Check for updates

RESEARCH ARTICLE

Studying accelerated cardiovascular ageing in Russian adults through a novel deep-learning ECG biomarker [version 1; peer review: awaiting peer review]

Ernest Diez Benavente [iD] [1,2], Francisco Jimenez-Lopez [3], Zachi I. Attia [3], Sofia Malyutina [4,5], Alexander Kudryavtsev [iD] [6,7], Andrew Ryabikov [4,5], Paul A. Friedman [3], Suraj Kapa [3], Michael Voevoda [iD] [4], Pablo Perel [iD] [8,9], Henrik Schirmer [iD] [10-12], Alun D. Hughes [iD] [13], Taane G. Clark [iD] [1,14], David A. Leon [iD] [7,14,15]

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK
[2]Department of Experimental Cardiology, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands
[3]Department of Cardiology, Mayo Clinic College of Medicine, Rochester, Minnesota, 55905, USA
[4]Research Institute of Internal and Preventive Medicine, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090, Russian Federation
[5]Novosibirsk State Medical University, Russian Ministry of Health, Novosibirsk, 630091, Russian Federation
[6]Northern State Medical University, Arkhangelsk, 163000, Russian Federation
[7]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, 9019, Norway
[8]World Heart Federation, Geneva, 1201, Switzerland
[9]Centre for Global Chronic Conditions, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK
[10]Department of Cardiology, Akershus University Hospital, Lorenskog, 1478, Norway
[11]Institute of Clinical Medicine, Camp Ahus, University of Oslo, Oslo, 0315, Norway
[12]Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, 9019, Norway
[13]MRC Unit for Lifelong Health and Aging, University College London, London, WC1E 6BT, UK
[14]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, UK
[15]International Laboratory for Population and Health, National Research University, Higher School of Echonomics, Moscow, 101000, Russian Federation

Open Peer Review

Reviewer Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background: A non-invasive, easy-to-access marker of accelerated cardiac ageing would provide novel insights into the mechanisms and aetiology of cardiovascular disease (CVD) as well as contribute to risk stratification of those who have not had a heart or circulatory event. Our hypothesis is that differences between an ECG-predicted and chronologic age of participants (δage) would reflect accelerated or decelerated cardiovascular ageing

Methods: A convolutional neural network model trained on over 700,000 ECGs from the Mayo Clinic in the U.S.A was used to predict the age of 4,542 participants in the Know Your Heart study conducted in two cities in Russia (2015-2018). Thereafter, δage was used in linear

regression models to assess associations with known CVD risk factors and markers of cardiac abnormalities.
Results: The biomarker δage (mean: +5.32 years) was strongly and positively associated with established risk factors for CVD: blood pressure, body mass index (BMI), total cholesterol and smoking. Additionally, δage had strong independent positive associations with markers of structural cardiac abnormalities: N-terminal pro b-type natriuretic peptide (NT-proBNP), high sensitivity cardiac troponin T (hs-cTnT) and pulse wave velocity, a valid marker of vascular ageing.
Conclusion: The difference between the ECG-age obtained from a convolutional neural network and chronologic age (δage) contains information about the level of exposure of an individual to established CVD risk factors and to markers of cardiac damage in a way that is consistent with it being a biomarker of accelerated cardiovascular (vascular) ageing. Further research is needed to explore whether these associations are seen in populations with different risks of CVD events, and to better understand the underlying mechanisms involved.

Keywords

Cardiovascular disease, Biomarkers, Aging, Machine Learning, Russian Federation

Corresponding author: Ernest Diez Benavente (Ernest.DiezBenavente@lshtm.ac.uk)

Author roles: Diez Benavente E: Conceptualization, Data Curation, Formal Analysis, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Jimenez-Lopez F**: Conceptualization, Supervision, Writing – Review & Editing; **Attia ZI**: Data Curation, Formal Analysis, Writing – Review & Editing; **Malyutina S**: Data Curation, Formal Analysis, Funding Acquisition, Writing – Review & Editing; **Kudryavtsev A**: Data Curation, Formal Analysis, Funding Acquisition, Project Administration, Writing – Review & Editing; **Ryabikov A**: Data Curation, Formal Analysis, Investigation, Writing – Review & Editing; **Friedman PA**: Supervision, Writing – Review & Editing; **Kapa S**: Supervision, Writing – Review & Editing; **Voevoda M**: Supervision, Writing – Review & Editing; **Perel P**: Conceptualization, Supervision, Writing – Review & Editing; **Schirmer H**: Conceptualization, Writing – Review & Editing; **Hughes AD**: Conceptualization, Writing – Review & Editing; **Clark TG**: Conceptualization, Supervision, Writing – Review & Editing; **Leon DA**: Conceptualization, Funding Acquisition, Supervision, Writing – Review & Editing

Introduction

The concept of biological age as distinct to chronologic age has attracted substantial interest. From a public health and clinical perspective, a valid and reliable measure of biological age would contribute to risk stratification and prioritisation of individuals or groups for treatments and interventions to reduce exposure to risk factors.

Various molecular biomarkers of biological age have been proposed including telomere length[1] and DNA-methylation patterns[2,3]. These markers require invasive procedures, including blood taking and processing of DNA samples. Moreover, attempting to determine biological age of a particular organ as opposed to techniques which use leucocyte telomere lengths to derive a general biological age for example[4] requires tissue-specific samples. These factors make them unsuitable for routine use in clinical practice, as well as making their collection and analysis in large population-based studies prohibitively expensive. Establishing a new biomarker which is non-invasive and easily deployable at large scale in all settings, including resource constrained ones would therefore have many advantages.

The concept of heart age has been previously used to assess the physiologic ageing of the cardiovascular system[5]. A range of heart age prediction tools have been developed and typically require a combination of biochemical (e.g. serum cholesterol levels (LDL, HDL)), diabetes diagnosis, etc.) and other inputs (e.g. chronologic age, body mass index (BMI), diastolic and systolic blood pressure (DBP and SBP), mean arterial pressure (MAP))[6]. Others incorporate imaging parameters such as coronary artery calcification, and carotid intima-media thickness into their risk prediction models[5,7,8].

The electrocardiogram (ECG) is a ubiquitous and key clinical tool to carry out cardiovascular disease (CVD) risk stratification, diagnosis and prognosis[9,10]. We have recently developed a deep-learning algorithm that has been applied to routinely collected digitised 12-lead ECGs from a large clinical population in the U.S.A that has achieved high-accuracy in the prediction of left ventricular systolic dysfunction (LVSD) both retrospectively[11] and prospectively[12], and atrial fibrillation while in normal sinus rhythm[13].

We have also developed a separate deep-learning model based on the same population to accurately predict biological age exclusively from digitised 12-lead ECG data[14]. Subtracting true chronologic from ECG-predicted age (ECG-age) yields a difference that we call δage (delta-age). This metric will be positive for individuals predicted to be biologically older than their chronologic age, or negative if predicted to be younger. In the US clinical population used to train the prediction algorithm, the predicted age had a mean absolute error of 7 years (SD: 5.6 years) compared to chronologic age[14]. A δage of 7 years or greater predicted a higher risk of CVD events as well as a higher prevalence of low ejection fraction, hypertension and coronary disease[14].

We suggest that δage may be regarded as a biologically proximal marker of the relative rate of cardiovascular ageing that reflect cumulative damage to date. We propose that it may have the potential for CVD risk stratification and case management as well as being a powerful tool for research to elucidate novel mechanisms of accelerated cardiac ageing and identify potential drivers of CVD differences across populations and between individuals.

In this study, we investigate how the novel biomarker ECG-derived δage relates to CVD risk profiles in a large cross-sectional study. We do this by studying the association of δage with established CVD risk factors as well as markers of age-related cardiovascular abnormalities in a cross-sectional study conducted in adults (35–69 years) from two Russian cities (Arkhangelsk and Novosibirsk; 2015–18).

Methods

Data sources and population

We used data obtained in a cross-sectional study in the Russian cities of Arkhangelsk and Novosibirsk. The study population included 4,542 men and women aged 35–69 years recruited from the general population between 2015 and 2018, as described in detail elsewhere[15]. All participants provided written informed consent as detailed in a previous publication describing the study protocol[15]. Ethical approval for the study was received from the ethics committees of the London School of Hygiene & Tropical Medicine (approval number 8808 received 24.02.2015; for substudy involving patients in treatment for alcohol problems approval number 12018; received 11/01/2017), Novosibirsk State Medical University (approval number 75 approval received 21/05/2015), the Institute of Preventative Medicine (no approval number; approval received 26/12/2014), Novosibirsk and the Northern State Medical University, Arkhangelsk (approval number 01/01-15 received 27/01/2015; for substudy involving patients in treatment for alcohol problems approval number 05/11-16 received 02/11/2016).

Information on smoking status, alcohol use and education level were obtained through computer assisted personal interview (CAPI) questionnaires at baseline interview[15]. Alcohol use was assessed using the Alcohol Use Disorders Identification Test (AUDIT) instrument[16]. Education was coded as 3 categories (incomplete secondary, secondary and higher education) as used in an earlier Russian study[17].

Participants were invited for a health check where a questionnaire and physical examination were performed in one of three study clinics[15]. Of those contacted and invited to take part, 47% attended the health check. A resting 10–25 seconds digital 12-lead ECG was obtained using the Cardiax device (IMED Ltd, Hungary) and the raw ECG signal was stored digitally[15]. Blood pressure and pulse rate were measured using the OMRON 705 IT automatic blood pressure monitor (OMRON Healthcare). The mean value of the second and third readings of systolic (SBP) and diastolic blood pressure (DBP) were used in the analysis. Mean arterial pressure (MAP) was calculated as:

$$MAP = \frac{SBP + 2 * DBP}{3}$$

Body mass index (BMI) was measured using the following formula:

$$BMI = \frac{w}{h^2}$$

where height (h; metres) was averaged across two measurements and obtained following previously defined methods[15] and weight (w; kg) was obtained using the TANITA BC 418 body composition analyser (TANITA, Europe GmbH) or a weighing scale when the TANITA could not be used[15]. Pulse wave velocity was obtained using the Vicorder device (Skidmore Medical Ltd, UK). Up to 8 pulse wave measurements were obtained based on quality of the measurements[15], and the pulse wave velocity measurement used in this study was the average of all available measurements.

During the course of the health check, blood was obtained from the participants who were requested not to eat or drink alcohol during the 4 hours immediately before their booked appointment. Serum samples were taken, centrifuged, and aliquoted into 1.8-mL barcoded cryovials which were then frozen at −80°C within 2 hours. These were then analysed in a single batch at the end of the fieldwork at an independent laboratory in Moscow. The lipid profile of each sample was measured and included high-density and low-density lipoproteins.

High-sensitivity cardiac troponin T (hs-cTnT) and N-terminal pro-B-type natriuretic peptide (NT-proBNP) were measured using a high-sensitivity electrochemiluminescence immunoassay (Roche Diagnostics GmbH, Hitachi, Japan) on a Cobas e411 analyser. The lower limit of detection for hs-cTnT test was 3 ng/L. Of the total 3,864 participants, 69 (1.79%) were below the limit of detection for Hs-cTnT which was 3 ng/L and were recoded to 2.9 ng/L[18]. Similarly, 22 participants (0.57%) were below the lower limit of detection for NT-proBNP which was 5 ng/L and their values were recoded to 4.9 ng/L[18].

Presence of plaques was estimated by high-resolution ultrasound using the Vivid q machine (GE HealthCare) with 6~13 MHz linear transducer in both side carotid arteries (common carotid artery, bifurcation, internal carotid artery). The still image for each identified plaque and cine loops in transverse and sagittal views was recorded digitally. The off-line reading of recorded images was performed with EchoPAC software (GE Vingmed Ultrasound, Horten, Norway) by an experienced reader (AR). Carotid plaque was defined as a focal structure encroaching into the arterial lumen at least 0.5 mm or 50% of the surrounding intima media thickness (IMT) value or IMT >1.5 mm as measured from the media-adventitia interface to the intima-lumen interface based on Mannheim Consensus[19]. Detailed methods are previously described elsewhere[19,20]. The current analysis uses the presence or absence of any plaques – a dichotomous variable.

ECG analysis of patients with multiple ECGs
During the course of the data collection, 402 randomly selected participants (202 in Arkhangelsk and 200 in Novosibirsk) were invited to retake the complete assessments performed in the study, including both baseline questionnaire and health-check measurements. The repeat interviews and examinations were performed one year after the first interview, with a two-week either side date difference allowed[21]. Out of the 402 total records, 369 repeat ECGs were available and were linked to the original ones taken a year earlier, both δage and a between-pair Pearson's correlation was calculated using R v3.4.4 base functions.

Overview of the AI model
A convolutional neural network (CNN) model using Keras with a Tensorflow 2 (Google, Mountain View, CA) backend was previously developed and validated. A total of 774,783 unique subjects with ECGs were used to develop the neural network: 399,750 in the training set, 99,977 in the internal validation set and 275,056 ECGs in the holdout testing set. The network contained stacked blocks of convolutional, max pooling, and batch normalization[22]. A detailed description of the network is described in our previously published paper[14]. The network had a single output (ECG-age) as continuous number. This model was not re-trained using the Russian data. The measure δage was obtained by subtracting the chronologic age from the ECG-age. In this way, a positive δage would represent an individual in which the ECG would predict an older age than the chronologic age (older by ECG), while a subject with a negative δage would represent an individual found to be younger by ECG.

Statistical methods
R v3.4.4. statistical software was used for statistical analysis. Continuous variables were standardised as SD units by subtracting the average and dividing by the standard deviation to allow comparable effects in regression analyses. Hs-cTnT and NTroBNP measurements were natural log transformed as they were not Gaussian distributed.

We used linear regression models to assess association of the risk factors as exposures using δage as the outcome. All models were adjusted for a priori confounders including chronologic age, sex and clinic where the measurement was taken, where the latter accounted for any systematic potential protocol and device calibration differences. Real chronologic age was also included in every model to remove the effects of any potential correlation between δage and chronologic age. A test for trend for smoking was performed by converting the ordered categories of number of cigarettes smoked into integer values. Basic models including one risk factor at a time were further adjusted for other risk factors plus education level.

We assessed whether any of the main effects varied by sex by undertaking an analysis of interactions. To assess the extent to which any results might be driven by the subset of people with pre-existing cardiovascular disease, sensitivity analysis was carried out by excluding any participant with either an ECG indicated myocardial infarction (MI) or self-reported history of stroke or MI and repeated the same linear regression models.

Results

A total of 4,780 resting ECGs of 15 seconds were collected from 4,378 participants in the main study sample (MSS) (402 ECGs were repeated for a randomly selected subset of MSS participants one year later in the repeat study sample (RSS)). After exclusion of participants with missing data for any variable considered, the final set consisted of 3,863 and 369 ECGs for the MSS and RSS, respectively. Overall, mean chronologic age was 54.6 years (SD 9.7) on the day of the first ECG measurement and 1,597 (41.3%) of them were men.

The ECG-age for all participants had a mean of 59.8 years (SD 10.6). The average δage of this study population was 5.32 years (SD 7.6) (Figure 1). This difference in predicted versus chronologic age was observed to be relatively constant for the age groups between 35 and 60 years old, with average δage converging towards the line of identity at older ages (Figure 2). The RSS measurements were correlated with their respective MSS measurements from the previous year (Pearson correlation $r = 0.72$) (Figure 3).

Linear regression analysis of known CVD risk factors on δage (adjusted for sex, chronologic age and clinic where the ECG was taken) showed that SBP, DBP, MAP, BMI, LDL/HDL ratio and smoking status were associated with an increase in δage (P<0.0001); DBP and BMI showed the strongest effects, which were almost identical (1 SD unit increase leading to + ~1.6 δage years) (Table 1). These effects were attenuated on mutual adjustment for confounders, but all variables remained statistically significant other than the AUDIT alcohol score (Table 2). Mutually adjusting smoking status and BMI revealed an increase in the effect of smoking. Further adjustment for education had minimal effect (results not shown).

In the fully adjusted model (Table 2), chronological age showed a negative association with δage, with a 2.66 (95% CI: 2.40, 2.92) years reduction in δage for a 10-year increase in chronologic age, and this effect was driven by the older age groups (Figure 2). This is consistent with the line of best fit moving closer to the line of identity with increasing chronological age as shown in Figure 2. δage differed between the two cities, in particular, the Novosibirsk population had a lower δage than that in Arkhangelsk (-1.21, 95% CI: -1.47, -0.54 years). Sex showed no significant effect on δage when studied by itself or in relation to individual CVD risk factors and no significant sex-interactions were observed. In the fully adjusted model (Table 2) female participants had higher δage years (0.68, 95% CI: 0.12, 1.24) compared to males.

Two markers of cardiac pathology (NTproBNP and hs-cTnT) showed positive associations between their log concentrations
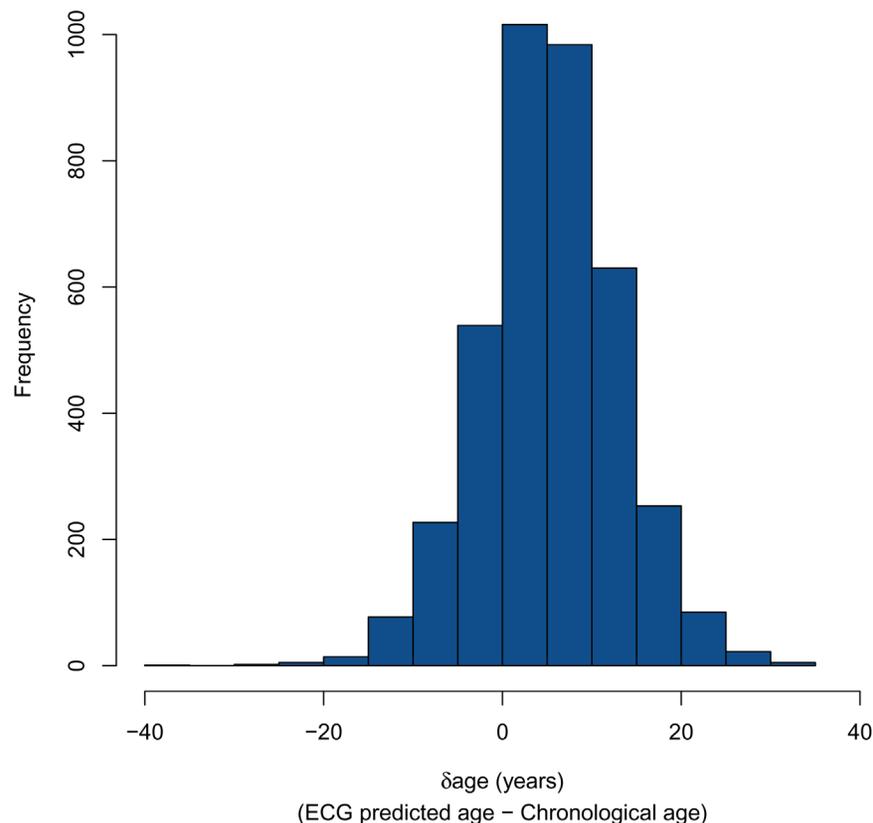
Figure 1. Histogram of δage (years) in Russian adults (n=3,864, 35–69 years) from two cities in Russia.
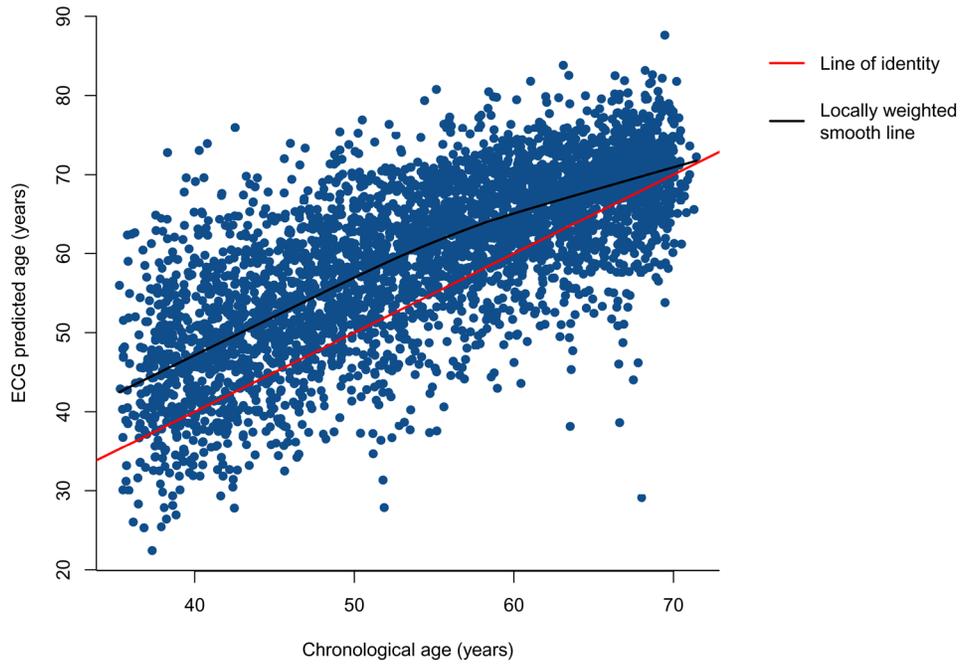
**Figure 2. Scatterplot of ECG predicted age against chronologic age in Russian adults (n=3,864, 35–69 years) from 2 cities in Russia.**
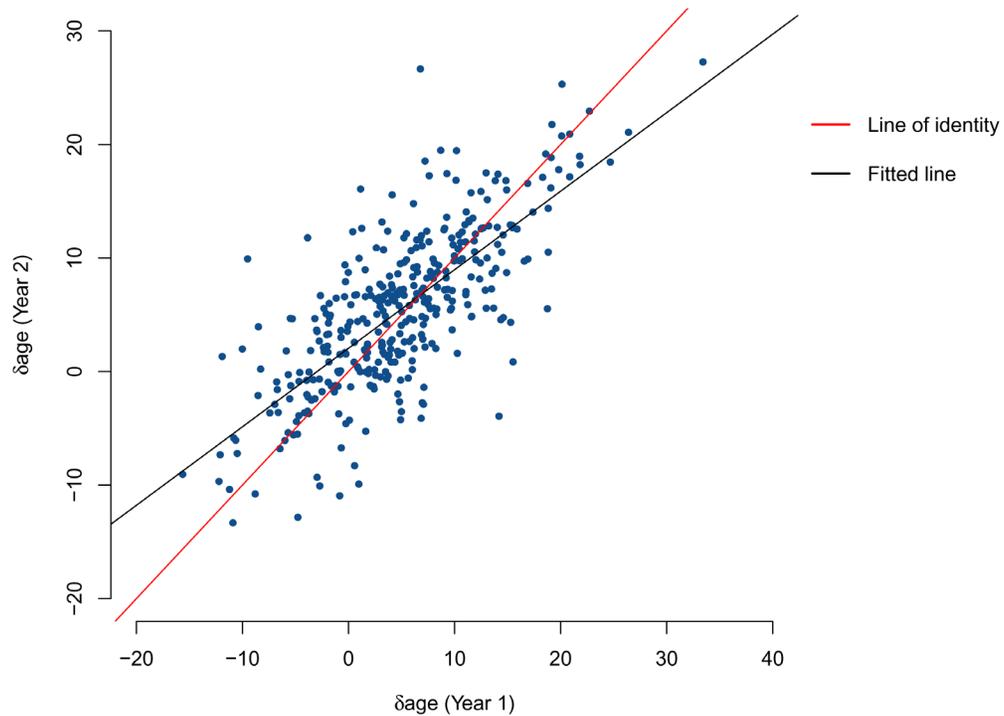

Figure 3. Scatterplot of δage for 369 records of the Main Study Subset participants (MSS, Year 1) against δage for linked Repeat Study Subset participants (RSS, Year 2), *Pearson correlation* (*r*) = 0.72.

Table 1. Associations of cardiovascular disease risk factors with δage (n = 3,863) adjusted for age, sex and clinic.

Exposures *		δage (years)	(95% CI)	p-value
Mean systolic blood pressure (SD units)		1.30	[1.05, 1.55]	<0.0001
Mean diastolic blood pressure (SD units)		1.62	[1.39, 1.86]	<0.0001
Mean arterial pressure (SD units)		1.56	[1.32, 1.80]	<0.0001
Mean pulse rate (SD units)		-0.12	[-0.25, 0.11]	0.30
Body mass index (SD units)		1.61	[1.38, 1.84]	<0.0001
Ratio LDL/HDL (SD units)		0.84	[0.6, 1.07]	<0.0001
AUDIT score (score units [0–40])		0.06	[-0.01, 0.13]	0.11
Smoking status	Never smoked [reference]	0	-	-
	Ex-smoker/Non-regular smoker	0.50	[-0.09, 1.09]	0.10
	Regular smoker (<=20 cigarettes)	0.67	[0.02, 1.31]	0.042
	Heavy smoker (>20 cigarettes)	1.56	[0.19, 2.92]	0.026
	P-value for trend	0.0032		

(*) Systolic Blood Pressure, Diastolic Blood Pressure, Mean Arterial Pressure, Mean Pulse Rate, Body Mass Index, and the ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL) were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable. Alcohol Use Disorders Identification Test (AUDIT).

Table 2. Mutually adjusted associations of established cardiovascular disease risk factors with δage (n = 3,863) additionally adjusted for age, sex and clinic.

Exposures *		δage (years)	(95% CI)	p-value
Mean systolic blood pressure (SD units)		1.01	[0.76, 1.26]	<0.0001
Body mass index (SD units)		1.34	[1.09, 1.58]	<0.0001
Ratio LDL/HDL (SD units)		0.35	[0.10, 0.59]	0.0052
AUDIT score (score units [0–40])		0.01	[-0.07, 0.08]	0.77
Smoking status	Never smoked [reference]	0	-	-
	Ex-smoker/Non-regular smoker	0.40	[-0.18, 0.98]	0.17
	Regular smoker (<=20 cigarettes)	0.79	[0.15, 1.42]	0.015
	Heavy smoker (>20 cigarettes)	1.22	[-0.12, 2.56]	0.074
	P-value for trend	0.0058		

(*) Systolic Blood Pressure, Body Mass Index, and the ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL) were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable. Alcohol Use Disorders Identification Test (AUDIT).

and δage (P<0.0001) (Table 3). Adjustment for other CVD risk factors marginally increased the strength of association of NTproBNP and attenuated the association of hs-cTnT with age. The associations of both of these biomarkers with δage remained significant after adjustment for risk factors (Table 3). These associations were marginally reduced by adjustment for education level.

Carotid plaque was associated with δage after adjustment for age, sex and clinic but this association was attenuated to null after further adjustment for CVD risk factors. Pulse wave velocity was positively associated with δage, although adjustment for CVD risk factors attenuated the effect appreciably (Table 3); this was largely accounted for by a shared association with systolic blood pressure.

Finally, we estimated the independent associations of MAP and pulse wave velocity on δage (Table 4). This highlighted the strong significant association of MAP and pulse wave velocity with δage with MAP showing an effect of 1.57 (95% CI: 1.33, 1.81) δage years increase per standard deviation unit increase of MAP (Table 4, **Model 1**). On mutual adjustment, the association between pulse wave velocity and δage was substantially reduced, while the association between MAP and δage only marginally reduced. Additional adjustment for other CVD risk factors further attenuated these associations.

However, the association with MAP remained statistically significant (P<0.0001).

Sensitivity analyses were performed to further explore the robustness of the effects observed. Following exclusion of all participants who self-reported having had a myocardial infarction or stroke or were found to have ECG-based evidence of myocardial infarction, associations with CVD risk factors changed only slightly. There was an increase of the adjusted association with smoking and the LDL/HDL ratio and a reduction

Table 3. Associations of markers of heart abnormalities and ageing biomarkers with δage adjusted for various factors (n = 3,843).

Exposures *	Model 1: Adjusted for Age, Sex and Clinic			Model 2: Model 1 + Significant Risk Factors from Table 2 (BMI, SBP, Ratio LDL/HDL and Smoking status)		
	δage (years)	(95% CI)	p-value	δage (years)	(95% CI)	p-value
Log Troponin T (SD units)	0.52	[0.26, 0.79]	<0.0001	0.28	[0.02, 0.54]	0.032
Log NTproBNP (SD units)	0.82	[0.57, 1.08]	<0.0001	0.88	[0.63, 1.14]	<0.0001
Presence of carotid plaque (yes/no)	0.70	[0.16, 1.24]	0.012	0.19	[-0.35, 0.72]	0.50
Mean pulse wave velocity (SD units)	0.81	[0.54, 1.08]	<0.0001	0.27	[-0.02, 0.55]	0.065
Mean pulse wave velocity (SD units) **	0.81	[0.54, 1.08]	<0.0001	0.38	[0.10, 0.67]	0.007

(*) Log Troponin T, Log NTproBNP and mean pulse wave velocity were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable.

(**) Not adjusted by Systolic Blood Pressure (SBP).

Ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL); Body Mass Index (BMI).

Table 4. Associations of mean arterial pressure and pulse wave velocity on δage adjusted for various factors (n = 3,843).

Exposure *	Model 1: Adjusted for Age, Sex and Clinic			Model 2: Model 1 + Mutual Adjustment			Model 3: Model 2 + Adjustment for BMI, Ratio LDL/HDL and smoking status		
	δage (years)	(95% CI)	p-value	δage (years)	(95% CI)	p-value	δage (years)	(95% CI)	p-value
Mean arterial pressure (SD units)	1.57	[1.33, 1.81]	< 0.0001	1.49	[1.23, 1.75]	<0.0001	1.34	[0.93, 1.45]	<0.0001
Mean pulse wave velocity (SD units)	0.81	[0.54, 1.08]	< 0.0001	0.26	[-0.02, 0.54]	0.072	0.20	[-0.08, 0.47]	0.17

(*) Mean arterial pressure and mean pulse wave velocity were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable. Ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL); Body Mass Index (BMI).

in the strength of association with hs-cTnT and NT-proBNP (Table 5–Table 7).

Discussion

In this study we have shown that the difference between the ECG-age obtained from a convolutional neural network and chronologic age, called here delta-age (δage), contains information about the level of exposure of an individual to established CVD risk factors and to markers of cardiac damage.

This novel biomarker is strongly associated in this Russian study population with established risk factors of CVD (systolic blood pressure, body mass index, smoking and serum cholesterol levels) in a direction that is intuitively consistent with δage being a marker of accelerated cardiovascular (vascular) ageing. In parallel we have found strong evidence that δage is associated with established markers of cardiac damage such as NTproBNP and hs-cTnT.

In order to explore the potential of this marker as a measure of accelerated cardiovascular ageing, we studied its association with pulse wave velocity, an established marker of vascular ageing. Carotid femoral pulse wave velocity is a proxy measurement for aortic stiffness[23]. Pulse wave velocity is known to be pressure dependent[24], which could account for the strong attenuation of the effect of the association of pulse wave velocity with δage we observe when arterial pressure is accounted for. Despite this, intrinsic stiffness independent of the

measured arterial pressure is a marker of ageing[25]. In this study, we have assessed this cumulative damage by considering both MAP and pulse wave velocity in the same model, this showed a positive association of pulse wave velocity with δage even after adjustment for MAP. Due to the complex nature of the interaction between pulse wave velocity and blood pressure, further research with well-powered cohorts would be required to disentangle the relationship of δage with traditional vascular markers of ageing.

The robustness of the δage biomarker and its ability to capture information about the intrinsic cardiovascular health of an individual can be observed in the reproducibility of our measurements taken one year apart in this study (Pearson's *r=0.72*), and across several years in a previous study[14]. Both of these results suggest that this measure is stable across time. In addition, δage captured the effect of complex interactions such as the negative confounding effect of smoking on BMI and the positive trend on δage observed with increasing amounts of cigarettes smoked. Of special interest would be to assess the effects on δage in an intervention study aimed at CVD health and risk improvement.

One of the most striking results from our study is to show how an algorithm that has been exclusively trained in a clinical population from the U.S.A provides such an informative picture about the CVD risk of a different population in Russia, suggesting that the information captured by the algorithm is

Table 5. Sensitivity analysis of the associations of cardiovascular disease risk factors with δage (n = 3,536) adjusted for age, sex and clinic. Participants were excluded if they presented with either self-reported myocardial infarction, self-reported stroke or ECG-based evidence of myocardial infarction.

Exposures *		δage (years)	(95% CI)	p-value
Mean systolic blood pressure (SD units)		1.28	[1.02, 1.54]	<0.0001
Mean diastolic blood pressure (SD units)		1.66	[1.42, 1.91]	<0.0001
Mean arterial pressure (SD units)		1.57	[1.32, 1.83]	<0.0001
Mean pulse rate (SD units)		-0.16	[-0.40, 0.08]	0.19
Body mass index (SD units)		1.61	[1.37, 1.86]	<0.0001
Ratio LDL/HDL (SD units)		0.89	[0.64, 1.14]	<0.0001
AUDIT score (score units [0-40])		0.07	[0.00, 0.15]	0.065
Smoking status	Never smoked [reference]	0	-	-
	Ex-smoker/ Non-regular smoker	0.50	[-0.12, 1.12]	0.12
	Regular smoker (<=20 cigarettes)	0.72	[0.04, 1.40]	0.037
	Heavy smoker (>20 cigarettes)	1.94	[0.53, 3.35]	0.0071
	P-value for trend	0.0029		

(*) Systolic Blood Pressure, Diastolic Blood Pressure, Mean Arterial Pressure, Mean Pulse Rate, Body Mass Index and the ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL) were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable. Alcohol Use Disorders Identification Test (AUDIT).

Table 6. Sensitivity analysis of the Mutually adjusted associations of established cardiovascular disease risk factors with δage (n = 3,563) additionally adjusted for age, sex and clinic. Participants were excluded if they presented with either self-reported myocardial infarction, self-reported stroke or ECG-based evidence of myocardial infarction.

Exposures *		δage (years)	(95% CI)	p-value
Mean systolic blood pressure (SD units)		0.98	[0.72, 1.24]	<0.0001
Body mass index (SD units)		1.32	[1.06, 1.58]	<0.0001
Ratio LDL/HDL (SD units)		0.41	[0.16, 0.67]	0.0016
AUDIT score (score units [0–40])		0.01	[-0.06, 0.09]	0.70
Smoking status	Never smoked [reference]	0	-	-
	Ex-smoker/ Non-regular smoker	0.40	[-0.21, 1.00]	0.19
	Regular smoker (<=20 cigarettes)	0.78	[0.11, 1.45]	0.021
	Heavy smoker (>20 cigarettes)	1.57	[0.19, 2.96]	0.026
	P-value for trend	0.0043		

(*) Systolic Blood Pressure, Diastolic Blood Pressure, Mean Arterial Pressure, Mean Pulse Rate, Body Mass Index and the ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL) were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable. Alcohol Use Disorders Identification Test (AUDIT).

Table 7. Sensitivity analysis on the effect of markers of heart abnormalities and ageing biomarkers on δage (n = 3,521). Participants were excluded if they presented with either self-reported myocardial infarction, self-reported stroke or ECG-based evidence of myocardial infarction.

	Model 1: Adjusted for Age, Sex and Clinic			Model 2: Model 1 + Significant Risk Factors from Table 1 (BMI, SBP, Ratio LDL/ HDL and Smoking status)		
Exposures *	δage (years)	(95% CI)	p-value	δage (years)	(95% CI)	p-value
Log Troponin T (SD units)	0.41	[0.13, 0.68]	0.0036	0.17	[-0.10, 0.44]	0.21
Log NTproBNP (SD units)	0.69	[0.42, 0.96]	<0.0001	0.76	[0.49, 1.03]	<0.0001
Presence of carotid plaque (yes/no)	0.59	[0.03, 1.15]	0.038	0.05	[-0.51, 0.60]	0.86
Mean pulse wave velocity (SD units)	0.82	[0.54, 1.11]	<0.0001	0.28	[-0.01, 0.58]	0.060
Mean pulse wave velocity (SD units) **	0.82	[0.54, 1.11]	<0.0001	0.60	[0.32, 0.88]	<0.0001

(*) Log Troponin T, Log NTproBNP and mean pulse wave velocity were transformed into SD units by subtracting the mean and dividing by the standard deviation so that their effects were comparable.

(**) Not adjusted for Systolic Blood Pressure (SBP).

Ratio of Low-Density Lipoproteins over High-Density Lipoproteins (Ratio LDL/HDL); Body Mass Index (BMI).

likely to be a generalizable marker of accelerated or decelerated cardiovascular ageing of an individual, and therefore robust and applicable across a number of healthcare settings.

In the context of this Russian study, it is striking to observe that the mean δage (5.32 years in the overall study) is consistently higher across all younger adults (ranging from 35 to 60 years) and the strength of the association with both risk factors and markers of cardiac abnormalities described in this study increases when including only these younger adults in the models (data not shown). Previous studies showed that reduced telomere length, a biomarker of ageing, is associated with an increase in risk of CVD and all-cause mortality in an adult Russian cohort[4]. Furthermore, the difference between the ECG-predicted age from this model and chronologic age (δage) has been shown to be a strong independent predictor of cardiovascular and all-cause mortality in complementary work done in a separate population[26]. Combining these results with the presented data on CVD risk factors and markers of cardiac damage one could hypothesise that this novel biomarker contains information which might help us understand the causes of the abnormally high CVD mortality in working-age adults in Russia and its relationship to the observed average δage, although similar mortality studies would need to be performed in Russia to confirm this.

A weakness of this study is that it was based on a population sample with suboptimal response rates. However, we are primarily focussed on understanding associations within the studied population rather than wanting to generalise to the target population of all residents of the two cities being studied. To this extent the potential lack of representativeness may not be that important. Other potential weaknesses include the fact that the devices and protocols used to measure ECGs were not the same in training and test samples (model design in the U.S.A and CVD risk analysis in Russia); however, despite that this non-random misclassification can add noise to our analysis, the results obtained from the CVD risk factors analysis suggest these measurements are robust with regards to expected direction of the associations. Finally, a strength of this study is the availability of a wide range of different CVD risk factors and markers of cardiovascular abnormalities.

Obtaining this novel proximal cardiovascular biomarker does not involve an added cost beyond performing a 12-lead ECG and therefore has the potential to be widely used in routine clinical settings, including in low- and middle-income countries and can be applied retrospectively to digitally-acquired or digitized ECGs. Currently, there are hand-held devices which can undertake clinically approved ECGs in order to diagnose cardiovascular phenotypes or illnesses such as atrial fibrillation and others. The use of these devices might open the door to even less costly real-time longitudinal assessments of cardiovascular health using deep-learning models[27].

Further research is needed to explore whether the δage biomarker has similar patterns of associations in other populations with different CVD contexts, as well as further explore and disentangle the determinants of the δage gap using established ECG features known to be affected by ageing and other CVD risk factors. Other research avenues might explore the genetic and epigenetic determinants of δage. For implementation in a clinical setting, it will be important to identify clinically relevant thresholds of this biomarker in order to establish CVD risk and streamline case management and prioritisation of patients.

Data availability

Data cannot be made available due to access needing to be reviewed by a steering committee upon request. All data used in this study can be requested through the Know Your Heart metadata portal (https://metadata.knowyourheart.science/) upon registration. Registered users are able to browse the variable-level metadata for the Know Your Heart study; however, the site does not provide access to the data themselves.

Bona fide researchers may apply to the study steering group (email: david.leon@lshtm.ac.uk) for an anonymised subset of the data. The primary criteria used by the steering group are scientific coherence of the proposed use of the data and the match between the stated aims of the proposed research and the variables requested. This means that applicants need to justify the sets of variables they wish to have. In some instances, the research topic of the application may be close to or replicates ongoing research either within the core team or by previous applicants. In these cases, this overlap would be communicated to the applicants with a suggestion for the way forward. More information can be found in the 'User Guide' document on the website.

References

1. Blackburn EH, Epel ES, Lin J: **Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection.** *Science.* 2015; **350**(6265): 1193–1198.
   PubMed Abstract | **Publisher Full Text**

2. Field AE, Robertson NA, Wang T, *et al.*: **DNA Methylation Clocks in Aging: Categories, Causes, and Consequences.** *Mol Cell.* 2018; **71**(6): 882–895.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Horvath S, Raj K: **DNA methylation-based biomarkers and the epigenetic clock theory of ageing.** *Nat Rev Genet.* 2018; **19**(6): 371–384.
   PubMed Abstract | **Publisher Full Text**

4. Stefler D, Malyutina S, Maximov V, *et al.*: **Leukocyte telomere length and risk of coronary heart disease and stroke mortality: prospective evidence from a Russian cohort.** *Sci Rep.* 2018; **8**(1): 16627.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Groenewegen KA, den Ruijter HM, Pasterkamp G, *et al.*: **Vascular age to determine cardiovascular disease risk: A systematic review of its concepts, definitions, and clinical applications.** *Eur J Prev Cardiol.* 2016; **23**(3): 264–274.
    PubMed Abstract | **Publisher Full Text**

6. D'Agostino RB Sr, Vasan RS, Pencina MJ, *et al.*: **General cardiovascular risk profile for use in primary care: the Framingham Heart Study.** *Circulation.* 2008; **117**(6): 743–753.
    PubMed Abstract | **Publisher Full Text**

7. Khalil Y, Mukete B, Durkin MJ, *et al.*: **A comparison of assessment of coronary calcium vs carotid intima media thickness for determination of vascular age and adjustment of the Framingham Risk Score.** *Prev Cardiol.* 2010; **13**(3): 117–121.
    PubMed Abstract | **Publisher Full Text**

8. Shaw LJ, Raggi P, Berman DS, *et al.*: **Coronary artery calcium as a measure of biologic age.** *Atherosclerosis.* 2006; **188**(1): 112–119.
    PubMed Abstract | **Publisher Full Text**

9. Voss A, Dietz R, Fiehring H, *et al.*: **High resolution ECG, heart rate variability and nonlinear dynamics: tools for high risk stratification.** In: *Proceedings of Computers in Cardiology Conference.*1993; 261–264.
    Publisher Full Text

10. Holmvang L, Lüscher MS, Clemmensen P, *et al.*: **Very early risk stratification using combined ECG and biochemical assessment in patients with unstable coronary artery disease (A thrombin inhibition in myocardial ischemia [TRIM] substudy). The TRIM Study Group.** *Circulation.* 1998; **98**(19): 2004–2009.
    PubMed Abstract | **Publisher Full Text**

11. Attia ZI, Kapa S, Lopez-Jimenez F, *et al.*: **Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram.** *Nat Med.* 2019; **25**(1): 70–74.
    PubMed Abstract | **Publisher Full Text**

12. Attia ZI, Kapa S, Yao X, *et al.*: **Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction.** *J Cardiovasc Electrophysiol.* 2019; **30**(5): 668–674.
    PubMed Abstract | **Publisher Full Text**

13. Attia ZI, Noseworthy PA, Lopez-Jimenez F, *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**(10201): 861–867.
    PubMed Abstract | **Publisher Full Text**

14. Attia ZI, Friedman PA, Noseworthy PA, *et al.*: **Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs.** *Circ Arrhythmia Electrophysiol.* 2019; **12**(9): e007284.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Cook S, Malyutina S, Kudryavtsev AV, *et al.*: **Know Your Heart: Rationale, design and conduct of a cross-sectional study of cardiovascular structure, function and risk factors in 4500 men and women aged 35-69 years from two Russian cities, 2015-18 [version 3; peer review: 3 approved].** *Wellcome Open Res.* 2018; **3**: 67.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Saunders JB, Aasland OG, Babor TF, *et al.*: **Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II.** *Addiction.* 1993; **88**(6): 791–804.
    PubMed Abstract | **Publisher Full Text**

17. Leon DA, Saburova L, Tomkins S, *et al.*: **Hazardous alcohol drinking and premature mortality in Russia: a population based case-control study.** *Lancet.* 2007; **369**(9578): 2001–2009.
    PubMed Abstract | **Publisher Full Text**

18. Iakunchykova O, Averina M, Kudryavtsev AV, *et al.*: **Evidence for a Direct Harmful Effect of Alcohol on Myocardial Health: A Large Cross-Sectional Study of Consumption Patterns and Cardiovascular Disease Risk Biomarkers From Northwest Russia, 2015 to 2017.** *J Am Heart Assoc.* 2020; **9**(1): e014491.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Touboul PJ, Hennerici MG, Meairs S, *et al.*: **Mannheim carotid intima-media thickness and plaque consensus (2004-2006-2011). An update on behalf of the advisory board of the 3rd, 4th and 5th watching the risk symposia, at the 13th, 15th and 20th European Stroke Conferences, Mannheim, Germany, 2004, Brussels, Belgium, 2006, and Hamburg, Germany, 2011.** *Cerebrovasc Dis.* 2012; **34**(4): 290–296.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Imahori Y, Frost C, Mathiesen EB, *et al.*: **Effect of adiposity on differences in carotid plaque burden in studies conducted in Norway and Russia: a cross-sectional analysis of two populations at very different risk of cardiovascular mortality.** *BMJ Open.* 2020; **10**(5): e036583.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

21. Morgan KE, Cook S, Leon DA, *et al.*: **Reflection on modern methods: calculating a sample size for a repeatability sub-study to correct for measurement error in a single continuous exposure.** *Int J Epidemiol.* 2019; **48**(5): 1721–1726.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Ioffe S, Szegedy C: **Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift**. In: Bach F, Blei D (eds) *Proceedings of the 32nd International Conference on Machine Learning.* Lille, France: PMLR, 2015; **37**: 448–456.
    Reference Source

23. Laurent S, Cockcroft J, Van Bortel L, *et al.*: **Expert consensus document on arterial stiffness: methodological issues and clinical applications.** *Eur Heart J.* 2006; **27**(21): 2588–2605.
    PubMed Abstract | **Publisher Full Text**

24. Spronck B, Heusinkveld MHG, Vanmolkot FH, *et al.*: **Pressure-dependence of arterial stiffness: potential clinical implications.** *J Hypertens.* 2015; **33**(2): 330–338.
    PubMed Abstract | **Publisher Full Text**

25. Benetos A, Laurent S, Hoeks AP, *et al.*: **Arterial alterations with aging and high blood pressure. A noninvasive study of carotid and femoral arteries.** *Arterioscler Thromb.* 1993; **13**(1): 90–97.
    PubMed Abstract | **Publisher Full Text**

26. Ladejobi AO, Medina-Inojosa J, Cohen MS, *et al.*: **The 12-Lead ECG as a Biomarker of Physiologic Age: A Validation Study**. *Press*.

27. Attia ZI, Jennifer D, John M, *et al.*: **Abstract 13447: Prospective Analysis of Utility of Signals From an Ecg-Enabled Stethoscope to Automatically Detect a Low Ejection Fraction Using Neural Network Techniques Trained From the Standard 12-Lead Ecg.** *Circulation.* 2019; **140**(Suppl_1): A13447.
    Reference Source