNP) across community cohorts to predict serum concentrations of the cardiovascular biomarker NT-proBNP from the surface ECG. It was validated in two external population-based cohorts with

robust generalization of results. The assessment of pNT-proBNP at baseline through AI-NT-proBNP showed predictive abilities comparable the biomarker NT-proBNP for prevalent and incident cardiovascular events, including heart failure, AF, stroke, MI, and cardiovascular death. After correction for confounders, pNT-proBNP revealed predictive value for prevalent and incident heart failure which was consistently comparable to mNT-proBNP. Furthermore, accuracy and reclassification improvement compared to baseline models were consistently similar for mNT-proBNP and pNT-proBNP. pNT-proBNP showed variable predictive value for incident AF or stroke compared to a baseline risk factor model and little to no additional value for prediction of cardiovascular death or MI compared to a baseline model in SHIP-START or SHIP-TREND. However, mNT-proBNP did not provide clinically relevant additional risk information compared to a baseline model including cardiovascular risk factors in this study either. Since pNT-proBNP revealed results for prediction of prevalent and incident cardiovascular disease which are comparable to the well-established biomarker NT-proBNP, we can assume that a model like AI-NT-proBNP can serve as a valuable auxiliary technology for cardiovascular risk prediction in population cohorts.

BNP and NT-proBNP are released from cardiomyocytes for volume and blood pressure homeostasis and to counteract cardiac stress upon different stimuli such as neuro-hormones, angiotensin II and endothelin I, and cytokines [32]. The predominant trigger is myocyte stretch. Parallel to the mechanisms leading to the secretion of this peptide, the electrophysiology of the cardiomyocytes seems to change, which leads to subtle changes of the ECG. Alterations of the ECG may thus confer important information, but remain less visible to the human eye. DL algorithms, however, appear to be able to identify such small electrocardiographic differences allowing for the derivation of disease specific information. Due to the inherent inexplicability of deep learning models, we cannot rule out the possibility that AI-NT-proBNP captures changes in ECGs which do not have secretion of NT-proBNP as a direct correlate, but instead, for example, cardiac fibrosis or conduction, or a combination of those. The analysis of these ECG changes likely impacts the predictive value of deep learning models like AI-NT-proBNP and could be the reason for higher AUROC values for pNT-proBNP compared to mNT-proBNP in some instances. This impact could be even more pronounced in models that are trained directly to predict specific cardiovascular diseases.

AI-NT-proBNP can predict serum levels of NT-proBNP with a decent R value. A larger training dataset with a wider range of NT-proBNP values could improve correlation between pNT-proBNP and mNT-proBNP, which could further

increase trust in the model and interpretability of its output. Yet, our study shows that pNT-proBNP can have similar value in cardiovascular risk assessment to mNT-proBNP in the general population, despite not having an R value close to 1, suggesting that aiming for a better correlation might not be necessary if identifying individuals at risk is the goal.

We chose NT-proBNP serum concentrations as a target variable when training an ECG deep learning model for several reasons. First, NT-proBNP has been shown to be highly predictive of multiple cardiovascular diseases. Thus, a model like AI-NT-proBNP can build on the long-standing clinical experience with this biomarker, increasing trustworthiness for clinical application and can be applied to several adverse cardiovascular outcomes. Furthermore, it makes training a model less dependent on complex disease definitions which can change over time. This is beneficial in contrast to models trained to directly predict specific cardiovascular outcomes, which might not be as useful for general cardiovascular risk prediction and cannot leverage on protracted evidence. Additionally, predicted NT-proBNP values are less abstract than disease risk probability. This increases interpretability of the model's output, which is important in clinical settings. Second, NT-proBNP serum concentrations are widely available in large datasets, which is a necessity for training reliable DL models. Incident cardiovascular diseases are rare in population-based datasets, hindering effective training if a model is developed to predict specific diseases in the general population. Additionally, a wide range of continuous values of this biomarker is available, resulting in a broader distribution of training data points and ultimately better performance of a model. We used classical statistical models as meta-models for prediction of adverse events to achieve a high performance despite comparatively low incidence of events in the general population. Third, a validation in external cohorts is more feasible with a standardized NT-proBNP laboratory measurement, increasing the credibility of this DL model.

Generalizability of ECG DL algorithms derived from patient data to populations outside the clinical setting has remained unsatisfactory as demonstrated for ECG-AF models [18, 25]. Therefore, we set out to generate a DL model applicable in the community by using data from population-based cohorts using one of the strongest cardiac biomarkers in screening for cardiovascular events in the community [12]. Other biomarkers with cardio-specificity and strong predictive value for cardiovascular events in the general population are cardiac troponins measured with high sensitivity assays, which have low intraindividual variability and therefore represent additional promising targets for deep learning models [33, 34].

We saw some variation in predictive value of pNT-proBNP in SHIP-START and SHIP-TREND, which underlines the heterogeneity of these cohorts. However, we did not observe a systematic shift, increasing our confidence in generalizability of AI-NT-proBNP.

pNT-proBNP, even when not adjusted for risk factors, demonstrates good predictive value in identifying individuals with prevalent heart failure using only the ECG. The availability of a standard 12-lead ECG and its automated analysis by a DL model could be particularly relevant as an initial screening step, possibly complementing other available screening measures such as point-of-care NT-proBNP testing kits. This augmentation of diagnostic value of the ECG can be especially crucial in middle- or low-income countries where laboratory infrastructure and trained personnel is sparse. A model like this could also be a relevant addition to the toolbox for CV risk assessment in industrialized countries, where other opportunities for fast and robust risk evaluation, for example by measuring natriuretic peptides, are not used often enough despite availability [35].

Limitations

Limitations of this study include the demographics. Considering that HCHS and SHIP both represent populations from a Western European country, AI-NT-proBNP might yield differing results in cohorts from other regions of the world. Furthermore, DL models are little explainable, which requires caution in the application of these models in high-risk settings, such as health care [36]. Whereas it remains largely unclear which ECG features contribute to the predictive ability of AI-NT-proBNP, robust validation can increase the trust in the reliability of the model. We could demonstrate sound external validation in two independent cohorts. Additionally, this model holds greatest potential in an outpatient-setting and could be used to assist trained healthcare professionals to initiate further diagnostics. Treatment decisions would not be directly based on the model, comparable to the biomarker NT-proBNP. Many of these limitations can be addressed by training AI-NT-proBNP on a larger and more diverse dataset.

To verify the ability of pNT-proBNP to be useful for CV risk assessment in the general population, multicenter studies comparing pNT-proBNP, mNT-proBNP and other strong biomarkers, namely troponins, are needed.

In conclusion, AI-NT-proBNP is a model that predicts an important cardiac biomarker and has the potential to be applied for simple, non-invasive screening for heart failure and cardiovascular disease risk in the general population. It may leverage the accessible diagnostic modality ECG, which

is fast, cheap, and reliable and could help identifying high-risk patients for referral to a specialist for further assessment and potential work-up. Such an approach could be applied widely and might be associated with improved cardiovascular disease outcomes [1].

Acknowledgments: We are grateful for the Hamburg City Health Study and SHIP participants, Epidemiological Study Centre staff, cooperation partners, patrons, and University Medical Centre Hamburg Deanery are acknowledged, as well as for fieldworkers, study physicians, and laboratory workers, and more. The German Centre for Cardiovascular Research's (DZHK) Shared Expertise program provided a platform for this study, for which we are grateful. We thank Jan A Kors (Erasmus MC Rotterdam) for his support with ECG conversion.

Research ethics: The studies were conducted in accordance with the Helsinki Declaration. The local Ethics Committee of the State of Hamburg Chamber of Medical Practitioners (PV5131) and the data protection commissioners of the University Medical Center Hamburg-Eppendorf and of the Free and Hanseatic City of Hamburg approved the HCHS. The Ethics Committee at the University Medicine Greifswald approved the SHIP studies (approval number BB 39/08).

Informed consent: All participants provided written, informed consent.

Author contributions: M.N., J.P.B. and R.B.S. conceptualized and designed the study; M.N., J.P.B., M.S.K., M.V., S.G., J.B., N.S., D.C., B.S., M.D. and R.B.S. analyzed and interpreted data; M.B., N.F., T.Z., S.F., S.B., M.D. and R.B.S. acquired data; M.N. and J.P.B. drafted the manuscript; all authors revised the manuscript and reviewed the data; M.N. and J.P.B. contributed equally.

Competing interests: The authors declare no competing interests regarding this study.

Research funding: The participating institutes and departments from the University Medical Center Hamburg-Eppendorf contribute with individual and scaled budgets to the overall funding. The study is also supported by the Innovative medicine initiative (IMI) under Grant No. 116074, by the Fondation Leducq under Grant Number 16 CVD 03, by the euCanSHare Grant Agreement No. 825903-euCanSHare H2020 and the DFG under project Grant TH1106/5-1; AA93/2-1. The license for the Food Frequency and Physical activity is provided by the DIFE. Technical equipment is provided by SIEMENS according to a contract for 12 years as well as by the Schiller AG on a loan basis for 6 years and by Topcon on a loan basis from 2017 until 2022. The Hamburg City Health Study is additionally supported by an unrestricted Grant (2017–2022) by Bayer. Project related analyses are supported by Amgen,

Astra Zeneca, BASF, Deutsche Gesetzliche Unfallversicherung (DGUV), DKFZ, DZHK, Novartis, Wolfgang Seefried Stiftung and Unilever. The study is further supported by donations from the “Förderverein zur Förderung der HCHS e.V.”, TEPE (2014) and Boston Scientific (2016). A current list of the supporter is available online at www.uke.de/hchs. This study is also supported by the German Centre of Cardiovascular Research (DZHK) with its Shared Expertise program (81X2710192). The Study of Health in Pomerania is part of the Community Medicine Research Network of the University Medicine Greifswald, which was funded by the German Federal Ministry for Education and Research, the Ministry for Education, Research and Cultural Affairs, and the Ministry for Social Affairs of the State Mecklenburg-West Pomerania. MN has received funding from the German Research Foundation (DFG) and the German Centre for Cardiovascular Research (DZHK e.V.) (81X2710192). MV has received funding from the German Centre for Cardiovascular Research (DZHK e.V.) (81X2400154). BS has received speaker fees from AstraZeneca and Abiomed; and research funding from the German Research Foundation, the Else Kröner-Fresenius-Stiftung and Abiomed. RBS has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under the grant agreement No 648131, from the European Union’s Horizon 2020 research and innovation programme under the grant agreement No 847770 (AFFECT-EU) and German Center for Cardiovascular Research (DZHK e.V.) (81Z1710103 and 81Z0710114); German Ministry of Research and Education (BMBF 01ZX1408A) and ERACoSysMed3 (031L0239). RBS has received lecture fees and advisory board fees from BMS/Pfizer outside this work. The remaining authors have nothing to disclose.

Data availability: All statistical analyses to reproduce the results reported in this paper will become open-source and available under MIT license at: <https://github.com/JanBrem/AI-NT-proBNP>.

References

- Zheng A, Cowan E, Mach L, Adam RD, Guha K, Cowburn PJ, et al. Characteristics and outcomes of patients with suspected heart failure referred in line with National Institute for Health and Care Excellence guidance. *Heart* 2020;106:1579–85.
- O’Neal WT, Sandesara PB, Claxton JS, MacLehose RF, Chen LY, Bengtson LGS, et al. Provider specialty, anticoagulation prescription patterns, and stroke risk in atrial fibrillation. *J Am Heart Assoc* 2018;7:e007943.
- Bottle A, Kim D, Aylin P, Cowie MR, Majeed A, Hayhoe B. Routes to diagnosis of heart failure: observational study using linked data in England. *Heart* 2018;104:600–5.
- Taylor CJ, Lay-Flurrie SL, Ordóñez-Mena JM, Goyder CR, Jones NR, Roalfe AK, et al. Natriuretic peptide level at heart failure diagnosis and risk of hospitalisation and death in England 2004–2018. *Heart* 2022;108:543–9.
- McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2021;42:3599–726.
- Wang TJ, Benjamin EJ, Wolf PA. Plasma natriuretic peptide levels and the risk of cardiovascular events and death. *N Engl J Med* 2004;350:655–63.
- Fonarow GC, Peacock WF, Phillips CO, Givertz MM, Lopatin M. Admission B-type natriuretic peptide levels and in-hospital mortality in acute decompensated heart failure. *J Am Coll Cardiol* 2007;49:1943–50.
- Lemos JAD, Omland T, Hall C. The prognostic value of B-type natriuretic peptide in patients with acute coronary syndromes. *N Engl J Med* 2001;345:1014–21.
- Redfield MM, Rodeheffer RJ, Jacobsen SJ, Mahoney DW, Bailey KR, Burnett JC. Plasma brain natriuretic peptide concentration: impact of age and gender. *J Am Coll Cardiol* 2002;40:976–82.
- Daniels LB, Maisel AS. Natriuretic peptides. *J Am Coll Cardiol* 2007;50:2357–68.
- Welsh P, Doolin O, Willeit P, Packard C, Macfarlane P, Cobbe S, et al. N-terminal pro-B-type natriuretic peptide and the prediction of primary cardiovascular events: results from 15-year follow-up of WOSCOPS. *Eur Heart J* 2013;34:443–50.
- Di Angelantonio E, Chowdhury R, Sarwar N, Ray KK, Gobin R, Saleheen D, et al. B-type natriuretic peptides and cardiovascular risk: systematic review and meta-analysis of 40 prospective studies. *Circulation* 2009;120:2177–87.
- Roalfe AK, Lay-Flurrie SL, Ordóñez-Mena JM, Goyder CR, Jones NR, Hobbs FDR, et al. Long term trends in natriuretic peptide testing for heart failure in UK primary care: a cohort study. *Eur Heart J* 2021;43:881–91.
- Iwaz JA, Maisel AS. Recent advances in point-of-care testing for natriuretic peptides: potential impact on heart failure diagnosis and management. *Expert Rev Mol Diagn* 2016;16:641–50.
- Kwon J, Lee SY, Jeon K, Lee Y, Kim K, Park J, et al. Deep learning-based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc* 2020;9:e014717.
- Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet* 2019;394:861–7.
- Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;25:70–4.
- Christopoulos G, Graff-Radford J, Lopez CL, Yao X, Attia ZI, Rabinstein AA, et al. Artificial intelligence–electrocardiography to predict incident atrial fibrillation: a population-based study. *Circ Arrhythm Electrophysiol* 2020;13:e009355.
- Jagodzinski A, Johansen C, Koch-Gromus U, Aarabi G, Adam G, Anders S, et al. Rationale and design of the Hamburg city health study. *Eur J Epidemiol* 2020;35:169–81.
- Völzke H, Schössow J, Schmidt CO, Jürgens C, Richter A, Werner A, et al. Cohort profile update: the study of health in Pomerania (SHIP). *Int J Epidemiol* 2022;51:e372–83.
- Braisch U, Koenig W, Rothenbacher D, Denking M, Friedrich N, Felix SB, et al. N-terminal pro brain natriuretic peptide reference values in community-dwelling older adults. *ESC Heart Fail* 2022;9:1703–12.

22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition [Internet]. arXiv; 2015. <http://arxiv.org/abs/1512.03385> [Accessed 19 Jul 2022].
23. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15:1929–58.
24. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 2009;150:604–12.
25. Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation* 2022;145:122–33.
26. Friedrich S, Groß S, König IR, Engelhardt S, Bahls M, Heinz J, et al. Applications of artificial intelligence/machine learning approaches in cardiovascular medicine: a systematic review with recommendations. *Eur Heart J – Digit Health* 2021;2:424–36.
27. Van Rossum G, Drake FL. Python 3 reference manual. Scotts Valley, CA: CreateSpace; 2009.
28. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. *Adv neural inf process syst* 2019. <http://arxiv.org/abs/1912.01703>.
29. Marcel S, Rodriguez Y. Torchvision the machine-vision package of torch. In: *Proc 18th ACM Int Conf Multimed* [Internet]. New York, NY, USA: Association for Computing Machinery; 2010:1485–8 pp.
30. Jones E, Oliphant T, Peterson P. others. SciPy: open source scientific tools for Python; 2001. Available from: <https://www.scipy.org>.
31. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng* 2007;9:90–5.
32. Goetze JP, Bruneau BG, Ramos HR, Ogawa T, de Bold MK, de Bold AJ. Cardiac natriuretic peptides. *Nat Rev Cardiol* 2020;17: 698–717.
33. Farmakis D, Mueller C, Apple FS. High-sensitivity cardiac troponin assays for cardiovascular risk stratification in the general population. *Eur Heart J* 2020;41:4050–6.
34. Clerico A, Zaninotto M, Passino C, Aspromonte N, Piepoli MF, Migliardi M, et al. Evidence on clinical relevance of cardiovascular risk evaluation in the general population using cardio-specific biomarkers. *Clin Chem Lab Med* 2021;59:79–90.
35. Hayhoe B, Kim D, Aylin PP, Majeed FA, Cowie MR, Bottle A. Adherence to guidelines in management of symptoms suggestive of heart failure in primary care. *Heart* 2019;105:678–85.
36. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health* 2021;3:e745–50.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/cclm-2023-0743>).