Longitudinal profiling of metabolic ageing trends in two population cohorts of young adults

Ville-Petteri Mäkinen[1,2,3]*, Mari Karsikas[3,4,5], Johannes Kettunen[3,4,5,6], Terho Lehtimäki[7], Mika Kähönen[8], Jorma Viikari[9,10], Markus Perola[6,11,12], Veikko Salomaa[6], Marjo-Riitta Järvelin[5,13,14,15], Olli T Raitakari[16,17,18], Mika Ala-Korpela[3,4,5,19]

[1]Computational and Systems Biology Program, Precision Medicine Theme, South Australian Health and Medical Research Institute, Adelaide, Australia;
[2]Australian Centre for Precision Health, University of South Australia, Adelaide, Australia;
[3]Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland;
[4]Biocenter Oulu, Oulu, Finland;
[5]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland
[6]Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland;
[7]Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland;
[8]Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland;
[9]Department of Medicine, University of Turku, Turku, Finland;
[10]Division of Medicine, Turku University Hospital, Turku, Finland;
[11]Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland;
[12]Estonian Genome Center, University of Tartu, Tartu, Estonia;
[13]Unit of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland;
[14]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK;
[15]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, UK;
[16]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland;
[17]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland;
[18]Centre for Population Health Research, University of Turku and Turku University Hospital;
[19]NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland.

*Corresponding author:
Ville-Petteri Mäkinen
South Australian Health and Medical Research Institute
North Terrace, SA 5000, Adelaide, Australia
Tel. +61 8 8128 4054
Email: ville-petteri.makinen@sahmri.com

Word count: 4,043

Abstract

Background

Quantification of metabolic changes over the human life course is essential to understanding ageing processes. Yet longitudinal metabolomics data are rare and long gaps between visits can introduce biases that mask true trends. We introduce new ways to process quantitative time-series population data and elucidate metabolic ageing trends in two large cohorts.

Methods

Eligible participants included 1,672 individuals from the Cardiovascular Risk in Young Finns Study and 3,117 from the Northern Finland Birth Cohort 1966. Up to three time points (ages 24-49) were analysed by nuclear magnetic resonance metabolomics and clinical biochemistry (236 measures). Temporal trends were quantified as median change per decade. Sample quality was verified by consistency of shared biomarkers between metabolomics and clinical assays. Batch effects between visits were mitigated by a new algorithm introduced in this report. The results below satisfy multiple testing threshold of $P < 0.0006$.

Results

Women gained more weight than men (+6.5% vs. +5.0%) but showed milder metabolic changes overall. Temporal sex differences were observed for C-reactive protein (women +5.1%, men +21.1%), glycine (women +5.2%, men +1.9%) and phenylalanine (women +0.6%, men +3.5%). In 566 individuals with ≥+3% weight gain versus 561 with weight change ≤−3%, divergent patterns were observed for insulin (+24% vs. −10%), very-low-density-lipoprotein triglycerides (+32% vs. −6%), high-density-lipoprotein$_2$ cholesterol (−6.5% vs. +4.7%), isoleucine (+5.7% vs. −6.0%) and C-reactive protein (+25% vs. −22%).

Conclusions

We report absolute and proportional trends for 236 metabolic measures as new reference material for overall age-associated and specific weight-driven changes in real-world populations.

Keywords

Ageing, metabolomics, batch effects, weight gain

Key messages

- We developed new statistical techniques that allowed us to investigate temporal trends in circulating biomarkers in their physical measurement units on population scale.
- We report previously unavailable quantitative trends for over 200 circulating metabolic measures, including lipoprotein subclass lipids, branched-chain amino acids and cardiometabolic biomarkers.
- We isolated associations between metabolic changes and weight gain in a carefully constructed design where weight loss and weight gain subgroups were matched at baseline.
- Our analyses revealed that weight change explains much of the metabolic variation (e.g. insulin, amino acids, lipoprotein subclass lipids), however, indicators such as waist-hip ratio, low-density-lipoprotein cholesterol and creatinine may reflect an overall ageing trend that affects all individuals.

Introduction

Increased longevity and falling birth rates across the world are creating an unprecedented situation in human history where the old will outnumber the young [1]. As age is the primary driver of chronic disease, understanding the nature of ageing within human populations is a prerequisite for mitigating the health consequences from shifting demographics. Age-driven change in metabolism is particularly topical due to the high burden of cardiometabolic diseases, the responsiveness of metabolism to personal and societal interventions, and recent advances in metabolomics profiling of morbidity and mortality [2–6]. There is, however, a knowledge gap: large-scale metabolomics studies of ageing in human populations have remained out of reach due to the rarity of long-term longitudinal data and technical challenges from long time gaps between sample collections [7].

A cross-sectional survey of young and old people is the easiest to organize, but it yields the weakest epidemiological evidence [8]; only longitudinal designs can provide well-grounded information on life course trajectories [9–15]. So far, most metabolomics ageing studies have been cross-sectional [16–20]. Given the caveats of cross-sectional modelling, the value of longitudinal data is well recognized [7,14,15,21–23]. Metabolic profiling across multiple time points produces more robust evidence and longitudinal designs allow for better detection and control of confounders [24]. On the other hand, new biases may arise when samples are collected from surveys separated by decades, including updates to collection protocols, storage effects and changes in biochemical assay methodology [9,10,24,25]. Hence new statistical methods that can remove the biases are essential for longitudinal designs.

In this study, we developed new methods to manage the biases and analysed two longitudinal cohorts that included altogether 4,789 participants with at least two time points that were a decade or more apart. To get an accurate picture of metabolic health and how it changed during follow-up, we investigated more than 200 variables that included anthropometric indicators, clinical biomarkers and nuclear magnetic resonance (NMR) metabolomics. We also used advanced sampling techniques to isolate metabolic associations with weight change and to characterize sex differences in temporal trends. Combined, our extensive results comprise a useful resource to understand how the metabolome changes due to aging in human populations.

Materials and Methods

Cardiovascular Risk in Young Finns Study (YFS)

The Cardiovascular Risk in Young Finns Study (YFS) is a population based prospective cohort study [26]. It was conducted at five medical schools in Finland (Turku, Helsinki, Kuopio, Tampere and Oulu), with the aim of studying the levels of cardiovascular risk factors in children and adolescents in different parts of the country. The baseline study in 1980 included 3,596 children and adolescents aged between 3 and 18 years. Results from clinical examination and fasting samples were used in the present study. Metabolomics data were available from three visits in 2001 (1,239 women and 1,007 men), 2007 (1,186 women and 974 men) and 2011 (1,112 women and 927 men).

Northern Finland Birth Cohort 1966 (NFBC1966)

The NFBC1966 is a longitudinal birth cohort established to study factors affecting preterm birth and consequent morbidity in the two northernmost provinces of Finland, Oulu and Lapland [27]. The NFBC1966 includes 12,231 births (12,058 alive) covering 96% of all eligible births in this region during January-December 1966. Data collections in 1997 (at age of 31) and 2012 (age 46) including clinical examination and fasting serum sampling was used in the present study. Metabolomics data were available from the 31-year (2,962 women and 2,749 men) and 46-year visits (3,237 women and 2,549 men).

Metabolomics and clinical biomarkers

A high-throughput NMR spectroscopy metabolomics platform was used to quantify over 200 lipid and metabolite measures from serum samples collected during the visits [28]. The platform applies a single experimental setup, which allows for simultaneous quantification of standard clinical lipids, 14 lipoprotein subclasses and individual lipids (triglycerides, phospholipids, free and esterified cholesterol) transported by these particles, multiple fatty acids, glucose and various glycolysis precursors, ketone bodies and amino acids in absolute concentration units. In addition, glucose, insulin, triglycerides, low-density-lipoprotein (LDL) cholesterol, high-density-lipoprotein (HDL) cholesterol, C-reactive protein and creatinine were assessed by standard clinical assays.

Sample quality control

To ensure the best possible estimates for age-associated changes in absolute concentrations, we employed a two-stage pre-processing protocol. First, we constructed multi-variate regression models of biomarkers that were available both from metabolomics and clinical assays (glucose, triglycerides, total cholesterol, HDL cholesterol and serum creatinine). Each of the five NMR measures was predicted from the combination of the five clinical biomarkers; our rationale was that

the biological correlations between the biomarkers [29,30] provides an additional source of consistency that would be broken by poor quality. The residuals from these models were then collected into a matrix of five columns (one column per model). The final quality score was defined as the first principal component of the residual matrix (Supplementary Figure S1). Hence samples with unusual and correlated residuals in multiple biomarkers were likely to get an extreme quality score. The cutoffs for acceptable deviation were set at the points where there was greater than 5% chance that the observed quality score was from the expected normal distribution. Consequently, 1,765 (9.8%) samples out of a total of 16,177 available in the YFS and NFBC1966 were excluded.

Calibration between visits

The second pre-processing step was aimed at bias between visits. We exploited longitudinal biomarker data in YFS to calibrate consecutive visits. We assumed *a priori* that subpopulations of the same sex, average age and body mass have identical average metabolic profiles. This means that if subsets of individuals from two consecutive visits have identical average features, we expect the subset averages of the biochemical data to be identical as well. Similarly, we expect that controlling for body mass and waist circumference simultaneously controls metabolically relevant lifestyle exposures at the population level. We chose body mass for matching due to its strong associations with metabolic measures and aging [12,31] and easy and standardized method of measurement.

Given consecutive visits A and B in the YFS cohort, a subset of participants was selected from visit A and another mutually exclusive subset from visit B. The selection was optimized so that the subsets had matching age, sex (225 men and 225 women) and body mass (Supplementary Table S1). We then defined a scaling factor $C = exp(mean(log(\text{subset of B}) - log(\text{subset of A})))$ for each metabolic variable that was not a ratio. It is plausible that batch effects manifest in the form of constant multipliers of concentrations since all the samples within a batch would have been collected, handled, stored and measured the same way. Lastly, the scale factor was applied to all data from visit B to equalize the measurement scale (Supplementary Figure S2). The procedure was applied first to the 2001 and 2007 visits, then to the 2007 and 2011 visits. The same procedure was not possible within the NFBC1966 due to the birth cohort design (i.e., no age spread). Instead, we calibrated the NFBC1966 data according to the YFS 2001 (NFBC1966 31-year visit) and YFS 2011 (NFBC1966 46-year visit).

Statistical analyses

Longitudinal associations were quantified based on pair-wise differences that were calculated separately for every individual between consecutive visits. Statistical significance was estimated by

10,000 permutation cycles and 95% confidence intervals (CI95) by 10,000 bootstrapping samples. The permutation analysis was summarized by the Z-score that indicates the distance between the observed difference and the simulated random differences in standard deviations of the simulated null distribution. Two-tailed P-values were calculated from the Z-scores by inverse standard normal distribution. These techniques were used throughout the study unless otherwise indicated.

To elucidate associations with body weight changes, we stratified men and non-pregnant women, respectively, into those who maintained their weight change within ±3% per decade and waist-hip-ratio (WHR) change between −0.05 and +0.015 (Stable), those who lost weight beyond the stable definition (Loss) and those who gained weight (Gain). To mitigate confounding from baseline status, we matched the weight gain and loss subsets according to baseline body mass index (BMI), WHR and clinical assays for insulin, glucose, C-reactive protein, triglycerides, LDL and HDL cholesterol (Supplementary Table S2).

Multiple testing

Principal component analysis (PCA) of the biochemical data revealed that the first 48 principal components explained 99% of the total variance when all data were pooled. These results were compatible with earlier work [32]. For consistency, we set the multiple testing threshold at the more conservative $P < 0.0006$ to match the previous paper (equivalent to Bonferroni adjustment for 83 independent tests at 5% type 1 error rate). All statistical analyses were conducted in the R environment (https://www.R-project.org).

Results

The age structure of the participants is illustrated in Figure1A and the basic characteristics are listed in Supplementary Table S3. The numbers of individuals included in different study settings after excluding participants with missing or low-quality samples are listed in Figure 1B. Women who were pregnant at the time of sample collection were excluded (N = 202 in NFBC1966, N = 107 in YFS). Mean BMI varied between 24.2 and 25.9 kg/m^2 across the cohorts and visits. Smoking was more prevalent in the first visit (21.2% in NFBC1966 and 24.7% in YFS) compared to the last visit (17.9% in NFBC1966 and 14.8% in YFS, $P = 5.8 \times 10^{-14}$ for the combined difference between visits over both cohorts). The prevalence of diabetes was ≤3.0% across the visits.

Sample quality

Biochemical sample quality was verified by comparing the clinical and NMR assays as described in Methods. Briefly, we used measures that were available from both clinical and NMR assays to determine if the aggregate discrepancy across multiple biomarker concentrations from the two sources was greater than could be expected by random measurement noise (Figure 2A,B). Importantly, the residuals from multiple biomarkers were analysed together to emphasize sample quality over isolated measurement errors in a single biomarker: if there is a consistent difference between NMR and clinical assays over multiple biomarkers, then it is likely that the biological material of the sample had been altered in some way rather than a technical error in a specific NMR quantity or biochemical assay.

To quantify the aggregate discrepancy, we developed a quality score that follows the standard normal noise distribution when all samples are of high quality (indicated as solid curves in Figure 2C-G). Samples that were deemed too far from the normal model were excluded. Of note, the 2012 collection of NFBC1966 was designed specifically to accommodate NMR metabolomics and the observed quality score distribution was almost identical with the predicted curve (Figure 1G).

Calibration between visits

Preliminary comparisons between the two follow-up periods within YFS revealed systematic bias between time points that was extreme (>3 median absolute deviations) and biologically implausible (Figure 3C). For this reason, we applied a calibration procedure according to non-biochemical characteristics (age, sex, BMI, weight, height and WHR) to remove bias between visits from the biochemical data (Figure 3A,B, technical details in Methods). We assumed that two subsets of people picked from the population who are identical in these characteristics (Supplementary Table

S1) would also have the same average metabolic profiles. This makes it possible to calculate a scaling factor between visits and use it to remove batch effects (Figure 3B).

The metabolic changes in calibrated concentrations were coherent and biologically meaningful, including consistent increases in glucose, triglycerides, cholesterol and glycoprotein acetyls, and consistent adjustments between NMR and clinical assays (Figure 3C). Please note that the results in the next section represent the combined trends over both YFS periods. The impact of calibration on temporal trends calculated this way are included in Supplementary Figure S3.

Age-associated changes

Temporal changes in selected metabolic measures are presented in Figure 4 and full statistics and confidence intervals for all measures are available in Supplementary Table S4. To strengthen conclusions about longitudinal changes, we used robust pair-wise median statistics (details in Methods) and developed further evaluation criteria to summarize the evidence obtained from four types of analyses. Firstly, we required that age-associations within the calibrated YFS data satisfied $P < 0.0006$. Secondly, we checked the concordance between calibrated and non-calibrated associations within the YFS. Thirdly, we checked the concordance between calibrated and non-calibrated associations within the NFBC1966. Lastly, we checked the concordance between longitudinal change and cross-sectional age correlation within the YFS. These four criteria were incorporated visually into Figure 4A-M. All findings mentioned in the main text represent calibrated median rates of change per decade in the combined YFS and NFBC1966 dataset and satisfy $P < 0.0006$ unless otherwise indicated.

The majority of the YFS participants (71%) and NFBC1966 participants (83%) gained weight during the follow-up period and the combined median rate was +4.0 kg (+5.9%) per decade. Accordingly, both BMI (+1.35 kg/m^2) and WHR (+0.048) increased. Simultaneously, clinical biomarkers deteriorated with increases in blood pressure (systolic +5 mmHg and diastolic +5 mmHg), glucose (+0.23 mmol/L, +4.6%), triglycerides (+0.16 mmol/L, +17%) and LDL cholesterol (+0.35 mmol/L, +12%).

The general pattern of increasing circulating lipids was reflected in NMR-based measures. These included lipoprotein measures related to apolipoprotein B such as medium LDL lipids (+0.086 mmol/L, +11%, Figure 4B,D), lipoprotein triglycerides such as VLDL-TG (+0.12 mmol/L, +23%, Figure 4F) and fatty acids such as omega-3 (+0.059 mmol/L, +11%, Figure 4H). The lipids in the large HDL subclass decreased (−0.055 mmol/L, −7.9%, Figure 4B), however, the association was not consistent across the evaluation criteria. We observed increases in most of the amino acids (Figure 4J) including

alanine (+0.0254 mmol/L, +6.3%), glutamine (+0.0074 mmol/L, +1.3%), glycine (+0.0103 mmol/L, +3.5%) and tyrosine (+0.0041 mmol/L, +7.8%).

Metabolic changes stratified by weight change

To isolate associations with weight change, we stratified the study participants into those who gained or lost weight, or who were stable (definitions in Methods). The weight loss and gain subgroups were matched at baseline (Supplementary Table S2). Body weight decreased −2.5 kg (−3.5%) in the former while increased +6.0 kg (+8.2%) per decade in the latter. Both subgroups showed an increase in waist-hip ratio, but weight gain was associated with a faster increase (+0.031 vs. +0.058 per decade, Figure 5A). Insulin decreased with weight loss but increased with weight gain (−10% vs. +24%, Figure 5K) and the pattern was similar for C-reactive protein (−22% vs. +25%, Figure 5L). Triglycerides, cholesterol, other lipids and most amino acids showed similar separations in the rate of change, albeit small increases rather than decreases were observed in the weight loss subgroup. The opposite pattern was observed for large HDL lipids (Figure 5B) and HDL cholesterol (Figure 5E).

In the stable subgroup (Supplementary Figure S4), clinical measures for total cholesterol (+0.35 mmol/L, +7.2% per decade), HDL cholesterol (+0.052 mmol/L, +4.5%), glucose (+0.20 mmol/L, +4.0%) and blood pressure increased (systolic +4.0 mmHg and diastolic +3.5 mmHg). The NMR data revealed increases in the two smallest VLDL, intermediate-density-lipoprotein (IDL) and all LDL subclass lipids and most fatty acids. Alanine increased (+3.1%), whereas histidine (−3.3%) and creatinine (−2.9 μmol/L, −4.1%) decreased.

Differences between sexes

To compare men and women, we visualized proportional changes with respect to the baseline (Figure 6A-L). Women gained more weight (+6.5% per decade vs. +5.0% in men), but experienced milder changes in multiple clinical biomarkers including insulin (+10.1% vs. +17.7%, Figure 6J), triglycerides (+14.9% vs. +20.1%, Figure 6F) and C-reactive protein (+5.1% vs. +21.1%, Figure 6K). Substantial divergence was observed in the absolute levels of circulating lipoprotein lipids (Supplementary Figure S5), however, most of these differences disappeared when normalized by baseline concentrations (Figure 6B,E-H). Diverging trends was observed in multiple amino acids (Figure 6I): glutamine (+1.9% vs. +0.8%) and glycine (+5.2% vs. +1.9%) increased more in women, whereas leucine (+0.6% vs. +3.4%) and phenylalanine (+0.6% vs. +3.5%) increased in men.

Interactions between sex and weight change

We used multi-variate linear regression to investigate interactions between sex and weight change (Supplementary Figure S6). For most metabolic measures, the baseline value (due to regression towards the mean) was the strongest regressor over baseline body mass, change in weight or sex (Supplementary Table S5). Interaction effects were observed for 35 measures and the regression coefficients revealed distinct patterns of associations. Changes in insulin were explained by weight change rather than sex (Figure 6V). Changes in VLDL and HDL lipids were mainly explained by weight change but with substantial sex effects (Figure 6M). Changes in isoleucine and leucine were associated with both weight change and sex (Figure 6U). Finally, changes in creatinine were mainly associated with sex (Figure 6W).

Discussion

In this study, we investigated ageing trends of metabolic measures in two large longitudinal cohorts of young and middle-aged adults. We developed new statistical techniques that allowed us to investigate how circulating biomarkers changed over time and we report previously unavailable quantitative trends for over 200 measures, including lipoprotein subclass lipids, branched-chain amino acids and cardiometabolic biomarkers. Furthermore, we isolated associations between metabolic changes and body mass in a carefully constructed design where weight loss and weight gain subgroups were matched at baseline. Our analyses revealed that weight change explains much of the metabolic variation especially regarding insulin, VLDL and HDL particles, amino acids and C-reactive protein. On the other hand, other indicators such as waist-hip ratio, blood pressure, LDL cholesterol and creatinine may represent an underlying stable trend that affects all individuals.

Confounding in longitudinal population data

Longitudinal designs are superior to cross-sectional ageing studies [7,9–13,24], but long gaps in sample collection (i.e., batch effects) can lead to severe confounding as we observed in the YFS. Batch correction is typically focused on non-biological variation that results from instrument drift and other technical variation within workflows, and sample stability that depends on the collection, handling, storage and human operators [25,33]. These effects can be minimized by rigorous workflows (e.g., the automated NMR pipeline has excellent reproducibility [28]), by including standardized control aliquots at regular intervals within a measurement series, by adding known concentrations of stable reference molecules or by applying statistical adjustment techniques [33].

Batch effects from study visits separated by long time gaps cannot be corrected by any of the previously mentioned statistical techniques if there are only two or three time points available. Furthermore, if the trait of interest (change in metabolite concentration between two measurements) and the source of the batch effect (differences between the two measurements due to non-biological factors) are perfectly correlated – which they are in this study – removing batch effects the usual way will also remove all biological trends from the concentration values.

New methodology for longitudinal studies

To address the challenge of batch effects that are correlated with the trait of interest, we developed an approach that first removes as much of the biological difference between the visits as possible (matched subsets in Figure 3A). This step allows for conventional batch correction to be applied as

if the traits and batches were uncorrelated (Figure 3B). The fundamental idea is widely applicable to studies that suffer from batch effects that are correlated with clinical outcomes.

Ageing trajectories and weight change

We estimated temporal rates for over 200 metabolic measures in large longitudinal population-based cohorts, which provides new understanding on the direction and magnitude of age-associated metabolic changes and how some of these changes may be driven by weight change. Reliable literature on longitudinal metabolic trajectories is scarce and direct comparisons are difficult, but the observed changes in insulin, glucose, various lipid measures and most of the amino acids are compatible with the previously reported patterns in young and middle-aged individuals [9,10,12,34]. We also observed consistent patterns between NMR-based and clinical assays. The combination of sample quality control, calibration and robust pair-wise median-based statistics that we developed proved to be a reliable framework to determine temporal trends in physical measurement units in omics studies that involve long time gaps.

When we controlled for weight change, waist-hip ratio and creatinine emerged as covariates of ageing (possibly pointing to loss of muscle vs. adipose [35]). We also observed consistent increases in cholesterol and blood pressure that are two classical cardiovascular risk factors. These shifts can be explained by modifiable life circumstances such as reduced physical activity [36,37], but they also fit with the genetic program of ageing and the molecular clock concept as observed in laboratory settings [38].

Strengths and weaknesses

Two independent cohorts of matching ethnicity, socioeconomics and time period provide us with robust data and high statistical power, but these strengths also mean that the results may not generalize outside the Northern European context. Furthermore, the results apply to adults under the age of 50 and further studies are needed to establish explicit links to late-life phenomena or how diet, exercise and genetics may influence the ageing trajectories of metabolite concentrations [13,36,39].

The combination of clinical and NMR assays gives us confidence that the data are coherent and of high technical quality across decades. We are also confident that the quality control method we introduced can detect non-informative samples accurately, as was demonstrated by the notable difference between the 2012 visit of NFBC1966 (short storage period, sample collection was designed to support metabolomics) and the other collections that were not specifically optimized for metabolomics studies (Figure 2). We used a two-layer approach to assess sample quality

(Supplementary Figure S1) since it gave us additional information on whether an outlier was due to a single extreme measurement in a single biomarker, or if all the biomarkers were systematically off by a small amount. This qualitative information may not be always necessary, and direct calculation of multi-variate metrics such as the Mahalanobis distance may be a more practical choice [40].

There is less clear evidence on how well the calibration technique works, although the consistency between statistical tests in Figure 4 is promising. Our matching procedure is likely to be highly effective against technical cohort effects. On the other hand, matching by body mass may miss some period effects such as diet trends and changing environmental exposures, however, studies that have these data available can use them for matching, which makes the calibration concept broadly applicable for epidemiological research. We have to wait for further studies with more numerous time points to fully assess the accuracy of the reported ageing trends and until such data become available, we urge caution when interpreting population-wide temporal shifts in absolute measurement units due to the inherent biases that come with samples separated by decades.

Conclusions

We solved technical and analytical challenges regarding longitudinal omics studies and the new techniques allowed new ways to study temporal trajectories in large-scale human populations. We found that weight gain drives most of the changes in systemic metabolism and the good news is that these changes are likely to be reversable by life-style adjustments as opposite trends were observed in the weight loss subgroup. This study also provides new reference information on how absolute concentrations of these and other indicators change over time in humans and what magnitude of changes are typical and achievable in real-world populations.

Ethics approval

The NFBC1966 was approved by the Northern Ostrobothnia Hospital District, Finland. The YFS was approved by the five universities with medical schools in Finland that were involved in the study (Turku, Helsinki, Tampere, Kuopio and Oulu). All participants gave written informed consent.

Data availability

The datasets used in the current study are available from the cohorts through application process for researchers who meet the criteria for access to confidential data (https://www.oulu.fi/nfbc/ and http://youngfinnsstudy.utu.fi). Regarding the YFS data the Ethics committee has concluded that under applicable law, the data from this study cannot be stored in public repositories or otherwise made publicly available. The data controller may permit access on case-by-case basis for scientific research, not however to individual participant level data, but aggregated statistical data, which cannot be traced back to the individual participants' data.

Supplementary data

Supplementary data are available at IJE online.

Author contributions

VPM and MAK conceived and designed the study. VPM conducted the statistical analyses and wrote the first draft. JK, TL, MK, JV, MP, VS, MRJ, OTR and MAK collected samples and produced the data. All authors reviewed the text and participated in the editing process.

Funding

This work was supported by Academy of Finland, Novo Nordisk foundation, Oulu Health and Wellfare Center, Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, ERDF European Regional Development Fund, EU Horizon 2020, EU Research Council and following foundations: Sigrid Juselius, Finnish Cardiovascular Research, Juho Vainio, Paavo Nurmi, Finnish Cultural, Tampere Tuberculosis, Emil Aaltonen, Yrjö Jahnsson, Signe and Arne Gyllenberg, and Finnish Diabetes Research. The Young Finns Study has been financially supported by the Academy of Finland grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of

Finnish Diabetes Association; This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 848146 for To Aition and grant agreement 755320 for TAXINOMISIS; European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation and Finnish Society of Clinical Chemistry.

Acknowledgements

MAK thanks Deborah Lawlor (University of Bristol, UK) for initial critical discussions on the caveats of cross-sectional epidemiological data. All authors acknowledge the continued contribution from the cohort participants and heath care professionals involved in the resource collection.

Conflict of interest

VS has consulted for Novo Nordisk and Sanofi and received modest honoraria from these companies. He also has ongoing research collaboration with Bayer Ltd (all unrelated to the present study). All other authors declare no competing interests.

References

1 Vollset SE, Goren E, Yuan C-W, *et al.* Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *Lancet Lond Engl* 2020;**396**:1285–306. doi:10.1016/S0140-6736(20)30677-2

2 Soininen P, Kangas AJ, Würtz P, *et al.* Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015;**8**:192–206. doi:10.1161/CIRCGENETICS.114.000216

3 de Diego I, Peleg S, Fuchs B. The role of lipids in aging-related metabolic changes. *Chem Phys Lipids* 2019;**222**:59–69. doi:10.1016/j.chemphyslip.2019.05.005

4 Gallois A, Mefford J, Ko A, *et al.* A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. *Nat Commun* 2019;**10**:4788. doi:10.1038/s41467-019-12703-7

5 Deelen J, Kettunen J, Fischer K, *et al.* A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun* 2019;**10**:3346. doi:10.1038/s41467-019-11311-9

6 Li X, Ploner A, Wang Y, *et al.* Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *eLife* 2020;**9**. doi:10.7554/eLife.51507

7 Mäkinen V-P, Ala-Korpela M. Metabolomics of aging requires large-scale longitudinal studies with replication. *Proc Natl Acad Sci U S A* 2016;**113**:E3470. doi:10.1073/pnas.1607062113

8 Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;**45**:1866–86. doi:10.1093/ije/dyw314

9 Wills AK, Lawlor DA, Matthews FE, *et al.* Life course trajectories of systolic blood pressure using longitudinal data from eight UK cohorts. *PLoS Med* 2011;**8**:e1000440. doi:10.1371/journal.pmed.1000440

10 Hopstock LA, Bønaa KH, Eggen AE, *et al.* Longitudinal and secular trends in total cholesterol levels and impact of lipid-lowering drug use among Norwegian women and men born in 1905-1977 in the population-based Tromsø Study 1979-2016. *BMJ Open* 2017;**7**:e015001. doi:10.1136/bmjopen-2016-015001

11 Wang Q, Ferreira DLS, Nelson SM, *et al.* Metabolic characterization of menopause: cross-sectional and longitudinal evidence. *BMC Med* 2018;**16**:17. doi:10.1186/s12916-018-1008-8

12 Würtz P, Wang Q, Kangas AJ, *et al.* Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med* 2014;**11**:e1001765. doi:10.1371/journal.pmed.1001765

13 Lehtovirta M, Pahkala K, Niinikoski H, *et al.* Effect of Dietary Counseling on a Comprehensive Metabolic Profile from Childhood to Adulthood. *J Pediatr* 2018;**195**:190-198.e3. doi:10.1016/j.jpeds.2017.11.057

14 Belsky DW, Moffitt TE, Cohen AA, *et al.* Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? *Am J Epidemiol* 2018;**187**:1220–30. doi:10.1093/aje/kwx346

15 Belsky DW, Caspi A, Houts R, *et al.* Quantification of biological aging in young adults. *Proc Natl Acad Sci U S A* 2015;**112**:E4104-4110. doi:10.1073/pnas.1506264112

16 Yu Z, Zhai G, Singmann P, *et al.* Human serum metabolic profiles are age dependent. *Aging Cell* 2012;**11**:960–7. doi:10.1111/j.1474-9726.2012.00865.x

17 Menni C, Kastenmüller G, Petersen AK, *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* 2013;**42**:1111–9. doi:10.1093/ije/dyt094

18 Auro K, Joensuu A, Fischer K, *et al.* A metabolic view on menopause and ageing. *Nat Commun* 2014;**5**:4708. doi:10.1038/ncomms5708

19 Chaleckis R, Murakami I, Takada J, *et al.* Individual variability in human blood metabolites identifies age-related differences. *Proc Natl Acad Sci U S A* 2016;**113**:4252–9. doi:10.1073/pnas.1603023113

20 Johnson LC, Parker K, Aguirre BF, *et al.* The plasma metabolome as a predictor of biological aging in humans. *GeroScience* 2019;**41**:895–906. doi:10.1007/s11357-019-00123-w

21 Moffitt TE, Belsky DW, Danese A, *et al.* The Longitudinal Study of Aging in Human Young Adults: Knowledge Gaps and Research Agenda. *J Gerontol A Biol Sci Med Sci* 2017;**72**:210–5. doi:10.1093/gerona/glw191

22 Chak CM, Lacruz ME, Adam J, *et al.* Ageing Investigation Using Two-Time-Point Metabolomics Data from KORA and CARLA Studies. *Metabolites* 2019;**9**. doi:10.3390/metabo9030044

23 Darst BF, Koscik RL, Hogan KJ, *et al.* Longitudinal plasma metabolomics of aging and sex. *Aging* 2019;**11**:1262–82. doi:10.18632/aging.101837

24 Wang Q, Würtz P, Auro K, *et al.* Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Med* 2016;**14**:205. doi:10.1186/s12916-016-0733-0

25 Santos Ferreira DL, Maple HJ, Goodwin M, *et al.* The Effect of Pre-Analytical Conditions on Blood Metabolomics in Epidemiological Studies. *Metabolites* 2019;**9**. doi:10.3390/metabo9040064

26 Raitakari OT, Juonala M, Rönnemaa T, *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *Int J Epidemiol* 2008;**37**:1220–6. doi:10.1093/ije/dym225

27 Rantakallio P. The longitudinal study of the Northern Finland birth cohort of 1966. *Paediatr Perinat Epidemiol* 1988;**2**:59–88. doi:10.1111/j.1365-3016.1988.tb00180.x

28 Soininen P, Kangas AJ, Würtz P, *et al.* High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *The Analyst* 2009;**134**:1781–5. doi:10.1039/b910205a

29 Bizzarri D, Reinders MJT, Beekman M, *et al.* 1H-NMR metabolomics-based surrogates to impute common clinical risk factors and endpoints. *eBioMedicine* 2022;**75**:103764. doi:10.1016/j.ebiom.2021.103764

30 Ala-Korpela M, Zhao S, Järvelin M-R, *et al.* Apt interpretation of comprehensive lipoprotein data in large-scale epidemiology: disclosure of fundamental structural and metabolic relationships. *Int J Epidemiol* 2021;:dyab156. doi:10.1093/ije/dyab156

31 Tam BT, Morais JA, Santosa S. Obesity and ageing: Two sides of the same coin. *Obes Rev* 2020;**21**. doi:10.1111/obr.12991

32 Wang Q, Jokelainen J, Auvinen J, *et al.* Insulin resistance and systemic metabolic changes in oral glucose tolerance test in 5340 individuals: an interventional study. *BMC Med* 2019;**17**:217. doi:10.1186/s12916-019-1440-4

33 Han W, Li L. Evaluating and minimizing batch effects in metabolomics. *Mass Spectrom Rev* Published Online First: 25 November 2020. doi:10.1002/mas.21672

34 Stancáková A, Civelek M, Saleem NK, *et al.* Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* 2012;**61**:1895–902. doi:10.2337/db11-1378

35 Srikanthan P, Karlamangla AS. Muscle mass index as a predictor of longevity in older adults. *Am J Med* 2014;**127**:547–53. doi:10.1016/j.amjmed.2014.02.007

36 Kujala UM, Mäkinen V-P, Heinonen I, *et al.* Long-term leisure-time physical activity and serum metabolome. *Circulation* 2013;**127**:340–8. doi:10.1161/CIRCULATIONAHA.112.105551

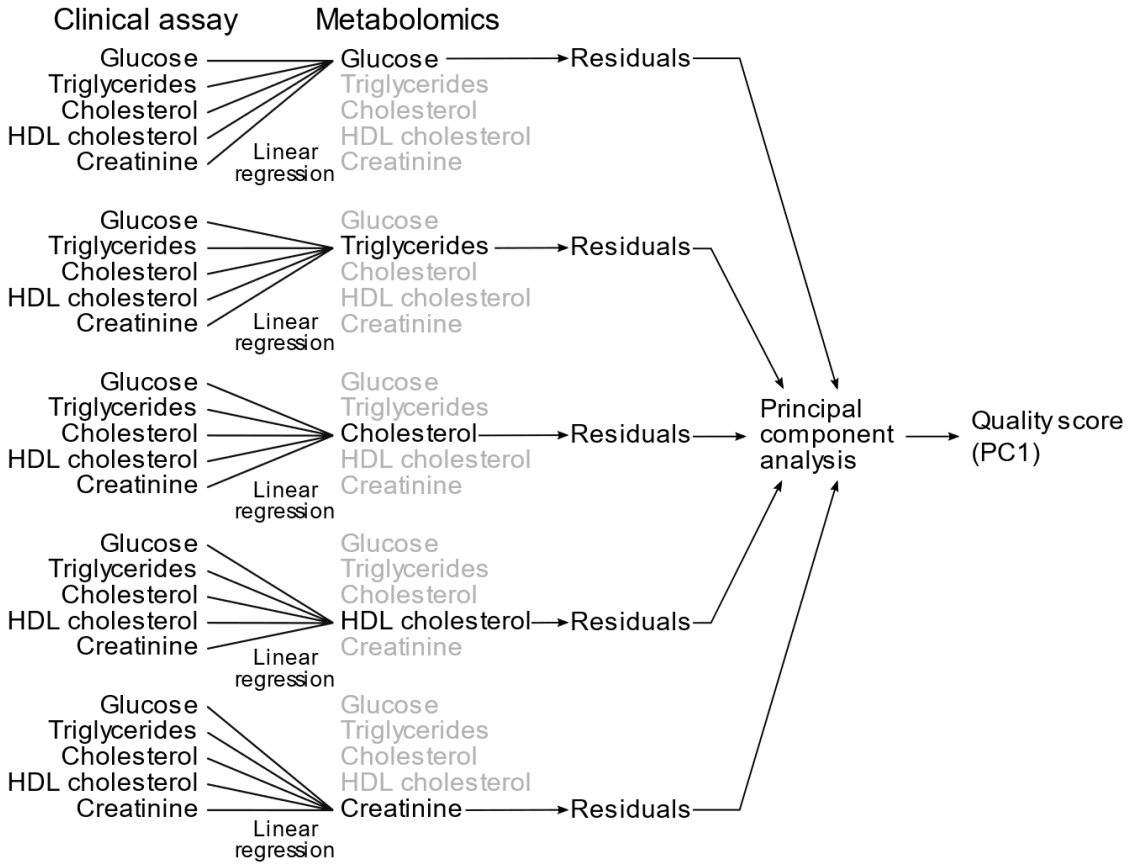
37 Bowden Davies KA, Pickles S, Sprung VS, *et al.* Reduced physical activity in young and older adults: metabolic and musculoskeletal implications. *Ther Adv Endocrinol Metab* 2019;**10**:2042018819888824. doi:10.1177/2042018819888824

38 Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* 2018;**19**:371–84. doi:10.1038/s41576-018-0004-3

39 Kettunen J, Demirkan A, Würtz P, *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016;**7**:11122. doi:10.1038/ncomms11122

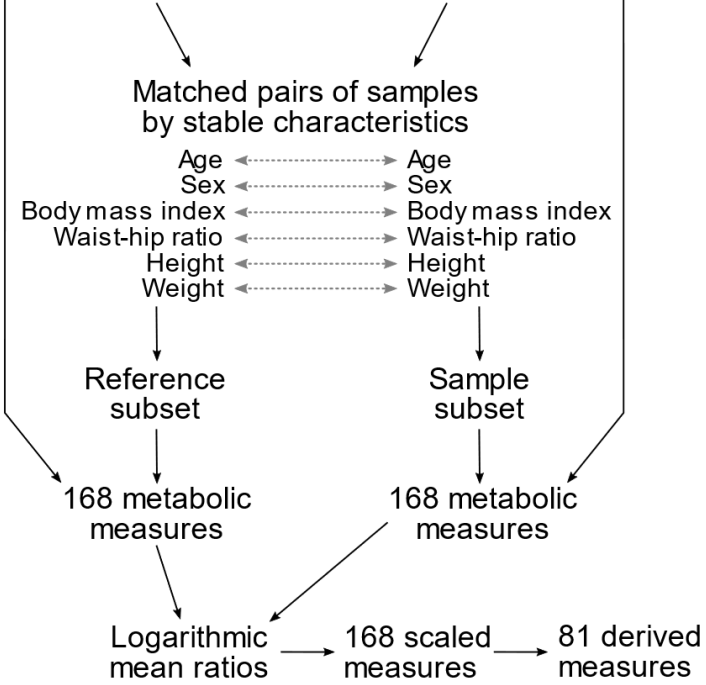
40 De Maesschalck R, Jouan-Rimbaud D, Massart DL. The Mahalanobis distance. *Chemom Intell Lab Syst* 2000;**50**:1–18. doi:10.1016/S0169-7439(99)00047-7

Supplementary Figure S1



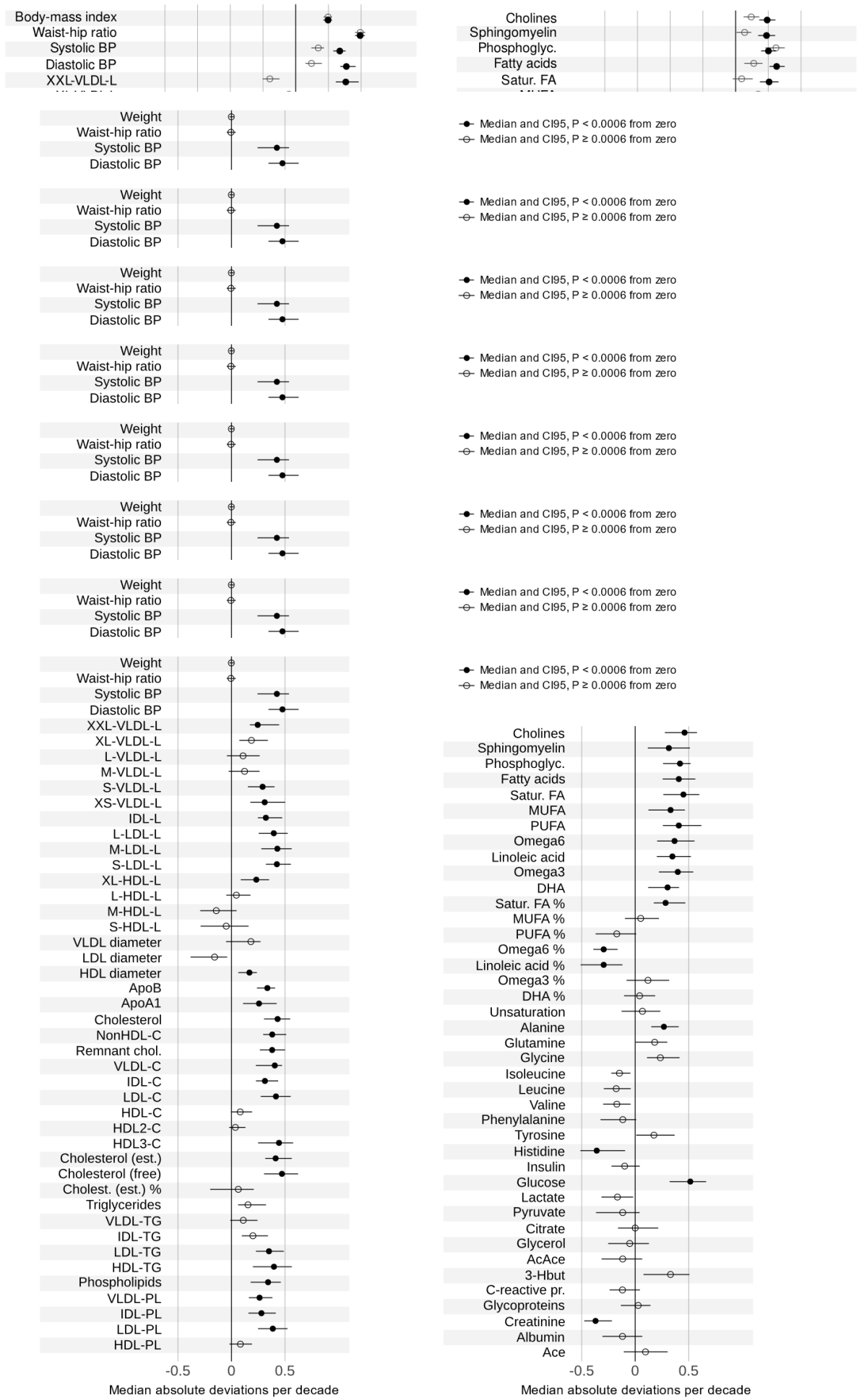
Consistency between shared biomarkers across NMR metabolomics and clinical assays was modelled by linear regression and principal component analysis (PCA). First, we predicted NMR measures from the clinical assays. The rationale behind the method goes as follows. Comparing just NMR and clinical glucose values will capture situations where the two measurements disagree, however, if the sample material itself was somehow compromised, both measurements would still be consistent. By including other biomarkers in the modelling, we were able to detect if there are biological inconsistencies in the clinical data, thus capturing issues related to the sample itself. In the second step, we collect the residuals from the linear regression models and then analyse them with principal component analysis. This step detects inconsistencies in the NMR data, that is, if the residuals across the NMR measures are repeatedly far from zero, their expected value.

Supplementary Figure S2



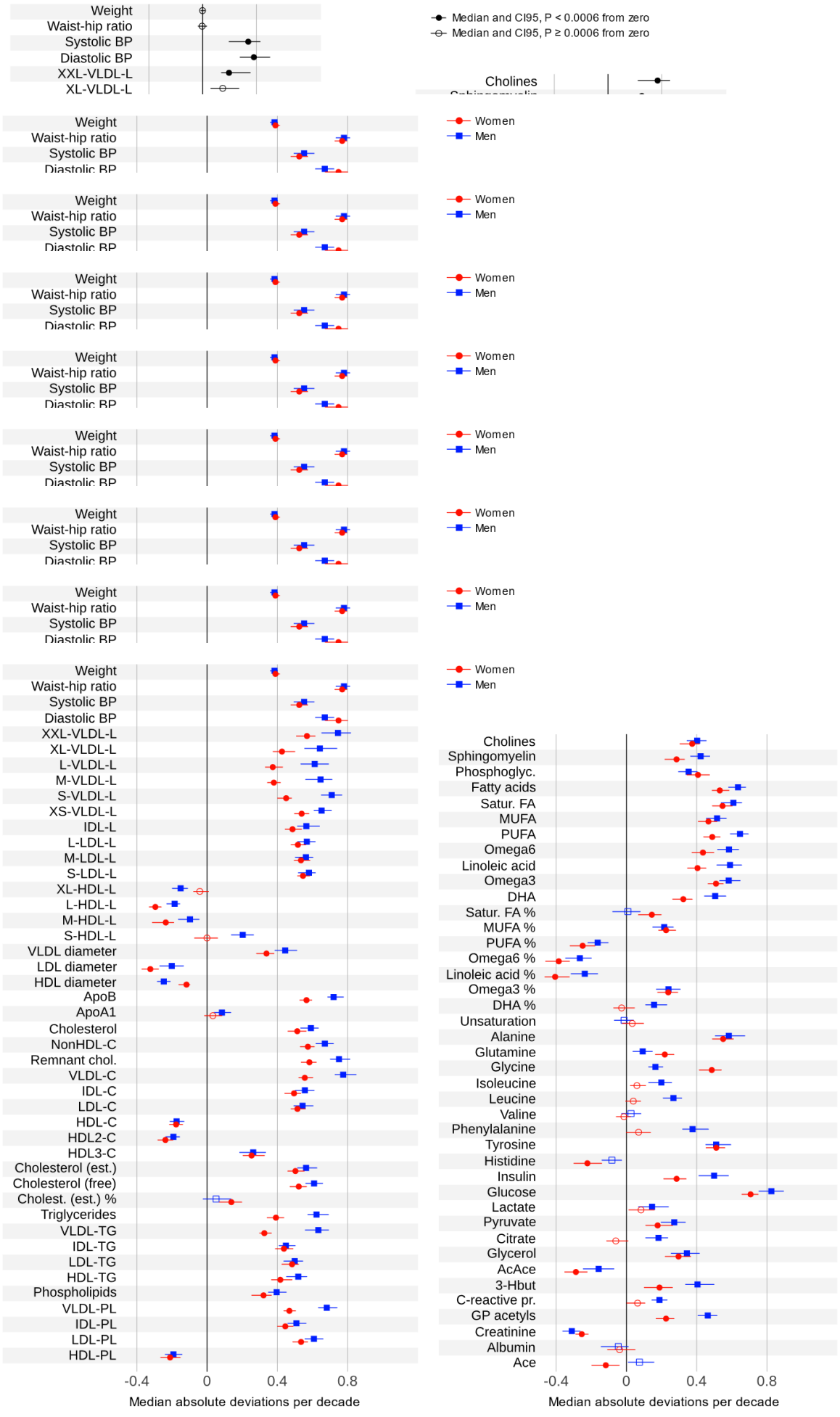
Calibration of metabolic measures between two cohort visits (batches) was achieved by matching according to basic characteristics and scaling based on mean values. First, we chose one of the visits as the reference dataset (e.g. YFS 2001) and another as the dataset to be calibrated (e.g. YFS 2007). Next, we identified pairs of individuals across the datasets that had the same age, sex and basic characteristics. Not all individuals would have a good match, therefore we chose the best matching 225 pairs of men and 225 pairs of women. We were thus able to define a *reference subset* of 450 individuals and a corresponding *sample subset* of 450 individuals. We then made the assumption that since these two subset had the same ages and basic characteristics, they also have the same average profile of the 168 metabolic measures that were directly measured in the study. Consequently, any observed differences between the profiles will be due to the batch effects rather than biology. To remove those batch effects, we calculated the ratio of the average metabolite concentration in the sample subset with respect to the reference subset. These scaling factors were then applied to the original dataset. Lastly, derived measures such as fatty acid ratios were updated using the calibrated concentration values.

Supplementary Figure S3



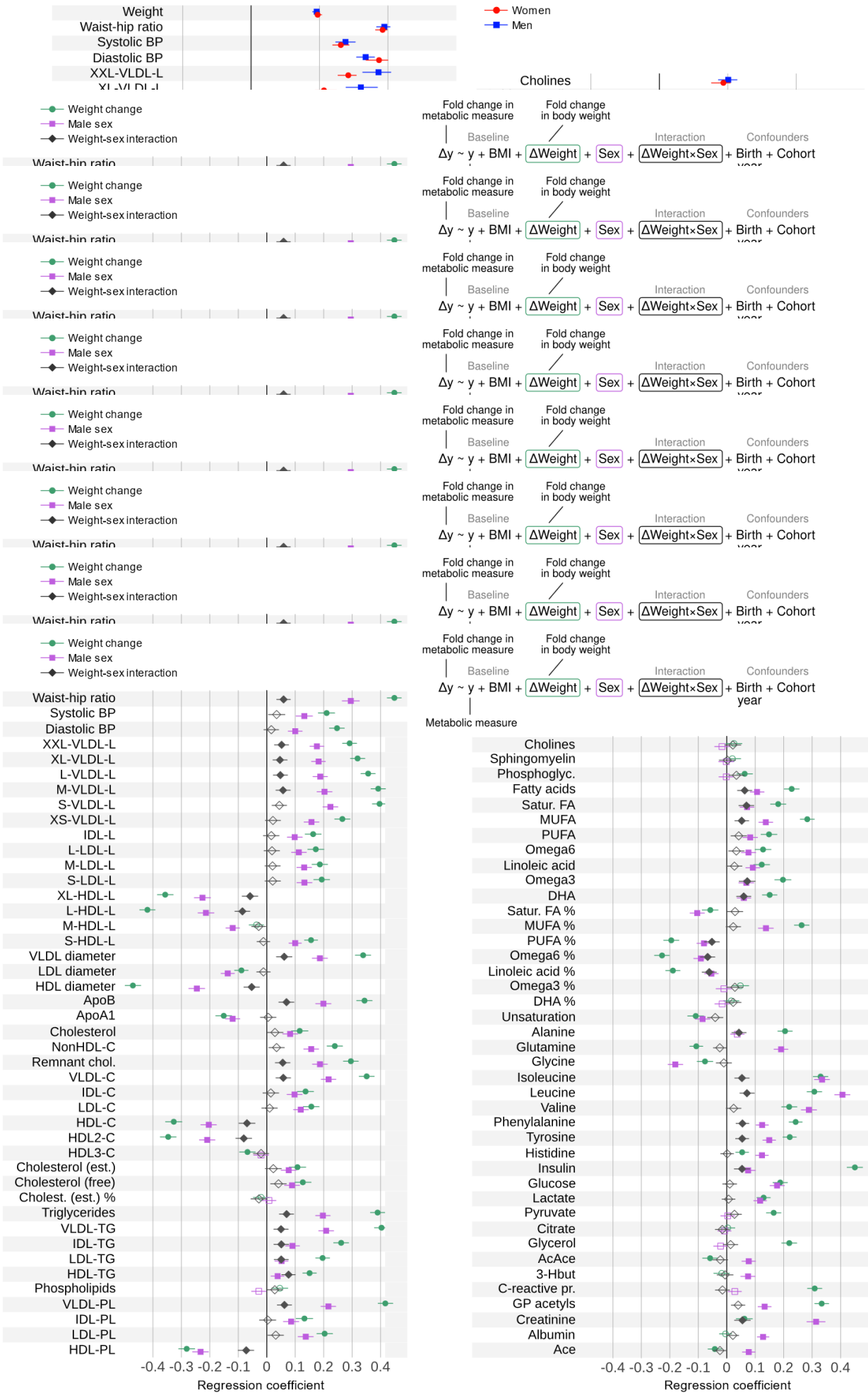
Supplement Figure: Rates of change for metabolic measures in the Stable weight subgroup. The results were calculated from the combined dataset of the Young Finns Study and the Northern Finland Birth Cohort 1966.

Supplement Figure S4



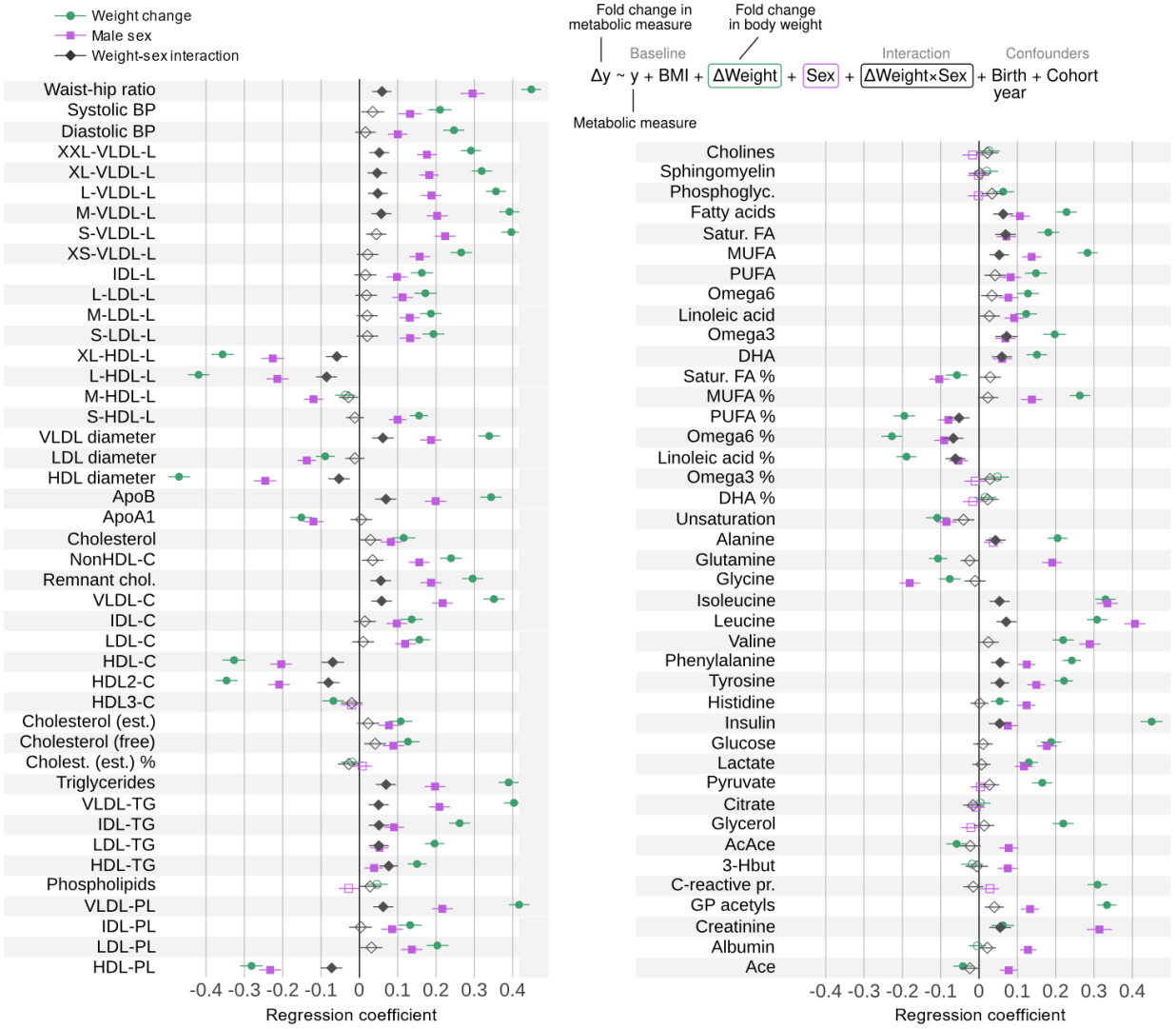
Supplement Figure: Absolute changes in metabolic measures in the combined dataset of the Young Finns Study and the Northern Finland Birth Cohort 1966.

Supplement Figure S5



Supplement Figure: Linear regression modelling of weight-sex interaction.

Supplement Figure S6



Supplement Figure: Linear regression modelling of weight-sex interaction.