The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences, 2024, **79**(2), 1–8
https://doi.org/10.1093/gerona/glad202
Advance access publication 29 August 2023
Original Article

Past or Present; Which Exposures Predict Metabolomic Aging Better? The Doetinchem Cohort Study

Annelot P. Smit, MSc,[1,2,*] Gerrie-Cor M. Herber, PhD,[1] Lieke M. Kuiper, MSc, MA,[1,3] Bette Loef, PhD,[1] H. Susan J. Picavet, PhD,[1] and W.M. Monique Verschuren, PhD[1,2]

[1]Center for Prevention, Lifestyle and Health, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
[2]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.
[3]Department of Internal Medicine, Erasmus MC, University Medical Center, Rotterdam, The Netherlands.

*Address correspondence to: Annelot P. Smit, MSc. E-mail: annelot.smit@rivm.nl

Decision Editor: Gustavo Duque, MD, PhD (Biological Sciences Section)

Abstract

People age differently. Differences in aging might be reflected by metabolites, also known as metabolomic aging. Predicting metabolomic aging is of interest in public health research. However, the added value of longitudinal over cross-sectional predictors of metabolomic aging is unknown. We studied exposome-related exposures as potential predictors of metabolomic aging, both cross-sectionally and longitudinally in men and women. We used data from 4 459 participants, aged 36–75 of Round 4 (2003–2008) of the long-running Doetinchem Cohort Study (DCS). Metabolomic age was calculated with the MetaboHealth algorithm. Cross-sectional exposures were demographic, biological, lifestyle, and environmental at Round 4. Longitudinal exposures were based on the average exposure over 15 years (Round 1 [1987–1991] to 4), and trend in these exposure over time. Random Forest was performed to identify model performance and important predictors. Prediction performances were similar for cross-sectional and longitudinal exposures in both men (R^2 6.8 and 5.8, respectively) and women (R^2 14.8 and 14.4, respectively). Biological and diet exposures were most predictive for metabolomic aging in both men and women. Other important predictors were smoking behavior for men and contraceptive use and menopausal status for women. Taking into account history of exposure levels (longitudinal) had no added value over cross-sectionally measured exposures in predicting metabolomic aging in the current study. However, the prediction performances of both models were rather low. The most important predictors for metabolomic aging were from the biological and lifestyle domain and differed slightly between men and women.

Keywords: Exposome, Human aging, Metabolomics, Sex differences

Worldwide life expectancy has increased steadily over the past century, which increased the proportion of older adults in the world population (1). Older adults experience a decrease in functionality and an increase in multimorbidity. However, there is large heterogeneity in health with aging: some can live an active life up to high ages, while others are already care-dependent at middle age (1,2). This heterogeneity indicates that it is important to distinguish chronological age, which is simply age in calendar years, from biological aging, which reflects decline in functioning.

To get insight into the process of aging, metabolites have been used to create metabolomic biomarker scores, also known as metabolomic aging scores (3,4). These metabolomic aging scores are associated with disease outcomes and mortality (3,4). In 2 prospective cohorts, Kuiper et al. demonstrated that the metabolomic aging biomarker "MetaboHealth," stood out as one of the most representative reflections of biological age (5). Investigating predictors of metabolomic aging might help to identify individuals who are likely to age faster compared to their peers. Insight into these predictors might help to target groups for preventive measures.

Health over the life course is determined by a multitude of risk factors to which people are exposed during their life (1,6). This could be assumed to also be the case for metabolomic aging. To take a full multienvironmental exposure into account, the exposome approach was proposed by Christopher Wild (7). The exposome approach is characterized by taking into account a wide variety of exposures from different domains, for example, lifestyle, environmental, demographic, and biological (7,8).

Prediction research into metabolomic aging is scarce. In a previous etiological study, metabolomic aging was associated with higher body mass index (BMI) and heavy drinking, but not with smoking and physical activity (9). However, this study only took into account a limited number of exposures. Furthermore, although exposures are likely to change over time (7,10), the added value of using longitudinal over cross-sectional exposures to predict metabolomic aging is unknown. This is of interest because aging is often seen as an accumulation of molecular and cellular damage due to a wide range of lifetime exposures. Compared to longitudinal exposure levels, cross-sectional exposure levels are more easy and less time-consuming to assess. For clinical practice, use of

Received: May 8 2023; Editorial Decision Date: August 10 2023.

cross-sectional exposure information would thus be more feasible than use of longitudinal exposure information in decision making and risk assessment. In addition, one would also like to know which exposures predict metabolomic aging. This information could help to recognize those who age at a fast pace. As men and women have different metabolite concentrations (11–13), it could be speculated that predictors of metabolomic aging differ for men and women.

In this study, we studied whether longitudinally measured exposures predicted metabolomic aging better than cross-sectionally measured predictors. For metabolomic aging, we used the metabolomics-based algorithm "MetaboHealth." We included a wide range of predictors with data available over a 15-year period to reflect an exposome approach. Additionally, we investigated which exposures were the most important predictors of metabolomic aging. We also investigated if predictors differed between men and women.

Method

Study Population

The Doetinchem Cohort Study (DCS) is a prospective cohort study that aims to gain insight into the impact of lifestyle and biological risk factors and health across the life span in Dutch adults (14,15). Between 1987 and 1991, 12 405 participants (aged 20–59 years) were examined (Round 1, R1). Of those participants, a random sample of 7 768 was reinvited to be examined in 1993–1997 (R2) and again in 1998–2002 (R3), 2003–2007 (R4), 2008–2012 (R5), and 2013–2017 (R6). The response rate in the first round was 62%, and the response rates in the subsequent rounds exceeded 75%. For the analyses, we excluded participants who had missing data on metabolomics (N = 56), all exposures (N = 2), and pregnant women because of physiological reasons (N = 3), which left 4 459 participants in the analyses.

Written informed consent from all participants was obtained. The DCS was approved by the Medical Ethics Committees of the Netherlands Organization of Applied Scientific Research.

Metabolomic Age

At Round 4, a total of 226 metabolomic markers were measured in stored blood samples by Nightingale Health Ltd, Helsinki, Finland, in ethylenediaminetetraacetic acid (EDTA) plasma via high-throughput nuclear magnetic resonance (NMR) in 2020. The methodology of NMR has been described previously (16,17).

In a past study, 12 cohort studies including a total of 44 000 individuals who had metabolomics data from the Nightingale platform available were used to develop a robust algorithm based on these 226 metabolic markers. This algorithm (ie, MetaboHealth) includes markers that were independently associated with mortality. In the present study, we used this MetaboHealth algorithm to calculate metabolomic aging (3). MetaboHealth can be calculated via a code available on the Github of D. Bizzari (https://github.com/DanieleBizzarri/MiMIR). The final algorithm included 14 of the 226 metabolites: total lipids in chylomicrons and extremely large very low-density lipoprotein (VLDL) particle, total lipids in small high-density lipoprotein, mean diameter for VLDL particles, ratio of polyunsaturated fatty acids to total fatty acids, glucose, lactate, histidine, isoleucine, leucine, valine, phenylalanine, acetoacetate, albumin, and glycoprotein acetyls. The MetaboHealth score ranged from −2 to 3 in most cohorts and a 1-unit increase in MetaboHealth was associated with a 2.73 times higher mortality risk (3). A higher score represents a higher metabolomic age.

Exposure Variables

We included exposure variables from the environmental, lifestyle, and biological domain (18), as well as demographic variables. All exposures are described below. Most exposures were repeatedly measured (R1–R4). An overview of all exposure variables, including in which round of the DCS the variable was collected, can be found in Supplementary Table 1.

Demographics

Self-reported demographics used were: sex (man/woman), educational level (low; medium; high), marital status (single; never married; married; widow/widower; and divorced), household composition (alone; with others), and working hours (hours/week).

Biological Exposures

BMI (measured weight [kg]/height [m^2]), waist circumference (centimeters), blood pressure (systolic/diastolic [mm/Hg]), and pulse rate (beats/minute) were assessed by trained staff according to standardized protocols. Menopausal status (premenopausal; perimenopausal; postmenopausal), contraceptive use (yes/no), cholesterol-lowering medication (yes/no), and antihypertensives use (yes/no) were self-reported.

Lifestyle Exposures

Lifestyle exposures were assessed via self-administered questionnaires. Alcohol use (no; every now and then, but less than 1 glass per week; yes; and amount among users) and smoking status (never; ex; current, and amount among users) were assessed. Smoking pack-years were calculated with the participant's number of years of smoking and the amount of cigarettes. For physical activity, 3 measures were included: occupational physical activity (sedentary job; standing job; manual work; heavy manual work; not applicable) (19), time spent on moderate–vigorous activity during leisure time <0.5 hour; 0.5–3.5 hours; ≥3.5 hours or more, of which <2 hours vigorous; ≥3.5 hours, of which ≥2 hours or more vigorous (19), and hours of doing sports in a week. The number of hours of sleep (≤5, 6, 7, 8, and ≥9 hours) was also assessed. Diet was measured with a 178-item validated food frequency questionnaire. A modified version of the Dutch Healthy Diet index (DHD) 2015 was calculated, ranging from 1 to 130, which indicates adherence to elements of the Dutch Nutrition Guideline (20). A higher score indicates higher adherence. The following food/nutrients intakes were added: energy intake (kcal per day), total, animal, and vegetable protein, fat, total carbohydrates, total fibers, and total water (all in g/day).

Environmental Exposures

Environmental exposures included factors like NO_2, noise, and "greenness" of the environment. Total NO_2 concentration (in μg/m^3), total $PM_{2.5}$ concentration (in μg/m^3), and total elemental carbon concentration (in μg/m^3) at the participants' home addresses for Rounds 1–3 were calculated via dispersion models based on concentration estimates of the Year 2000 and for Round 4 were calculated as the average of the Years 2000 and 2010 (21). Rail traffic noise levels in 2016 for the entire 24-hour period at home address (in dB) and

road traffic noise levels in 2016 for the entire 24-hour period at home address (in dB) were calculated by Standard Model Instrumentation for Noise Assessments (22). Greenness was measured using the normalized difference vegetation index (2010 data) in buffers of 300 and 1 000 m around the participants' home address using Landsat 5 Thematic Mapper (United States Geological Service). Both vegetation indexes were on a scale from −1 to 1. A higher score indicates more greenness.

Statistical Analyses

Our first results showed that female-specific exposures were important predictors for metabolomic aging in the total population (Supplementary Figure 1). We therefore stratified our analyses by sex and recalculated metabolomic age, that is, MetaboHealth, for men and women separately because metabolomic age is scaled based on the input sample. In this study, metabolomic age was adjusted for chronological age, because we were interested whether a respondent is metabolomically "older" (positive value) or "younger" (negative value) than his/her peers of the same chronological age. We therefore calculated the raw residuals of the linear regression of metabolomic age and chronological age for all participants of the same sex.

To study which exposures were most predictive of metabolomic aging at Round 4, we followed the statistical analyses steps described previously (23). We investigated the predictors for metabolomic aging in a Random Forest (RF) for both men and women. Additionally, these RFs were performed with cross-sectionally as well as longitudinally measured predictors. In the cross-sectional RF models exposures of Round 4 were included. In the longitudinal RF models, we included the average levels and the trend of the exposures over Rounds 1–4. We describe below how we calculated the average and trend of exposures and further steps of the RF models. All analyses were performed in R (version 4.2.0) (R Foundation, Vienna, Austria).

1. For the longitudinal models we calculated the Area-Under-Exposure (AUE) and Trend-Of-Exposure (TOE) for each exposure. The AUE represented the average exposure level from Rounds 1–4. The TOE represented the average trend over time from Rounds 1–4. To provide further clarification on these 2 metrics: in general, aging is seen as an accumulation of deficits, that are build up over the life course due to a wide range of exposures in interaction with genetic makeup. From that perspective one would expect cumulative exposure (eg, AUE) to contribute to metabolomic age. Furthermore, the trend of exposure was used, to study if given a similar AUE, distinguishing between a pattern of increasing or decreasing exposure would be of added explanatory value. Further details of this method are explained in Loef et al. (23).
2. We used the R-package "randomForest" (24) to analyze which cross-sectional and longitudinal exposures contributed most to the prediction of the age-adjusted metabolomic age score (25). RF is a machine learning technique that uses decision trees to investigate predictors of a specific outcome, in our case metabolomic age at Round 4. This technique is nonparametric; therefore, we made no assumptions about the distributions of exposures or the outcome. Moreover, RF can deal with correlated variables because each tree is built on a different bootstrap sample. We tuned the following parameters to optimize the RF models: size of random sample of exposures used at each split (mtry), number of trees (ntree), minimum number of observations in the final nodes (nodesize), and maximum number of terminal nodes (maxnodes) (26,27). In addition, we based our optimal prediction performance on the root mean square error (RMSE), explained variance (R^2), and mean absolute error (MAE). We randomly divided our data sets in 80% training set and 20% test data set and performed a 5-fold cross-validation to overcome overfitting or selection bias issues. This cross-validation was performed with the R-package caret2 (28). We used the RMSE, R^2, and MAE of the optimally tuned cross-sectional and longitudinal RF models for the test data set. Moreover, we compared these models with the RMSE, R^2, and MAE of null model, which predicts the training data set mean when no exposures are included.
3. To investigate the most important predictors for metabolomic aging, we examined the variable importance ranking with the optimally tuned RF for men and women. The order of the listed predictors is based on the increase in the mean square error when the specific predictor is excluded from the RF model, keeping all other variables fixed. We performed a post hoc cross-validation to investigate which predictors together are responsible for the optimal prediction performance, also called the parsimonious model. To investigate this relation we plotted the RMSE for each number of exposures selected. Flattening of the RMSE curve means less added value of the combination of variables on the prediction performance. The number of exposures selected was based on the flattening in RMSE when adding an extra predictor to the parsimonious model.

Results

Table 1 shows a selection of characteristics of the included demographic, biological, lifestyle, and environmental exposures stratified by sex, both cross-sectionally as well as longitudinally. The complete table can be found in Supplementary Table 2. At Round 4, our study population comprised 4 459 participants. The age-adjusted MetaboHealth score, which reflects whether the participant is "metabolomically" older/younger than his peers, ranged from −1.5 to 2.3. The participants had a mean age of 55 years (range 36–76 years), 48% of them were men, and they had a mean metabolomic age score of −0.03. At Round 4, men were on average 1 year older, were more often highly educated, used less often blood pressure medication, drank more glasses of alcohol per day, had a higher number of smoking pack-years, and had less adherence to the DHD-15 than women. The average crude MetaboHealth score, BMI, NO_2 levels, and elemental carbon levels were similar between men and women at Round 4. The same tendencies were seen for the longitudinal exposures. Over a 15-year period, men used on average less often blood pressure medication and had a lower adherence to the DHD-15, while the number of glasses of alcohol per day and smoking pack-years was higher than in women. The average crude MetaboHealth score, BMI, NO_2 levels, and elemental carbon levels were similar between men and women from Rounds 1–4. With respect to the trend over a 15-year period,

Table 1. Selection of Longitudinal and Cross-Sectional Characteristics Stratified by Sex

Characteristic	Men (n = 2 114)		Women (n = 2 345)	
	Mean/%	*SD*/*n*	Mean/%	*SD*/*n*
MetaboHealth score*—Round 4	−0.1	0.4	0.1	0.4
Chronological age—Round 4	56.1	9.8	55.1	9.9
Demographic				
Educational level—Round 4				
Low	7.2	152	8.7	205
Medium	66.0	1,396	71.4	1,674
High	26.5	561	19.6	460
Biological				
Body mass index (kg/m^2)				
Average R1–4	25.9	3.0	25.3	4.0
Trend R1–4	0.7	0.7	0.8	0.9
Round 4	26.8	3.4	26.3	4.6
Contraceptive use				
Average R1–4; % time yes	NA	NA	6.5	13.8
Trend R1–4; % from no to yes	NA	NA	20.0	468
Round 4; % yes	NA	NA	9.1	213
Blood pressure medication				
Average R1–4; % time yes	7.9	20.7	8.9	21.9
Trend R1–4; % from no to yes	3.7	78	4.6	108
Round 4 % yes	15.6	329	17.3	405
Lifestyle exposures				
Glasses of alcohol (per day)				
Average R1–4	1.6	1.5	0.6	0.8
Trend R1–4	0.0	0.5	0.1	0.2
Round 4	1.6	1.6	0.7	0.9
Smoking pack-years				
Average R1–4	12.2	13.9	7.3	10.1
Trend R1–4	1.2	3.4	0.9	2.6
Round 4	13.2	15.8	8.3	12.6
Dutch Healthy Diet index				
Average R1–4	62.2	12.1	68.6	11.0
Trend R1–4	0.9	6.6	1.6	6.4
Round 4	63.4	14.0	70.5	12.9
Environmental				
Total NO_2 (μg/m^3)				
Average R1–4	28.7	1.8	28.7	1.8
Trend R1–4	−1.1	0.7	−1.1	0.7
Round 4	26.1	2.1	26.1	2.1

Notes: NA = not applicable; *SD* = standard deviation.
*Crude MetaboHealth score calculated on the input sample including both men and women.

men and women only differed in the trend of adherence to the DHD-15; adherence increased less in men than women from R1–4.

Prediction Performance of Metabolomic Aging Models

We tuned the RF in the training data set to investigate which cross-sectional and longitudinal exposures are most important for predicting the participants' metabolomic age at Round 4. In the parsimonious cross-sectional test data sets, the RMSE (0.40 vs 0.41 for men; 0.36 vs 0.39 for women) and MAE (0.31 vs 0.32 for men; 0.29 vs 0.31 for women) were almost similar compared to the null models (ie, a model without any exposures). Additionally, in the parsimonious cross-sectional test data sets, the R^2 slightly improved compared to the null model (6.8% vs 0.00% for men; 14.8% vs 0.00% for women). Similar results were found for the parsimonious longitudinal models, which is displayed in Table 2. Because the RMSE, R^2, and MAE were similar for the cross-sectional and longitudinal models, we decided to display only the predictors of the cross-sectional models.

Predictors of Metabolomic Aging

Figure 1 shows the top 20 ranked predictors of metabolomic aging for the tuned cross-sectional RF model in men and women separately on the whole data set. In both men and women, the top 10 predictors consisted mostly of biological exposures and lifestyle exposures, and included both waist circumference in the top 5 ranked predictors. In men, the top 5 consisted in addition of daily number of cigarettes, intake of carbohydrates, intake of vegetable protein, and intake of kilocalories. In women BMI, menopausal status, contraceptive use, and intake of fat ranked within the top 5. The variable importance plots of the longitudinal model are displayed in Supplementary Figure 2 and showed similar predictors in the top 20 as the cross-sectional model.

To investigate the number of exposures required for a best performing, yet parsimonious model, we performed cross-validation with stepwise inclusion of exposures and plotted the RMSE per total number of exposures. Figure 2 shows the results of the RMSE plot of the cross-validation of the cross-sectional RF model in both men and women in the training data set. For men, the RMSE decreased from 1 to 10 exposures and flattened out at 10 selected exposures

Table 2. Number of Included Exposures, RMSE, R^2, and MAE of the Cross-Sectional as Well as Longitudinal in the Test Data Set of the RF Models in Men and Women

Model	Number of Exposures	RMSE	R^2 (%)	MAE
Men				
Null	0	0.41	0	0.32
Cross-sectional	10	0.40	6.8	0.31
	36	0.40	7.0	0.31
Longitudinal	17	0.40	5.8	0.32
	76	0.40	8.8	0.31
Women				
Null	0	0.39	0	0.31
Cross-sectional	15	0.36	14.8	0.29
	38	0.37	14.0	0.29
Longitudinal	13	0.36	14.4	0.29
	81	0.37	13.6	0.29

Notes: MAE = mean absolute error; RF = Random Forest; RMSE = root mean square error.

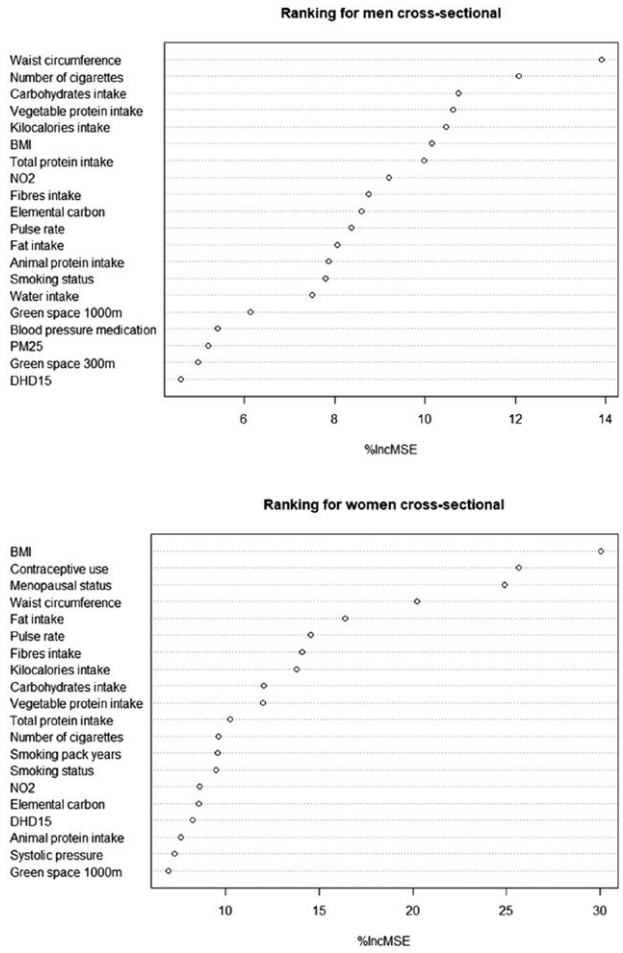

Figure 1. Variable importance ranking of Random Forest (RF) with MetaboHealth as outcome and cross-sectionally measured exposures of Round 4 in men and women. The x-axis shows the percentage increase in mean square error (MSE) when a particular predictor is removed from the RF model.

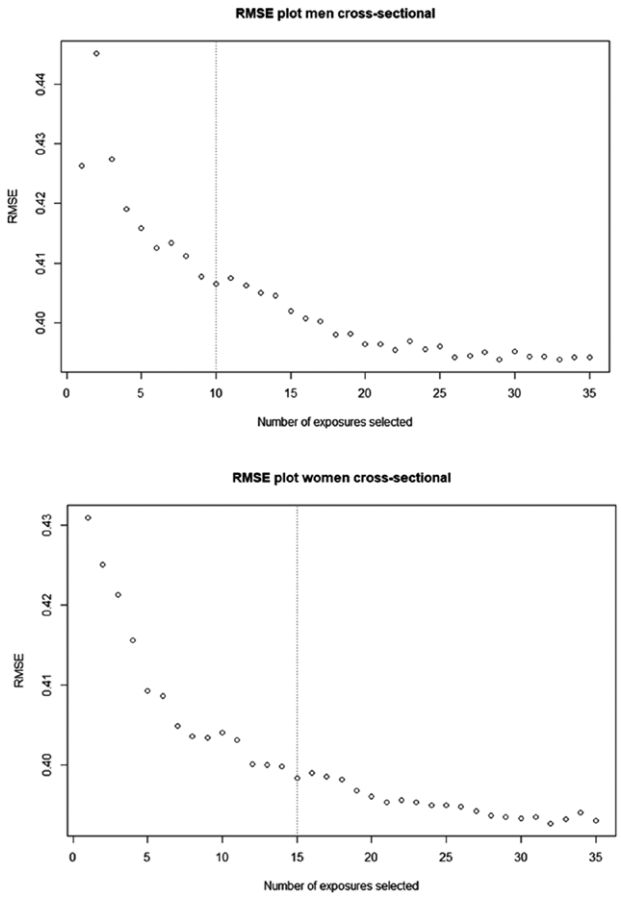

Figure 2. Root mean square error (RMSE) plot of prediction performance of *X* number of selected exposures in Random Forest model with MetaboHealth Round 4 (R4) as outcome and cross-sectionally measured exposures of R4 in the training data set in men and women. The dotted gray line reflects the optimal number of selected exposures ($X = 10$ men, $X = 15$ women).

(dotted gray line). For women, the RMSE decreased from 1 to 10 exposures and flattened out at 15 selected exposures (dotted gray line). Table 2 displays the RMSE, R^2, MAE, and optimal number of exposures included of all models in the test data set.

Discussion

We showed that cross-sectionally measured exposures and longitudinally measured exposures predicted metabolomic aging similarly in both men and women. However, the prediction performance of both models was rather low. The most important predictors for metabolomic aging came from the biological and lifestyle domain and differed between men and women. These findings are relevant to gain insight into which predictors can be used to identify those who age at a fast pace based on metabolites.

To our knowledge, this study is the first that has compared cross-sectional and longitudinal models in predicting metabolomic aging. For the exposures we included cross-sectionally measured exposures as well as averages and trends of longitudinally measured exposures. The prediction performances of both cross-sectional models and longitudinal models were similar in both men and women. Yet, the R^2 of the models for men were lower than for women. The lack of an added value of longitudinal exposures might be a result of the metabolites, such as blood lipids and its subfractions, that were included in the algorithm of MetaboHealth and can respond to lifestyle changes within weeks (29–31). Even though the general thought is that cumulative levels of life-long exposures influence health in later life (18,32), our results suggest that taking into account average history of exposure levels (longitudinal approach) compared to cross-sectionally measured predictors has little to no added value in predicting metabolomic aging. This finding is beneficial for clinical practice as current exposures are advantageous because they are more easily accessible than longitudinal exposures. However, future studies are needed to improve better prediction of metabolomic aging before this conclusion can be extended to clinical practice.

Although our cross-sectional and longitudinal models similarly predicted metabolomic aging, these models only slightly predicted metabolomic aging better than the null models. For application of the algorithm in preventive or clinical practice, we have to consider the performances of the models in accurately predicting metabolomic age. In the cross-sectional and longitudinal models of men, the RMSE was similar as the null models and the R^2 was 5.8% to 6.8% higher than the null models. In the cross-sectional and longitudinal models of women, the RMSE was 7.7% lower than

the null models and the R^2 was 14.4% to 14.8% higher than the null models. We expected a greater improvement, as we took into account an exposome approach when we selected our predictors. The relatively small improvement might be due to: (i) lack of accurate assessment of exposures either because of measurement error or because exposures were not measured at a critical period during the course life (33), (ii) the lack of information on important exposures for metabolomic aging, or (iii) the fact that chronological age might be the major predictor of metabolomic aging given the modest effects of lifestyle exposures and environmental exposures on health in general (34). Future studies are needed to investigate whether the prediction performance can be improved. Yet, if these studies fail to improve prediction performance, then this measure is not accurate enough to be used in clinical practice.

The most important predictors for metabolomic aging came from the lifestyle domain and biological domain in both men and women. The fact that especially lifestyle and biological exposures were the most prominent predictors for metabolomic aging might be related to the fact that the metabolites included in the MetaboHealth algorithm are mainly (lipo)protein-related (3). Some of the top-ranked predictors for metabolomic aging differed between men and women. In men, smoking behavior was more important as predictor for metabolomic aging than in women. In a previous risk assessment study, smoking behavior was not associated with metabolomic aging based on an alternative metabolomic aging score (9); however, this study did not stratify for sex while smoking behavior affects healthy aging differently for men and for women (35,36). Furthermore, we found that contraceptive use and menopausal status were the strongest predictors for metabolomic aging in women. Because menopausal status and contraceptive use are likely to be correlated, we investigated whether women who had an early menopause (menopause before the age of 45 years) significantly differed from women that had a normal menopause with respect to metabolomic aging (ie, age-adjusted metabolomic age). We found no differences in metabolomic aging between these groups. Menopausal status and contraceptive use are known to change metabolite concentrations (37–39). Given that our metabolomic aging biomarker encompasses various lipid fractions, the identification of menopausal status as an important predictor may be attributed, in part, to alterations such as the adverse shifts observed in cholesterol levels among postmenopausal women (39). Until now, no studies have investigated the importance of female-specific exposures like menopausal status and contraceptive use to predict metabolomic aging. Therefore, our results cast a new light on female-specific predictors for metabolomic aging.

The top-ranked predictors for metabolomic aging did not include exposures like physical activity, education, and alcohol use. In previous risk assessment studies, these exposures were associated with metabolomic aging and/or healthy aging (6,9,40). However, associations between determinants and outcomes do not directly imply predictive value (41). The goal of an association model is to assess the relation between exposures and outcomes and thus for example find group differences or regression slopes. In prediction models, however, the goal is to build a as good as possible model with exposures (predictors) to *predict* the outcome. Physical activity, education, and alcohol might be associated with metabolomic aging, but in our study did not have predictive value when also other exposures are taken into account.

A strength of our study is the use of an exposome approach to investigate predictors of the metabolomic aging. Additionally, because we had data over a period of 15 years we were able to investigate the added value of using longitudinal over cross-sectional exposures in predicting metabolomic aging. Limitations of our study were that even though we included a wide variety of exposures from different domains, we lacked information of exposures such as factors belonging to the psychological domain, potential contaminants, and the working environment. Inclusion of these exposures would possibly have improved the prediction performance of our models. In addition, only 2 exposures from the environmental domain (ie, elemental carbon and NO_2) were in the top-ranked predictors for men. This might be explained by the fact that the current study was conducted in 1 town in the Netherlands, leading to limited variation in these exposures. Lastly, even though we found that contraceptive use was an important predictor of metabolomic aging, we lacked information on the type of contraceptive used. Therefore, we could not investigate whether type of contraceptive use differentially affects metabolomic aging.

Early recognition of the heterogeneity in health with aging could be unraveled by metabolites-driven aging scores, like MetaboHealth. In this study, we showed that taking into account history of exposure levels (longitudinal approach) over cross-sectionally measured predictors had little to no added value in predicting metabolomic age. The most important predictors for metabolomic aging were mainly from the biological and lifestyle domain and differed between men and women. However, the prediction performances of both models were rather modest. Before this approach can be used in clinical practice, future research is needed to replicate our findings and increase model performance to predict metabolomic aging.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences* online.

Funding

The Doetinchem Cohort Study was financially supported by the Ministry of Health, Welfare and Sport of the Netherlands and the National Institute for Public Health and the Environment. This research was funded by the Strategic Program RIVM (Dutch National Institute for Public Health and the Environment), grant number: S/010003/03/CP, and by ZonMW (research grant 457001001 for the VOILA project).

Conflict of Interest

None.

Acknowledgments

We thank all the participants for their participation in the study, and the fieldworkers of the Municipal Health Service in Doetinchem for their contribution to the data collection for this study.

References

1. World Health Organization. *World Report on Ageing and Health*. Geneva, Switzerland: World Health Organization; 2015.
2. Lowsky DJ, Olshansky SJ, Bhattacharya J, Goldman DP. Heterogeneity in healthy aging. *J Gerontol A Biol Sci Med Sci*. 2014;69(6):640–649. https://doi.org/10.1093/gerona/glt162
3. Deelen J, Kettunen J, Fischer K, et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun*. 2019;10(1):3346. https://doi.org/10.1038/s41467-019-11311-9
4. Rutledge J, Oh H, Wyss-Coray T. Measuring biological age using omics data. *Nat Rev Genet*. 2022;23(12):715–727. https://doi.org/10.1038/s41576-022-00511-7
5. Kuiper LM, Polinder-Bos HA, Bizzarri D, et al. Epigenetic and metabolomic biomarkers for biological age: a comparative analysis of mortality and frailty risk. *J Gerontol A Biol Sci Med Sci*. 2023:glad137. https://doi.org/10.1093/gerona/glad137
6. Abud T, Kounidas G, Martin KR, Werth M, Cooper K, Myint PK. Determinants of healthy ageing: a systematic review of contemporary literature. *Aging Clin Exp Res*. 2022;34(6):1215–1223. https://doi.org/10.1007/s40520-021-02049-w
7. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*. 2005;14(8):1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456
8. Vrijheid M. The exposome: a new paradigm to study the impact of environment on health. *Thorax*. 2014;69(9):876–878. https://doi.org/10.1136/thoraxjnl-2013-204949
9. Robinson O, Chadeau Hyam M, Karaman I, et al. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. *Aging Cell*. 2020;19(6):e13149. https://doi.org/10.1111/acel.13149
10. Mäkinen V-P, Ala-Korpela M. Metabolomics of aging requires large-scale longitudinal studies with replication. *Proc Natl Acad Sci USA*. 2016;113(25):E3470–E3470. https://doi.org/10.1073/pnas.1607062113
11. Mäkinen VP, Karsikas M, Kettunen J, et al. Longitudinal profiling of metabolic ageing trends in two population cohorts of young adults. *Int J Epidemiol*. 2022;51:1970–1983. https://doi.org/10.1093/ije/dyac062
12. Bell JA, Santos Ferreira DL, Fraser A, et al. Sex differences in systemic metabolites at four life stages: cohort study with repeated metabolomics. *BMC Med*. 2021;19(1):58. https://doi.org/10.1186/s12916-021-01929-2
13. Darst BF, Koscik RL, Hogan KJ, Johnson SC, Engelman CD. Longitudinal plasma metabolomics of aging and sex. *Aging (Albany NY)*. 2019;11(4):1262–1282. https://doi.org/10.18632/aging.101837
14. Picavet HSJ, Blokstra A, Spijkerman AM, Verschuren WM. Cohort profile update: the Doetinchem Cohort Study 1987–2017: lifestyle, health and chronic diseases in a life course and ageing perspective. *Int J Epidemiol*. 2017;46(6):1751–1751g. https://doi.org/10.1093/ije/dyx103
15. Verschuren WM, Blokstra A, Picavet HS, Smit HA. Cohort profile: the Doetinchem Cohort Study. *Int J Epidemiol*. 2008;37(6):1236–1241. https://doi.org/10.1093/ije/dym292
16. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am J Epidemiol*. 2017;186(9):1084–1096. https://doi.org/10.1093/aje/kwx016
17. Soininen P, Kangas AJ, Würtz P, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst*. 2009;134(9):1781–1785. https://doi.org/10.1039/b910205a
18. Wild CP. The exposome: from concept to utility. *Int J Epidemiol*. 2012;41(1):24–32. https://doi.org/10.1093/ije/dyr236
19. Pols MA, Peeters P, Ocke MC, Slimani N, Bueno-de-Mesquita HB, Collette H. Estimation of reproducibility and relative validity of the questions included in the EPIC Physical Activity Questionnaire. *Int J Epidemiol*. 1997;26(suppl_1):S181. https://doi.org/10.1093/ije/26.suppl_1.s181
20. Looman M, Feskens EJ, de Rijk M, et al. Development and evaluation of the Dutch Healthy Diet index 2015. *Public Health Nutr*. 2017;20(13):2289–2299. https://doi.org/10.1017/S136898001700091X
21. Velders GJM, Maas RJM, Geilenkirchen GP, et al. Effects of European emission reductions on air quality in the Netherlands and the associated health effects. *Atmos Environ*. 2020;221:117109. https://doi.org/10.1016/j.atmosenv.2019.117109
22. Schreurs EM, Jabben J, Verheijen ENG. *STAMINA—Model Description. Standard Model Instrumentation for Noise Assessments. STAMINA—Model Beschrijving Standard Model Instrumentation for Noise Assessments*. Bilthoven, Netherlands : Rijksinstituut voor Volksgezondheid en Milieu RIVM; 2010.
23. Loef B, Wong A, Janssen NAH, et al. Using random forest to identify longitudinal predictors of health in a 30-year cohort study. *Sci Rep*. 2022;12(1):10372. https://doi.org/10.1038/s41598-022-14632-w
24. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18–22.
25. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
26. Probst P, Wright MN, Boulesteix A-L. Hyperparameters and tuning strategies for random forest. *WIREs Data Min Knowl Discovery*. 2019;9(3):e1301. https://doi.org/10.1002/widm.1301
27. Scornet E, Coeurjolly JF, Leclercq-Samson A. Tuning parameters in random forests. *Esaim: Proceedings*. 2017;60:144–162. https://doi.org/10.1051/proc/201760144
28. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1–26. https://doi.org/10.18637/jss.v028.i05
29. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. *Proteomics*. 2006;6(17):4716–4723. https://doi.org/10.1002/pmic.200600106
30. Weckwerth W, Morgenthal K. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov Today*. 2005;10(22):1551–1558. https://doi.org/10.1016/S1359-6446(05)03609-3
31. Schranner D, Kastenmüller G, Schönfelder M, Römisch-Margl W, Wackerhage H. Metabolite concentration changes in humans after a bout of exercise: a systematic review of exercise metabolomics studies. *Sports Med Open*. 2020;6(1):11. https://doi.org/10.1186/s40798-020-0238-4
32. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. *J Epidemiol Community Health*. 2003;57(10):778–783. https://doi.org/10.1136/jech.57.10.778
33. Siroux V, Agier L, Slama R. The exposome concept: a challenge and a potential driver for environmental health research. *Eur Respir Rev*. 2016;25(140):124–129. https://doi.org/10.1183/16000617.0034-2016
34. Patel CJ, Rehkopf DH, Leppert JT, et al. Systematic evaluation of environmental and behavioural factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. *Int J Epidemiol*. 2013;42(6):1795–1810. https://doi.org/10.1093/ije/dyt208
35. Beltrán-Sánchez H, Finch CE, Crimmins EM. Twentieth century surge of excess adult male mortality. *Proc Natl Acad Sci USA*. 2015;112(29):8993–8998. https://doi.org/10.1073/pnas.1421942112
36. Preston SH, Wang H. Sex mortality differences in the United States: the role of cohort smoking patterns. *Demography*. 2006;43(4):631–646. https://doi.org/10.1353/dem.2006.0037
37. Auro K, Joensuu A, Fischer K, et al. A metabolic view on menopause and ageing. *Nat Commun*. 2014;5:4708. https://doi.org/10.1038/ncomms5708

38. Ruoppolo M, Campesi I, Scolamiero E, et al. Serum metabolomic profiles suggest influence of sex and oral contraceptive use. *Am J Transl Res*. 2014;6(5):614–624. PMID: 25360225; PMCID: PMC4212935.
39. Ambikairajah A, Walsh E, Cherbuin N. Lipid profile differences during menopause: a review with meta-analysis. *Menopause*. 2019;26(11):1327–1333. https://doi.org/10.1097/GME.0000000000001403
40. McLaughlin SJ, Kim S, Li LW, Zhang J. Educational differences in trajectories and determinants of healthy ageing in midlife and older Americans. *Maturitas*. 2020;134:21–28. https://doi.org/10.1016/j.maturitas.2020.01.002
41. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry*. 2020;77(5):534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671