

1 **Capturing population differences in rates of vascular aging using a deep learning**
2 **electrocardiogram algorithm: a cross-sectional study**

3 Ernest Diez Benavente ^{1,2}

4 Francisco Lopez-Jimenez ³

5 Olena Iakunchykova ⁴

6 Sofia Malyutina ^{5,6}

7 Alexander Kudryavtsev ^{4,7}

8 Andrew Ryabikov ^{5,6}

9 Paul A. Friedman ³

10 Suraj Kapa ³

11 Pablo Perel ^{8,9}

12 Tom Wilsgaard ⁴

13 Henrik Schirmer ^{10,11,12}

14 Taane G. Clark ^{1,13}

15 Zachi I. Attia ³

16 David A. Leon ^{4,8,14}

17

18 ¹ Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine,
19 London, UK

20 ² Department of Experimental Cardiology, University Medical Center Utrecht, Utrecht,
21 Netherlands

22 ³ Department of Cardiology, Mayo Clinic College of Medicine, Rochester, MN, USA

23 ⁴ Department of Community Medicine, UiT Arctic University of Norway, Tromsø, Norway

24 ⁵ Research Institute of Internal and Preventive Medicine, Branch of Institute of Cytology and
25 Genetics, Siberian Branch of the Russia Academy of Sciences, Novosibirsk, 630090, Russian
26 Federation

27 ⁶ Novosibirsk State Medical University, Russian Ministry of Health, Novosibirsk, 630091, Russian
28 Federation

29 ⁷ Northern State Medical University, Arkhangelsk, Russian Federation

30 ⁸ World Heart Federation, Geneva, Switzerland.

31 ⁹ Centre for Global Chronic Conditions, London School of Hygiene & Tropical Medicine, London,
32 UK.

33 ¹⁰ Department of Cardiology, Akershus University Hospital, Lørenskog, Norway

34 ¹¹ Institute of Clinical Medicine, Campus Ahus, University of Oslo, Norway

35 ¹² Department of Clinical Medicine, UiT The University of Norway, Tromsø, Norway

36 ¹³ Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical
37 Medicine, London, UK

38 ¹⁴ International Laboratory for Population and Health, National Research University, Higher
39 School of Economics, Moscow, Russia

40 * Corresponding author: Dr Ernest Diez Benavente, Department of Experimental Cardiology,
41 University Medical Center Utrecht, Utrecht, Netherlands, telephone: +447749700604,
42 email: e.diezbenavente@umcutrecht.nl

43

44

45 **Abstract** (292 words)

46 Background: Cardiovascular event rates increase with age in all populations. This is thought to be
47 the result of multiple underlying molecular and cellular processes that lead to cumulative vascular
48 damage. Apart from arterial stiffness based on pulse wave velocity there are few other non-
49 invasive measures of this process of vascular aging. We have developed a potential biomarker of
50 vascular aging using deep-learning to predict age from a standard 12-lead electrocardiogram
51 (ECG). The difference between ECG predicted and chronological age (δ -age) can be interpreted
52 as a measure of vascular aging.

53 Methods: We use data collected in two cross-sectional studies of adults aged 40-69 years in
54 Norway and Russia to test the hypothesis that mean levels of δ -age, derived from a deep-
55 learning model trained on a US population, correspond to the known large differences in
56 cardiovascular mortality between the two countries.

57 Findings: Substantial differences were found in mean δ -age between populations: Russia-USA
58 (+5.2 years; 0.7, 10 IQR) and Norway-USA (-2.6 years; -7, 2 IQR). These differences were only
59 marginally explained when accounting for differences in established cardiovascular disease risk
60 factors.

61 Interpretation: δ -age may be an important biomarker of fundamental differences in cardiovascular
62 disease risk between populations as well as between individuals.

63 **Keywords:** Cardiovascular disease, Biomarkers, Aging, Machine Learning, Electrocardiogram

64

65

66 **Background**

67

68 Fatal and non-fatal cardiovascular disease (CVD) event rates, such as stroke and myocardial
69 infarction, increase steeply with age. This relationship is observed in all populations, independent
70 of mean level of risk, and is the basis for regarding CVD as intimately connected with biological
71 aging. Over the past decade much has been written about the concept of vascular aging
72 associated with increases in factors such as arterial stiffness¹ and the progressive build-up of
73 atherosclerotic lesions;² both of which have well-established associations with cardiovascular
74 event rates.^{3,4} It has been postulated that there are shared molecular mechanisms of aging that
75 promote macrovascular and microvascular pathologies⁵ including oxidative stress, mitochondrial
76 dysfunction and chronic low-grade inflammation.⁶ However, there are few population level
77 summary measures that attempt to quantify the cumulative overall effect of these various
78 underlying processes on vascular aging.

79 We have developed a novel potential biomarker of vascular aging. This arose from work to
80 predict age from a deep learning analysis of 12-lead ECGs.⁷ It was noted that while predicted
81 ECG-age was highly correlated with chronological age there were deviations between the two.⁷
82 This led to the proposal that the direction and magnitude of the difference between ECG-age and
83 chronological age (delta-age (δ -age)) may be a measure of relative vascular aging: if δ -age is
84 positive, it would suggest a higher cumulative vascular damage and therefore, higher rate of
85 vascular aging relative to the reference population from which the age algorithm was derived.
86 This has been supported by our subsequent work that has shown δ -age to be prospectively
87 associated with mortality,⁸ cardiovascular events⁹ and cross-sectionally with a range of
88 established CVD risk factors including smoking and blood pressure.¹⁰ There is also a strong
89 positive association between δ -age and pulse-wave velocity,¹⁰ an established marker of vascular
90 stiffness, which has been regarded as one of the few direct markers of vascular aging.

91 In contrast to the substantial clinical and experimental work around the underlying concept of
92 vascular aging in individuals, little attention has been given to exploring the idea that populations
93 with different rates of CVD might be regarded as having different rates of vascular aging. It is
94 theoretically plausible that populations with high compared to low CVD event rates may have
95 different rates of vascular aging. To address this hypothesis we compare mean δ -age from
96 several populations that have contrasting rates of CVD mortality and life expectancy.

97

98 **Results**

99

100 We compared δ -age in two recent population-based studies of adults aged 40-69 years: the
101 Know Your Heart (KYH; 2015-18) study in Russia and the 7 study (T7; 2015-16) in Norway. To
102 predict ECG-age we employed a convolutional neural network (CNN) model trained on 700,000
103 ECGs from the Mayo Clinic (USA) (see Figure 1a). By definition the predicted age for the Russian
104 and Norwegian populations was relative to the reference clinical population in the USA. Rates of
105 CVD mortality in Russia are over seven times greater than in Norway in the 40-69 year age
106 group. Given these differences in CVD mortality we hypothesised that the mean difference
107 between ECG predicted and chronological age (δ -age) for Russia would be positive, while the
108 equivalent mean value of δ -aged for Norway would be negative. We also investigated how far any
109 differences in mean delta-age between the Russian and Norwegian study populations could be
110 explained by differences in known CVD risk factors.

111 Digitised ECGs were available for 5,459 T7 participants and 3,973 KYH participants aged
112 between 40 and 69 years. After exclusion of records with missing data ($N_{T7}=74$, $N_{KYH}=486$) a final
113 set of 5,385 T7 and 3,487 KYH participants were retained for analysis. The average age of the
114 participants was 59 ± 8.3 years (54.7% female) for T7 and 56 ± 8.6 years (58.4% female) for
115 KYH (Table S1). Mean δ -age was +5.2 years (0.7, 10 IQR) for the KYH (Russia) study
116 population, (Figure 1b) and -2.6 (-7, 2 IQR) for T7 (Norway) study population (Figure 1c). The

117 mean difference in δ -age between KYH and T7 studies (δ -age gap_{KYH-T7}) was +7.8 years (95%
118 CI: 7.5, 8.1) showing that the Russian KYH participants had an average ECG-age that was nearly
119 8 years older than Norwegian T7 participants (Table 1). Having adjusted for both chronological
120 age and sex, the gap in δ -age between the two study populations (KYH-T7) was 7.0 years (95%
121 CI: 6.7, 7.2; $p < 0.0001$).

122 The *loess* smoothed fitted line for KYH (Figure 1b) was above the line of identity at all ages. For
123 T7 (Figure 1c) the fitted line was below the line of identity from age 50 years. However, the δ -age
124 gap_{KYH-T7} was positive across the entire population, as shown by the fitted line for KYH being
125 above that of T7 for the entire range of chronological age shown, the gap being slightly larger at
126 older ages. There were small differences in δ -age between men and women in KYH (women
127 were predicted 0.3 years younger on average, $p = 0.28$) and T7 (0.3 years, $p = 0.15$).

128 The distribution of δ -age in the two studies (Figure 1d) showed a marked displacement of the T7
129 population to the left of the KYH population. In contrast there is far less separation of the
130 distribution of established CVD predictors in the two study populations (Figures 1e-g), although
131 differences in the mean values are slightly higher in KYH than in T7. The degree of separation of
132 these various distributions may be quantified in terms of a dissimilarity index that does not
133 depend upon the shape of the distribution.¹¹ This distribution has a range from zero (identical) to
134 1 (no overlap). The dissimilarity index for δ -age was 0.61 (95% CI: 0.59 – 0.62), compared with
135 0.13 (95% CI: 0.11 – 0.16) for SBP, 0.21 (95% CI: 0.19 – 0.23) for BMI, and 0.20 (95% CI: 0.18
136 – 0.22) for LDL/HDL ratio, reinforcing the notion of δ -age having increased power to cross-
137 sectionally differentiate between populations presenting higher CVD risk than established CVD
138 risk factors.

139 The established CVD risk factors SBP, BMI, ratio LDL/HDL and smoking were associated with δ -
140 age in both studies (Table S2). Adjustment for these cross-sectionally measured risk factors
141 along with sex and age reduced the gap in δ -age between studies to 6.5 years (95% CI: 6.2, 6.8;
142 $p < 0.0001$) (Table 1).

143 Finally, we adjusted for a larger set of established CVD-related risk factors which also showed
144 significant differences between KYH and T7 (Table S1). These were diastolic blood pressure,
145 pulse rate, haemoglobin A1c (HbA1c) and education level. This further adjustment reduced the
146 gap in δ -age between studies to six years difference (95% CI: 5.7, 6.3; $p < 0.0001$), with a slight
147 reduction in the gap when further adjusting for history of hypertension and diabetes and the final
148 model accounted for 25.8% of the gap in δ -age between studies (Table 1).

149

150 Discussion

151

152 These results confirm our hypothesis that mean difference between predicted age and
153 chronological age (δ -age) in the study populations in Russia was substantially different to the
154 difference in the Norwegian study, consistent with the much higher CVD mortality rates in Russia
155 compared to Norway. Moreover, relative to the US clinical reference population mean δ -age was
156 higher in KYH (Russia) but lower in T7 (Norway), in line with the intermediate position of CVD
157 mortality in the US between Russia and Norway. However, while the reference population in the
158 US was from a clinical series, the Russian and Norwegian studies were population-based studies.
159 As discussed elsewhere, CVD mortality rates in the Norwegian and Russian cities from which the
160 study participants were drawn were very close to their respective national averages.¹²

161 Showing that mean δ -age is so much higher in the Russian compared to Norwegian study
162 population strongly suggests that δ -age may be a robust biomarker of relative vascular aging
163 between populations as well as being a predictor of mortality and associated with CVD risk
164 factors within populations. While CVD mortality is not the only determinant of differences in life
165 expectancy between countries, it is notable that the difference in δ -age between Russia and
166 Norway (7.8 years) is similar to the difference in life expectancy at birth between the two
167 countries (8.8 years in 2016).

168 Adjustment for a wide range of CVD risk factors attenuated the difference between KYH and T7
169 by just under 20% (from 6-95 to 5-64 years). This is consistent with recent work that suggests
170 that a core set of established CVD risk factors were able to explain only a quarter of the observed
171 differences in CVD mortality between Russia and Norway at ages 40-69 years.¹³ There are a
172 number of explanations for the inability of cross-sectional risk factor profiles to explain differences
173 in CVD mortality *per se* that may also apply to the moderate attenuation of differences in δ -age.
174 These explanations include within person variability in the risk factor levels from single cross-
175 sectional measures such that they provide inadequate assessments of long-term risk factor levels
176 integrated over the life course.¹³

177 As we have demonstrated for a selected set of established CVD risk factors, there is far greater
178 overlap in their distributions between the Russian and Norwegian studies compared with that for
179 δ -age. It is therefore not that surprising that they have limited scope to explain differences in δ -
180 age between the populations.

181 The appreciable difference between T7 and KYH in mean δ -age is reflected in the shift to the left
182 of the whole T7 distribution of δ -age relative to KYH. How can we interpret this? In contrast to
183 established predictive risk factors, the information from the ECG captured by the CNN generated
184 algorithm for predicted age could be regarded as a direct expression of the accumulation of age-
185 related damage or impairment to the vascular tree and its effects on cardiac structure and
186 function. From this perspective, δ -age may reflect a net integrated effect on vascular aging that is
187 the result of many different processes and cumulative exposures to both known and unknown risk
188 factors. In this sense this integrated measure of age-related changes in cardiac structure and
189 function may be far more proximal to the ultimate causes of CVD events and mortality than are
190 single or even multiple measures of predictive risk factors.

191 The specific ECG features used by the δ -age algorithm remain unclear. Recent advances have
192 been able to highlight the segments of an ECG that a CNN algorithm uses to make predictions¹⁴.
193 However, the unresolved functioning of the algorithm does not prevent the objective evaluation of

194 its potential utility, especially when studying CVD risk at a population level, where the benefit
195 obtained from the ability to measure CVD risk trends is likely to outweigh any biases at the
196 individual level which could have a more direct effect if this measure was used in a clinical
197 setting. Through the use of epidemiological approaches, we have identified the ability of δ -age to
198 replicate between population patterns of CVD risk and mortality suggesting that it captures
199 fundamental average differences in cardiac structure/function within populations.

200 We believe this biomarker may enable us to close some of the knowledge gap between
201 established risk factors and predictors and proximal measures of CVD risk at a population level
202 (Figure 1h). Further studies would require population level data from more diverse populations, in
203 order to ascertain the ability of δ -age to reproduce country-level CVD event trends as well as
204 using cohorts with both raw ECG and CVD events follow-up data. Additionally, ECG signals could
205 be used outcomes in GWAS studies to determine potential genetic contributions to δ -age and
206 provide insights into new biological pathways relevant to vascular aging. Finally, the use of this
207 marker if further validated may enable novel studies of CVD risk in low- and middle-income
208 countries where resources for more complex phenotyping are scarce.

209 **Materials and Methods**

210

211 We used data obtained in two population studies, the Tromsø 7 (T7), the 7th survey of the
212 Tromsø Study in Tromsø, Norway¹⁵ and the Know Your Heart (KYH) study, a cross-sectional
213 survey of a random sample of the population of two Russian cities: Arkhangelsk and
214 Novosibirsk.¹² A total of 21,083 men and women aged 40-99 years were recruited from the
215 general population, as described in detail elsewhere.¹⁶ The KYH study participants were
216 comprised of 4,542 men and women aged 35-69 years recruited from the general population
217 between 2015 and 2018 as described in detail elsewhere.¹²

218

219 For T7 65% of those contacted and invited to take part attended the health check. A randomly
220 selected set of 13,028 participants (9,925 randomly selected and 3,103 selected based on
221 participation in previous rounds) were invited for a second visit where a resting 10 seconds digital
222 12-lead ECG was obtained using the Schiller device AT104 PC and the raw ECG signal was
223 stored digitally. For KYH of those contacted and invited to take part 47% attended the health
224 check. A resting 10-25 seconds digital 12-lead ECG was obtained using the Cardiax device
225 (IMED Ltd, Hungary) and the raw ECG signal was stored digitally.

226

227 A convolutional neural network (CNN) model using Keras with a Tensorflow (Google, Mountain
228 View, CA) backend was previously developed and validated. A total of 774,783 unique subjects
229 with ECGs were used to develop the neural network: 399,750 in the training set, 99,977 in the
230 internal validation set and 275,056 ECGs in the holdout testing set. The network contained
231 stacked blocks of convolutional, max pooling, and batch normalization.¹⁷ A detailed description of
232 the network is described in our previously published paper.⁷

233

234 We used linear regression models to assess association of the risk factors as exposures using
235 δ age as the outcome. All models were adjusted for a priori confounders including chronologic age
236 and sex. True chronologic age was included in every model to remove the effects of any potential
237 correlation between δ -age and chronologic age. A test for trend for smoking was performed by
238 converting the ordered categories of number of cigarettes smoked into integer values. Basic
239 models including one risk factor at a time were further adjusted for potential confounders.

240 Density plots were generated for δ -age, SBP, BMI and Ratio LDL/HDL. In order to assess the
241 overlap of the density plots for each variable in the populations we used the R package
242 *overlapping* with a 100 bootstraps per variable.^{11,18}

243 For further details see Supplementary Methods.

244

245 **Acknowledgments**

246

247 We would like to thank all the team that undertook the field work in the Know Your Heart and
248 Tromsø 7 studies and all the participants and George Davey Smith for helpful comments on an
249 earlier draft. Part of DAL's contribution was undertaken within the framework of the Basic
250 Research Program of the National Research University Higher School of Economics.

251

252 **Funding**

253

254 TGC received funding from the MRC UK (Grant no. MR/K000551/1, MR/M01360X/1,
255 MR/N010469/1, MR/R020973/1) and BBSRC UK (BB/R013063/1). Know Your Heart (KYH) is a
256 component of the International Project on Cardiovascular Disease in Russia (IPCDR) funded by a
257 Wellcome Trust Strategic Award [100217], UiT The Arctic University of Norway (UiT), Norwegian
258 Institute of Public Health, and Norwegian Ministry of Health and Social Affairs. SM, AR are
259 supported by Russian Academy Sciences (AAAA-A17-117112850280-2). The funding bodies had
260 no role in the design of the study, data collection, analysis, interpretation of data, or in writing the
261 manuscript.

262

263 **Author Contributions:** EDB, PP, DAL and FJL conceived the study, DAL, TGC, HS and TW
264 designed the study, EDB performed the statistical analysis under the supervision of DAL and
265 TGC. ZIA contributed to the analysis and interpretation of the machine learning analysis under
266 the supervision of FJL, SK and PAF. AK contributed to the acquisition, analysis and interpretation
267 of the data from Arkhangelsk in the Know Your Heart Study, SM and AR contributed to the
268 acquisition, analysis and interpretation of data in the Know your Heart study, HS, OY and TW
269 contributed to the acquisition, analysis and interpretation of data in the Tromsø 7 study. EDB
270 wrote the first draft of the manuscript. All authors commented and edited on various versions of
271 the draft manuscripts and gave final approval to the submitted version and agree to be
272 accountable for all aspects of work ensuring integrity and accuracy.

273 **Competing Interest Statement:** The authors declared no conflicts of interest.

274

275 **References**

276

- 277 1 Donato AJ, Machin DR, Lesniewski LA. Mechanisms of Dysfunction in the Aging
278 Vasculature and Role in Age-Related Disease. *Circ Res* 2018; **123**: 825–48.
- 279 2 de Groot E, Hovingh GK, Wiegman A, *et al.* Measurement of arterial wall thickness as a
280 surrogate marker for atherosclerosis. *Circulation* 2004; **109**: III33-8.
- 281 3 Zhong Q, Hu M-J, Cui Y-J, *et al.* Carotid-Femoral Pulse Wave Velocity in the Prediction of
282 Cardiovascular Events and Mortality: An Updated Systematic Review and Meta-Analysis.
283 *Angiology* 2018; **69**: 617–29.
- 284 4 Bamberg F, Sommer WH, Hoffmann V, *et al.* Meta-analysis and systematic review of the
285 long-term predictive value of assessment of coronary atherosclerosis by contrast-
286 enhanced coronary computed tomography angiography. *J Am Coll Cardiol* 2011; **57**:
287 2426–36.
- 288 5 Ungvari Z, Tarantini S, Sorond F, Merkely B, Csiszar A. Mechanisms of Vascular Aging, A
289 Geroscience Perspective: JACC Focus Seminar. *J Am Coll Cardiol* 2020; **75**: 931–41.
- 290 6 Zoltan U, Stefano T, J. DA, Veronica G, Anna C. Mechanisms of Vascular Aging. *Circ Res*
291 2018; **123**: 849–67.
- 292 7 Attia ZI, Friedman PA, Noseworthy PA, *et al.* Age and Sex Estimation Using Artificial
293 Intelligence From Standard 12-Lead ECGs. *Circ Arrhythmia Electrophysiol* 2019; **12**:
294 e007284.
- 295 8 Adetola L, Jose M-I, Michal S, *et al.* ECG-derived age and survival: validationg the

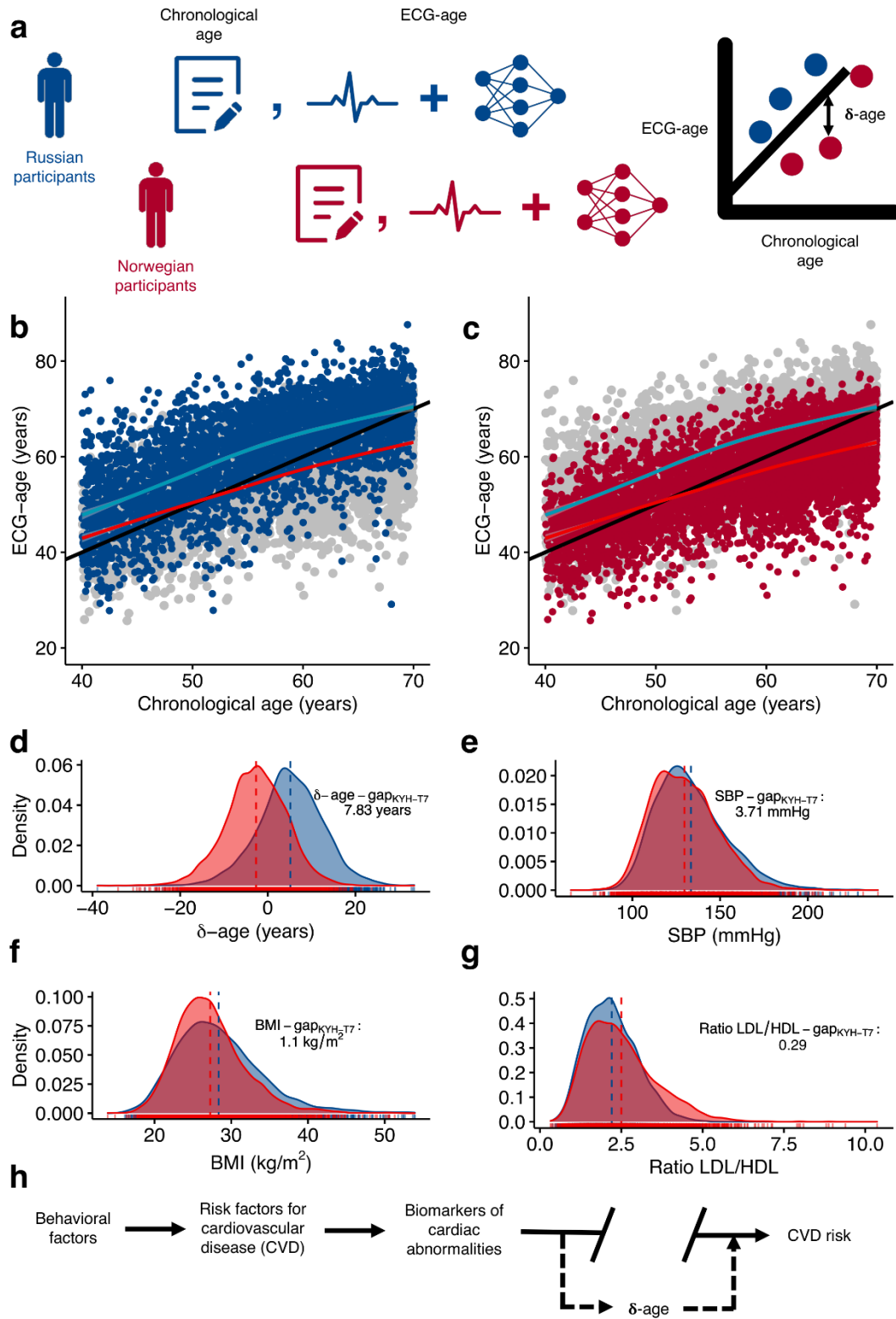
- 296 concept of physiologic age detected by ECG using artificial intelligence. *J Am Coll Cardiol*
297 2020; **75**: 3469.
- 298 9 Medina-Inojosa J, Ladejobi A, Attia Z, *et al*. The association of artificial intelligence-
299 enabled electrocardiogram-derived age (physiologic age) with atherosclerotic
300 cardiovascular events in the community. *Eur Heart J* 2020; **41**.
301 DOI:10.1093/ehjci/ehaa946.2905.
- 302 10 Diez Benavente E, Jimenez-Lopez F, Attia ZI, *et al*. Studying accelerated cardiovascular
303 ageing in Russian adults through a novel deep-learning ECG biomarker [version 1; peer
304 review: awaiting peer review]. *Wellcome Open Res* 2021; **6**.
305 DOI:10.12688/wellcomeopenres.16499.1.
- 306 11 Pastore M, Calcagni A. Measuring Distribution Similarities Between Samples: A
307 Distribution-Free Overlapping Index . *Front. Psychol.* . 2019; **10**: 1089.
- 308 12 Cook S, Malyutina S, Kudryavtsev A V, *et al*. Know Your Heart: Rationale, design and
309 conduct of a cross-sectional study of cardiovascular structure, function and risk factors in
310 4500 men and women aged 35-69 years from two Russian cities, 2015-18 [version 3; peer
311 review: 3 approved]. *Wellcome Open Res* 2018; **3**.
312 DOI:10.12688/wellcomeopenres.14619.3.
- 313 13 Trias-Llimós S, Pennells L, Tverdal A, *et al*. Quantifying the contribution of established risk
314 factors to cardiovascular mortality differences between Russia and Norway. *Sci Rep* 2020;
315 **10**: 20796.
- 316 14 van de Leur RR, Blom LJ, Gavves E, *et al*. Automatic Triage of 12-Lead ECGs Using
317 Deep Convolutional Neural Networks. *J Am Heart Assoc* 2020; **9**: e015138.
- 318 15 Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njølstad I. Cohort profile: the
319 Tromso Study. *Int J Epidemiol* 2012; **41**: 961–7.

- 320 16 Wynn R, Oyeyemi SO, Budrionis A, Marco-Ruiz L, Yigzaw KY, Bellika JG. Electronic
321 Health Use in a Representative Sample of 18,497 Respondents in Norway (The Seventh
322 Tromsø Study - Part 1): Population-Based Questionnaire Study. *JMIR Med Inf* 2020; **8**:
323 e13106.
- 324 17 Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by
325 Reducing Internal Covariate Shift. In: Bach F, Blei D, eds. Proceedings of the 32nd
326 International Conference on Machine Learning. Lille, France: PMLR, 2015: 448–56.
- 327 18 Inman HF, Bradley EL. The overlapping coefficient as a measure of agreement between
328 probability distributions and point estimation of the overlap of two normal densities.
329 *Commun Stat - Theory Methods* 1989; **18**: 3851–74.

It is made available under a [CC-BY 4.0 International license](#) .

330

331 **Figures**



332
333

334 **Figure 1. A novel digital ECG biomarker of vascular aging for population level assessment**
335 **of CVD risk. a)** Schematic followed to obtain ECG-age and form participants of both Tromsø 7
336 (T7) and Know Your Heart (KYH) studies, chronological age was obtained via questionnaire data
337 and ECG-age was obtained by performing an, at least, ten second resting 12-lead ECG and
338 processing the ECG raw data through a CNN model trained on a USA clinical population to
339 predict age. δ -age was obtained by subtracting the chronologic age to the ECG-age. **b)**
340 Scatterplot of ECG-age versus chronological age for 3,487 participants form KYH (Russia) in blue
341 and 4,704 T7 participants from Norway in grey. Locally weighted smooth line was presented in
342 light blue for KYH and light red for Norway. **c)** Scatterplot of ECG-age versus chronological age,
343 in this panel dark red dots represent T7 participants (Norway) and grey dots represent KYH
344 participants (Russia). This panel presents the same data as panel **b)** with different colouring for
345 clarity. **d)** Density plot of δ -age for Russia and Norway, δ -age-gap is not adjusted. **e)** Density plot
346 of systolic blood pressure (SBP) for Russia and Norway, SBP-gap is not adjusted. **f)** Density plot
347 of body mass index (BMI) for Russia and Norway, BMI-gap is not adjusted. **g)** Density plot of ratio
348 of low-density lipoprotein (LDL) cholesterol over high density lipoprotein (HDL) cholesterol for
349 Russia and Norway, Ratio LDL/HDL-gap is not adjusted. **h)** Proposed theoretical framework
350 depicting δ -age in the causal pathway of cardiovascular disease (CVD) as a proximal biomarker
351 of CVD risk which stores information about both behavioural and risk factors as an accumulated
352 exposure over time. Note: The line of identity in panels **b)** and **c)** can be seen as a proxy for the
353 American population, representing an average δ -age of +0.44 years.⁷
354
355

356 **Tables**
357

358 **Table 1. Table of explanatory linear models for the absolute difference in δ -age between**
359 **the T7 and KYH studies (δ -age-gap_{KYH-T7}).** The second column (δ -age-gap_{KYH-T7} explained) is
360 the difference between the unadjusted δ -age-gap_{KYH-T7} and the δ -age-gap_{KYH-T7} of each adjusted
361 model, reported as a percentage of the unadjusted δ -age-gap_{KYH-T7}.

362

Model	δ -age-gap _{KYH-T7} (years)	95% CI	p-value	δ -age-gap _{KYH-T7} explained (%)
Model 1: Not adjusted	7.8	[7.5, 8.1]	<0.0001	-
Model 2: Chronological age and sex	7	[6.7, 7.2]	<0.0001	7.8
Model 3: Model 2, SBP, BMI, Smoking and Ratio LDL/HDL adjusted	6.5	[6.2, 6.8]	<0.0001	16.5
Model 4: Model 3, DBP, Pulse rate, HbA1c, Education	6	[5.7, 6.3]	<0.0001	23.2
Model 5: Model 4, history of hypertension and history of diabetes adjusted	5.8	[5.5, 6.2]	<0.0001	25.8

363
364

