Multi-omics in research: epidemiology, methodology, and advanced data analysis

Faquih, T.O.

Chapter 4
Robust metabolomic age prediction based on a wide selection of metabolites

Key words: metabolomics, ridge regression, prediction, aging, sleep, cardiometabolic disease, depression

Tariq Faquih[1], Astrid van Hylckama Vlieg[1], Praveen Surendran[2,3,4,5], Ruifang Li-Gao[1,6], Renée de Mutsert[1], Frits R. Rosendaal[1], Raymond Noordam[7], Diana van Heemst[7], Ko Willems van Dijk[8,9,10], Dennis Mook-Kanamori[1,11]

[1] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; T.O.Faquih@lumc.nl (T.O.F.); R.Li@lumc.nl (R.L.-G.); R.de_Mutsert@lumc.nl (R.d.M); F.R.Rosendaal@lumc.nl (F.R.R.); A.van_Hylckama_Vlieg@lumc.nl (A.v.H.V.); D.O.Mook@lumc.nl (D.O.M.-K.)
[2] British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; ps629@medschl.cam.ac.uk (P.S.)
[3] British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK; ps629@medschl.cam.ac.uk (P.S.)
[4] Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK; ps629@medschl.cam.ac.uk (P.S.)
[5] Rutherford Fund Fellow, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; ps629@medschl.cam.ac.uk (P.S.)
[6] Metabolon, Inc. Morrisville, North Carolina, United State of America; R.Li@lumc.nl (R.L.-G.)
[7] Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands; R.Noordam@lumc.nl (R.N.); D.van_Heemst@lumc.nl (D.v.H.)
[8] Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; K.Willems_van_Dijk@lumc.nl (K.W.v.D)
[9] Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands; K.Willems_van_Dijk@lumc.nl (K.W.v.D)
[10] Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands; K.Willems_van_Dijk@lumc.nl (K.W.v.D)
[11] Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands; D.O.Mook@lumc.nl (D.O.M.-K.)

Manuscript in preparation

1 ABSTRACT

Chronological age is a major risk factor for numerous diseases. However, the biological aging process is complex and often does not align with chronological age. Metabolite levels are thought to reflect the integrated effects of both genetic and environmental factors, including age, and thus may provide a signature for biological age. Here, we set out to develop a rigorous metabolomic age prediction model by applying ridge regression and bootstrapping with 826 metabolites (of which 678 endogenous and 148 xenobiotic) measured in 11977 individuals (age range: 18-75 years old) from the INTERVAL study (Cambridge, UK). Subsequently, the metabolomic age prediction model was applied in the Netherlands Epidemiology of Obesity Study (NEO) (n=599) to quantify the difference between metabolomic and chronological age (Δ age). We assessed the influence of cardiovascular disease (CVD), hypertension, type 2 diabetes, obesity/body mass index (BMI), depression, and sleep duration on the Δ age. The metabolomic age models using both endogenous and xenobiotic metabolites demonstrated high correlation with chronological age (R^2=0.82). In NEO, CVD associated with increased Δ age by approximately 12 years and obese BMI (45kg/m^2) associated with increased Δ age by approximately 8.5 years, respectively. In summary, we developed robust models for predicting metabolomic age in a large relatively healthy population with a wide age range. We further demonstrated that Δ age can potentially reflect the effects of CVD and obesity/BMI in the NEO study.

2 INTRODUCTION

Chronological age is a major risk factor for a multitude of diseases (1, 2). The biology of aging is a complex and multifactorial process that is influenced by lifestyle and environmental factors (3-5). However, it is evident that the rate of aging varies between individuals, wherein some individuals are able to live to an older chronological age without age-related diseases and disability compared with individuals in the same age group (3).This suggests that chronological age does not fully align with the biological aging process. Thus, several studies aimed to capture the signature of biological changes due to aging by predicting age using biological factors. Such studies used DNA methylation (6) and proteins (7).

Metabolomic profiling aims to identify small molecules that are mostly substrates and products of cell metabolism (metabolites). The number of metabolomics studies has increased in recent years due to major technological advances and availability of commercial and noncommercial analyses platforms. In addition, the current platforms have improved capability to detect and quantify large numbers of endogenous and xenobiotic metabolites. Since individual metabolomic profiles reflect the influences of both genetic and acquired factors, they are thought to provide a holistic representation of biological processes, such as aging (8, 9). Furthermore, metabolomic profiles are strongly affected by chronological age (4, 10) and sex (11, 12), they have been used to develop prediction models of chronological age (i.e., the metabolomic age) (13-16). However, predicting metabolomic age using metabolomics had limited success or faced methodological limitations due to several reasons. First, some studies used an insufficiently small sample size to predict age using hundreds or even thousands of metabolite predictors (10, 17). The inclusion of larger number of predictors than the sample often causes large overfitting and bias in such models. Second, studies can be limited by the age distribution of the cohort study. This restricts the model to a specific age range which affects the generalizability of the model in other studies and other age groups (10). Third, lack of generalizability can also be caused if the model was developed in cohorts with an oversampling of individuals with specific disease outcomes or a specific population (7, 13). Fourth, even when the previous considerations are addressed, oversimplified statistical methodology and application can lead to a flawed, biased, and an overfitted model—such as the case with stepwise selection models(18). Finally, the model's validity and generalizability are seldom examined in external studies or different populations via external validation (19-21).

In this study, we aimed to develop a model to predict metabolomic age in a large healthy population with a widespread age range, using a single untargeted metabolomics platform. To attain this goal, we developed a prediction model using data from the INTERVAL study (22) (University of Cambridge, Cambridge, UK). Metabolomic profiles were available in 11977 participants as measured using Metabolon's (Durham, North Carolina, USA) untargeted metabolomics platform. These measurements included a broad range of endogenous and xenobiotic metabolites (n=1,363) from various biochemical pathways, thereby enabling the capture of metabolites related to a vast range of ageing effects. The model was subsequently applied to the Netherlands Epidemiology of Obesity Study (NEO) (n=599) to determine the effects of six health-related phenotypes associated with aging (23, 24): cardiovascular disease, hypertension, type 2 diabetes, obesity and body weight, depression, and sleep duration on the difference between metabolomic age and chronological age (14, 15).

3 METHODS

3.1 INTERVAL Study

The INTERVAL study is a prospective cohort study of approximately 50,000 participants nested within a pragmatic randomized sample of blood donors (22). Between 2012 and 2014, blood donors, aged 18 years and older, were consented and recruited from 25 National Health Service Blood and Transplant (NHSBT) static donor centers across the UK. Individuals with major disease (myocardial infarction, stroke, cancer etc.) as well as those who reported being unwell or having had recent illness or infection or did not fulfill the other criteria required for blood donation (22, 25) were ineligible for the study. Therefore, participants included in the study were predominantly healthy. Participants completed online questionnaires addressing basic lifestyle and health-related information, including self-reported height and weight, ethnicity, current smoking status, alcohol consumption, doctor-diagnosed anemia, use of medications (hormone replacement therapy, iron supplements) and menopausal status (22). Untargeted metabolomic data were available in 11979 individuals (age range 18 - 75). Two individuals had incorrect or missing height/weight values and were therefore excluded from the study. Thus, the final sample size for the current study was 11977.

3.2 Untargeted metabolomic measurements

Untargeted metabolomic measurements were quantified at Metabolon Inc. (Durham, North Carolina, USA) using Metabolon™ Discovery HD4 platform. In brief, this process involves four independent ultra-high-performance liquid chromatography mass spectrometry (UHPLC-MS/MS) platforms (26, 27). Two using positive ionization reverse phase chromatography, one using negative ionization reverse phase chromatography, and one using hydrophilic interaction liquid chromatography negative ionization (27). Known metabolites were annotated at Metabolon Inc. with chemical names, super pathways, sub pathways, biochemical properties, and compound identifiers from various metabolite databases. Metabolomic measurements in the INTERVAL study were conducted in three batches (n=4087, 4566, and 3326). Subsequent harmonization and quality checks were performed between the batches by Metabolon. All metabolite measurements were scaled to a median of 1.

3.3 Selection of Predictors

We aimed to select metabolites consistently and reliably measured by the Metabolon platform. In total 1411 metabolites were measured in the INTERVAL study. First, we removed metabolites completely missing in at least one of the batches (n=175). Second, we excluded metabolites without annotation/unnamed (n=258), keeping only endogenous and xenobiotic metabolites. These metabolites were excluded as they are inconsistently measured by the platform and are highly variable between batches and studies. Moreover, the lack of full annotation increases the uncertainty that we are using the same metabolite across batches and studies and leaves no secondary information for further verification. Third, we excluded metabolites measured in less than 100 individuals (n=30). Fourth, we excluded metabolites that were not measured in the NEO study (n=122). The final set of metabolites included 826 metabolites, with 678 endogenous and 148 xenobiotic metabolites (Figure 1).

3.4 Missing value imputation

Missing values were imputed using the pipeline as described in our previous work (28). In brief, endogenous metabolites were imputed by multiple imputation using chained equations method to generate five imputed datasets (m=5). For each metabolite with missing values, we used the

outcome variable (i.e., age), 5-10 highly correlated related metabolites, body mass index (BMI), center number where the blood samples were collected, and the batch number to impute the missing values. Xenobiotic metabolites were imputed to zero to account for true missingness.

3.5 Model Development

For the development of the model, we used ridge regression (29) to reduce potential overfitting. Cross validation was performed (n=10) in each imputed dataset to calculate the optimal shrinkage term (lambda). Subsequently the mean of the lambda values was used to develop the model on the stacked imputed datasets (i.e., all 5 datasets combined as one). Accordingly, the weight of the observations in the stacked dataset was set to 1/m = 1/5 = 0.2 (20). As the outcome (age) is a continuous variable, we assessed the fit of the model by deriving the R^2. As an additional sensitivity test, we used generalized additive model (GAM) to examine and calculate the R^2 for the nonlinear correlation. Internal validation was performed using bootstrapping (b=100). Bootstrapping results were used for the optimization of R^2 and the calculation of the mean squared error (MSE) and the mean absolute error (MAE) of model. Two models were developed using two sets of the selected metabolite predictors. First, we used the full set of endogenous and xenobiotic metabolites (n=826) to develop "model A". Second, we used only the endogenous metabolites (n=678) to develop "model B". As previous studies found that metabolomic profiles (12, 30) and aging (31-33) are influenced by the sex of individuals, we included sex as an additional predictor in both models.

3.6 Sample size considerations

The primary database used to create the prediction model was the INTERVAL study (n= 11977) with 826 predictors. we used the formulas described by Riley et al. (34) to confirm that our sample size (n) and number of predictor (p) are sufficient to minimize overfitting and provide high precision. First, we used n and p to check that the calculated Copas global shrinkage factor(35, 36) is above the recommended 0.9 threshold (34). Based on this calculation the estimated shrinkage factor was 0.95 if the adjusted R^2 of the model was assumed to be 0.7. Second, we calculated the sample size required to ensure a small difference between the R^2 and the adjusted R^2 for the development model. Assuming the adjusted R^2 was 0.7 again and a small desired R^2 difference (R^2_{diff}=0.025), then the sample size required to achieve this should be at least n=9913. Third, we checked the sample size required for precise residual standard deviation of the model. Accordingly, we found the multiplicative margin of error (MMOE) to be less than 10% (MMOE =1.3%) using our n and p in the INTERVAL study. Finally, we checked the precision of the mean predicted outcome value (predicted age) of the model. We used n and p for the INTERVAL study and assumed that predicted age would have a mean of 45 and a variance of 35. Accordingly, the upper and lower bounds were approximately 45.34 and 44.65 respectively. Thus, the MMOE for the mean predicted outcome was less 1% (MMOE= 45.34 /45 = 1.007 = 0.7%). Therefore, the sample size of the INTERVAL study was optimal to minimize overfitting, optimism, and provide a precise estimation of the residual standard deviation and mean predicted values.

3.7 Metabolomic Age and Health-Related Phenotypes

3.7.1 NEO Study

The Netherlands Epidemiology of Obesity (NEO) study is a population-based, prospective cohort study of individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. Men and women aged between 45 and 65 years with a self-reported BMI of 27 kg/m^2

or higher, living in the greater area of Leiden (in the West of the Netherlands) were eligible to participate in the NEO study. In addition, all inhabitants aged between 45 and 65 years from one municipality (Leiderdorp) were invited, irrespective of their BMI. Recruitment of participants started in September 2008 and completed at the end of September 2012. In total, 6,671 participants have been included. Participants were invited to come to the NEO study center of the LUMC for one baseline study visit after an overnight fast. A blood sample of 108 mL was taken from the participants after an overnight fast of at least 10 hours. Untargeted metabolomics measurements are available in the 599 individuals from Leiderdorp sub-population (single batch). Therefore, only 599 individuals were included in this study. All metabolite measurements were scaled to a median of 1. Missing values in the metabolites measurements was conducted using the same method in the INTERVAL study using a single imputation.

3.7.2 Health-Related phenotypes in the NEO study

Several health-related phenotypes, including cardiometabolic risk factors, are associated with both metabolomic profiles (12, 37) and aging (23, 24). In the NEO study, individual patient data was available for the following health-related phenotypes: cardiovascular disease (CVD), type 2 diabetes (T2D), hypertension, BMI, depressive symptoms, and total sleep duration. At baseline of the NEO study, 21 individuals had self-reported cardiovascular disease, which was defined as a composite trait composed of myocardial infarction, angina, congestive heart failure, stroke, and peripheral vascular disease. In addition, 89 individuals had type 2 diabetes, based on impaired fasting glycaemia (6.1-7.0 mmol/L), fasting plasma glucose higher or equal to 7.0 mmol/L, or self-reported diabetes mellitus 1 or 2 medication. Hypertension was characterized in 213 individuals, who had systolic blood pressure >=140mmHg or diastolic blood pressure >=90mmHg. BMI was calculated based on physical examination measurements and 79 participants had a BMI in the obese range (>35 kg/m^2). Depression score (1-84) was derived from the Inventory of Depressive Symptomatology (IDS) questionnaire. We defined current depression status as a score of 14 and above (23). Using the IDS questionnaire, 111 participants were characterized as depressed at baseline. Total Sleep duration was derived from self-reported questionnaire; specifically, from the question: “on an average day, how much sleep do you get?”. We further categorized sleep duration using the 5th percentile to define the shortest sleep duration, 5th to 20th as short, 20th to 80th as medium, 80th to 95th as long, and 95th and above as the longest sleep duration. Mean sleep duration was 7 hours of which the majority (71%) reported a total sleep duration between 6 and 8 while short sleep duration (5.5-6 hours) accounted for 15.9%.

3.7.3 Estimating the Effects of Health-Related Phenotypes on Metabolomic Age

The developed prediction model was applied to derive metabolomic age for 599 individuals in the NEO study with metabolomics data. To assess the effects of health-related phenotypes on the difference between metabolomic age and chronological age—henceforth referred to as Δ age—with the available 6 health-related phenotypes. For a better interpretation of the BMI effect on the Δ age, we centered it on the median healthy BMI range (BMI = 22 kg/m^2) as defined by the World Health Organization(38). We used a simple linear regression analysis for each of the six clinical phenotypes as exposures separately. The Δ age was the outcome variable in each analysis.

4 RESULTS

4.1 Population characteristics

Characteristics of the total INTERVAL study population and age subgroups are summarized in Table 1. In total 11977 individuals were included and had a normal age distribution with a mean age of 45 years and a range of 18 – 75 years. The number of men and women was approximately equal (50.2% were men). BMI was largely in the recommended range of 18.5-24.9 kg/m^2 (38) with a mean BMI of 22.8 kg/m^2 and was similar in all age groups.

Characteristics of the NEO study population are summarized in Table 2. The mean age was 56 years and ranged from 45 to 66. Mean BMI was 25.9, slightly over the recommended BMI (38) and was similar in men and women.

Table 1. Characteristics of the INTERVAL study population.

	Total	18 to 25	25 to 35	35 to 45	45 to 55	55 to 65	65 to 75
N	11977	1178	2245	2152	2867	2602	811
Age (years), mean (range)	45 (18-75)						
Men, n (%)	6019 (50.2%)	485 (41.2%)	965 (42.3%)	1042 (48.4%)	1560 (54.4%)	1480 (56.9%)	546 (67.3%)
BMI (kg/m^2), mean (SD)	22.8 (4.2)	21.3 (4.0)	22.0 (4.2)	23.3 (4.4)	23.5 (4.3)	23.0 (3.9)	22.6 (3.5)

Table 2. Characteristics of the NEO study population.

	Total
n	599
Age (years), mean (range)	56 (45-65)
Men, n (%)	284 (47.7)
BMI (kg/m^2), mean (SD)	25.9 (4.0)
Obesity, n (%)	79 (13.2)
CVD, n (%)	21 (3.5)
T2D, n (%)	89 (14.9)
Hypertension, n (%)	213 (35.6)
Depression, n (%)	111 (19)
Total Sleep Duration(hours), mean (SD)	7 (0.98)
Shortest (<5.5), n (%)	40 (6.7)
Short (5.5 - 6), n (%)	95 (15.9)
Medium (6 - 8), n (%)	425 (71.0)
Long (8 - 8.5), n (%)	16 (27)
Longest (>8.5), n (%)	21 (3.5)

4.2 Metabolomic Age Prediction

Prediction models A (endogenous plus xenobiotic metabolites) and B (endogenous metabolites only) were developed in the INTERVAL study using ridge regression. The workflow including metabolite selection, missing value imputation, and analyses are summarized in Figure 1. Internal validation using bootstrapping (b=100) and optimization provided an R^2 of 0.83 (MSE=31, MAE=4.4) for model A, and 0.82 (MSE=33.7, MAE=4.6) for model B. GAM R^2 was slightly higher for both models, 0.85 for model A and 0.84 for model B (Figure 2, Supplementary Table 1). Full tables with the intercept, sex, and metabolite coefficients for model A and B are provided in supplementary Table 2. This table also contains the mean values for the metabolites from the INTERVAL study.

Figure 1: Flowchart of the selection of predictor metabolites and the development steps for the metabolomic age prediction model in the INTERVAL study

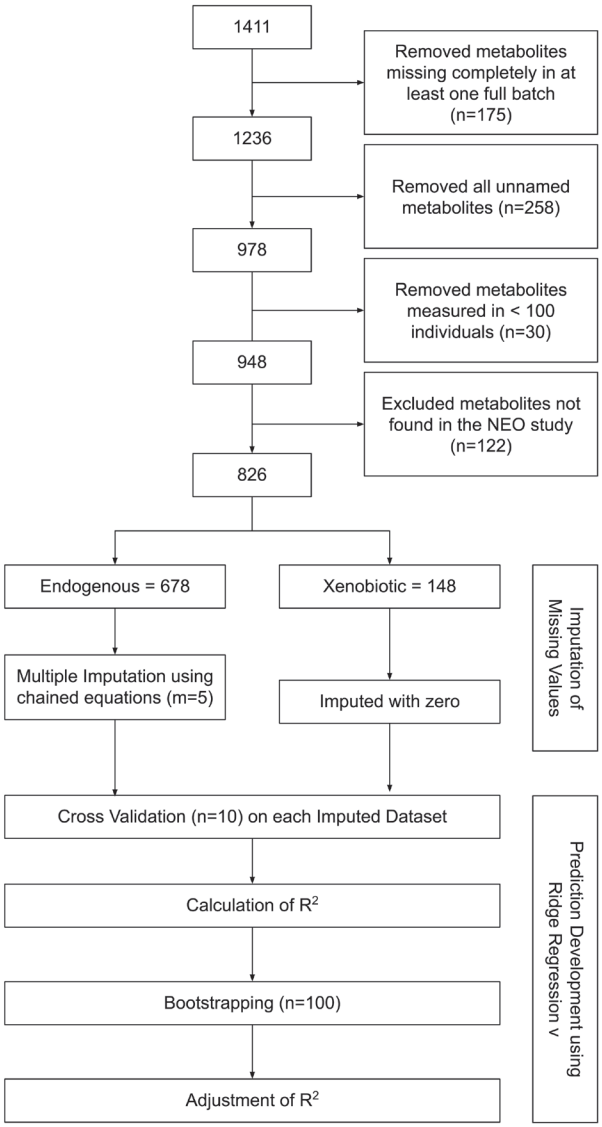
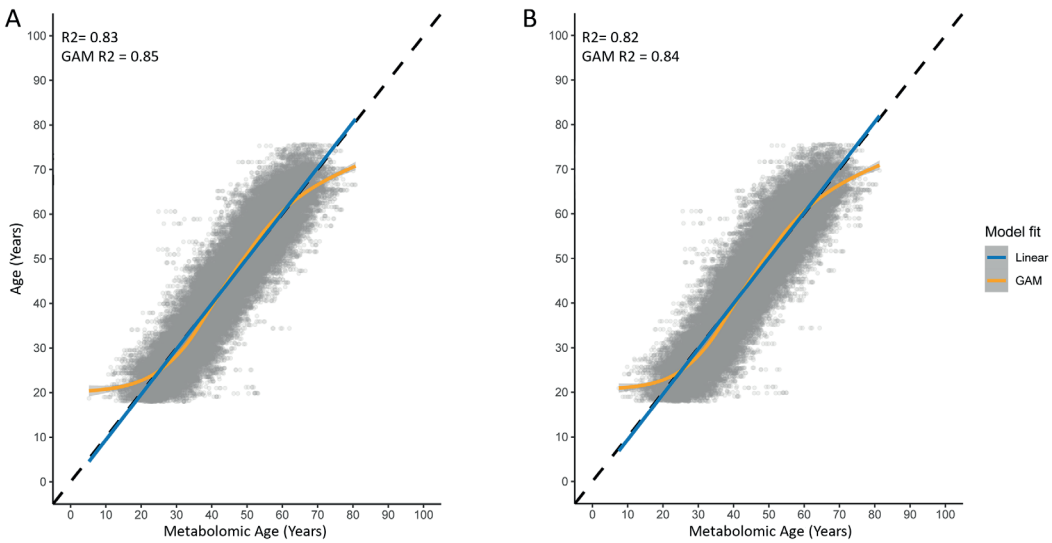


Figure 2: Correlation plots of the Metabolomic age (predicted age) on the horizontal axis and the chronological age on the vertical axis for model A (A) and model B (B). The data used is the stacked imputed datasets in the INTERVAL study. Abbreviations: GAM, generalized additive model.



4.3 Chronological age versus Metabolomic Age

The coefficient estimates from model A and B from the INTERVAL study were subsequently applied to the NEO data to predict metabolomic age. We first performed calibration in the large in the NEO data to readjust the intercept (model A from 51.97 to 50.83; model B from 51.85 to 49.59). Subsequently, the effect of the phenotypes CVD, BMI/obesity, T2D, hypertension, depression, and total sleep duration on Δ age were assessed, for both model A and model B. First, we plotted Δ age against chronological age and highlighted individuals with each respective phenotype for model A and model B (Figure 3 A-F; and Figure 4 A-F). Overall, model A showed higher variance in Δ age (σ^2=170) compared with model B (σ^2=84.1), notably for individuals with a negative health trait (i.e., with CVD or depression etc.).

Figure 3: Scatter plot representing the age difference (Δ age)—as predicted using model A—on the y-axis and chronological age on the x-axis in the NEO study. Each subplot highlights individuals with specific phenotypes.

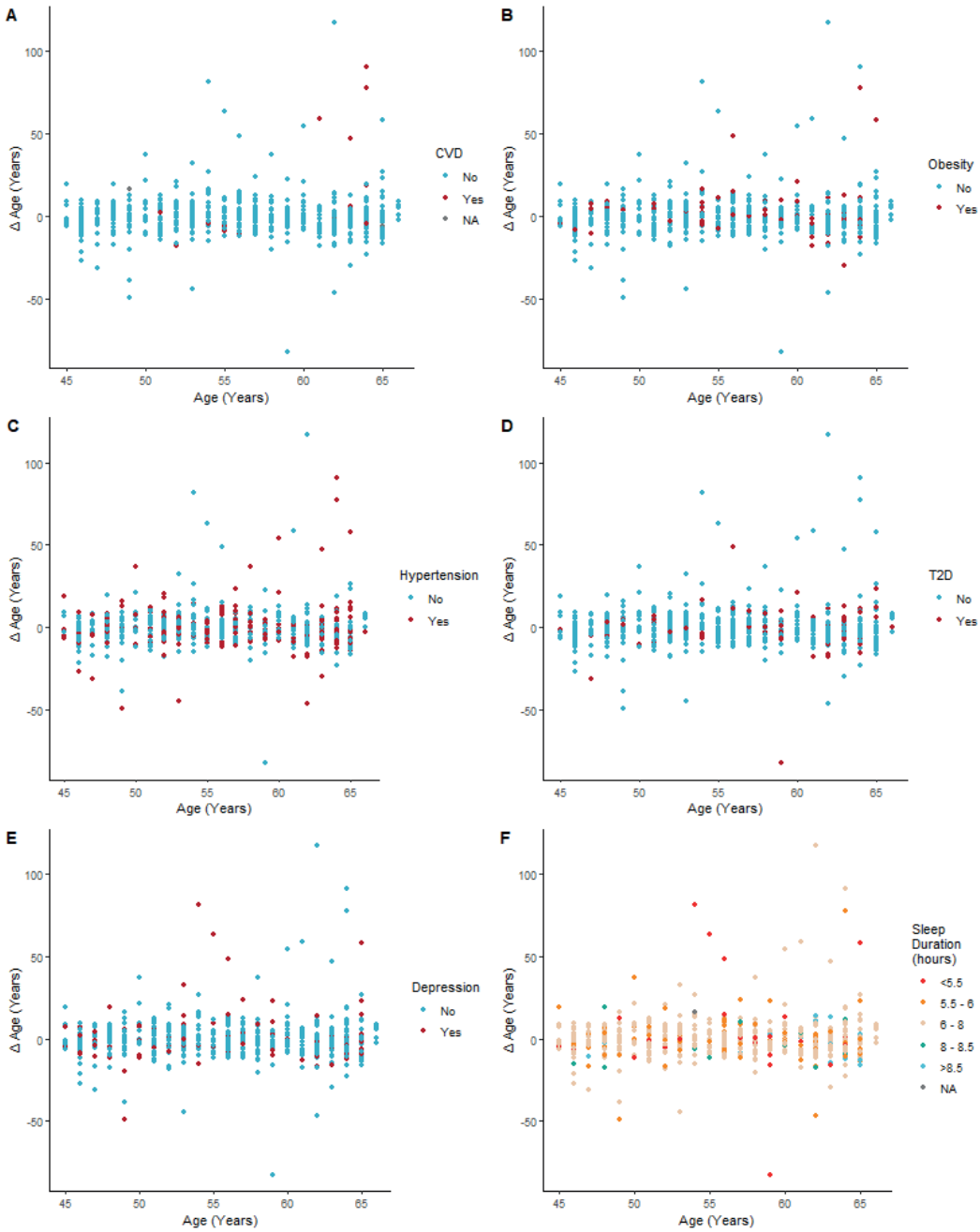
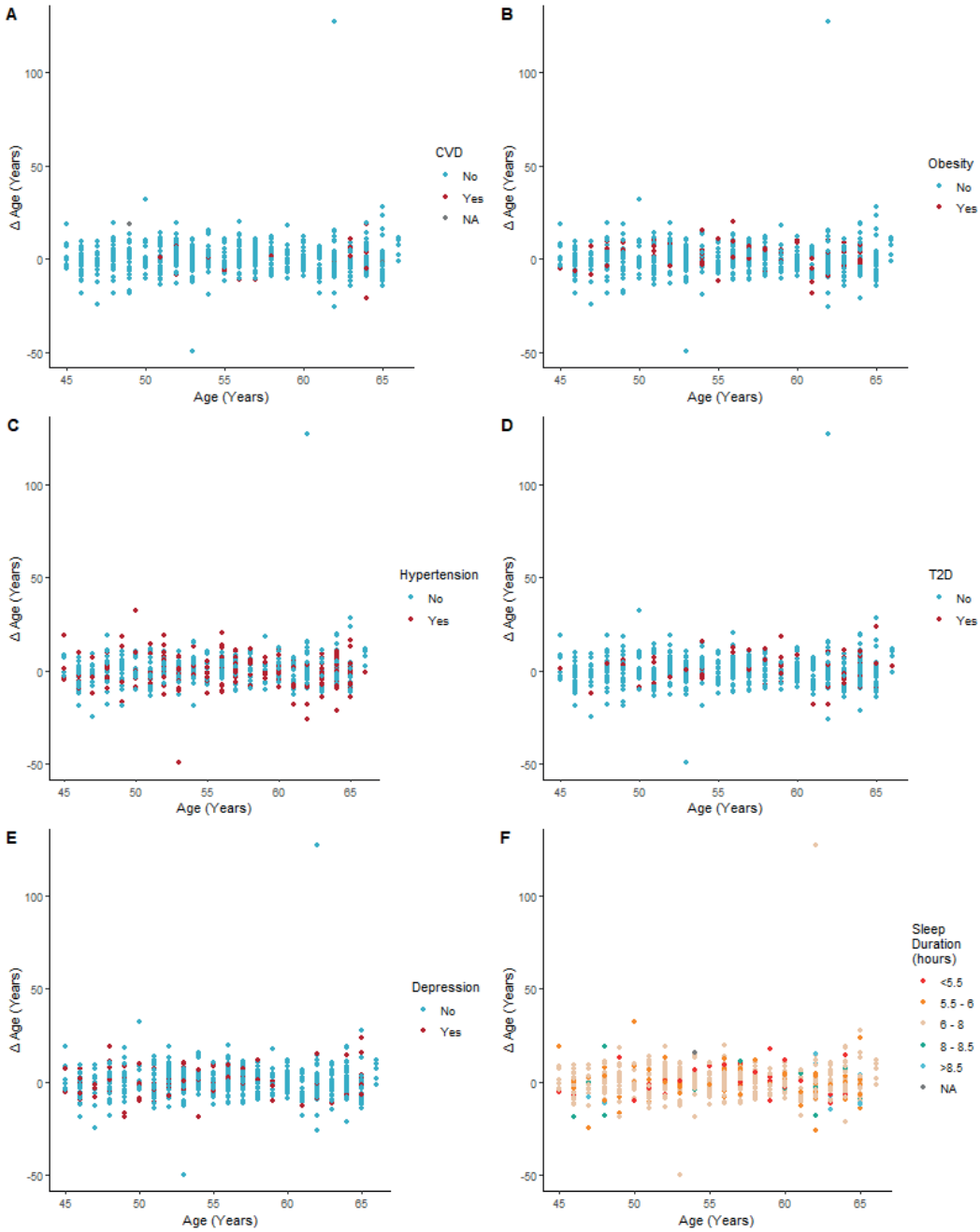


Figure 4: Scatter plot representing the age difference (Δ age)—as predicted using model B—on the y-axis and chronological age on the x-axis in the NEO study. Each subplot highlights individuals with specific phenotypes.



Second, we assessed the effects of the health-related phenotypes on the Δ age using linear regression. These results are shown in Figure 5 and Supplementary Table 3. In both models, BMI was associated with an increase of Δ age (beta [95% confidence intervals (CI)]: model A = 0.37 [0.1 – 0.64]; model B = 0.25 [0.06 – 0.43] years/ kg/m2). Obesity was also associated with an increase of Δ age but with wide CI in both models (model A = 3.31 [0.14 – 6.49]; model B = 1.89 [-0.27 – 4.06]). Thereafter, using the estimates from models A and B, we calculated the age at BMI's 22, 25, 35, and 45 kg/m2 (Figure 6). Overall, model A showed higher Δ age estimations at all BMI levels. At 35 kg/m2, the projected Δ age increased by 3.7 and 2.4 years in models A and B, respectively, as compared with 25 kg/m2. Whereas at BMI of 45 kg/m2, the projected Δ age rose by 8.5 and 5.6 years in models A and B, respectively, compared with BMI of 25 kg/m2.

CVD was strongly associated with an increased Δ age in model A (12.13 years [6.34 – 17.92]) but not model B (-0.89 years [-4.89 – 3.09]). No associations were found in either models for T2D, hypertension, depression, or total sleep duration.

Figure 5: A forest plot of estimates of phenotypes with the Δ age using model A (A) and model B (B).

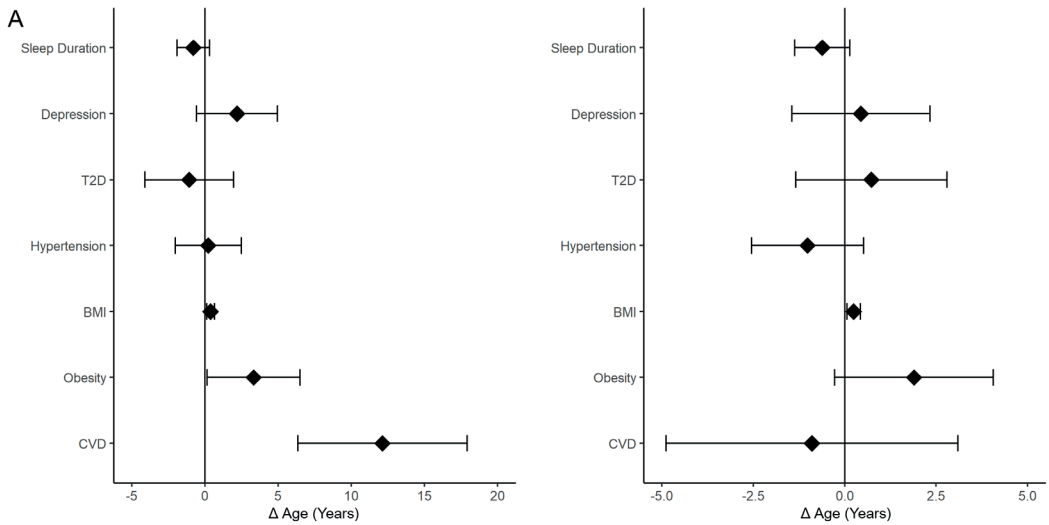
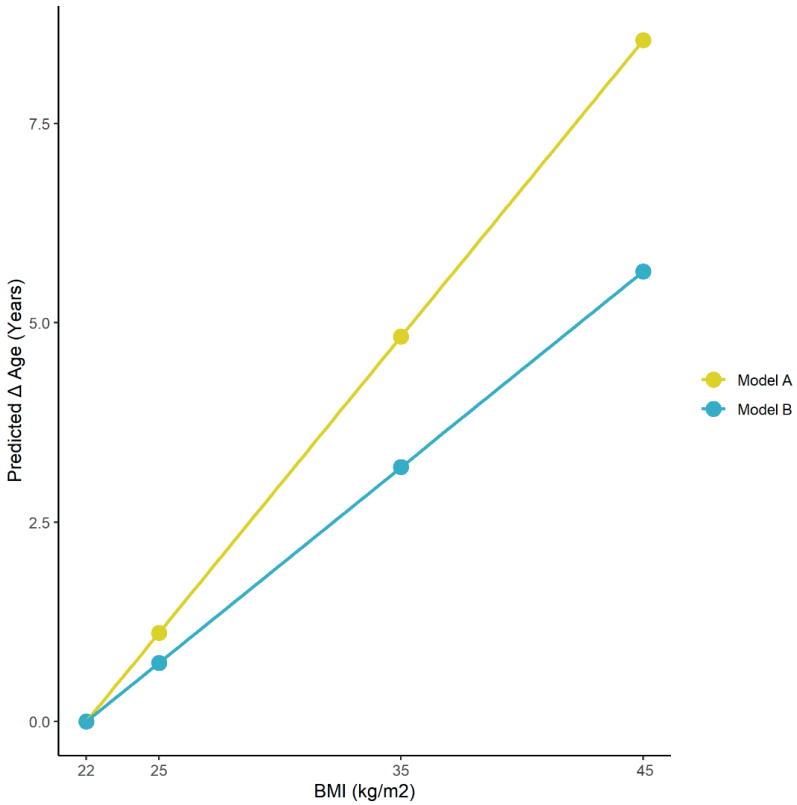


Figure 6: Predicted Δ age at 22,25,35, and 45 kg/m2 based on the estimates from model A and model B of BMI in the NEO study



5 DISCUSSION

In this study, we developed a prediction model for metabolomic age based on metabolite measurements, including a wide range of endogenous and xenobiotic metabolites belonging to a variety of biochemical pathways. The metabolomic measurements were performed in a single study using the same metabolomic platform and harmonized for within study and between batch variation. Importantly, we used multiple imputation, ridge regression, and bootstrapping(20) to develop and internally validate the metabolomic age prediction models. We developed two models, with the first (model A) using both endogenous and xenobiotic metabolites (n=826) and the second (model B) using the endogenous metabolites only (n=678). Both models had high adjusted R^2 (model A = 0.83; Model B = 0.82), however, model B had slightly higher MSE and MAE, indicating higher error for the predicted values.

5.1 Ridge Regression for Metabolomic Age

Our metabolomic age model is based on ridge regression which generates shrinkage factors that shrink metabolite coefficients with weak influence on the model to small values. Unlike methods such as LASSO and elastic net regression, this shrinkage never causes the coefficients to reach zero. Therefore, unlike those methods, ridge regression does not perform selection of predictors/metabolites due to the shrinkage term. Thus, the model consistently includes the

same metabolites during the cross validation and internal validation by bootstrapping as well as during redevelopment and external validation in studies such as NEO (20). Ridge regression thus ensures rigorousness and consideration of all the selected metabolite predictors. In our model we have found the nonlinear GAM R^2 to be slightly higher than linear R^2 in both models. Therefore, predicting metabolomic age from this model in a nonlinear form, by adding a quadratic term or using a spline function, can further expand the model flexibility.

5.2 The effect of health traits on metabolomic age

After developing the metabolomic age prediction models, we used the NEO study for two purposes. First, we used the NEO study to assess the viability of the prediction models. This was done by estimating the metabolomic age in the NEO study and readjusted the intercepts of the models accordingly (calibration in the large). The readjustment of the intercepts was minimal in the NEO metabolomic age models. Second, we used the models in NEO to examine the influence of six health phenotypes on the Δ age. We expected that the cases with negative health phenotypes had a higher Δ age (BMI/obesity, T2D, hypertension, CVD, depression, and short sleep). On the other hand, we expected no effect or a lower Δ age for individuals who did not suffer from these phenotypes or reported longer sleep durations.

Overall, Model A showed wider Δ age distributions compared with model B. As expected, our results showed that BMI, obesity, and particularly CVD had a strong effect on the Δ age using model A. Despite the equivalent R^2 for model A and B in the INTERVAL study, only BMI was associated with the Δ age estimated from model B. Since model A includes xenobiotic metabolites, medication related metabolites are likely important for capturing the effects of CVD and obesity on metabolomic age. The effect of BMI on age is pronounced in our estimation of Δ age using BMI's of 25, 35 and 45 kg/m^2. Indeed, at an obese BMI of 45 kg/m^2 the predicted increase in Δ age was 8.5 and 5.6 years for models A and B, respectively.

None of the other health traits; sleep duration, depression, T2D, or hypertension affected Δ age in either model. This was unexpected as for example T2D and hypertension are associated with CVD risk and have increased incidence rates in older individuals (24, 39). In addition, a previous study on metabolic age has shown that metabolites were a strong predictor of T2D and CVD (14). One possible reason for this is the exclusion of insulin dependent T2D individuals and those with a history of major disease outcomes such as myocardial infarction and stroke in the INTERVAL study due to the blood donor eligibility criteria in UK (40). This exclusion criteria may have affected the ability of metabolomic age, particularly in model B, to reflect these phenotypes in the Δ age in the NEO study, which did not have such exclusions. However, Δ age from model A showed a clear association with CVD. Therefore, other factors, such as the small sample size or unknown factors, could be involved possibly in addition to the exclusion criteria in the INTERVAL study. Further investigation is required in a larger sample size to verify the metabolomic age and the factors affecting the models. Depression was similarly not associated with the metabolomic age developed in a previous study. In contrast, they found an association between depression and their epigenetic, proteomic, and transcriptomic age predictors in the same population (7). Regarding the sleep duration results, some studies reported a U-shaped association between sleep and health outcomes (41). In this study we only examined the linear association in the NEO study. Thus, further examination of different association patterns for sleep are needed in future studies. Finally, we avoided examining etiological associations between the metabolites with Δ age and the six phenotypes as this was beyond the scope of paper. However, further exploration of etiological associations for specific metabolites with metabolomic age and health related phenotypes could be of interest for future studies.

5.3 Strengths and Limitations

A major strength of our current study is the development of robust metabolomic age prediction models based on a large number of metabolites measured in a large cohort with a wide age distribution. This sample size was confirmed to fit the required criteria for developing a model with low overfitting (34). Furthermore, the INTERVAL study included relatively healthy blood donors as participants, making it a suitable study to develop a metabolomic age without being affected with selection bias based on specific disease-related inclusion criteria. Unlike previous metabolomic age predictors (alternatively referred to as metabolomic "clocks"), we developed our model using the latest Metabolon metabolomics platform that measures a large selection of metabolites. A benefit of this platform is the inclusion of xenobiotics, such as those derived from medication or pollution as well as endogenous metabolites. Thus, we were able to include metabolites originating from internal and external sources. The xenobiotic metabolites represent some of the acquired environmental and lifestyle exposures of individuals that could play a role in biological aging.

Regarding the limitations, because the INTERVAL study does not have patient data for health-related outcomes, we could not assess the effects of different disease outcomes and health phenotypes on the metabolomic age. A second limitation was the small sample size included from of the NEO study relative to the number of predictors (n<p) and the narrower age range. Therefore, applying the models in the NEO study lead to loss in power and affected the reliability and performance of the models (20). Another limitation was the low number of events in some of the selected phenotypes in the NEO study. This was the case for CVD (n=21), T2D (n=89) and short/long sleep duration. It is possible that the larger number of metabolites from endogenous and xenobiotics may capture the metabolomic phenotype of some of the health traits more accurately in larger sample size. In addition, the results for sleep duration could have been affected by the sleep variable used. In NEO sleep duration was derived from a self-reported questionnaire, which is prone to recall bias. For future studies, sleep duration and quality quantification can be improved by using objectively measured actigraphy data.

5.4 External Validation for Metabolomic Age and Future Applications

Prediction models may suffer from overfitting due to selection bias, small sample size, methodology limitations, and lack of internal validation and calibration during development. However, another common issue regarding prediction modelling, such as the case with metabolomic age, is the challenge to externally validate them. This a common issue with prediction models in general. Indeed, few studies perform external validation of prediction models (19, 42). The reasons for the lack of external validation include the of difficulty applying and reproducing the prediction model method, lack of the full prediction variables to develop the model in the new dataset, or a lack of an appropriate effective sample size for external validation. Without external validation the quality of the models cannot be properly assessed, and a model could still be overfitted despite presenting good results during internal validation (20, 21, 43). We took advantage of the sample size and wide age range to use a stringent ridge regression and internal validation method to develop the metabolomic age model. The resulting model demonstrated a high R^2 for the metabolomic age and, as shown in NEO, was influenced by CVD and BMI. Accordingly, future robust external validation, using weak and moderate calibrations (43) would be valuable for the metabolomic age models presented in this paper. Furthermore, examination of the association between different phenotypes, such as those used in the NEO study, in the external validation would be valuable to provide more statistical power to assess their influence on metabolomic age and Δ age. Indeed, this could improve and verify the assessment of the phenotypes known

to influence metabolomic profiles (14, 24, 39, 44) and predicted metabolomic age as reported in our study and previous studies.

Several age prediction models have been developed that utilize different biological measurements such as targeted metabolomics (7, 14), proteomics (7), and DNA methylation (6). Here, we used the Metabolon untargeted metabolomics platform for the prediction of metabolomic age. In addition to our aim of addressing the primary issues with the development of metabolomic age models, the Metabolon platform expanded on the range of metabolites that can potentially be a better predictor of age. For example, previous metabolomic age studies that could not measure or include xenobiotic metabolites. In our study, we were able to include this additional group of metabolites in model A. We found that model A's Δ age, but not model B's Δ age, was greatly increased by CVD and obesity. This apparent additional predictive value of xenobiotics can be further investigated in future studies. Furthermore, Model A may also be used in tandem with metabolomic age from targeted platforms, and biological clocks of different biological molecules measurements similar to the work by Jansen et al. (7), to possibly improve or compare their predictive performance and their ability of capturing the effects of health-related phenotypes.

6 CONCLUSIONS

We developed metabolomic age prediction models in a large relatively healthy population using a wide array of endogenous and xenobiotic metabolites. In model A with the endogenous and xenobiotic metabolites and in model B with endogenous metabolites only, the R^2 of the linear fit was 0.82 and 0.83, respectively. The hypotheses that the predicted metabolomic age reflects metabolomic age and that health-related phenotypes increase metabolomic age was subsequently tested in the NEO study. These analyses revealed that obesity and CVD increased metabolomic age only in the model A, indicating possibly higher predictive value from external influences as reflected by the xenobiotic metabolites. We provided the full list of metabolites and their coefficients for both models. This data can enable other researchers to replicate our metabolomic age prediction model, externally validate it in their own studies with different disease outcomes and combine them with other age prediction models.

7 REFERENCES

1. North BJ, Sinclair DA. The intersection between aging and cardiovascular disease. Circ Res. 2012;110(8):1097-108.
2. Broglio SP, Eckner JT, Paulson HL, Kutcher JS. Cognitive decline and aging: the role of concussive and subconcussive impacts. 2012;40(3):138.
3. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell. 2013;153(6):1194-217.
4. Hoffman JM, Lyu Y, Pletcher SD, Promislow DEL. Proteomics and metabolomics in ageing research: from biomarkers to systems biology. Essays Biochem. 2017;61(3):379-88.
5. Brooks-Wilson AR. Genetics of healthy aging and longevity. 2013;132(12):1323-38.
6. Horvath S. DNA methylation age of human tissues and cell types. Genome Biology. 2013;14(10):3156.
7. Jansen R, Han LK, Verhoeven JE, Aberg KA, van den Oord EC, Milaneschi Y, et al. An integrative study of five biological clocks in somatic and mental health. eLife. 2021;10.
8. Rattray NJW, Deziel NC, Wallach JD, Khan SA, Vasiliou V, Ioannidis JPA, et al. Beyond genomics: understanding exposotypes through metabolomics. Hum Genomics. 2018;12(1):4.
9. Alonso A, Marsal S, Julia A. Analytical methods in untargeted metabolomics: state of the art in 2015. Front Bioeng Biotechnol. 2015;3:23.
10. Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data. Nature Reviews Genetics. 2022.
11. Martin FJ, Montoliu I, Kussmann M. Metabonomics of ageing - Towards understanding metabolism of a long and healthy life. Mech Ageing Dev. 2017;165(Pt B):171-9.
12. Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, et al. Human serum metabolic profiles are age dependent. Aging Cell. 2012;11(6):960-7.
13. Macdonald-Dunlop E, Taba N, Klarić L, Frkatović A, Walker R, Hayward C, et al. A catalogue of omics biological ageing clocks reveals substantial commonality and associations with disease risk. Aging. 2022;14(2):623-59.
14. van den Akker EB, Trompet S, Barkey Wolf JJH, Beekman M, Suchiman HED, Deelen J, et al. Metabolic Age Based on the BBMRI-NL (1)H-NMR Metabolomics Repository as Biomarker of Age-related Disease. Circulation Genomic and precision medicine. 2020;13(5):541-7.
15. Hertel J, Friedrich N, Wittfeld K, Pietzner M, Budde K, Van der Auwera S, et al. Measuring Biological Age via Metabonomics: The Metabolic Age Score. J Proteome Res. 2016;15(2):400-10.
16. Rist MJ, Roth A, Frommherz L, Weinert CH, Kruger R, Merz B, et al. Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. PLoS One. 2017;12(8):e0183228.
17. Hwangbo N, Zhang X, Raftery D, Gu H, Hu S-C, Montine TJ, et al. A Metabolomic Aging Clock Using Human Cerebrospinal Fluid. The Journals of Gerontology: Series A. 2021;77(4):744-54.
18. Smith G. Step away from stepwise. Journal of Big Data. 2018;5(1):32.
19. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? Clinical Kidney Journal. 2020;14(1):49-58.
20. Steyerberg EW. Clinical Prediction Models. 2nd ed. Cham, Switzerland: Springer International Publishing; 2019 2019.
21. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research:: A clinical example. Journal of Clinical Epidemiology. 2003;56(9):826-32.
22. Moore C, Sambrook J, Walker M, Tolkien Z, Kaptoge S, Allen D, et al. The INTERVAL trial to determine whether intervals between blood donations can be safely and acceptably decreased to optimise blood supply: study protocol for a randomised controlled trial. Trials. 2014;15:363-.
23. Emami M, Agbaedeng TA, Thomas G, Middeldorp ME, Thiyagarajah A, Wong CX, et al. Accelerated

Biological Aging Secondary to Cardiometabolic Risk Factors Is a Predictor of Cardiovascular Mortality: A Systematic Review and Meta-analysis. Canadian Journal of Cardiology. 2022;38(3):365-75.
24. Buford TW. Hypertension and aging. Ageing research reviews. 2016;26:96-111.
25. NHS Blood Donation Who can give blood 2022 [updated 2022/10/07/. Available from: https://www.blood.co.uk/who-can-give-blood.
26. Evans A, Bridgewater B, Liu Q, Mitchell M, Robinson R, Dai H, et al. High resolution mass spectrometry improves data quantity and quality as compared to unit mass resolution mass spectrometry in high-throughput profiling metabolomics. 2014;4(2):1.
27. Rhee EP, Waikar SS, Rebholz CM, Zheng Z, Perichon R, Clish CB, et al. Variability of Two Metabolomic Platforms in CKD. Clinical Journal of the American Society of Nephrology. 2019;14(1):40.
28. Faquih T, van Smeden M, Luo J, le Cessie S, Kastenmüller G, Krumsiek J, et al. A Workflow for Missing Values Imputation of Untargeted Metabolomics Data. Metabolites. 2020;10(12).
29. Hoerl AE, Kennard RW. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. 1970;12(1):55-67.
30. Saner C, Harcourt BE, Pandey A, Ellul S, McCallum Z, Kao K-T, et al. Sex and puberty-related differences in metabolomic profiles associated with adiposity measures in youth with obesity. Metabolomics. 2019;15(5):75.
31. Hägg S, Jylhävä J. Sex differences in biological aging with a focus on human studies. eLife. 2021;10:e63425.
32. Nakamura E, Miyao KJTJoGSABS, Sciences M. Sex differences in human biological aging. 2008;63(9):936-44.
33. McCrory C, Fiorito G, McLoughlin S, Polidoro S, Cheallaigh CN, Bourke N, et al. Epigenetic clocks and allostatic load reveal potential sex-specific drivers of biological aging. 2020;75(3):495-503.
34. Riley RD, Snell KIE, Ensor J, Burke DL, Harrell FE, Jr., Moons KGM, et al. Minimum sample size for developing a multivariable prediction model: Part I - Continuous outcomes. Stat Med. 2019;38(7):1262-75.
35. Copas JB. Regression, prediction and shrinkage. Journal of the Royal Statistical Society: Series B. 1983;45(3):311-35.
36. Copas JB. Using regression models for prediction: shrinkage and regression to the mean. Statistical methods in medical research. 1997;6(2):167-83.
37. Regan JA, Shah SH. Obesity Genomics and Metabolomics: a Nexus of Cardiometabolic Risk. Current cardiology reports. 2020;22(12):174.
38. WHO WHO. A healthy lifestyle - WHO recommendations. World Health Organization: WHO. 2010.
39. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007–2017. Cardiovascular Diabetology. 2018;17(1):83.
40. Watts M. Can People with Diabetes Give Blood? 2022 [updated 2022/09/08/. Available from: https://www.diabetes.co.uk/can-people-with-diabetes-give-blood.html.
41. Knutson KL, Turek FW. The U-shaped association between sleep and health: the 2 peaks do not mean the same thing. Sleep. 2006;29(7):878-9.
42. Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of Clinical Epidemiology. 2015;68(1):25-34.
43. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Medicine. 2019;17(1):230.
44. Ahola-Olli AV, Mustelin L, Kalimeri M, Kettunen J, Jokelainen J, Auvinen J, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia. 2019;62(12):2298-309.

8 FUNDING

The NEO study is supported by the participating Departments, the Division, and the Board of Directors of the Leiden University Medical Centre, and by the Leiden University, Research Profile Area 'Vascular and Regenerative Medicine'. The analyses of metabolites are funded by the VENI grant (ZonMW-VENI Grant 916.14.023) of D.O.M.-K., D.v.H. and R.N. were supported by a grant of the VELUX Stiftung [grant number 1156]. T.O.F. was supported by the King Abdullah Scholarship Program and King Faisal Specialist Hospital & Research Center [No. 1012879283].

9 CONFLICTS OF INTEREST

R.L.-G. is a part-time clinical research consultant for Metabolon, Inc. All other co- authors have no conflicts of interest to declare.

10 ACKNOWLEDGMENTS AND DISCLOSURES

The authors of the NEO study thank all participants, all participating general practitioners for inviting eligible participants, all research nurses for data collection, and the NEO study group: Pat van Beelen, Petra Noordijk, and Ingeborg de Jonge for coordination, laboratory, and data management.

11 AUTHOR CONTRIBUTIONS

T.O.F.- conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing-original draft. R.L.-G.- validation, writing – review & editing. P.S. – supervision, conceptualization, project administration, resources, funding acquisition, writing – review & editing - R.d.M.- project administration, resources, funding acquisition, writing – review & editing. R.N. and D.V.H.- funding acquisition, writing – review & editing. F.R.R.- funding acquisition. A.v.H.V. and K.W.v.D- conceptualization, supervision, writing – review & editing. D.O.M.-K.- conceptualization, supervision, funding acquisition, writing – review & editing.