Supplementary data contents

Appendix 1: Detailed description of the RODAM study population (Ghanaians)......................................2

Appendix 2: Power calculations at the beginning of RODAM study (Ghanaians) 5

Appendix 3: Comparing baseline characteristics of participants in the current epigenetic study to the general RODAM study population (Ghanaians)...7

Appendix 4: Description of data collection, processing, and storage in RODAM study (Ghanaians). 8

Appendix 5: Description of post-hoc analyses 12

Appendix 6: Description of Indian Migration study (Indians) 14

Appendix 7: Descriptions of PURE-SA-NW study (native South Africans) 17

Appendix 8: Lifestyle factors, cardiometabolic traits and epigenetic age acceleration in the total RODAM study population (migrant and non-migrant Ghanaians combined)...20

Appendix 9: Sensitivity linear regression analyses in the RODAM study (Ghanaians) on fasting blood glucose by excluding participants taking anti-diabetic medications..21

Appendix 10: Baseline characteristics of participants in IMS study (Indians)..22

Appendix 11: Pearson's correlations between age measures and between age acceleration measures in IMS study (Indians). 24

Appendix 12: Migration status and EAA Measures in IMS study..25

Appendix 13: Forest plots of regression model β (and 95% confidence intervals) relating EAA Measures to cardiometabolic related traits in the IMS study. 27

Appendix 14: Lifestyle factors, cardiometabolic traits and epigenetic age acceleration in migrant vs non-migrant Indians (IMS study). 29

Appendix 15: Baseline characteristics of participants in PURE-SA-NW study (native South Africans). ..32

Appendix 16: Pearson's correlations between age measures and between age acceleration measures in PURE-SA-NW study (native South Africans). 34

Appendix 17: Forest plots of model β (and 95% confidence intervals) relating EAA Measures to lifestyle factors in PURE-SA-NW study. 35

Appendix 18: Lifestyle factors, cardiometabolic traits and epigenetic age acceleration in the PURE-SA-NW study (native South Africans)...37

Appendix 19: Summary of results post-hoc analyses 38

Appendix 20: STROBE Statement—Items to be included when reporting observational studies in a conference abstract 40

Appendix 21: Reporting checklist for genetic association study...41

Appendix 22: Data Analysis Plan for current analyses 48

Appendix 1. Detailed description of the RODAM study population (Ghanaians).

Study design.

The Research on Obesity and Diabetes among African Migrants (RODAM) project—an European Commission funded project is a cross-sectional study that aims to understand the reasons for the high prevalence of obesity and T2D among African migrants by (1) studying the complex interplay between environmental exposures and genetics and their relative contributions to the high prevalence of obesity and T2D; (2) to identify specific risk factors within these broad categories to guide intervention programmes and (3) to provide a basic knowledge for improving diagnosis and treatment.[1] The full study protocol has been published elsewhere.[1]

Inclusion Criteria

Assessing the role of gene–environmental interactions in risk factors, such as obesity and T2D, among migrant populations requires a highly standardized approach and relatively homogeneous migrating and non-migrating populations.[1] For this reason, we concentrated on a relatively homogenous Sub-Saharan African (SSA) migrant population (Akan ethnic group) to enable comparisons of mechanism of obesity and T2D between migrants living in different European countries and their compatriots living in rural and urban SSA.[2] Ghanaians are one of the largest SSA migrant groups in Europe, with the majority living in the main cities in United Kingdom (UK), the Netherlands, and Germany.[3-5] Consequently, adult Ghanaians (aged ≥25 years) were recruited in rural and urban Ghana, and in the cities of Amsterdam, Berlin, and London.

Recruitment of the study participants

Engagement of Ghanaian community

Previous work in the Netherlands and a feasibility study among African communities in the UK showed that involvement of the community leaders enhanced study participation and helped prevent low response rates relating to language barriers and lack of understanding about the relevance of the study.[6] The RODAM project, therefore, involved the Ghanaian community leaders in all sites. This included working with religious communities (e.g., churches and mosques), endorsement from local key figures and establishing relationships with healthcare organizations that served these groups. In addition, the project team provided information about the study via local media aimed at the Ghanaian population (e.g., Ghanaian radio and TV stations).[1]

Owing to differences in population registration systems across European countries as well as in Ghana, different approaches were needed for the recruitment of the study populations across different locations. For example, there exists a population register in the Netherlands where the Ghanaian migrants were identified and randomly selected for the study. In Ghana, the UK and Germany, the situation is quite different as there are no population registers that would allow for easy identification of these populations. It was important, however, to adopt the recruitment strategies that were comparatively as similar as possible across different locations. Described below are the various recruitment strategies in each site.[1]

Recruitment strategy in Ghana

In Ghana, two cities (Kumasi and Obuasi) and 15 villages in the Ashanti region served as the urban and rural recruitment sites. The initial sampling frame was the list of enumeration areas (EAs) in the Ashanti region from the 2010 census. A multistage random sampling procedure was adopted to arrive at the sampling of 30 EAs. EAs were stratified, weighted and a random sample of rural and urban EAs was selected. There are over 2000 urban EAs and more than 1000 rural EAs. The first stage was to group the districts into two main categories: districts with a high number of urban (Kumasi and Obuasi) areas and districts with a high number of rural EAs. The next stage of sampling was to put the EAs together in each of the categories and take a weighted random sample of 10 for Kumasi and 5 for Obuasi, respectively. The procedure was repeated for the rural EAs by adding all the EAs in the selected districts and weighted from the first stage together after which a simple random sample procedure was adopted to select the total number of rural EAs (15) required for the study. Letters were sent to all selected health and

community authorities to notify them of the start of the study. We sent team members to the various communities to stay among them. Once within the community, the team then organized mini clinics in the field for a period of 1–2 weeks depending on the sampled population and responsiveness of respondents.[1]

Recruitment strategy in the Netherlands

In the Netherlands, Ghanaian participants were randomly drawn from Amsterdam Municipal Health register. This register contains data on country of birth of citizens and their parents, thus allowing for sampling based on the Dutch standard indicator for ethnic origin. All selected participants aged ≥25 years were sent a written invitation combined with written information regarding the study and a opting out response card. Participants were reminded by phone or by home visit after 2 weeks if there is no response. After a positive response, an appointment for physical examination at a local health center was made over the phone followed by a confirmation letter of the appointment, and a digital or paper version of the questionnaire (depending on the preference of the participant) that was sent to the participant's home address.[1]

Recruitment strategy in the United Kingdom

The UK has no population register for migrant groups. Consequently, Ghanaian organizations served as the sampling frame. Lists of these organizations were obtained from the Ghanaian Embassy and the Association of Ghanaian Churches in the UK in the boroughs known to have the greatest concentration of Ghanaians. Lists of all members of their organizations, if available, were also requested, from which several all participants aged ≥25 years were invited to participate in the study. The selected participants were sent a written invitation combined with written information regarding the study and an opting out response card. Participants were sent a confirmation letter of the appointment for a physical examination at a local health center, church, or community center, including a digital or paper version of the questionnaire (depending on the preference of the participant) if they agreed to participate in the study.[1]

Recruitment strategy in Germany

In Berlin, a list of Ghanaian individuals (born in Ghana, or Ghanaian passport holders) was provided by the registration office and was supplemented with contact details of members of Ghanaian organizations and churches in Berlin. From this combined list, all participants aged ≥25 years were invited to participate in the study. In addition, a written invitation combined with written information regarding the study and a response card were sent to the selected participants. Participants were reminded after 2 weeks if there is no response. After a positive response, the participants were contacted by phone to schedule date and location of the interview with a trained research assistant or opt for the digital online version. After the completion of the questionnaire, a date for physical examination was then scheduled.[1]

REFERENCES

1. Agyemang C, Beune E, Meeks K, et al. Rationale and cross-sectional study design of the Research on Obesity and type 2 Diabetes among African Migrants: the RODAM study. 2015; **4**(3).
2. Organization WH. Health of migrants: the way forward: report of a global consultation. Health of migrants: the way forward: report of a global consultation; 2010: 112-.
3. Robards J, Berrington AJDR. The fertility of recent migrants to England and Wales. 2016; **34**: 1037-52.
4. Nieswand BJPdMzAuD, Bundesamt für Migration und Flüchtlinge, Deutsche Gesellschaft für Internationale Zusammenarbeit N, Eschborn. Der Migrations-Entwicklungs-Nexus in Afrika. Diskurswandel und Diasporaformation. 2011: 400-25.
5. Bethlehem J. 080u08urvey sampling ey sampli. 2009.
6. Agyemang C, Nicolaou M, Boateng L, Dijkshoorn H, van de Born B-J, Stronks KJEjopc. Prevalence, awareness, treatment, and control of hypertension among Ghanaian population in Amsterdam, the Netherlands: the GHAIA study. 2013; **20**(6): 938-46.

Appendix.2. Power calculations at the beginning of RODAM study (Ghanaians).

In the RODAM study, the aim was to sample 6250 individuals comprising five subgroups of 1250 each from five locations (rural Ghana, urban Ghana, Amsterdam, Berlin, and London). Subsets were to be selected for each of the phenotypic, genetic, and epigenetic studies. In consideration of these three types of data, three distinct types of statistical power calculations were evaluated. The three statistical power calculations are showed here to show how the RODAM numbers in the general populations and within the epigenetic analyses were arrived at.

Phenotypic association: For the phenotypic association analysis, we assumed a prevalence of (type 2 diabetes) T2D of <5% in rural Ghana, 6–7% in urban Ghana and >12% in Europe. For obesity, we assumed a prevalence of <5% in rural Ghana, 17% in urban Ghana and 30% in Europe.[1] In general, we aimed for a power of 0.90 with α=0.05 (including Bonferroni correction). Using these parameters, a sample size of approximately 1230 is needed in the rural Ghana, urban Ghana, Amsterdam, London, and Berlin subsets to detect a difference between the group proportions of 5%. T2D was defined as fasting plasma glucose ≥7 mmol/L, or pre-existing antidiabetic medication or glycated hemoglobin (HbA1c) ≥6.5%. Generalized obesity was defined as body mass index (BMI) ≥30 kg/m^2, and central obesity as a waist circumstance >102 cm in men or >88 cm in women.[1]

Genetic association: The RODAM study aimed to genotype a substantial part of the total sample among patients with diabetes (cases) and non-diabetes (controls). The genotyping would be based on a standard single nucleotide polymorphism array platform suitable for performing genome-wide association study as well as candidate gene analysis or related approaches. Using the statistical power calculator on binary traits for a case–control study design (S Purcell; http://pngu.mgh.harvard.edu/), assuming an allele frequency of 0.25, a prevalence of 16%, a relative risk of 1.3 and 1.6 (Aa and AA rep.), a D' of 1, the number of cases of 1000 (1:2 case–control ratio) and a type I error rate of 0.05, approximately 540 cases were considered sufficient to obtain a power of 0.90.[1] Although the latter necessary sample size was covered by our study, it should be noted that correction for multiple tests, that is, if more than one genetic variant was tested, was not considered here. Nevertheless, candidate gene or related association approaches within this single study setup have sufficient statistical power to detect moderate and low effect sizes.

Epigenetic association: To explore the influence of epigenetic factors on the main outcomes, we aimed to assess the differences in methylation levels among Ghanaians living in rural and urban Ghana and Europe and their relative contributions to the differences to obesity and T2D that may be observed. With a study power of 0.80 and α=0.05 (and SD±0.10%), at least 64 people per group are needed to detect a mean percentage difference in methylation density of 5%.[1] When multiple correction was considered, about 300 individuals per group of participants were needed. Epigenome-wide analysis would be performed. Association analysis of the DNA methylation profiles and obesity and T2D and related phenotypes would subsequently be performed. Promising loci would be validated using next-generation bisulfite sequencing.

REFERENCES

1. Agyemang C, Beune E, Meeks K, et al. Rationale and cross-sectional study design of the Research on Obesity and type 2 Diabetes among African Migrants: the RODAM study. 2015; **4**(3).

Appendix 3. Comparing baseline characteristics of participants in the current epigenetic study to the general RODAM study population (Ghanaians).

	Epigenetics study (Current analysis)		Overall RODAM study population	
	Non-migrants[1] (n=347)	Migrants[2] (n=365)	Non-migrants[1] (n=2566)	Migrants[2] (n=3332)
Demographic factors				
Chronological Age[3]	52.36(9.84)	49.9(9.76)	47.64(12.90)	46.51(11.02)
Female, n (%)	243(70.02)	166(45.47)	1714(66.80)	1955(58.71)
Education, n (%)				
No education	160 (46.10)	79(21.64)	1227 (50.23)	652(21.50)
Primary school	136(39.19)	154(42.21)	865(35.20)	1137(37.49)
Secondary school	38(10.95)	78(21.36)	247(10.11)	753(24.83)
Tertiary	13(3.74)	54(14.79)	103(4.22)	491(16.19)
Behavior related factors				
Any Alcohol consumption, n (%)	121(34.90)	152(41.60)	891(34.72)	1166(36.01)
Smoking, n (%)				
Current	5(1.44)	18(4.93)	38(1.60)	124(4.10)
Never	309(89.04)	316(86.57)	2246(92.00)	2651(88.01)
Past	33(9.51)	31(8.49)	157(6.42)	239(7.92)
Physical activity levels, n (%) [4]				
Low	131(37.75)	137(37.53)	712(29.26)	782(31.69)
Moderate	65(18.73)	86(23.56)	449(18.48)	534(21.64)
High	151(43.51)	142(38.90)	1269(52.22)	1152(46.68)
Total Energy intake, Kcal/day[5]	2386.24(824.96)	2894.10 (1113.31)	2601(1130.07)	2852(1206.01)
Length of stay for migrants, years[6]	NA	20.39(12.30–25.35)	NA	17.27 (9.97–23.90)
One carbon metabolism nutrient intake, mean (SD)[7]				
Vitamin B2 (riboflavin, mg/day)	1.29(0.51)	2.24(1.60)	1.33(0.53)	2.49(1.82)
Vitamin B6 (mg/day)	2.32(0.80)	3.01(1.29)	2.49(1.04)	3.14(1.35)
Vitamin B9 (folate; µg /day)	311.56(113.22)	445.31(203.87)	338.5(149.8)	466.9(217.6)
Vitamin B12 (cyanocobalamin; µg /day)[6]	5.92(3.81–10.17)	12.82(7.19–34.43)	5.49(3.92–7.46)	12.03(6.99–32.72)
Cardio-metabolic Factors, mean (SD)				
Body mass index (kg/m^2)	24.74(5.59)	28.18(4.92)	24.98(5.34)	28.75(4.95)
Systolic blood pressure (mmHg)	130.91(22.27)	136.77(18.44)	125.7(20.56)	134.2(17.89)
Diastolic blood pressure (mmHg)	80.8(12.32)	84.85(11.44)	78.34(12.10)	83.81(11.23)
Fasting blood glucose (mmol/L)[6]	5.13(4.73–6.18)	5.26(4.87–7.61)	5.04(4.69–5.40)	5.08(4.69–5.59)

[1]Ghanaians living in rural and urban Ghana were categorised as non-migrants, [2] Ghanaians living in Amsterdam, Berlin and London were categorised as migrants. [3]Age provided by the participant during questionnaire interviews, clock, [4] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria, [5] Total energy intakes obtained by food frequency questionnaires in Kcal/day, [6] Data presented as medians, interquartile range, [7] One carbon metabolism nutrient intake obtained via food frequency questionnaires. Abbreviations: NA= Not applicable.

Appendix 4. Description of data collection, processing, and storage in RODAM study (Ghanaians).

Data collection

Data collection was composed of questionnaire/interviews, physical examination, and biological samples.[1] The methods for data collection in RODAM were identical in all locations following standard operating procedures and applying standardized tools. Questionnaires were adapted to the local circumstances in Ghana, the Netherlands, the UK, and Germany, where needed. For example, for all sites, modification of questions with respect to educational system or social security system was made. Before data collection, a 2-day training course in the Netherlands was organized for all those involved in data collection, on the overall project's procedures. The training involved the administration of different questionnaires, physical examinations, processing of blood and urine samples in the laboratory and transport and storage of samples. The work package leaders, in turn, trained all recruitment team members in each site on all aspects of the study. Interviewers were recruited among Ghanaian-speaking residents in Europe, were introduced to the aims and procedures of the study and were instructed and trained on the use of the questionnaires. Research assistants also received the online tutorial for the Oracle Clinical data entry system. The performance of each interviewer or research assistant was monitored during the initial interviews and physical examinations. Feedback was provided and changes were initiated if needed. The RODAM coordinator monitored data collection at all study sites to ensure standardization of methods.

Questionnaire/interviews

All the participants who agreed to participate in the study received an appointment for a structured interview. The interviews were conducted by trained interviewers of Ghanaian background and lasted for about 60–120 min. To increase participation rate, the participants were given a range of options including interviews in participants' homes and digital or paper version of the questionnaire, depending on the preference of the participant. The interviewers conducted the interviews in the preferred language of the respondent either in English, German, Dutch or Ghanaian languages. The interview was based on a structured health questionnaire and contained questions on a large range of topics including demographics (age, sex, education), migration-related factors (age at first migration, duration of residence), health behavior (dietary behavior, physical activity, alcohol consumption, and tobacco smoking and use of anti-diabetic medications). Appropriate validated instruments were used for questionnaires. For example, physical activity was measured using the WHO Global Physical Activity Questionnaire (GPAQ) V.2.43.[2] Dietary behavior was determined by a Food Propensity Questionnaire (FPQ),[3] specifically developed in RODAM to include Ghana-specific foods. In addition, a 24 h dietary recall questionnaire was administered to a subset of the study population in each site (n=5*100). Dietary intake of Vitamin B2 (mg/day), Vitamin B6 (mg/day), Vitamin B9 (mcg/day), Vitamin B12 (mcg/day) and total energy intake (Kcal/day) was obtained from the FFQs.

Physical measurements

All participants who complete the questionnaire were invited for physical examination in the local research clinic or in a health center. At the start of the visit, the study and the procedures involved were explained to each participant and informed consent was signed if not already carried out so at home. After informed consent was given, physical measurements were made, and fasting blood samples were collected. Physical examinations were performed with validated devices according to standardized operational procedures. Physical examinations comprised assessment of anthropometrics including weight and height measurements. The portable stadiometer SECA 217 was used for height measurement, the SECA 877 for weight measurement. Blood pressure was measured three times using validated semiautomated device (The Microlife WatchBP home) with appropriate cuffs in a sitting position after at least 5 min rest. Each participant received a summary of his/her main results accompanied by an explanation and the recommendation to contact his/her GP if the results are abnormal.

Blood samples

Fasting venous blood samples were collected by trained research assistants in all sites. All blood samples were manually processed and aliquoted immediately after collection by a trained technician or research assistant according to standard operational procedures, and then temporarily stored at the local research location. Immediate

processing and cryopreservation at the research location had the advantage of preserving any highly labile molecules in the samples. Standardized procedures ensured that each sample were collected, handled, processed, transported, and stored in the same way across the sites. The samples were then transported to the respective local laboratories (Durrer Center for Cardiogenetic Research at the AMC, Amsterdam; Kwame Nkrumah University of Science and Technology, KCCR, Kumasi, Faculty of Infectious and Tropical Diseases, LSHTM, London & Institute of Tropical Medicine and International Health, Berlin), where samples were checked, registered, and stored at −80°C. These samples included EDTA whole blood, heparin plasma and serum. In addition, on the spot, fasting plasma glucose level were assessed by validated hand-held device (Accu-Chek Performa metre+Accu-Chek Inform II test strip (Roche, Germany) in all sites to provide accurate glucose determination as blood glucose concentration tends to decline over time in blood samples.

Transfer of biological material to dedicated centers for biochemical analyses and genotyping.

From the local research centers, the two aliquoted samples and a 2 mL EDTA were then transported to Berlin for biochemical analyses including glucose metabolism (fasting blood glucose). Another 4 mL EDTA whole blood sample was transported to Nottingham for DNA extraction and genotyping. Shipping of the samples from European sites was carried out using Styrofoam boxes filled with dry ice and from Ghana in dry shippers filled with liquid nitrogen. Numerous factors including temperature, packaging, courier, sample type, import/export requirements, seasons, costs, and transit time/ship days can affect biological specimen integrity during transportation domestically and internationally. Hence, staff involved with shipping the specimens were trained to minimize factors that might affect the integrity of the specimens. In addition, training regarding legal or regulatory aspects of shipment of specimens such as Material Transfer Agreement (MTA) were given. Each center maintained a shipment log to record the receipt and dissemination of shipments sent from the center. Each shipment entry is given a unique shipment number.

DNAm processing, profiling, and quality control

DNA extraction and methylation profiling were performed on whole-blood samples by Source Bioscience, Nottingham, UK. Bisulfite DNA treatment was achieved using the Zymo EZ DNA Methylation kit, and the quality of the conversion was determined by high-resolution melting analyses. The converted DNA was amplified and hybridized on the Illumina Human Methylation 450 K array, which measures DNA methylation levels of approximately 485 000 CpG sites. The samples were randomly divided over nine bisulfite conversion and hybridization batches.

Raw 450K data were processed for primary quality control using "R" *(version 3.2.2)* and MethylAid (*version 1.4.0*).[4] Methyl Aid detects bad quality samples using sample dependent and sample-independent control CpG sites present on the 450K array itself. Methyl Aid threshold values included: Methylated and Unmethylated intensities of 10.5, overall quality control = 11.75, bisulfite control = 12.75, hybridization control of 12.50, and a detection P-value of 0.95. Based on these thresholds, twelve samples were considered outliers. Genotyping data (not reported here) revealed eight samples with a sex discordance compared with the phenotype data that were subsequently excluded.

Functional normalization was applied using the "*preprocessFunnorm* function" of the "R" *minfi* package to normalize raw 450K data.[5] This method has been shown to outperform other methods. Principal component analysis (PCA) on the normalized dataset annotated for sex, recruitment site, self-reported ethnic group within Ghana, bisulfite batch, hybridization batch and plate position revealed three additional gender discordant samples and some stratification on plate position. No other outliers were observed in the PCA. Principal component 1 and 2 explained 30% of the variance. Sex discordant samples detected by genetic and/or epigenetic analyses were removed. Plate position was added as covariate in the statistical model. This resulted in a sample size of 712 for the current analyses. All nonspecific CpG sites were removed as well as CpG sites annotated to chromosomes X and Y. Removal of these CpG sites resulted in a set of 429,459 CpG sites which was used to identify differentially methylated positions (DMP) and differentially methylated regions (DMRs) in linear regression analysis. Cell

composition of whole blood samples is a source of variability in DNA methylation and has thereby the potential to cause confounding. We therefore estimated cell distributions using the method proposed by *Houseman et al.*[6]

Data linkage and management

Each participant was given a unique study number that linked all the data collected on them in the survey and laboratory results. The data were entered through the LimeSurvey and Oracle Clinical. Data are stored centrally at the Amsterdam University Medical Centre, Location AMC, on a SQL server. This server referred to 'data warehouse' is secured and engineered and maintained by an external party within the Amsterdam University Medical Centre, Location AMC, namely the Clinical Research Unit (CRU). Check programs were used to detect any problems such as duplicate identifiers. Clean data are currently available only to study scientific staff and approved collaborators. As part of a major initiative to make programme data more accessible to other scientists, anonymized data for the study will be submitted to the European Genome Archive (EGA).

REFERENCES

1. Agyemang C, Beune E, Meeks K, et al. Rationale and cross-sectional study design of the Research on Obesity and type 2 Diabetes among African Migrants: the RODAM study. 2015; **4**(3).

2. Armstrong T, Bull FJJoPH. Development of the world health organization global physical activity questionnaire (GPAQ). 2006; **14**(2): 66-70.

3. Subar AF, Dodd KW, Guenther PM, et al. The food propensity questionnaire: concept, development, and validation for use as a covariate in a model to estimate usual food intake. 2006; **106**(10): 1556-63.

4. Van Iterson M, Tobi EW, Slieker RC, et al. MethylAid: visual and interactive quality control of large Illumina 450k datasets. 2014; **30**(23): 3435-7.

5. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. 2014; **30**(10): 1363-9.

6. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. 2012; **13**(1): 1-16.

Appendix.5. Description of post-hoc analyses

Sensitivity analyses in RODAM study

We performed sensitivity analyses to check the robustness our findings. First, we assessed whether associations between cardiometabolic traits and EAA measures detected in migrants or non-migrants were also apparent in the total study population (migrants and non-migrants combined). This would indicate that our findings were not substantially influenced by large differences in baseline characteristics between the migrant and non-migrant groups, as the effects in the combined RODAM population would be largely free of influences of sub-group differences in characteristics.

Second, we assessed whether associations between FBG and EAA were influenced by usage of anti-diabetic medications. Anti-diabetic medications such as metformin and insulin have been associated with lower levels of EAA.[1] This sensitivity analyses were achieved by excluding participants taking anti-diabetes medications in the linear regression models.

Replication in independent cohorts

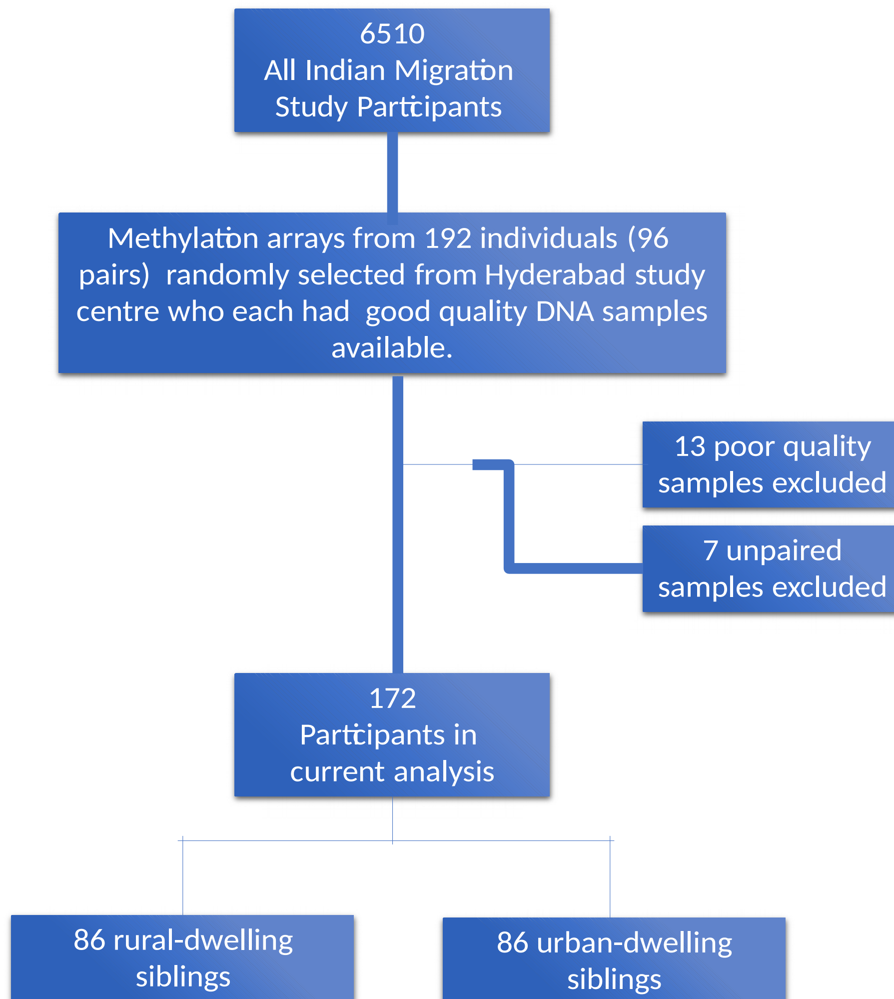
We sought to replicate our findings in independent cohorts of migrant and non-migrant populations from Africa or other low- or middle-income countries (LMIC) where urbanization and health transitions are prominent. As such, we performed replication analysis in the Indian Migration Study (IMS) and in Prospective Urban and Rural Epidemiology study's South African, North West province cohort (PURE-SA-NW).[2,3] The IMS sub-sample comprised of Infinium® HumanMethylation450 BeadChip data from 86 rural-urban Indian sibling pairs. Migrant urban siblings were factory workers in Hyderabad, or their spouses. Their non-migrant, rural dwelling siblings were sex and age matched (within 5 years) with their urban sibling.[2] The PURE-SA-NW study sub-sample (with DNAm data based on Infinium MethylationEPIC BeadChip) comprised of 120 native Batswana men from North West province in South Africa aged 45 to 88 years.[3] While we performed a full replication in the IMS (all statistical analyses), the replication in the PURE-SA-NW was partial for the following reasons: 1) participants from the PURE-SA-NW study were considered as non-migrants as there was no comparison group that had migrated out of South Africa to a HIC, 2) there was absence of overweight/obese individuals in PURE-SA-NW study which limited our replication of cardiometabolic factors (BMI, BP and FBG). In replication analyses, emphasis was placed on consistency in direction of effects due to smaller sample sizes in replication cohorts (limited statistical power). Detailed study descriptions and replication procedures are provided in Appendices 6 & 7, respectively.

REFERENCES

1. Li M, Duan L, Cai Y, Hao B, Chen J, Liu H. Effect of Metformin on the Epigenetic Age of Peripheral Blood in Patients With Diabetes Mellitus (DM). 2021.
2. Ebrahim S, Kinra S, Bowen L, et al. The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. *PLoS Med* 2010; **7**(4): e1000268.
3. Cronjé HT, Elliott HR, Nienaber-Rousseau C, Pieters M. Replication and expansion of epigenome-wide association literature in a black South African population. *Clinical Epigenetics* 2020; **12**(1): 1-13.

Appendix .6. Description of Indian Migration study (Indians)

Flow chart of participation



Study design

The design of the Indian Migration Study has been previously described previously.(1, 2) In summary, the IMS was conducted in four factories located in cities across India: Lucknow (Northern), Nagpur (Central), Hyderabad (Southern) and Bangalore (Southern). Factory workers and their co-resident spouses were recruited to the study if they were rural – urban migrants. Each participant identified a non-migrant sibling living in the rural village of origin who was also invited to join the study. If migrants had multiple siblings, they were asked to invite the one closest to them in age and of the same sex. The fieldwork for this study took place between 2005 and 2007. DNA methylation data was generated on a subset of IMS participants (n=192 samples, 96 pairs) recruited from Hyderabad. Sibling pairs were same sex and were age matched within 5 years of age.

Ethical approval and consent to participate.

Ethical approval for the IMS was obtained from an Indian central institutional review board (All India Institute of Medical Sciences (AIIMS), New Delhi, India (Reference Number: A-60/4/8/2004), as well as institutional review boards at each of the study sites. Written informed consent (witnessed thumbprint if illiterate) was obtained from the participants to utilize their unidentified biological samples in future for genetic studies.

DNA methylation data

DNA from IMS participants was analyzed using HumanMethylation450 BeadChips (Illumina, San Diego, CA, USA). Approximately 500ng of DNA was bisulphite modified using EZ DNA methylation kits (Zymo Research, Orange, CA, USA). The manufacturer's protocol was followed using the alternative incubation conditions recommended when using Illumina BeadChips. Quality control and processing was conducted according to the *meffil* pipeline which has been previously described.(3) Using this pipeline, samples were normalized using 18 control probe PCs derived from the technical probes informed by *meffil* scree plots. During normalization, BeadChip ID was included as a random effect. Following normalization, 172 samples and 478,580 probes were included in the dataset.

Measurements

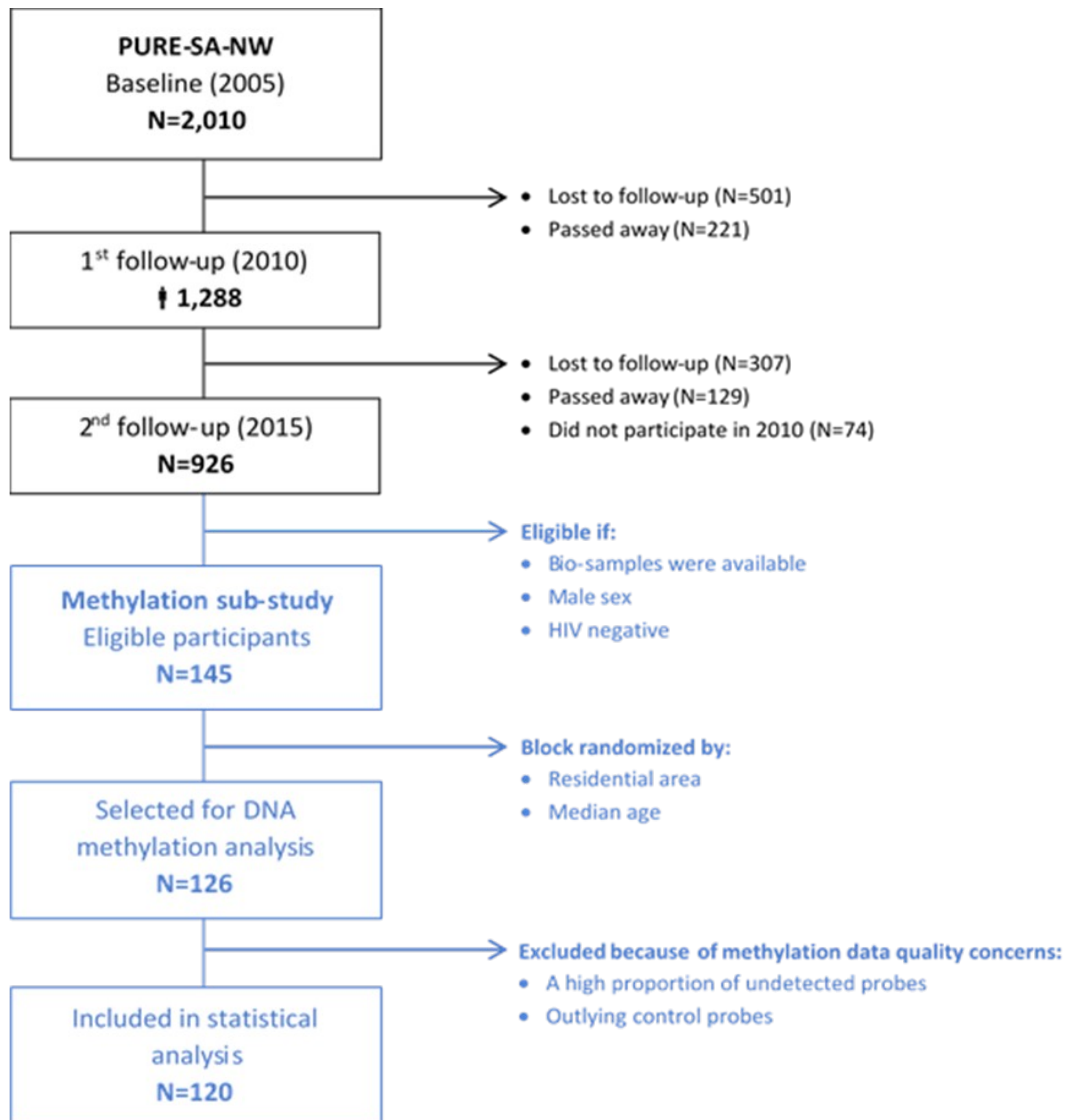
Fasting blood samples were collected and separated into plasma and serum. Fasting plasma glucose was measured on the day of collection at each of the four study centers as described previously.(1) Blood pressure was measured using an Omron M5-I automatic machine while sitting. Participants rested for 5 minutes before recording. The right upper arm and an appropriately sized cuff was used.

Participants were interviewed using a structured questionnaire to obtain information about tobacco use, alcohol consumption and physical activity. Current smokers were those who smoked within the last 6 months. Former smokers reported previous smoking but not in the last 6 months. For alcohol, individuals were dichotomized into those who reported any drinking vs those who reported that they had never consumed alcohol. Assessment of physical activity has been described in detail previously.(4) Physical Activity Levels were used to group individuals into categories using previously described cut-points.(4) Micro-nutrient levels were estimated from Food Frequency questionnaire data as described previously.(5) Height was measured using a portable plastic stadiometer. Body weight was measured using digital weighing scales with 100g accuracy. BMI was calculated as a function of weight (kg)/height (m^2).

REFERENCES

1. Ebrahim S, Kinra S, Bowen L, Andersen E, Ben-Shlomo Y, Lyngdoh T, et al. The effect of rural-to-urban migration on obesity and diabetes in India: a cross-sectional study. PLoS Med. 2010;7(4):e1000268.
2. Lyngdoh T, Kinra S, Shlomo YB, Reddy S, Prabhakaran D, Smith GD, et al. Sib-recruitment for studying migration and its impact on obesity and diabetes. Emerging Themes in Epidemiology. 2006;3(1):2.
3. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics. 2018;34(23):3983-9.
4. Sullivan R, Kinra S, Ekelund U, Bharathi A, Vaz M, Kurpad A, et al. Socio-demographic patterning of physical activity across migrant groups in India: results from the Indian Migration Study. PLoS One. 2011;6(10):e24898.
5. Shridhar K, Dhillon PK, Bowen L, Kinra S, Bharathi AV, Prabhakaran D, et al. Nutritional profile of Indian vegetarian diets–the Indian Migration Study (IMS). Nutrition journal. 2014;13(1):55.

Appendix.7. Descriptions of PURE-SA-NW study (native South Africans)



Flow chart of participation

Black proportion represents the larger PURE-SA-NW study that the methylation sub-study cohort was nested in. Blue proportion represents the PURE-SA-NW methylation sub-study cohort presented in this paper. Abbreviations: HIV: Human immunodeficiency virus, PURE-SA-NW: Prospective Urban and Rural Epidemiology Study-South African arm – North West cohort

Study design

The PURE-SA-NW is a sub-cohort of the international prospective urban rural epidemiology (PURE) study, comprising of self-identified Batswana adults residing in the North West province of South Africa. Detailed descriptions of the international PURE and PURE-SA-NW cohorts have been published previously.(1, 2) The PURE-SA-NW study data were collected in 2005, 2010 and 2015. For the sub-cohort investigated in the current investigation, 126 participants were randomly selected for DNA methylation analyses from a group of 990 individuals who took part in the 2015 PURE-SA-NW data collection. Eligibility depended on the following inclusion criteria: availability of bio-samples, testing negative for the human immunodeficiency virus at the time of data collection and male sex. These criteria were incorporated to eliminate confounding by sex and CD4 cell counts.(3)

Ethics approval and consent to participate.

Ethical approval for the 2015 data collection of the PURE-SA-NW study was granted by the Health Research Ethics Committee of the North-West University (NWU-00016-10-A1, NWU-00119-17-A1). All participants provided written informed consent, including consent for genetic/epigenetic analysis. All procedures described were performed in accordance with the revised version of the Helsinki Declaration of 1975.

DNA methylation data

DNA was isolated using QIAGEN Flexigene DNA extraction kits (QIAGEN® Valencia, CA, USA) according to the manufacturer's protocol with minor modifications. (8) Quantification was conducted with PicoGreen® dsDNA quantitation assay (Invitrogen™, Carlsbad, CA, USA). Five hundred nanograms of DNA from each participant was bisulfite-converted using the Zymo EZ DNAm™ kit (Zymo Research, Irvine, CA, USA), followed by genome-wide DNAm profiling on the Illumina Infinium MethylationEPIC BeadChip, according to the manufacturer's protocol (Illumina®, San Diego, CA, USA). Samples were randomized across slides to minimize the possibility of confounding by batch. Quality control and processing was conducted according to the *meffil* pipeline which has been previously described. (3) Using this pipeline, samples were normalized using 18 control probe principal components (PCs) derived from the technical probes informed by *meffil* scree plots. Principal component analysis of the control probes identified 12 PCs to be included in the functional normalization. In addition, slide was specified as a random effect to be included to address batch variance. Sample cell fractions (B-cells, CD4 and CD8 T-cells, neutrophils, monocytes and natural killer cells) were estimated using the IDOL optimized L-DMR library for whole blood samples.(3) Following quality control and normalization, 120 samples and 857,516 probes were included in the final dataset.(3)

Measurements

Height and weight were quantified using a stadiometer and an electronic scale. BMI was calculated as weight per unit height squared (kg/m^2). Alcohol consumption (current, former, or never) and smoking status (current, former or never) were self-reported, using a standardized questionnaire. Alcohol consumption was dichotomized as any or no consumption to ensure comparability to the RODAM data. Fasting blood samples were collected and handled as described previously.(4) Fasting blood glucose was quantified using the Cobas® Integra 400 (Roche® Clinical System, Roche Diagnostics, Indianapolis, IN, USA). Systolic and diastolic blood pressure were measured using the OMRON M6 device (Omron Healthcare, Kyoto, Japan). Two measurements were taken after a 10-minute rest period. Data from the second measurement were used for analysis.

REFERENCES

1. Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. American heart journal. 2009;158(1):1-7. e1.
2. De Lange Z, Pieters M, Jerling JC, Kruger A, Rijken DC. Plasma clot lysis time and its association with cardiovascular risk factors in black Africans. PLoS one. 2012;7(11):e48881.
3. Cronjé HT, Elliott HR, Nienaber-Rousseau C, Pieters M. Replication and expansion of epigenome-wide association literature in a black South African population. Clinical Epigenetics. 2020;12(1):1-13.
4. Pieters M, Kotze RC, Jerling JC, Kruger A, Ariëns RA. Evidence that fibrinogen γ′ regulates plasma clot structure and lysis and relationship to cardiovascular risk factors in black Africans. Blood, The Journal of the American Society of Hematology. 2013;121(16):3254-60.

Appendix 8. Lifestyle factors, cardiometabolic traits and epigenetic age acceleration in the total RODAM study population (migrant and non-migrant Ghanaians combined).

	IEAA (Horvath)[1] (N=712)		EEAA (Hannum)[2] (N=712)		PhenoAgeAccel[3] (N=712)		PhenoAgeAccel[4] (N=712)	
	Crude[5]	Fully adjusted [6]	Crude[5]	Fully adjusted [6]	Crude[5]	Fully adjusted[6]	Crude[5]	Fully adjusted[6]
	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Lifestyle factors								
Tobacco smoking								
Never	ref	ref	ref	ref	ref	ref	ref	ref
Past	–0.96 (–1.53;3.48)	–1.21 (–4.19;1.77)	–1.69 (–3.88;0.49)	1.65 (–0.89;4.19)	0.14 (–2.73;3.00)	–0.49 (–3.85;2.87)	–1.30 (–3.07;0.46)	0.01 (–0.01;0.02)
Current	–0.12 (–1.43;1.67)	–0.33 (–3.69;3.05)	0.29 (–1.07;1.64)	1.81 (–1.07;4.67)	–0.80 (–2.57;0.97)	–0.51 (–4.31;3.29)	–3.84 (–5.39;2.30)	–0.01 (–0.05;0.0)
Physical activity levels [7]								
Low	ref	ref	ref	ref	ref	ref	ref	ref
Moderate	–0.52 (–1.72;0.68)	–0.25 (–1.64;1.15)	–0.69 (–1.74;0.35)	–0.28 (–1.48;0.90)	0.07 (–1.30;1.44)	0.60 (–0.97;2.17)	0.12 (–0.65;0.88)	–0.16 (–0.46;0.14)
High	–0.89 (–1.89;0.10)	–0.35 (–1.49;0.81)	–0.65 (–1.52;0.22)	–0.39 (–1.37;0.60)	–0.65 (–1.79;0.49)	–0.49 (–1.79;0.80)	–0.02 (–0.65;0.61)	0.23 (–0.35;0.82)
One carbon metabolism nutrients [7]								
Vit B2	–0.48 (–0.82;0.14)	–0.34 (–0.93;0.25)	–0.41 (–0.71; –0.11)	–0.29 (–0.79;0.21)	–0.56 (–0.95; –0.17)	–0.79 (–1.46; –0.13)	–0.02 (–0.24;0.19)	–0.16 (–0.47; –0.14)
Vit B9	–0.00 (–0.01;0.01)	0.02 (–0.01;0.03)	–0.01 (–0.02; –0.001)	–0.01 (–0.02; –0.001)	–0.01 (–0.02; –0.001)	–0.02 (–0.03; –0.01)	0.001 (–0.002;0.003)	–0.001 (–0.005;0.002)
Cardiometabolic factors								
BMI	–0.11 (–0.19;0.03)	–0.09 (–0.17; –0.01)	–0.14 (–0.21; –0.08)	–0.12 (–0.19; –0.05)	–0.26 (–0.35; –0.17)	–0.26 (–0.35; –0.17)	–0.13 (–0.18;0.08)	–0.08 (–0.13; –0.04)
FBG	0.17 (0.04;0.30)	0.14 (0.01;0.28)	0.19 (0.07;0.31)	0.17 (0.05;0.29)	0.28 (0.13;0.43)	0.26 (0.11;0.41)	0.18 (0.11;0.27)	0.16 (0.09;0.24)

[1]Intrinsic epigenetic age acceleration (within each cell) obtained using Horvath clock, [2]Extrinsic epigenetic age acceleration (between different cells) using Hannum clock, [3] Epigenetic age acceleration incorporating clinical traits obtained using the PhenoAge clock, [4]Epigenetic age acceleration incorporating plasma proteins obtained using GrimAge clock, [5]crude linear regression model, [6] fully adjusted linear regression model for age, sex, education, smoking, physical activity, alcohol intake, vitamin intake, total energy intake, and duration of stay in host countries for migrants respectively, [7] Physical activity categories according to the GPAQ questionnaire. Vit= Vitamin, Vitamin intake was measured in mg/day for Vitamin B2, while vitamin B12 was measured in mcg/day, BMI= body mass index (kg/m^2), FBG= fasting blood glucose (mmol/L). Highlighted results are the statistically significant results in the adjusted linear regression models. CI= confidence interval

Appendix 9. Sensitivity linear regression analyses in the RODAM study (Ghanaians) on fasting blood glucose by excluding participants taking anti-diabetic medications.

	Non–migrants[1] (N=310)		Migrants[2] (N=342)	
	Crude	**Adjusted**	**Crude**	**Adjusted**
	Beta (95%CI)	**Beta (95% CI)**	**Beta (95% CI)**	**Beta (95%CI)**
IEAA				
FBG, mmol/L	0.08(–0.08;0.25)	0.06(–0.11;0.23)	0.30(0.00;0.61)	0.33(0.02;0.64)
EEAA				
FBG, mmol/L	0.07(–0.07;0.21)	0.03(–0.11;0.18)	0.30(0.03;0.57)	0.28(0.01;0.55)
PhenoAgeAccel				
FBG, mmol/L	0.17(–0.02;0.36)	0.15(–0.04;0.35)	0.35(0.01;0.68)	0.40(0.06;0.75)
GrimAgeAccel				
FBG, mmol/L	0.15(0.05;0.25)	0.13(0.02;0.23)	0.27(0.07; 0.47)	0.17(0.02; 0.32)

[1] Participants from rural and urban Ghana, [2] Migrant Ghanaians living in Europe (Amsterdam, Berlin, London), [3] Fully adjusted for age, sex, education, alcohol consumption, smoking, physical activity, total energy, BMI, and duration of stay in the host country among migrants, Abbreviations: FBG=fasting blood glucose, IEAA= Intrinsic epigenetic age acceleration obtained via Horvath clock, EEAA= extrinsic epigenetic age acceleration obtained via Hannum clock, PhenoAgeAccel= epigenetic age acceleration obtained via PhenoAge clock, GrimAgeAccel= epigenetic age acceleration obtained via Grim Age clock. A total of 60 participants taking diabetic medication were excluded, CI= confidence interval.

Appendix 10. Baseline characteristics of participants in IMS study (Indians).

	Non-migrants[1] (n=86)	Migrants[2] (n=86)
Aging estimators, mean (SD)		
Chronological Age[3]	45.91(8.17)	46.16(7.29)
DNAm HorvathAge[4]	51.48(7.72)	50.46(9.21)
DNAm HannumAge[5]	60.76(8.16)	61.07(8.79)
DNAm PhenoAge[6]	50.80(10.33)	52.62(10.17)
DNAm GrimAge[7]	53.55(9.32)	52.03(7.41)
IEAA[8]	0.30(5.27)	-0.30(6.1)
EEAA[9]	-0.26(4.37)	0.36(5.3)
PhenoAgeAccel[10]	-0.83(7.67)	0.83(8.48)
GrimAgeAccel[11]	0.88(4.5)	-0.88(3.85)
Other Demographic factors		
Female, n (%)	38(44.19)	38(44.19)
Education, n (%)		
Primary school	54 (62.79)	27(31.39)
Secondary school	38(10.95)	78(21.36)
Tertiary	24(27.90)	44(51.26)
Behavior related factors		
Any Alcohol consumption, n (%)	26(30.23)	27(31.39)
Smoking, n (%)		
Current	17(19.86)	7(8.14)
Never	69(80.23)	76(88.4)
Past	0	3(3.4)
Physical activity levels, n (%) [12]		
Low	40(46.51)	35(40.69)
Moderate	30(34.88)	36(41.86)
High	12(13.95)	10(11.62)
Total Energy intake, Kcal/day[13]	2467.38(852.06)	2741.71(928.10)
Length of stay for migrants, years	NA	26.01(9.81)
One carbon metabolism nutrient intake, mean (SD)[14]		
Vitamin B2 (riboflavin, mg/day)	1.48(0.59)	1.89(0.80)
Vitamin B6 (mg/day)	1.87(0.76)	2.18(0.92)
Vitamin B9 (folate; µg /day)	266.19(107.75)	328.32(139.14)
Vitamin B12 (cyanocobalamin; µg /day)[15]	4.05(6.07)	5.53(6.55)
Cardio-metabolic Factors, mean (SD)		
Body mass index (kg/m^2)	23.00(4.1)	25.84(3.44)
Systolic blood pressure (mmHg)	125.7(20.24)	124.44(15.99)
Diastolic blood pressure (mmHg)	81.64(11.72)	79.2(9.12)
Fasting blood glucose (mmol/L)[15]	4.97(0.12)	5.12(0.69)

[1]Rural non-migrant siblings in India. [2] Siblings that migrated from rural to urban areas in India, [3]Age provided by the participant during questionnaire interviews, [4]DNA methylation age obtained using the

Horvath clock, [5]DNA methylation age obtained using the Hannum clock, [6] DNA methylation age obtained using the PhenoAge clock, [7]DNA methylation age obtained using GrimAge clock, [8]Intrinsic epigenetic age acceleration (within each cell) obtained using Horvath clock,[9] Extrinsic epigenetic age acceleration (between different cells) using Hannum clock, [10] Epigenetic age acceleration incorporating clinical traits obtained using the PhenoAge clock, [11]Epigenetic age acceleration incorporating plasma proteins obtained using GrimAge clock, [12] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria, [13] Total energy intakes obtained by food frequency questionnaires in Kcal/day, [14] One carbon metabolism nutrients intake obtained via food frequency questionnaires, [15] Data presented as medians, interquartile range. Categorical variables do not always add up to column totals due to missing values. Abbreviations: NA= Not applicable.

Appendix 11. Pearson's correlations between age measures and between age acceleration measures in IMS study (Indians).

a) Correlations between age measures:

	Chronological Age[1]	DNAm HorvathAge[2]	DNAm HannumAge[3]	DNAm PhenoAge[4]	DNAm GrimAge[5]
Chronological Age	1.00				
DNAm HorvathAge	0.65	1.00			
DNAm HannumAge	0.79	0.71	1.00		
DNAm PhenoAge	0.61	0.65	0.66	1.00	
DNAm GrimAge	0.86	0.60	0.72	0.52	1.00

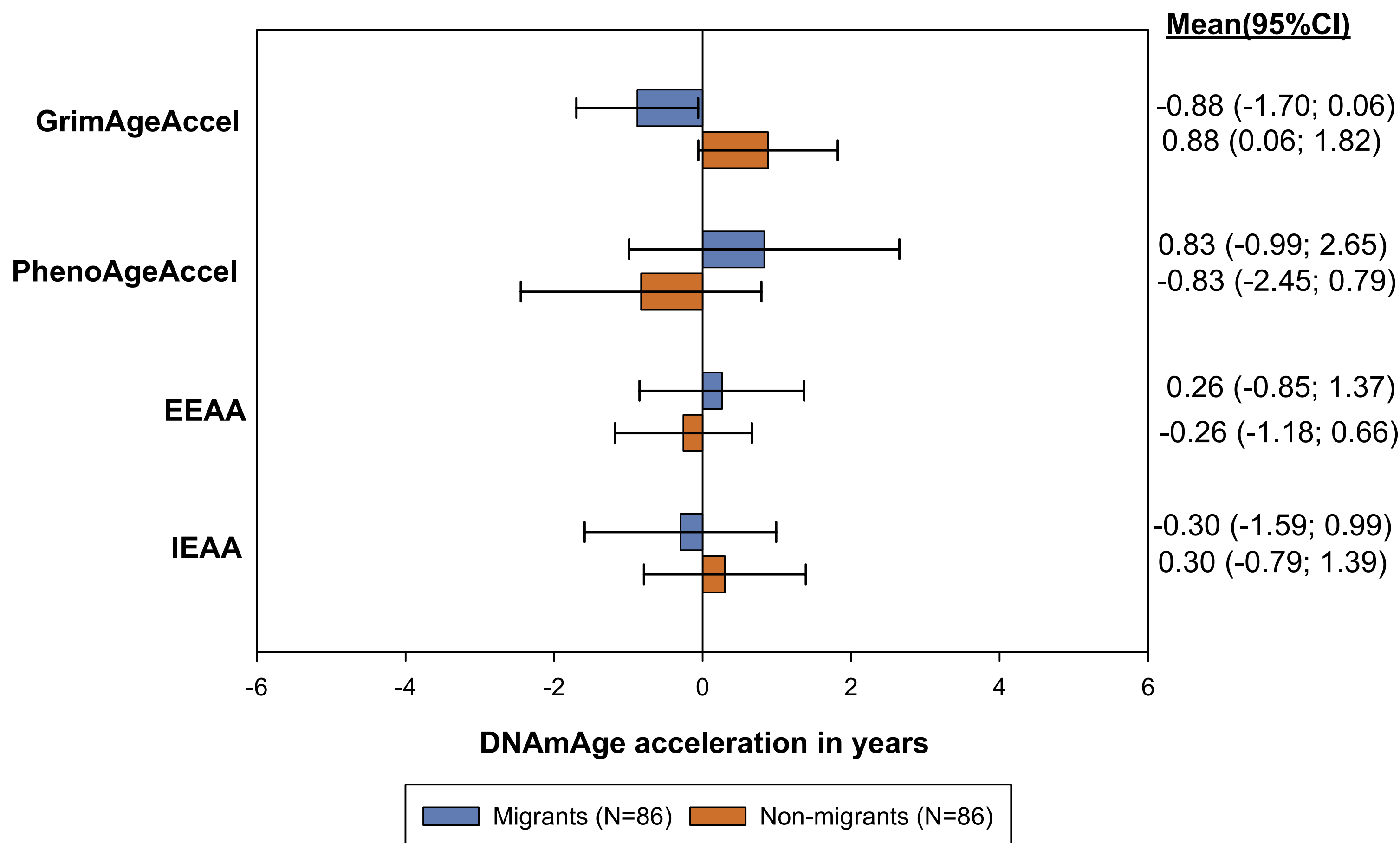
b) Correlations between DNAmAge acceleration measures:

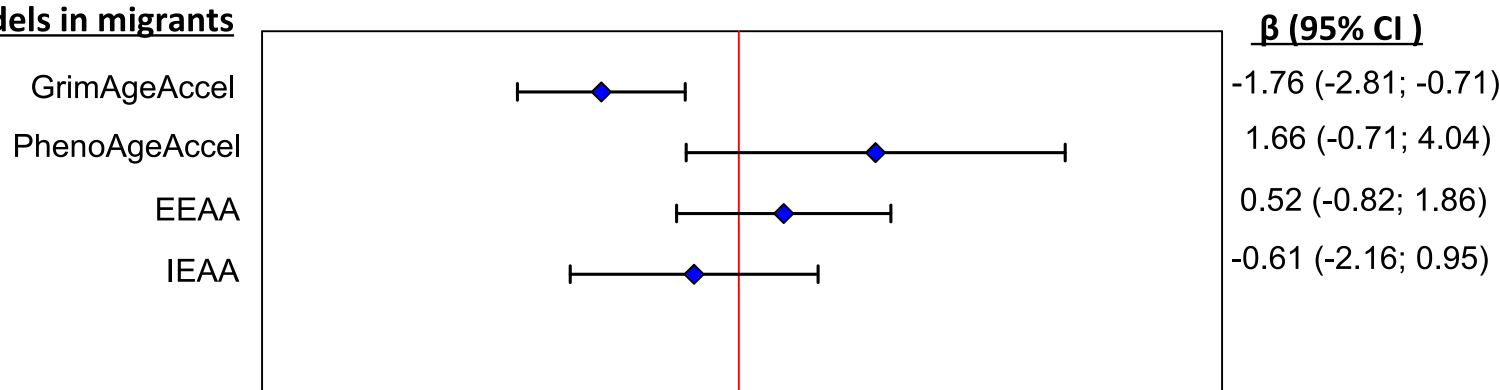
	IEAA (Horvath)[6]	EEAA (Hannum)[7]	PhenoAgeAccel[8]	GrimAgeAccel[9]
IEAA (Horvath)	1.00			
EEAA (Hannum)	0.41	1.00		
PhenoAgeAccel	0.40	0.28	1.00	
GrimAgeAccel	0.07	0.06	-0.01	1.00

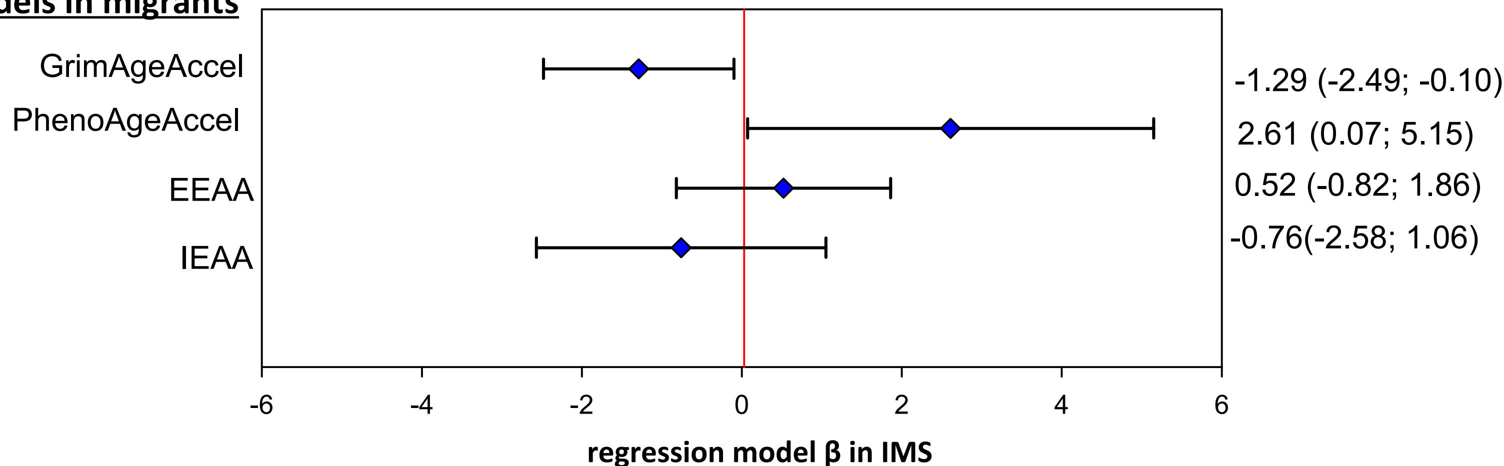
[1]Chronological age obtained from interview questionnaires, [2]Horvath Age obtained using the Horvath clock, [3]Hannum age obtained using Hannum clock, [4] DNAm PhenoAge obtained using the PhenoAge clock, [5] DNAm Grim Age obtained using the Grim Age clock. [6] Intrinsic epigenetic age acceleration obtained via the Hannum clock, [7]Extrinsic epigenetic age acceleration obtained via Hannum clock, [8] PhenoAge Acceleration obtained via the PhenoAge clock, [9] GrimAge acceleration obtained via the GrimAge clock.

Appendix.12: Migration status and EAA Measures in IMS study. The plot depicts mean (and 95% confidence intervals) for epigenetic age acceleration (IEAA, EEAA, PhenoAgeAccel and GrimAgeAccel) in migrants compared to non-migrants. Additionally, the plot also depicts regression model β with 95% confidence intervals for the associations between migration status and four EAA Measures in RODAM study. Red line = reference line for confidence intervals. Abbreviations: IEAA= Intrinsic epigenetic age acceleration, EEAA= extrinsic epigenetic age acceleration, PhenoAge Accel= Pheno Age Acceleration, GrimAge Accel= Grim Age Acceleration.

a) Mean DNAmAge acceleration (years) in migrant vs non-migrants in IMS (Indians).







b) Beta coefficients for linear regression models of DNAmAge accelerations versus migration status. Outcome = DNAmAge acceleration measures ; predictor= migration status; reference group = non-migrants. Model 2 adjusted for age, sex and education. Red line= reference line for confidence intervals.

Appendix.13: Forest plots of regression model β (and 95% confidence intervals) relating EAA Measures to cardiometabolic related traits in the IMS study. The plot depicts regression model β with 95% confidence intervals. N=86 for migrants and 86 for non-migrants. Red line = reference line for confidence intervals. Abbreviations: IEAA= Intrinsic epigenetic age acceleration, EEAA= extrinsic epigenetic age acceleration, PhenoAge Accel= Pheno Age Acceleration, GrimAge Accel= Grim Age Acceleration, ref= reference, active= levels of physical activity, BMI= body mass index (kg/m2), FBG= fasting blood glucose (mmol/L), Vit= Vitamin, BP=blood pressure (mmHg), Length stay= duration of stay in Europe for migrants, adj=final models adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, and duration of stay in host countries for migrants respectively, crude= un-adjusted models. Vitamin intake was measured in mg/day for Vitamin B6 and Vitamin B12, while vitamin B9 and Vitamin B12 were measured in mcg/day. Reference (comparison) groups for modifiable risk factors: smoking = non-smokers, alcohol consumption = no (never) alcohol consumption, physical activity= less physically active.

IEEA obtained via Horvath Clock (IMS)
EEEA obtained via Hannum Clock (IMS)
PhenoAgeAccel (IMS)
GrimAgeAccel (IMS)
Non-migrants
Migrants
Modifiable Lifestyle factors
Any alcohol. crude
Any alcohol.adj
Moderate active.crude
Moderate active.adj
Highly active.crude
Highly active.adj
Current smokers.crude
Current smokers. adj
Past smokers.crude
Past smokers.adj
Length stay.crude
Length stay.adj
One carbon metabolism nutrients
Vit B2.crude
Vit B2.adj
Vit B6.crude
Vit B6.adj
Vit B9.crude
Vit B9.adj
Vit B12.crude
Vit B12.adj
Cardiometabolic factors
BMI.crude
BMI.adj
Systolic BP.crude
Systolic BP.adj
Diastolic BP.crude
Diastolic BP.adj
FBG.crude
FBG.adj

Appendix 14. Lifestyle factors, cardiometabolic traits and epigenetic age acceleration in migrant vs non-migrant Indians (IMS study).

	IEAA (Horvath)[1]		EEAA (Hannum)[2]		PhenoAgeAccel[3]		GrimAgeAccel[4]	
	Non–migrants[5] (N=86)	Migrants[6] (N=86)	Non–migrants[5] (N=86)	Migrants[6] (N=86)	Non–migrants[5] (N=86)	Migrants[6] (N=86)	Non–migrants[5] (N=86)	Migrants[6] (N=86)
	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Lifestyle factors								
Alcohol consumption								
No	ref	ref	ref	ref	ref	ref	ref	ref
Yes (crude)	0.97 (–1.49;3.44)	0.22 (–2.62;3.05)	–0.62 (–2.67;1.42)	–0.07 (–2.53;2.39)	–0.16 (–3.76;3.43)	–1.58 (–5.51;2.33)	2.79 (0.77;4.82)	1.06 (–0.76;2.78)
Yes (adjusted)	–0.06 (–3.01;2.89)	0.08 (–0.16;0.32)	–1.73 (–4.10;0.63)	–1.02 (–4.15;2.10)	–0.19 (–4.59;4.19)	–0.14 (–4.91;4.63)	0.98 (–0.86;2.85)	–0.20 (–0.32; –0.09)
Physical activity levels[7]								
Low	ref	ref	ref	ref	ref	ref	ref	ref
Moderate (crude)	–3.04 (–7.69;1.64)	5.91 (–0.49;12.30)	–2.96 (–6.84;0.92)	4.75 (–0.79;10.31)	–3.45 (–10.40;3.49)	2.22 (–6.77;11.22)	3.43 (–0.64;7.52)	–0.13 (–4.13;3.89)
Moderate (adjusted)	–2.87 (–7.96;2.21)	3.67 (–3.91;11.25)	–4.28 (–8.36; –0.19)	4.52 (–2.25;11.31)	–3.77 (–11.34;3.79)	2.08 (–8.25;12.42)	0.92 (–2.28;4.12)	0.61 (–3.07;4.29)
High (crude)	–4.17 (–9.39;1.04)	6.27 (–0.90;13.45)	–4.52 (–8.86; –0.17)	5.64 (–0.59;11.88)	–4.01 (–11.78;3.76)	3.76 (–6.32;13.86)	2.27 (–2.29;6.83)	–3.09 (–7.59;1.40)
High (adjusted)	–3.44 (–9.13;2.24)	4.68 (–3.29;12.66)	–4.38 (–8.95;0.18)	4.93 (–2.20;12.07)	–4.59 (–13.05;3.86)	4.63 (–6.24;15.52)	0.95 (–2.62;4.53)	–0.72 (–4.61;3.14)
Tobacco smoking								
Never	ref	ref	ref	ref	ref	ref	ref	ref
Past (crude)	NA	–4.85 (–11.95;2.26)	NA	–0.99 (–7.23;5.23)	NA	–8.01 (–17.89;1.86)	NA	3.89(–0.09;7.88)
Past (adjusted)	NA	–5.42 (–13.48;2.63)	NA	–0.17 (–7.39;7.03)	NA	–7.67 (–18.66;3.32)	NA	4.21 (0.29;8.13)
Current (crude)	0.26 (–2.59;3.11)	2.01 (–2.76;6.77)	1.02 (–1.33;3.38)	2.13 (–2.04;6.32)	–1.23 (–5.37;2.91)	–2.21 (–8.83;4.42)	6.39 (4.39;8.39)	6.43 (3.75;9.11)
Current (adjusted)	0.72 (–3.35;4.80)	2.48 (–2.86;7.84)	0.90 (–2.37;4.17)	2.20 (–2.58;7.00)	–4.46 (–10.53;1.61)	0.55 (–6.75;7.86)	4.58 (2.01;7.14)	4.68 (2.08;7.29)

Length of stay in Europe among migrants								
Length of stay (crude)	NA	0.03 (–0.11;0.16)	NA	0.03 (–0.09;0.14)	NA	–0.02 (–0.21;0.17)	NA	–0.12 (–0.21;0.04)
Length of stay (adjusted)	NA	–0.11 (–0.34;0.11)	NA	–0.02 (–0.22;0.18)	NA	–0.09 (–0.41;0.22)	NA	–0.07 (–0.18;0.04)
One carbon metabolism nutrients [7]								
Vit B2 (crude)	–0.76 (–2.70;1.18)	0.93 (–0.70;2.56)	0.33 (–1.28;1.94)	1.03 (–0.38;2.45)	–1.31 (–4.13;1.50)	–0.42 (–2.71;1.86)	0.43 (–1.23;2.08)	–1.03 (–2.05;0.02)
Vit B2 (adjusted)	–1.68 (–6.79;3.42)	–0.34 (–5.08;4.41)	2.11 (–1.98;6.20)	–1.29 (–5.49;2.90)	–1.11 (–8.71;6.48)	0.26 (–6.18;6.69)	–1.46 (–4.62;1.70)	1.09 (–1.22;3.41)
Vit B6 (crude)	0.63(–0.87;2.14)	0.17 (–1.25;4.41)	0.43 (–0.81;1.68)	0.68 (–0.54;1.92)	–0.57 (–2.77;1.62)	–1.21 (–3.18;0.76)	1.39 (0.14;2.65)	–1.02 (–1.89; –0.14)
Vit B6 (adjusted)	1.04(–1.50;3.57)	–4.29 (–9.48;0.90)	0.02 (–2.01;2.05)	–2.96 (–7.55;1.63)	0.58 (–3.19;4.36)	–8.41 (–15.46; –1.37)	0.46 (–1.11;2.03)	–0.08 (–2.61; –2.46)
Vit B9 (crude)	–0.001 (–0.02;0.01)	0.003 (–0.01;0.02)	–0.002 (–0.01;0.01)	0.003 (–0.005;0.01)	–0.01 (–0.02;0.01)	–0.001 (–0.01;0.01)	–0.002 (–0.01;0.01)	–0.01 (–0.01;0.003)
Vit B9 (adjusted)	–0.001 (–0.02;0.02)	0.001 (–0.02;0.03)	–0.01 (–0.02;0.002)	–0.0001 (–0.02;0.02)	–0.02 (–0.05;0.01)	0.02 (–0.01;0.05)	–0.001 (–0.02;0.01)	–0.01 (–0.02;0.003)
Vit B12 (crude)	–0.02 (–0.22;0.16)	0.17 (–0.02;0.37)	–0.04 (–0.20;0.11)	0.16 (–0.003;0.34)	0.06 (–0.21;0.33)	–0.08 (–0.36;0.21)	0.03 (–0.13;0.19)	–0.04 (–0.17;0.08)
Vit B12 (adjusted)	0.06 (–0.22;0.35)	0.25 (–0.07;0.57)	–0.07 (–0.30;0.16)	0.23 (–0.06;0.52)	0.17 (–0.24;0.61)	0.03 (–0.41;0.47)	0.06 (–0.12; 0.24)	0.02 (–0.14;0.18)
Cardiometabolic factors								
BMI (crude)	–0.23 (–0.50;0.04)	0.30 (–0.07;0.68)	–0.24 (–0.47; –0.02)	0.18 (–0.15;0.51)	–0.24 (–0.47; –0.02)	0.18 (–0.14;0.51)	–0.05 (–0.28;0.19)	–0.03 (–0.27;0.21)
BMI (adjusted)	–0.12 (–0.48;0.23)	0.23 (–0.29;0.76)	–0.16 (–0.45;0.12)	0.12 (–0.35;0.59)	–0.02 (–0.56;0.52)	0.44 (–0.29;1.16)	0.19 (–0.03;0.42)	0.03 (–0.24;0.30)
SBP (crude)	0.006 (–0.05;0.06)	0.04 (–0.07;0.08)	0.006 (–0.05;0.06)	0.01 (–0.08;0.09)	0.01 (–0.05;0.06)	0.005 (–0.09;0.09)	0.00 (–0.04;0.05)	0.07 (0.01;0.13)
SBP (adjusted)	0.001 (–0.07;0.07)	–0.03 (–0.12;0.07)	–0.002 (–0.05;0.06)	–0.02 (–0.11;0.07)	–0.002 (–0.06;0.06)	–0.02 (–0.11;0.07)	0.02 (–0.02;0.06)	0.07 (0.02;0.12)
DBP (crude)	0.003 (–0.09;0.10)	–0.01 (–0.15;0.14)	0.003 (–0.09;0.10)	–0.01 (–0.15;0.14)	0.003 (0.09;0.10)	–0.005 (–0.15;0.14)	–0.02 (–0.09;0.07)	0.12 (0.03;0.20)
DBP (adjusted)	–0.01 (–0.12;0.11)	–0.04 (–0.20;0.13)	–0.01 (–0.10;0.08)	–0.02 (–0.17;0.13)	–0.01 (–0.11;0.09)	–0.02 (–0.17;0.13)	0.01 (–0.06;0.08)	0.12 (0.04;0.19)
FBG (crude)	–2.27 (–4.67;0.11)	1.60 (–0.27;3.48)	–1.43 (–3.42;0.57)	1.66 (0.05;3.29)	–1.42 (–3.42;0.56)	1.67 (0.05;3.29)	–0.93 (–2.99;1.13)	–0.06 (–1.26;1.15)
FBG (adjusted)	–2.29 (–5.12;0.54)	1.49 (–0.52;3.52)	–0.83 (–3.18;1.51)	1.45 (–0.33;3.25)	–0.83 (–3.18;1.50)	1.46 (–0.34;3.25)	–1.66 (–3.45;0.13)	–0.35 (–1.38;0.68)

[1]Intrinsic epigenetic age acceleration (within each cell) obtained using Horvath clock, [2]Extrinsic epigenetic age acceleration (between different cells) using Hannum clock, [3]Epigenetic age acceleration incorporating clinical traits obtained using the PhenoAge clock, [4]Epigenetic age acceleration incorporating plasma proteins obtained using Grim Age clock, [5] Rural non–migrant siblings in India. [6] Siblings that migrated from rural to urban areas in India, [7] Levels of physical activity categorised according to GPAQ criteria. Abbreviations : Crude = crude linear regression model, adjusted= fully adjusted linear regression model for age, sex,

education, smoking, physical activity, alcohol intake, total energy intake, and duration of stay in host countries for migrants respectively, Length stay= duration of stay in Europe for migrants, Vit= Vitamin, Vitamin intake was measured in mg/day for Vitamin B2 and Vitamin B6, while vitamin B9 and Vitamin B12 were measured in mcg/day, BMI= body mass index (kg/m^2), SBP= systolic blood pressure (mmHg), DBP= diastolic blood pressure (mmHg), FBG= fasting blood glucose (mmol/L).

Appendix 15. Baseline characteristics of participants in PURE-SA-NW study (native South Africans).

	Replication in PURE-SA-NW[1]
	Total (n=120)
Aging estimators, mean (SD)	
Chronological Age[2]	62.9(9.97)
DNAm HorvathAge[3]	56.2(8.29)
DNAm HannumAge[4]	51.4(7.89)
DNAm PhenoAge[5]	47.1(8.76)
DNAm GrimAge[6]	64.3(8.95)
IEAA[7]	4.4e-17(6.64)
EEAA[8]	5.7e-17(5.62)
PhenoAgeAccel[9]	2.26e-17(2.26)
GrimAgeAccel[10]	1.8e-16(5.32)
Other Demographic factors	
Female, n (%)	0.00
Education, n (%)	
No education	26(21.7)
Primary school	6 (55.0)
Secondary school	2 (23.3)
Tertiary	-
Behavior related factors	
Any Alcohol consumption, n (%)	58(48.3)
Smoking, n (%)	
Current	61(50.8)
Never	56(46.7)
Past	3(2.5)
Physical activity levels, n (%) [11]	2.41 (0.94)
Total Energy intake, Kcal/day[12]	2905.14(1428.22)
One carbon metabolism nutrient intake, mean (SD)[13]	
Vitamin B2 (riboflavin, mg/day)	1.79 (1.21;2.58)
Vitamin B6 (mg/day)	3.28 (2.17;4.93)
Vitamin B9 (folate; µg /day)	560 (336;807)
Vitamin B12 (cyanocobalamin; µg /day)	4.33 (2.52;8.64)
Cardio-metabolic Factors, mean (SD)	
Body mass index (kg/m^{2})	21.2 (18:7;25.2)
Systolic blood pressure (mmHg)	136.5 (121.8;147.0)
Diastolic blood pressure (mmHg)	83.0 (77.0;94)
Fasting blood glucose (mmol/L)	5.18 (4.65;5.55)

[1] Prospective Urban and Rural Epidemiology study's South African, North West province cohort (PURE-SA-NW), [2]Age provided by the participant during questionnaire interviews, [3]DNA methylation age obtained using the Horvath clock, [4]DNA methylation age obtained using the Hannum clock, [5]

DNA methylation age obtained using the PhenoAge clock, [6] DNA methylation age obtained using GrimAge clock, [7]Intrinsic epigenetic age acceleration (within each cell) obtained using Horvath clock, [8] Extrinsic epigenetic age acceleration (between different cells) using Hannum clock, [9] Epigenetic age acceleration incorporating clinical traits obtained using the PhenoAge clock, [10]Epigenetic age acceleration incorporating plasma proteins obtained using GrimAge clock, [11] Physical activity as a continuous variable using the adapted physical activity score, [12] Total energy intakes obtained by food frequency questionnaires in Kcal/day, [13] One carbon metabolism nutrients intake obtained via food frequency questionnaires. Categorical variables do not always add up to column totals due to missing values.

Appendix.16. Pearson's correlations between age measures and between age acceleration measures in PURE-SA-NW study (native South Africans).

a) Correlations between age measures:

	Chronological Age[1]	DNAm HorvathAge[2]	DNAm HannumAge[3]	DNAm PhenoAge[4]	DNAm GrimAge[5]
Chronological Age	1.00				
DNAm HorvathAge	0.58	1.00			
DNAm HannumAge	0.68	0.59	1.00		
DNAm PhenoAge	0.51	0.55	0.41	1.00	
DNAm GrimAge	0.80	0.53	0.58	0.52	1.00

b) Correlations between DNAm age acceleration measures:

	IEAA (Horvath)[6]	EEAA (Hannum)[7]	PhenoAgeAccel[8]	GrimAgeAccel[9]
IEAA (Horvath)	1.00			
EEAA (Hannum)	0.34	1.00		
PhenoAgeAccel	0.35	0.13	1.00	
GrimAgeAccel	0.10	0.06	0.21	1.00

[1]Chronological age obtained from interview questionnaires, [2]Horvath Age obtained using the Horvath clock, [3]Hannum age obtained using Hannum clock, [4]PhenoAge obtained using the PhenoAge clock, [5] Grim Age obtained using the Grim Age clock. [6] Intrinsic epigenetic age acceleration obtained via the Hannum clock, [7]Extrinsic epigenetic age acceleration obtained via Hannum clock, [8] PhenoAge Acceleration obtained via the PhenoAge clock, [9]GrimAge acceleration obtained via the GrimAge clock.

Appendix.17: Forest plots of model β (and 95% confidence intervals) relating EAA Measures to lifestyle factors in PURE-SA-NW study. The plot depicts regression model β with 95% confidence intervals. N=120. Red line = reference line for confidence intervals. Abbreviations: IEAA= Intrinsic epigenetic age acceleration, EEAA= extrinsic epigenetic age acceleration, PhenoAge Accel= Pheno Age Acceleration, GrimAge Accel= Grim Age Acceleration, ref= reference, active= levels of physical activity, Vit= Vitamin, adj=final models adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake respectively, crude= un-adjusted models. Vitamin intake was measured in mg/day for Vitamin B6 and Vitamin B12, while vitamin B9 and Vitamin B12 were measured in mcg/day. Reference (comparison) groups for modifiable risk factors: smoking = non-smokers, alcohol consumption = no (never) alcohol consumption, physical activity= less physically active.

IEEA obtained via Horvath Clock (PURE-SA-NW study)
EEEA obtained via Hannum Clock (PURE-SA-NW study)
PhenoAgeAccel (PURE-SA-NW study)
GrimAgeAccel (PURE-SA-NW study)
Modifiable Lifestyle factors
Any alcohol. crude
Any alcohol.adj
Physical activity.crude
Physical activity.adj
Current smokers.crude
Current smokers. adj
-8
-6
-4
-2
0
2
4
6
8
One carbon metabolism nutrients
Vit B2.crude
Vit B2.adj
Vit B6.crude
Vit B6.adj
Vit B9.crude
Vit B9.adj
Vit B12.crude
Vit B12.adj
-15
-10
-5
0
5
10
15

Appendix .18. Lifestyle factors, cardiometabolic traits and epigenetic age acceleration in the PURE-SA-NW study (native South Africans).

	IEAA (Horvath)[1] (N=120)		**EEAA (Hannum)[2] (N=120)**		**PhenoAgeAccel[3] (N=120)**		**Grim AgeAccel[4] (N=120)**	
	Crude[5]	**Fully adjusted [6]**	**Crude[5]**	**Fully adjusted [6]**	**Crude[5]**	**Fully adjusted[6]**	**Crude[5]**	**Fully adjusted[6]**
	β (95% CI)	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Lifestyle factors								
Alcohol consumption								
No	ref	ref	ref	ref	ref	ref	ref	ref
Yes	–1.13 (–3.56;1.30)	–1.80 (–5.21;1.61)	0.12 (–2.00;2.25)	–0.56 (–3.55;2.44)	0.44 (–2.42;3.31)	2.74 (–1.16;6.65)	3.46 (1.59;5.33)	0.38 (–2.25;3.01)
Physical activity levels[7]								
WPA[8]	0.88 (–0.40;2.16)	0.99 (–0.46;2.45)	0.61 (–0.66;1.89)	–0.61 (–2.09;0.86)	–0.61 (–2.09;0.86)	–1.38 (–3.04;0.29)	0.92 (–0.11;1.95)	0.52 (–0.61;1.64)
Tobacco smoking								
Never	ref	ref	ref	ref	ref	ref	ref	ref
Current	0.34 (–2.06;2.75)	1.04 (–2.45;4.53)	0.83 (–1.24;2.91)	–0.16 (–3.22;2.91)	–0.84 (–3.64;1.95)	–0.51 (–4.51;3.49)	4.59 (2.81;6.37)	3.60 (0.91;6.30)
One carbon metabolism nutrients [9]								
Vit B2 (mg/day)	1.34 (–0.81;3.51)	3.28 (–1.41;7.98)	1.07 (–3.09;5.22)	1.87 (–0.70;4.43)	1.87 (–0.70;4.43)	3.69 (–1.72;9.10)	0.77 (–1.04;2.59)	–0.002 (–3.65;3.64)
Vit B6 (mg/day)	0.02 (–2.06;2.10)	1.55 (–2.92;6.02)	–1.11 (–5.07;2.84)	1.73 (–0.71; 4.18)	1.73 (–0.71; 4.18)	7.25 (2.10;12.40)	0.29 (–1.44;2.02)	–1.74 (–5.21;1.73)
Vit B9 (mcg/day)	–0.09 (–2.09;1.89)	–1.30 (–4.88;2.27)	–1.81 (–3.47; –0.15)	–2.04 (–5.21;1.12)	–0.62 (–2.99;1.74)	–6.74 (–10.86; –2.61)	1.55 (–0.08;3.19)	2.50 (–0.26;5.25)
Vit B12 (mcg/day)	0.78 (–0.22;1.780)	0.26 (–1.08;1.61)	0.85 (–0.34;2.04)	0.54 (–0.67;1.75)	0.54 (–0.67;1.75)	–0.28 (–1.83;1.27)	–0.24 (–1.09;0.61)	–0.44 (–1.49;0.60)

[1]Intrinsic epigenetic age acceleration (within each cell) obtained using Horvath clock, [2]Extrinsic epigenetic age acceleration (between different cells) using Hannum clock, [3] Epigenetic age acceleration incorporating clinical traits obtained using the PhenoAge clock, [4]Epigenetic age acceleration incorporating plasma proteins obtained using GrimAge clock, [5]crude linear regression model, [6] fully adjusted linear regression model for age, education, smoking, physical activity, alcohol intake, vitamin intake and total energy intake respectively,[7] Physical activity categories according to the GPAQ questionnaire, [8] Physical activity as a continuous weighted physical activity. [9] Once carbon metabolism nutrients obtained via food frequency questionnaires, Vit= Vitamin, Vitamin intake Highlighted results are the statistically significant results in both the crude and adjusted linear regression models. CI= confidence interval.

Appendix. 19. Summary of results post-hoc analyses

1.) Associations of cardiometabolic traits and EAA in combined migrants and non-migrants' population.

In adjusted linear regression models, we found that tobacco smoking and physical activity levels were not associated with all EAA measures in the combined sample of migrants and non-migrants. Higher vitamin B2 (riboflavin) intake was negatively associated with PhenoAgeAccel. Higher Vitamin B9 (folate) intake was negatively associated with EEAA, PhenoAgeAccel and GrimAgeAccel. Higher BMI was negatively associated with all EAA measures. Higher FBG levels were positively associated with all EAA measures (Appendix.8).

2.) Exclusion of participants taking anti-diabetic medications in FBG linear regression models.

In adjusted linear regression models with participants taking anti-diabetic medications excluded, we found that higher FBG levels were still positively associated with all EAA measures among migrants, and only with GrimAgeAccel among non-migrants.

3.) Replication in independent cohorts

Full replication in the IMS cohort:

Descriptive characteristics

A total of 172 IMS participants were included in the analysis. Of these, 86 were migrants and 86 were non-migrants. Mean chronological age of migrants was (46 ± 8) years and non-migrants was (46 ± 7) years. Mean DNAm HorvathAge and DNAm GrimAge were higher in migrants compared to non-migrants. Mean DNAm HannumAge and DNAm PhenoAge were lower in migrants compared to non-migrants. Migrants had minimal baseline differences with non-migrants except for education and dietary intake (higher in migrants) due to matching by age and sex (Appendix.10).

Correlation among DNAmAge estimators

Chronological age positively correlated with all four DNAmAges; r = 0.65 for DNAm HorvathAge, 0.79 for DNAm HannumAge, 0.61 for DNAm PhenoAge and 0.86 for DNAm GrimAge (Appendix.11). Correlations among EAA measurements were weaker with correlation coefficients ranging from -0.01 (PhenoAgeAccel vs GrimAgeAccel) to 0.41 (IEAA and EEAA).

Migration and EAA measures

We found that migrants have lower IEAA and GrimAgeAccel compared to non-migrants (IEAA = –0.30 vs 0.30; GrimAgeAccel = –0.88 vs 0.88, Appendix.10). We also found that migrants have higher EEAA and PhenoAgeAccel compared to non-migrants (EEAA = 0.26 vs –0.26; PhenoAgeAccel = 0.83 vs –0.83, Appendix.10). Migration status was negatively associated with GrimAgeAccel, and positively with PhenoAgeAccel after adjusting for age, sex, and education. However, duration of stay in the host country among migrants was not associated with any EAA measure (Appendix.12).

Lifestyle factors and EAA among migrants and non-migrants

Alcohol consumption, physical activity, and smoking (Appendix.13&14): In adjusted linear regression models, alcohol consumption was negatively associated with GrimAgeAccel in migrants but not non-migrants. Moderate physical activity levels were negatively associated with EEAA among non-migrants but not migrants, while higher levels of physical activity were negatively associated with EEAA among non-migrants but not migrants (crude models). Current and past tobacco smoking was positively associated with GrimAgeAccel among migrants and non-migrants.

One-*carbon metabolism nutrients (Appendix.13&14):* In crude linear regression models, higher Vitamin B6 (pyridoxine) intake was negatively associated with GrimAgeAccel in migrants but not non-migrants. Higher Vitamin B9 (folate) intake was negatively associated with all EAA measures (consistency in direction of effects in all four EAA measures). Such associations were not observed for Vitamin B2 (riboflavin) and Vitamin B12 (cobalamin) in either migrants or non-migrants.

Cardiometabolic factors and EAA among migrants and non-migrants

BMI: higher BMI was negatively associated with EEAA and PhenoAgeAccel among non-migrants in crude linear regression models. Such associations were not observed among migrants (Appendix.13&14).

FBG: higher FBG was positively associated with EEAA and PhenoAgeAccel among migrants in crude linear regression models. Such associations were not apparent among non-migrants (Appendix.13&14).

BP: Both systolic and diastolic blood pressure were positively associated with GrimAgeAccel after adjusting for confounders measures in migrants. Such associations were not apparent among non-migrants (Appendix.13&14).

Partial replication PURE-SA-NW cohort:

<u>Descriptive characteristics</u>

A total of 120 PURE-SA-NW participants were included in the analysis. All participants were men. Mean chronological age of migrants was (63 ± 10) years. In general, more had finished primary school, were smokers, had lower BMI, and FBG (Appendix.15).

<u>Correlation among DNAmAge estimators</u>

Chronological age positively correlated with all four DNAmAges; r = 0.58 for DNAm HorvathAge, 0.68 for DNAm HannumAge, 0.51 for DNAm PhenoAge and 0.80 for DNAm GrimAge (Appendix.16). Correlations among EAA measurements were weaker with correlation coefficients ranging from 0.06 (EEAA vs GrimAgeAccel) to 0.35 (IEAA and PhenoAgeAccel).

<u>Lifestyle factors and EAA among migrants and non-migrants</u>

Alcohol consumption, physical activity, and smoking (Appendix.17&18): In adjusted linear regression models, alcohol consumption was not associated with any EAA measure. Current tobacco smoking was positively associated with GrimAgeAccel. Physical activity levels were not associated with any EAA measures.

One-carbon metabolism nutrients (Appendix.17&18): In adjusted linear regression models, higher Vitamin B6 (pyridoxine) intake was positively associated with GrimAgeAccel. Higher Vitamin B9 (folate) intake was negatively associated with EEAA in crude linear regression models, and with PhenoAgeAccel in adjusted models. Such associations were not observed for Vitamin B2 (riboflavin) and Vitamin B12 (cobalamin).

STROBE Statement
Strengthening the reporting of observational studies in epidemiology

Appendix. 20. STROBE Statement—Items to be included when reporting observational studies in a conference abstract

Item	Recommendation
Title	Indicate the study's design with a commonly used term in the title (e.g., cohort, case-control, cross sectional) **This information is available on cover page**
Authors	Contact details for the corresponding author. **This information is available on the cover page**
Study design	Description of the study design (e.g. cohort, case-control, cross sectional) **This information is available in methods**
Objective	Specific objectives or hypothesis **This information is available in background**
Methods	
Setting	Description of setting, follow-up dates or dates at which the outcome events occurred or at which the outcomes were present, as well as any points or ranges on other time scales for the outcomes (e.g., prevalence at age 18, 1998-2007). **This information is available in method**
Participants	*Cohort study*—Give the most important eligibility criteria, and the most important sources and methods of selection of participants. Briefly describe the methods of follow-up *Case-control study*—Give the major eligibility criteria, and the major sources and methods of case ascertainment and control selection *Cross-sectional study*—Give the eligibility criteria, and the major sources and methods of selection of participants **This information is available in methods**
	Cohort study—For matched studies, give matching and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	Clearly define primary outcome for this report. **This information is available in methods**
Statistical methods	Describe statistical methods, including those used to control for confounding **This information is available in methods**
Results	
Participants	Report Number of participants at the beginning and end of the study **This information is available in methods**
Main results	Report estimates of associations. If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **This information is available in results** Report appropriate measures of variability and uncertainty (e.g., odds ratios with confidence intervals **This information is available in results**
Conclusions	General interpretation of study results **This information is available in interpretation.**

Appendix 21. Reporting checklist for genetic association study.

Based on the STREGA guidelines.

Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman M, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N; STrengthening the REporting of Genetic Association Studies. STrengthening the REporting of Genetic Association Studies (STREGA): An Extension of the STROBE Statement.

		Reporting Item	Page Number
Title and abstract			
Title	#1a	Indicate the study's design with a commonly used term in the title or the abstract	Cover page
Abstract	#1b	Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract section
Background/rationale			
	#2	Explain the scientific background and rationale for the investigation being reported	Introduction, page 1
Objectives			
	#3	State specific objectives, including any prespecified hypotheses. State if the study is the first report of a genetic association, a replication effort, or both.	Introduction, page 1
Study design			
	#4	Present key elements of study design early in the paper	Methods, study population, page 2
Setting			
	#5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Methods, study population, page 2

Eligibility criteria

	Item	Checklist item	Page
	#6a	Cohort study – Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study – Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross-sectional study – Give the eligibility criteria, and the sources and methods of selection of participants. Give information on the criteria and methods for selection of subsets of participants from a larger study, when relevant.	Methods, study population, page 2
	#6b	Cohort study – For matched studies, give matching criteria and number of exposed and unexposed. Case-control study – For matched studies, give matching criteria and the number of controls per case.	n/a

Variables

	Item	Checklist item	Page
	#7a	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Methods, page 2&3
	#7b	Clearly define genetic exposures (genetic variants) using a widely used nomenclature system. Identify variables likely to be associated with population stratification (confounding by ethnic origin).	n/a

Data sources/measurement

	Item	Checklist item	Page
	#8a	For each variable of interest give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one	Methods, page 2 &3

		group. Give information separately for for exposed and unexposed groups if applicable.	
	#8b	Describe laboratory methods, including source and storage of DNA, genotyping methods and platforms (including the allele calling algorithm used, and its version), error rates and call rates. State the laboratory / centre where genotyping was done. Describe comparability of laboratory methods if there is more than one group. Specify whether genotypes were assigned using all of the data from the study simultaneously or in smaller batches.	Methods, page 2 & 3
Bias			
	#9a	Describe any efforts to address potential sources of bias	Discussion, limitations, page 8
	#9b	Describe any efforts to address potential sources of bias	Discussion, limitations, page 8
Study size			
	#10	Explain how the study size was arrived at	Methods, sample justification, page 2
Quantitative variables			
	#11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why. If applicable, describe how effects of treatment were dealt with.	Methods, page 2 & 3
Statistical methods			
	#12a	Describe all statistical methods, including those used to control for confounding. State	Methods, statistical

		software version used and options (or settings) chosen.	analyses, page 3
	#12b	Describe any methods used to examine subgroups and interactions	Methods, post hoc analyses, page 3
	#12c	Explain how missing data were addressed	Methods, statistical analyses, page 3
	#12d	If applicable, explain how loss to follow-up was addressed	n/a
	#12e	Describe any sensitivity analyses	Methods, post-hoc analyses, page 3
	#12f	State whether Hardy-Weinberg equilibrium was considered and, if so, how.	n/a
	#12g	Describe any methods used for inferring genotypes or haplotypes	n/a
	#12h	Describe any methods used to assess or address population stratification.	n/a
	#12i	Describe any methods used to address multiple comparisons or to control risk of false positive findings.	Methods, page 4
	#12j	Describe any methods used to address and correct for relatedness among subjects	n/a
Participants			
	#13a	Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed. Give information separately for for exposed and unexposed groups if applicable. Report numbers of individuals in whom genotyping was attempted and	Methods, study population,page 2

		numbers of individuals in whom genotyping was successful.	
	#13b	Give reasons for non-participation at each stage	n/a
	#13c	Consider use of a flow diagram	Methods, study population, page 2
Descriptive data			
	#14a	Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders. Give information separately for exposed and unexposed groups if applicable. Consider giving information by genotype	Results, baseline characteristics, page 4
	#14b	Indicate number of participants with missing data for each variable of interest	n/a
	#14c	Cohort study – Summarize follow-up time, e.g. average and total amount.	n/a
Outcome data			
	#15	Cohort study Report numbers of outcome events or summary measures over time.Give information separately for exposed and unexposed groups if applicable. Report outcomes (phenotypes) for each genotype category over time Case-control study – Report numbers in each exposure category, or summary measures of exposure.Give information separately for cases and controls . Report numbers in each genotype category. Cross-sectional study – Report numbers of outcome events or summary measures. Give information separately for exposed and unexposed groups if applicable. Report outcomes (phenotypes) for each genotype category	Results, migration and EAA measures, page 4

Main results

	#16a	Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Results, page 4&5
	#16b	Report category boundaries when continuous variables were categorized	n/a
	#16c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	n/a
	#16d	Report results of any adjustments for multiple comparisons	Results, post-hoc findings, page 5

Other analyses

	#17a	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Results, post-hoc findings, page 5
	#17b	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Results, post-hoc findings, page 5
	#17c	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	Results, post-hoc findings, page 5

Key results

	#18	Summarise key results with reference to study objectives	Discussion, key findings, page 5

Limitations

	#19	Discuss limitations of the study, considering sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	Discussion, strengths and limitations, page 8

Interpretation			
	#20	Give a cautious overall interpretation considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	Discussion, page 5-7
Generalisability			
	#21	Discuss the generalisability (external validity) of the study results	Discussion, page 7 & 8
Funding			
	#22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Abstract and declarations

- The STREGA checklist is distributed under the terms of the Creative Commons Attribution License CC-BY. This checklist was completed on 21. February 2021 using https://www.goodreports.org/, a tool made by the EQUATOR Network in collaboration with Penelope.ai

Appendix 22. Data Analysis Plan – quantitative studies

Template version August 2018. AMC Department of Public Health

1. Working title

Associations of lifestyle and cardiometabolic factors with epigenetic age acceleration (EAA) among migrant and non-migrant Ghanaians.

2. Anticipated authors (in expected order for publication)

Felix Chilunga, Peter Henneman, [other RODAM committee members], [external contributors], Charles Agyemang.

3. Problem statement and relevance

- 3.1. Relevance: which specific public health problem is so important that it needs further study?

Migration from low-and middle-income countries (LMIC) to high income countries (HIC), predominantly for economic reasons, has been increasing in the last few decades with around five million migrants moving to Organization for Economic Co-operation and Development (OECD) countries each year since 2015.(1) While migration associated with economic growth is likely to raise the average levels of individual well-being, there are concerns about potential deleterious effects on health.

Migrants from LMICs to high income countries HICs are disproportionately affected by cardiometabolic diseases (i.e., diabetes, stroke, etc.) compared to their host populations, as well to populations from their countries of origin.(2, 3) This burden of cardiometabolic diseases (CMD) has been observed to increase with duration of stay in host countries.(4) Adoption of sedentary lifestyle and unhealthy diets, psychosocial stress and limited access to health care are thought to be the major driving forces. However, mechanisms underlying the disproportionally high CMD burden among migrants have not been fully elucidated.

Epigenetic changes through gene-environmental interactions could be a potential factor mediating the change in disease burden among migrants, but data is lacking. One of the ways of studying epigenetic changes is by means of epigenetic age acceleration (EAA).(5) EAA represents instances where body tissue has aged faster than normally would at the same chronological age.(5) Basically, chronological aging is associated with corresponding changes in DNA methylation in the same individual. As such, DNA methylation (DNAm) can be used to accurately predict the chronological age of an individual. The set of cytosine-phosphate-guanine sites (CpGs) coupled to mathematical algorithms that estimate chronological age (in years) from a DNA source are termed epigenetic clocks.(5) When quantified in blood, EAA can be divided into two forms: intrinsic and extrinsic. Intrinsic epigenetic age acceleration (IEAA) captures biological aging within each cell independently of proportions of naïve or senescent cytotoxic T cells, while extrinsic epigenetic age acceleration (EEAA) quantifies epigenetic aging in immune-related components. Several epigenetic clocks have been developed to measure IEAA and EEAA. (5) The most used are Horvath Age which measures IEAA, Hannum Age which measures EEAA, and Pheno Age which measures EEAA and includes clinical characteristics such as chronological age, albumin, creatinine, glucose, CRP, lymphocyte percentage, mean cell volume, red blood cell distribution width, alkaline phosphatase, and white blood cell count in its algorithm.(6-8)

Since DNA methylation is influenced by environment factors, epigenetic clocks can capture accelerated tissue aging (EAA) from environmental factors such as education, lifestyle and neighbourhood environment.(9) Moreover, EEA has also been associated with the CMDs and risk of mortality in general. Studying EEA changes among migrants might offers clues to the aetiology to the disproportionally high burden of CMD among migrants, and potentially unveil opportunities for anti-aging treatments.(10)

- 3.2. Knowledge need: what is a key question that should we answer to address this problem?

What are the differential associations between lifestyle and cardiometabolic factors and EAA among migrant Africans in Europe and non-migrant Africans?

- 3.3. Previous studies: why is this question yet not sufficiently addresses in the scientific literature?

Previous studies in populations from HIC (e.g., Europeans) have unveiled associations between EAA and multiple factors such as education, lifestyle factors, cardiometabolic factors, multiple chronic conditions, and all-cause mortality, among others.(9, 11) In general, the studies have shown that higher education levels and higher intake of vegetables are negatively associated with EAA, while unhealthy diets, tobacco smoking, obesity, diabetes, coronary heart disease are positively associated with EAA.(5, 9) Despite this abundant evidence in HIC, data in populations from LMIC are lacking. To date, no study has investigated the role of CMD in the changing burden of CMD among Africans or any other LMICs. DNA methylation which underlies EEA is influenced by the genotype, as well as the environment in general.(12) The genotype of European descent populations (in which most studies have been conducted) is dissimilar to that of populations from LMIC (e.g., Africans).(13) Moreover, the lifelong environmental exposures are different between populations from HIC as compared to populations from LMIC.(14) For instance, populations from LMIC are more exposed to poor early life environments (maternal malnutrition, low birth weight etc.) and infectious diseases than those from HIC, which might potentially contribute to accelerated aging of body tissue, which is in turn compounded by lifestyle changes in adulthood such as poor diet, tobacco smoking and physical inactivity.

4. General objective (in one sentence)

- To investigate associations of lifestyle and cardiometabolic factors with EAA among migrant Africans in Europe in comparison to non-migrant Africans.

5. Research questions (about three)

- What are the differences in EAA between migrant Africans in Europe compared to non-migrant Africans?
- What are the associations of lifestyle and cardiometabolic factors with EAA among migrant Africans in Europe compared to non-migrant Africans?

6. Key hypothesis (if any)

Considering that migrant African populations have higher rates of CMD than their non-migrant counterparts, we hypothesize that changes in lifestyle upon migration lead to accelerated biological aging of tissues (higher EAA), which in turn leads to higher CMD incidence. As such, we expect migrants to exhibit faster EAA than their non-

migrant counterparts. In addition, we expect lifestyle and cardiometabolic factors to be more positively associated with EEA in migrants than in non-migrants.

7. Design, study population and variables (max 3 paragraphs)

A cross-sectional analysis of the baseline data from the Research on Diabetes and Obesity Among Migrants (RODAM) study, collected between 2012 and 2015, will be performed.(15) All Ghanaian participants of the RODAM study with epigenetic data will be included in the analyses. This sample consists of 736 participants (267 with diabetes and 427 controls). These participants were sampled based on an initial power of 0.80 and alpha of 0.05 that was needed to detect 5% DNAm differences among participants with diabetes (n=300) compared to controls without diabetes (n=300). After data collection, the final epigenetic sample had proportions of participants with diabetes equally distributed (50-50) among migrants Ghanaians and non-migrant Ghanaians.

Variables of interest for this study are:

- Sex: based on self-report
- Chronological age: based on self-report.
- Education levels: based on self-report.
- Lifestyle factors:
    - tobacco smoking: based on self-report (yes/no)
    - alcohol consumption: based on food frequency questionnaire.
    - physical activity levels: according to the GPAQ criteria
    - intake of one nutrient metabolism nutrients from food frequency questionnaire (Vitamin B2, Vitamin B6, Vitamin B9, Vitamin B12)
- Cardiometabolic factors:
    - body mass index (BMI)
    - diastolic blood pressure: as a mean of three measurements at rest
    - systolic blood pressure: as a mean of three measurements at rest
    - fasting blood glucose (FBG): from fasted blood samples
- Other descriptive characteristics:
    - use of antidiabetic medications.
    - duration of stay in host country.

Missing data:

We will impute any missing data if the missing variables are less than 30% of the total sample. We will assess if missing data can be explained by the observed data (missing at random, MAR). If MAR, multiple imputation will be applied (n = 5; discriminant fully conditional specification (FCS) method). FCS is also known as multiple

imputation by chained equations (MICE) and uses separate conditional univariate imputation models specified for each incomplete variable, with other variables as predictors.

Analytic techniques (max 2 paragraphs)

1) Analysis Software:

   All analyses will be conducted in R studio through the central digital workplace.

2) Calculation of DNAm Ages and DNAm Age Acceleration:

   We will calculate the three DNAmAges with the help of the Bioconductor package *Methylclock* from normalized DNAm data. HorvathAge will be estimated using 353 CpGs as specified in *Horvath et al*. IEAA will be obtained as residuals from the regression model that regresses HorvathAge on chronological age and blood cell counts. HannumAge will be estimated using 71 CpGs as specified in *Hannum et al*. EEAA will be obtained as residuals from a regression model that regresses HannumAge aggregated with three blood cell components (naïve cytotoxic T cells, exhausted cytotoxic T cells, and plasmablasts) on chronological age. PhenoAge will be estimated using 513 CpGs as specified in *Levine et al.* PhenoAge acceleration will be obtained as residuals of the regression models that regress PhenoAge on chronological age without adjusting for blood cell counts. Cell counts will be estimated using a method developed by *Houseman et al*, implemented using *Methylclock* package.

3) Descriptive analyses

   Since we are interested in migrant and non-migrant Ghanaian groups, baseline data (DNAmAges, demographics, lifestyle factors and cardiometabolic factors) will be presented by these categories. Baseline characteristics will be checked for normality in case of numerical data and will be presented by frequencies and percentages (categorical data), arithmetic means and standard deviations (normally distributed numerical data), and medians and interquartile ranges (not-normally distributed numerical data).

4) Correlations between DNAmAges and between EAA measures

   Pearson's correlations will be performed between chronological age and DNAmAges, as well as between the EAA measures to ascertain the usability of these DNAmAge measures in this study. High positive correlation between DNAmAges and chronological age will enable utility of the DNAmAge measures in the study. We will assume high correlations at an $r > 0.70$.

5) Differences in DNAmAges and EAA between migrants and non-migrants

   Linear regressions will be performed between EAA measures and migration status.

   - Outcome variables: Three measures of EAA (Horvath Age, Hannum Age, Pheno Age).
   - Predictor variable: migration status (migrants vs non-migrants).
   - Covariates: chronological age, sex, and education.

- We will not adjust for lifestyle and cardiometabolic factors to incorporate these effects in the outcome.
- Reporting: we will report the beta coefficients and the corresponding 95% confidence intervals.

6) Associations of lifestyle and cardiometabolic factors with EAA

Linear regressions were performed between EAA measures (outcome variable) and CMD traits.

- Outcome variables: Three measures of EAA (Horvath Age, Hannum Age, Pheno Age).
- Predictor variables: smoking status, alcohol consumption, physical activity, vitamin intake, duration of stay in host country (migrants), BMI, FBP and BP independently.
- Covariates:
  - Model with lifestyle predictors (e.g., smoking) will be adjusted for demographics (chronological age, sex, and education), and other lifestyle factors (e.g., alcohol consumption, physical activity, vitamin intake).
  - Model with cardiometabolic factors (e.g., BMI) will be adjusted for demographics (chronological age, sex, and education), and other lifestyle factors (e.g., tobacco smoking, alcohol consumption, physical activity, vitamin intake).
- Reporting: we will report the beta coefficients and the corresponding 95% confidence intervals.

7) Sensitivity analyses:

a. We will repeat the linear regression analyses between lifestyle factors and cardiometabolic factors and EAA in the total RODAM population (migrants and non-migrants together). This will ensure that the effects detected between migrants and non-migrants are also apparent in the total population independent of differences between groups, ensuring validity even with minimal differences between groups.

b. We will repeat the linear regression analyses between lifestyle factors and cardiometabolic factors and EAA in a sub-group without participants taking anti-diabetic medication. This will ensure that the effects detected between migrants and non-migrants are not influenced by intake of diabetic medication which can decrease EEA.

8) Replication in independent cohorts.

We will further validate our findings by means of replication in independent cohorts. Migration cohorts are few and far between, as are cohorts from Africa with epigenetic data. As such, we will replicate our study in any cohorts from Africa or other LMIC where the epidemiological transitions are taking place. Even though the cohorts will not be likely to be like RODAM, presence of corresponding effects in such cohorts will strengthen our findings. Lack of replication will not rule out our findings due to differences in ethnicity, environments and study designs etc. The cohorts to consider include (based on literature search):

I. Indian Migration cohort: DNA methylation data on rural- urban migrants siblings vs rural non-migrant siblings as published by *ELLIOTT, Hannah R., et al. Migration and DNA methylation: a comparison of methylation patterns in type 2 diabetes susceptibility genes between Indians and Europeans. Journal of diabetes research & clinical metabolism, 2013, 2: 6.*

II. PURE- South African sub-cohort: DNA methylation data on native black South Africans (Batswana) from the North west province as published by *CRONJÉ, H. Toinét, et al. Replication and expansion of epigenome-wide association literature in a black South African population. Clinical epigenetics, 2020, 12.1: 1-13.*

III. Bangladesh cohort: DNA methylation data on young Bangladeshi adults as published by *FINER, S., et al. Is famine exposure during developmental life in rural Bangladesh associated with a metabolic and epigenetic signature in young adulthood? A historical cohort studies. BMJ open, 2016, 6.11.*

IV. Gambia cohort: DNA methylation data on rural African women as published by *DOMINGUEZ-SALAS, Paula, et al. DNA methylation potential: dietary intake and blood concentrations of one-carbon metabolites and cofactors in rural African women. The American journal of clinical nutrition, 2013, 97.6: 1217-1227.*

V. Bangladesh cohort 2: DNA methylation in whole blood among Bangladeshi men and women as published by *JANSEN, Rick J., et al. The effect of age on DNA methylation in whole blood among Bangladeshi men and women. BMC genomics, 2019, 20.1: 1-14.*

Data from all cohorts is not freely available online. As such, we will contact all cohorts to inquire about presence of lifestyle and cardiometabolic measurements, and willingness to share the data. Cohorts with the needed variables and willing to share their data will be incorporated in our study.

9) Multiple testing correction:

Lifestyle factors and cardiometabolic factors are largely dependent in many pathways. For instance, physical activity, poor diet, alcohol consumption, BMI and diabetes are interlinked in similar pathways. As such, we will utilise alternative methods to minimise false positive findings as opposed to the regular multiple testing correction methods (false discovery rate, Bonferroni correction etc.). First, we will ensure that statistically significant results have a similar trend across all DNAmAge measures in that migrant or non-migrant group (i.e., direction of effects). Second, we will ensure that the statistically significant results are also detectable in the total RODAM population (migrants and non-migrants together; same direction of effects). Third, we will ensure that the statistically significant results should also be replicable in independent cohorts (same direction of effects and statistically significance).

8. Check for quality and relevance (optional)

9.1. Are you sure that you can answer the research questions? If not, what are the main risks?

A potential risk of our study is the complexity of the results and subsequently the interpretation of the results, since we will compare two groups (migrants and non-migrants) and many lifestyles and cardiometabolic factors. Furthermore, there is a chance that our findings will not be replicable in other cohorts. By minimizing the false positives through replications, the most significant and true findings should stand out.

9.2. Has the analysis particular strengths? Is it innovative? If not, what can you add to literature?

Epigenetics has not been understudied in the context of epidemiological transitions. This will be one of the first studies to investigate EAA and CMD among Africans or LMICs.

9.3. Will the results be of interest to readers in other countries? If, not why publish internationally?

With increasing rates of migration to Europe and other HIC, most countries will be interested to know whether the high burden of CMD is associated with higher EAA, and whether this lays a foundation for future anti-aging treatments in the group. Moreover, two thirds of the world population live in LMIC, therefore inclusion of a comparison group from LMIC (non-migrants) will also interest the readers from this region, as there is potential to know whether EAA is also influencing CMD burden in this group.

9.4. Could you regard any of the potential results as 'negative'? Are you sure that you would publish these as well?

The finding of no associations between lifestyle and cardiometabolic factors and EEA will be regarded as a 'negative' result. This might imply that the unhealthy phenotypes seen among migrants are not through epigenetic mechanisms, especially EAA. Therefore, publication of the results is desirable, irrespective of the outcome.

10. References (only key publications)

1. Organisation for Economic Co-operation and Development. International Migration Outlook 2019 2019 [Available from: https://www.oecd-ilibrary.org/content/publication/c3e35eec-en.

2. Yusuf S, Reddy S, Ôunpuu S, Anand S. Global burden of cardiovascular diseases: Part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. Circulation. 2001;104(23):2855-64.

3. Agyemang C, Meeks K, Beune E, Owusu-Dabo E, Mockenhaupt FP, Addo J, et al. Obesity and type 2 diabetes in sub-Saharan Africans–Is the burden in today's Africa similar to African migrants in Europe? The RODAM study. BMC medicine. 2016;14(1):166.

4. Boateng D, Agyemang C, Beune E, Meeks K, Smeeth L, Schulze M, et al. Migration and cardiovascular disease risk among Ghanaian populations in Europe: the RODAM study (research on obesity and diabetes among African migrants). Circulation: Cardiovascular Quality and Outcomes. 2017;10(11):e004013.

5. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nature Reviews Genetics. 2018;19(6):371.

6. Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):3156.

7. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Molecular cell. 2013;49(2):359-67.

8. Levine ME, Hosgood HD, Chen B, Absher D, Assimes T, Horvath S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. Aging (Albany NY). 2015;7(9):690.

9. Fiorito G, McCrory C, Robinson O, Carmeli C, Rosales CO, Zhang Y, et al. Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. Aging (Albany NY). 2019;11(7):2045.

10. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10(4):573.

11. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome biology. 2016;17(1):171.

12. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011;3(3):267-77.

13. Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, et al. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics. 2011;6(5):623-9.

14. Gluckman PD, Hanson MA. The developmental origins of health and disease. Early life origins of health and disease: Springer; 2006. p. 1-7.

15. Agyemang C, Beune E, Meeks K, Owusu-Dabo E, Agyei-Baffour P, Aikins Ad-G, et al. Rationale and cross-sectional study design of the Research on Obesity and type 2 Diabetes among African Migrants: the RODAM study. BMJ open. 2015;4(3):e004877.

Table.1. Baseline characteristics of participants

	Total	Non-migrants[1]	Migrants[2]	P-value[3]
Aging components, mean (SD)				
Chronological Age[4]				
HorvathAge[5]				
HannumAge[6]				
PhenoAge[7]				
IEAA[8]				
EEAA[9]				
PhenoAA[10]				
Other Demographic factors				
Female, n (%)				
Education, n (%)				
No education				
Primary school				
Secondary school				
Tertiary				
Behavior related factors				
Any Alcohol consumption, n (%)				
Smoking, n (%)				
Current				
Never				
Past				
Physical activity levels, n (%)[11]				
Low				
Moderate				
High				
Length of stay for migrants, years				
One carbon metabolism nutrient, mean (SD)[12]				
Vitamin B2 (riboflavin, mg/day)				
Vitamin B6 (mg/day)				
Vitamin B9 (folate; µg /day)				
Vitamin B12 (cyanocobalamin; µg /day)				

Cardio-metabolic Factors, mean (SD)				
Body mass index (kg/m^{2})				
Systolic blood pressure (mmHg)				
Diastolic blood pressure (mmHg)				
Fasting blood glucose (mmol/L)				

[1] Ghanaians living in Amsterdam, Berlin and London were categorised as migrants, [2] Ghanaians living in rural and urban Ghana were categorised as non-migrants,[3] P-value for difference in means, [4] Age provided by the participant during questionnaire interviews, [5] DNA methylation age obtained using the Horvath clock ,[6] DNA methylation age obtained using the Hannum clock, [7] DNA methylation age obtained using the Levine clock, [8] Intrinsic epigenetic age acceleration (within each cell) obtained using Horvath clock, [9] Extrinsic epigenetic age acceleration (between different cells) using Hannum clock, [10] Epigenetic age acceleration incorporating clinical traits obtained using the PhenoAge clock, [11] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria,[12] Once carbon metabolism nutrients obtained via food frequency questionnaires.

Table 2. Associations between lifestyle factors and cardiometabolic traits with Intrinsic epigenetic age acceleration (IEAA) in migrant vs non-migrant Ghanaians populations.

	Non-migrants[1]					**Migrants[2]**				
	Totals	**Crude**		**Fully adjusted[3]**		**Totals**	**Crude**		**Fully adjusted[3]**	
	N (%)	**β (95%CI)**	**P-value[4]**	**β (95%CI)**	**P-value[4]**	**N (%)**	**β (95%CI)**	**P-value[4]**	**β (95%CI)**	**P-value[4]**
Lifestyle factors										
Alcohol consumption										
No										
Yes										
Smoking										
Never										
Past										
Current										
Physical activity levels [5]										
Low										
Moderate										
High										
Length of stay (migrants), yrs.										

One carbon metabolism nutrients [6]										
Vit B2 (riboflavin; mg/day)										
Vit B6 (mg/day)										
Vit B9 (Folate; mcg/day)										
Vit B12 (Cobalamin; mcg/day)										
Cardiometabolic traits										
BMI, kg/m^2										
FBG, mmol/L										
SBP, mmHg										
DBP, mmHg										

[1] Participants from rural and urban Ghana, [2] Migrant Ghanaians living in Europe(Amsterdam, Berlin, London) [3] Fully adjusted for age, sex, education, alcohol consumption, smoking, physical activity, dietary patterns, and BMI (for fasting blood glucose and blood pressure), [4] P is statistically significant level at alpha<0.05 [5] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria,[5] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria. [6] [1]one carbon metabolites obtained via food frequency questionnaires. Abbreviations: BMI=Body mass index, FBG=fasting blood glucose, SBP =systolic blood pressure, DBP =diastolic blood pressure.

Table 3. Associations between lifestyle factors and cardiometabolic traits with Extrinsic epigenetic age acceleration (EEAA) in migrant vs non-migrant Ghanaians populations.

	Non-migrants[1]					Migrants[2]				
	Totals	Crude		Fully adjusted[3]		Totals	Crude		Fully adjusted[3]	
	N (%)	β (95%CI)	P-value[4]	β (95%CI)	P-value[4]	N (%)	β (95%CI)	P-value[4]	β (95%CI)	P-value[4]
Lifestyle factors										
Alcohol consumption										
No										
Yes										
Smoking										
Never										
Past										
Current										
Physical activity levels [5]										
Low										
Moderate										
High										
Length of stay (migrants), yrs.										

One carbon metabolism nutrients [6]										
Vit B2 (riboflavin; mg/day)										
Vit B6 (mg/day)										
Vit B9 (Folate; mcg/day)										
Vit B12 (Cobalamin; mcg/day)										
Cardiometabolic traits										
BMI, kg/m^2										
FBG, mmol/L										
SBP, mmHg										
DBP, mmHg										

[1] Participants from rural and urban Ghana, [2] Migrant Ghanaians living in Europe(Amsterdam, Berlin, London) [3] Fully adjusted for age, sex, education, alcohol consumption, smoking, physical activity, dietary patterns, and BMI (for fasting blood glucose and blood pressure), [4] P is statistically significant level at alpha<0.05 [5]Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria,[5] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria. [6] [1]one carbon metabolites obtained via food frequency questionnaires. Abbreviations: BMI=Body mass index, FBG=fasting blood glucose, SBP =systolic blood pressure, DBP =diastolic blood pressure.

Table 4. Associations between lifestyle factors and cardiometabolic traits with PhenoAge acceleration in migrant vs non-migrant Ghanaians populations.

	Non-migrants[1]					**Migrants**[2]				
	Totals	**Crude**		**Fully adjusted**[3]		**Totals**	**Crude**		**Fully adjusted**[3]	
	N (%)	**β (95%CI)**	**P-value**[4]	**β (95%CI)**	**P-value**[4]	**N (%)**	**β (95%CI)**	**P-value**[4]	**β (95%CI)**	**P-value**[4]
Lifestyle factors										
Alcohol consumption										
No										
Yes										
Smoking										
Never										
Past										
Current										
Physical activity levels [5]										
Low										
Moderate										
High										
Length of stay (migrants), yrs.										

One carbon metabolism nutrients [6]										
Vit B2 (riboflavin; mg/day)										
Vit B6 (mg/day)										
Vit B9 (Folate; mcg/day)										
Vit B12 (Cobalamin; mcg/day)										
Cardiometabolic traits										
BMI, kg/m^2										
FBG, mmol/L										
SBP, mmHg										
DBP, mmHg										

[1] Participants from rural and urban Ghana, [2] Migrant Ghanaians living in Europe(Amsterdam, Berlin, London) [3] Fully adjusted for age, sex, education, alcohol consumption, smoking, physical activity, dietary patterns, and BMI (for fasting blood glucose and blood pressure), [4] P is statistically significant level at alpha<0.05 [5]Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria,[5] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria. [6] [1]one carbon metabolites obtained via food frequency questionnaires. Abbreviations: BMI=Body mass index, FBG=fasting blood glucose, SBP =systolic blood pressure, DBP =diastolic blood pressure

Table.5. Associations between lifestyle factors and cardiometabolic traits with DNAm measures of age acceleration in total RODAM study.

	Totals	IEEA [1]				EEAA[2]				PhenoAge Accelaration[3]			
		Crude		Fully adjusted[4]		Crude		Fully adjusted[4]		Crude		Fully adjusted[4]	
	N (%)	β (95%CI)	P-value[5]	β (95%CI)	P-value[5]	β (95%CI)	P-value[5]	β (95%CI)	P-value[5]	β (95%CI)	P-value[5]	β (95%CI)	P-value[5]
Lifestyle factors													
Alcohol consumption													
No													
Yes													
Smoking													
Never													
Past													
Current													
Physical activity [6]													
Low													
Moderate													
High													

One carbon metabolism nutrient[11]													
Vit B2 (riboflavin; mg/day)													
Vit B6 (mg/day)													
Vit B9 (Folate; mcg/day)													
Vit B12 (Cobalamin; mcg/day)													
Cardiometabolic traits													
BMI, kg/m^2													
FBG, mmol/L													
SBP, mmHg													
DBP, mmHg													

[1] Extrinsic epigenetic age acceleration, [2] Intrinsic epigenetic age acceleration (Amsterdam, Berlin, London), [3] PhenoAge acceleration , [4] Fully adjusted for age, sex, education, alcohol consumption, smoking, physical activity, dietary patterns, and BMI (for fasting blood glucose and blood pressure), [5]P is statistically significant level at alpha<0.05 5Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria, [6] Physical activity categorised according to the global physical activity questionnaire (GPAQ) criteria. [7] one carbon metabolites obtained via food frequency questionnaires. Abbreviations: BMI=Body mass index, FBG=fasting blood glucose, SBP =systolic blood pressure, DBP =diastolic blood pressure