Baltimore Longitudinal Study of Aging: The Associations Between Individual's Allostatic Loads and Energy Expenditure

Columbia Aging Center

Runze Cui

2023-12-04

Contents

Introduction

Background

Psychosomatic research has increasingly focused on exploring the relationship between psychological factors and major physiological regulatory systems in humans such as the parasympathetic nervous system (PNS) and the hypothalamic-pituitary-adrenal (HPA) axis (Wiley et al., 2016). As the field of psychosomatic medicine has advanced, it is increasingly common for multiple biomarkers to be assessed. Although many biomarkers exist for each physiological system, there is little consensus on how to integrate these biomarkers to assess the state and functioning of physiological systems and multisystem physiological dysregulation.

Physiological dysregulation can be quantified using the allostatic load (AL) index, a multi-system metric of physiological dysregulation with strong theoretical foundation and prior evidence that it predicts the loss of functional capacity over time. The additional energetic cost of living from physiological dysregulation - the "load" in allostatic load - can be quantified as excess energy expenditure (EE) [1], supported by cellular experiments demonstrating that glucocorticoid-induced cellular allostatic load increases EE by approximately 60%, in parallel with accelerated cellular aging (Bobba-Alves et al., 2023). The central hypothesis contends that individuals exhibiting higher allostatic load will manifest correspondingly elevated levels of energy expenditure. More specifically, we anticipate a positive correlation between allostatic load and energy expenditure. Additionally, We predict this correlation to persist across genders, and remain significant after adjusting for variables such as age, body mass index (BMI), and existing medical conditions.

Data Information

Initiated in 1958 by Dr. Nathan Shock, the Baltimore Longitudinal Study of Aging (BLSA)[2], conducted under the auspices of the National Institute on Aging, represents a groundbreaking endeavor in gerontological research. Characterized by its longitudinal approach, the BLSA systematically investigates the multifaceted dimensions of aging, encompassing physical, cognitive, and psychological variables. Distinct in its inclusive sampling across a broad spectrum of adult ages, the study offers an in-depth analysis of the aging trajectory commencing from middle adulthood. Through rigorous and extensive participant evaluations, the BLSA has been instrumental in deconstructing prevalent aging stereotypes, notably the alleged inescapable decline in cognitive faculties. The study's profound contributions have notably advanced scientific discourse on aging, simultaneously informing and shaping pertinent public health policies.

In this study, we recorded a total of six observations of energy expenditure (EE, measured in kcal/min) for each participant, from 12:30 PM to 15:30 PM. We then calculated the hourly energy expenditure (HEE) by

[1] Energy expenditure can be measured by doubly-labeled water techniques

[2] Official Webpage: National Institute on Aging: Baltimore Longitudinal Study of Aging, https://www.blsa.nih.gov/

multiplying each EE value by 30 minutes and subsequently dividing by two hours, to serve as our continuous response variable. Our dataset also includes 19 continuous covariates (`age`, `BMI`, `albumin`, `adiponectin`, `cholestrol`, `crp`, `creatinine`, `sbp`, `dbp`, `fast_glucose`, `hdl`, `ldl`, `hba1c`, `sodium`, `triglycerides`, `tsh`, `leptin`, `ucic_acid`, `vb12`) and 3 categorical variables (`sex`, `race`, `edu_level`) along with the cluster's `id`[3]. In subsequent analyses, we employed 18 of these continuous biomarker variables to estimate the allostatic load score using a single factor model[4]. This estimation was achieved by examining and visualizing the distribution of each biomarker, then determining whether the subject's biomarker levels at each time point were within normal or abnormal ranges, based on clinical or percentile thresholds (if clinial cutoff is not available). The allostatic load score can also be computed using the R package `pscore`, developed by Joshua F. Wiley from Monash University's School of Psychological Sciences[5].

The raw dataset for our study is extensive, encompassing detailed demographic information, including binary indicators for subjects with family incomes exceeding $10,000, as well as various uncommonly used biomarkers. We removed these variables. Additionally, considering that our study primarily aims to investigate changes over time, we decided to exclude participants who have only a single recorded visit. The rationale behind this decision is that single-visit data does not contribute to the whole longitudinal progression. This adjustment ensures that our analysis remains aligned with our research goals and is conducted on a dataset that is both manageable and relevant to our objectives. Finally, some clusters have missing data for only specific types of biomarkers. This could be due to the measurement process. For example, physician forgot to record or measure certain biomarker values for some individuals. Or some subjects are reluctant to disclose certain biomarker values in this study because they have certain diseases that can be revealed by these biomarker values. However, this would affect the consistency of our estimation of AL scores. To avoid these measurement biases, we also remove these subjects from our dataset.

Methods

Exploratory Data Analysis (EDA)

Summary Statistics Table Findings: We first construct a table summarizing the sample on the basis of their covariates at baseline separated by sex. Since the full table is too long, we put it into the **Appendix: Table** section at the end. As we mentioned in the introduction part, after getting rid of individuals who have missing data on certain variable measures, the dataset has actually unbalanced design. More discussion of the reasons of unbalanced data structure and missing issues are provided subsequently.

[3]See the descriptive statistics table in the Methods section for details

[4]More complex models are discussed in the paper "Modeling Multisystem Physiological Dysregulation" which could further improve the model performance

[5]Details about how to use: https://github.com/JWiley/score-project

Distributions of Variable for Each Individual: Distribution checking could help us familiarize the potential pattern of our variables, to understand whether the variables satisfy the normality assumption for subsequent model fitting, and whether the variables contain anomalous outliers [6]. We then use histograms and bar charts to visualize the distribution of continuous and categorical variables, respectively.

Figure 1.1: Histograms of Averaged Continuous Variable's Value For Each Individual

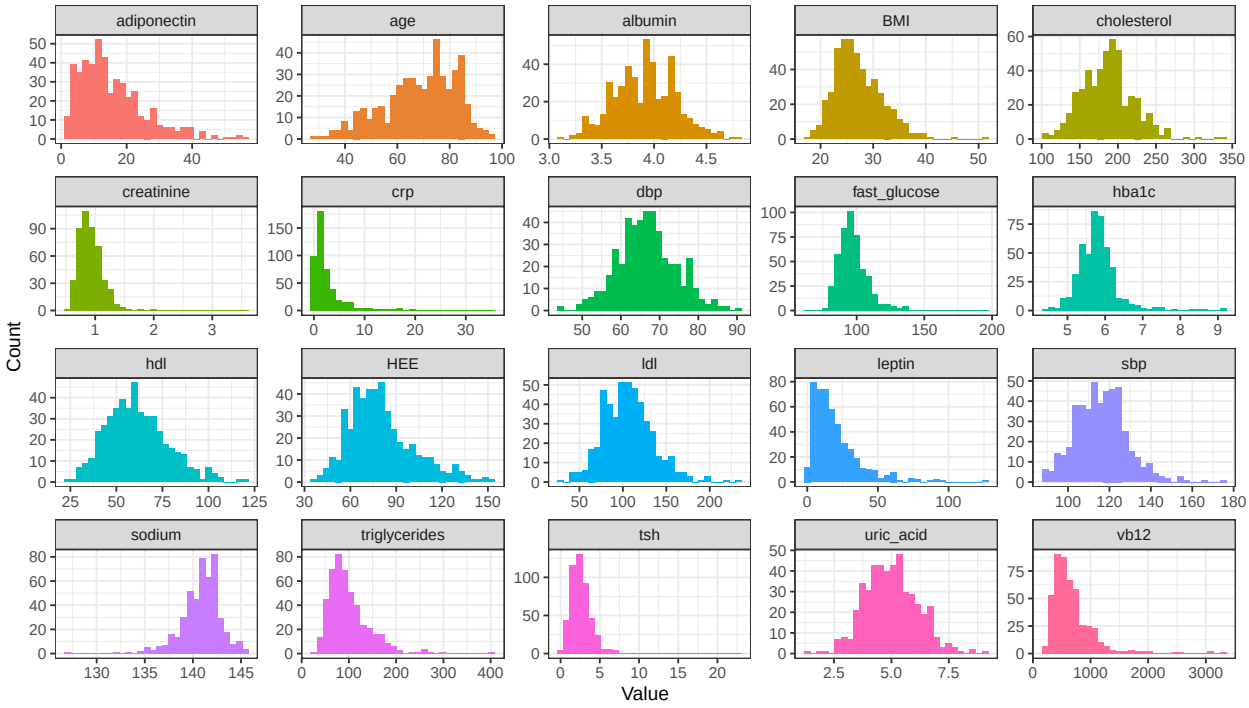
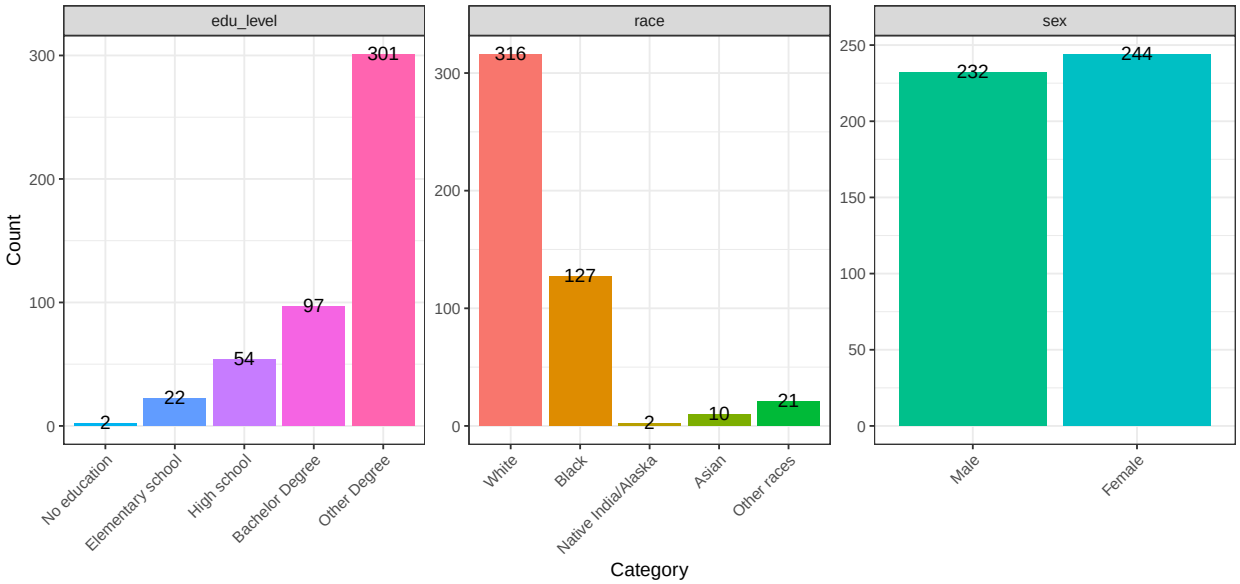


Figure 1.2: Bar Charts of Categorical variable For Each Individual



[6] the normality assumption and outlier checking are only applicable to continuous variables

For certain biomarkers, such as `leptin`, `adiponectin`, `creatinine`, `cholesterol`, `crp`, `triglycerides`, `TSH`and `VB12`, their distributions are not symmetric and contain multiple outliers. It is known that the marginal model does not require a strict assumption of normal distribution. However, when estimating the AL score using percentile thresholds[7] and applying linear mixed models, a normality assumption is still necessary. These issues can be addressed through data pre-processing such as log-transformation and the removal of outliers before the comprehensive data analysis.

To get rid of unusual outliers, we set up a criteria such that: $\begin{cases} X_i \geq \mu + 5\sigma \rightarrow \text{NA} \\ X_i \leq \mu - 5\sigma \rightarrow \text{NA} \end{cases}$

Allostatic Load Score Calculation

Use the clinical or percentile cutoff to recode the normal/abnormal for each biomarker: We used the clinical cutoffs from Mayo Clinic Official Website[8] to to determine whether each biomarker in each person's record was either healthy or unhealthy. These binary variables help us calculate a composite score to quantify the level of allostatic load at each individual's visit. Information about clinical thresholds stratified by sex are as follows:

Table 2: Mayo Clinic Biomarker Clinical Cutoffs

Biomarker	Male Normal Range	Female Normal Range
Adiponectin	change based on BMI (ug/mL)	change based on BMI (ug/mL)
Fasting Glucose	70 - 100 (mg/dl)	70 - 100 (mg/dl)
Albumin	3.4 - 5.4 (g/dL)	3.4 - 5.4 (g/dL)
BMI	18.5 - 30	18.5 - 30
Cholesterol	0 - 200 (mg/dL)	0 - 200 (mg/dL)
Systolic pressure	90 - 120 (mmHg)	90 - 120 (mmHg)
Diastolic pressure	60 - 80 (mmHg)	60 - 80 (mmHg)
Creatinine	0.74 - 1.35 (mg/dL)	0.59 - 1.04 (mg/dL)
CRP	0 - 10 (ug/mL)	0 - 10 (ug/mL)
HBA1C (%)	4.1 - 5.7 (%)	4.1 - 5.7 (%)
HDL	>= 60 (mg/dL)	>= 60 (mg/dL)
LDL	100 - 129 (mg/dL)	100 - 129 (mg/dL)
leptin	change based on BMI (ng/mL)	change based on BMI (ng/mL)
sodium	135 - 145 (mmol/L)	135 - 145 (mmol/L)
Triglycerides	0 - 150 (mg/dL)	0 - 150 (mg/dL)
TSH	0.3 - 4.2	0.3 - 4.2
uric acid	3.7 - 8 (mg/dL)	2.7 - 6.1 (mg/dL)
VIT 12	180 - 914 (pg/mL)	180 - 914 (pg/mL)

[7]If we use clinical threshold to estimate AL score, the variables should not be transformed due to different scale
[8]Mayo Clinic Official Webpage: https://www.mayoclinic.org/

In this study, we do not need to consider percentile cutoffs because Mayo clinical cutoffs are available for all biomarkers. In addition, for certain biomarkers whose thresholds fluctuate with an individual's "body mass index" ("BMI"), such as "adiponectin" and "leptin," we need to create customized R functions[9] to generate its normal/abnormal binary indicators.

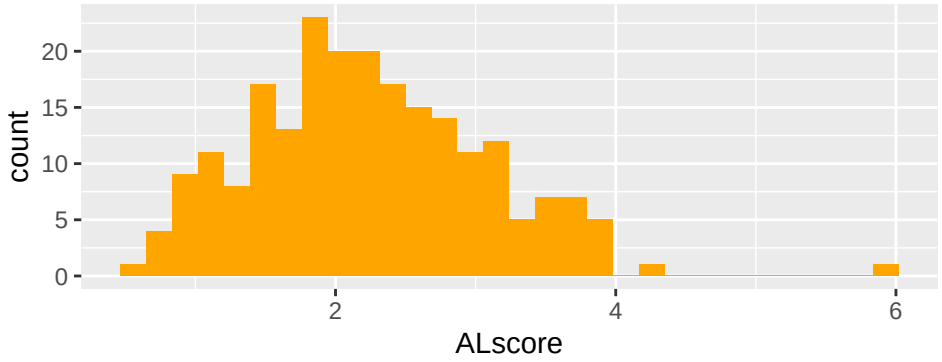
Compute the AL score as one of our covariates for further model building: Remember the allostatic load can be estimated by single-factor model based on our biomarkers. For each record, we can use the number of abnormal divided by the total non-missing biomarker number to quantify the allostatic load level (AL score). Specifically, larger AL score indicates higher allostatic loads and risk of physiological dysregulation.

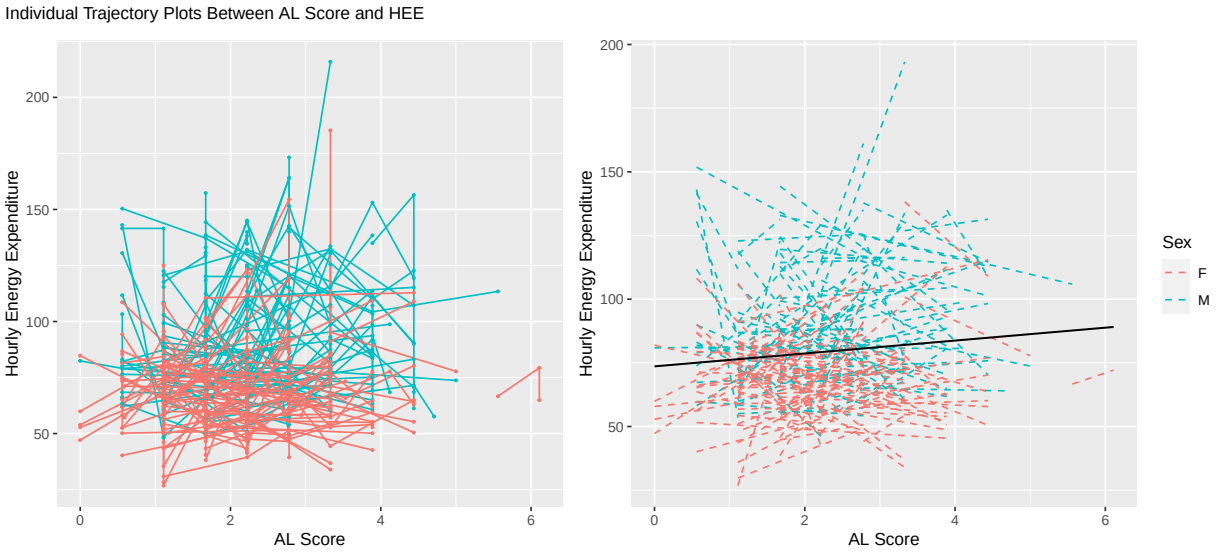
Addressing missing data: There are a variety of statistical methods for dealing with missing data in longitudinal studies, and which method is supposed to be applied depending on the pattern of our missing data. Since the BLSA is a cohort study, and all participants are enrolled voluntarily. The reasons that our participate drop out from the study would not depend on our outcome and/or covariates of interest. So, we assumed the existence of a missing completely at random (MCAR) pattern, and we can use some complete-case version of methods[10] for the estimation.

```
## # A tibble: 6 x 8
## # Rowwise:
##   id      age race  sex   edu_level   HEE   BMI ALscore
##   <fct> <dbl> <fct> <fct> <fct>     <dbl> <dbl>   <dbl>
## 1 180      85 0     M     4          87.3  22.9    1.11
## 2 180      87 0     M     4          65.4  23.3    0.56
## 3 180      91 0     M     4          82.2  24.7    0.56
## 4 276      83 0     M     4          76.4  34.6    2.22
## 5 276      84 0     M     4         121.   34.8    2.78
## 6 276      88 0     M     4          60.3  31.7    2.78
```

Distribution of BMI, AL score and initial exploration: After obtaining the AL scores, we need to pre-check the distributions, means and even dependency structures before fitting the model. Individual trajectory plots can help us understand the underlying trends between the response variable HEE and covariates such as BMI and Al score. In addition, it is possible to visualize how all clusters change over time in general or even by sex.

[9] R Code is available at the Appendix: Code section at the end
[10] Such as complete-case Generalized Estimating Equations method





The histogram of AL score is symmetric bell-shaped, which means it's normal distributed and we do not need

any further transformation. There is a tendency for HEE with increasing AL score for the whole sample. Assuming equal increases in AL score, male subjects tended to have a faster and more significant increase in HEE than female subjects. Some previous research has already proved this statement that: "Males tend to have a higher EE than females due to their higher muscle mass, which requires more energy for maintenance." (Zunzer et al., 2013). However, HEE decreases significantly as subjects age, and the decreasing trend is more obvious in female than male group. This is also not surprised since older people would have relatively lower total metabolic rate and less daily exercise, resulting in less energy expenditure at the end.

Fitting Marginal Models By Applying Generalized Estimating Equations (GEE)

As we mentioned, we assume the missing data follows MCAR pattern. We try to understand the underlying relationship between `HEE` and `ALscore`, `age`, and `BMI` using Complete-case GEE method:

$$\sum_{k=1}^{K} I(\mathbf{R}_k = 1)\mathbf{D}_k^T \mathbf{V}_k^{-1}(\mathbf{Y}_k - \mu_k) = 0$$

Specifically, $I(\mathbf{R}_k = 1) = 0/1$ indicates whether the k^{th} subject has complete data, $\mathbf{D}_k = \frac{\partial \mu_k(\beta)}{\partial \beta}$ and $\mathbf{V}_k = \mathbf{S}_k^{1/2}\mathbf{R}_k\mathbf{S}_k^{1/2}$. In other words, we only use the subjects who do not have missing data[11]. If missingness is MCAR, we know the complete-case analysis is consistent and valid[12] by proving:

$$\mathbf{E}_{\mathbf{Y},\mathbf{R}}[I(\mathbf{R}_k = 1)\mathbf{D}_k^T \mathbf{V}_k^{-1}(\mathbf{Y}_k - \mu_k)] = \mathbf{D}_k^T \mathbf{V}_k^{-1} \mathrm{P}(\mathbf{R}_k = 1)\mathbf{E}_{\mathbf{Y}}[(\mathbf{Y}_k - \mu_k)] = 0$$

Discussion the association between GEE and WLS method: The GEE estimation method is versatile and can handle various types of response variables, such as count and binary data. While, GEE method will perform the same as Weight least squared (WLS) method, if our sample size is not small and strongly skewed distribution with continuous response of interest. (Miller, M. E. et al., 1997). This can be shown by the β estimating equations comparison. The WLS method's minimization solves:

$$\sum_{k=1}^{K} \mathbf{X}_k^T \mathbf{W}_k(\mathbf{Y}_k - \mu_k) = 0$$

Obviously, this would be the same as GEE if the response variable $\mathbf{Y}_k$ is continuous. In this section, We fit a same mean model using `geeglm()` function with 3 different correlation structures: working independent, working exchangeable, and working auto-regressive 1, and compared their estimates, robust standard errors, Wald test statistics and p-values[13].

[11] Only use participants who visited 3 times in this study. Subjects who only have twice visits will be droped
[12] Point estimate is consistent and efficient (relatively small robust SE)
[13] The results table and interpretations are available in Results section

Fitting Linear Mixed Effects Model (LMM)

If the missingness is indeed MCAR and the LMM is correctly specified, then consistent and valid results can be obtained using only the complete data. However, sample size, generalizability[14], and adherence to other modeling assumptions must be considered to ensure the reliability and applicability of the results of the study. There are many different cases for the LMM. In this study, we fit 8 models with the same mean structure, comparing their AICs and log-likelihood ratios to select a better-fitting model. The mean structures are the same as those we previously used in the Complete-case GEE.

Table 3: LMM Model Comparison

Model	AIC	Log.likelihood
Independent, homoskedastic errors	5950.819	-2970.410
Random intercepts + independent, homoskedastic errors	5909.350	-2948.675
Random intercepts/slopes + independent, homoskedastic errors	5906.109	-2945.055
Random intercepts + auto-regressive errors	5911.296	-2948.648
Random intercepts + exponential spatial errors	5911.296	-2948.648
Random intercepts + exponential spatial and independent, homoskedastic errors	5913.297	-2948.649
Random intercepts/slope + exponential spatial errors	5907.532	-2944.766
Random intercepts/slope + exponential spatial and independent, homoskedastic errors	5909.532	-2944.766

Given the comparative table above, the random intercepts/slopes model with independent and homoskedastic errors has smallest AIC and large log likelihood value. We state it is better fitted than other cases and use it as our final linear mixed effect model, mathematically expressing as:

$$\begin{aligned} Y_{ki} &= \mathbf{X}_{ki}^T\beta + \gamma_{0k} + T_{ki}\gamma_{1k} + \epsilon_{ki} \\ &= \beta_0 + \beta_1 X_{\text{ALscore},ki} + \beta_2 X_{\text{age},ki} + \beta_3 X_{(\text{sex=M}),ki} + \gamma_{0k} + \gamma_{1k} T_{\text{age},ki} + \epsilon_{ki} \end{aligned}$$

$$\gamma_i = \begin{bmatrix} \gamma_{0i} \\ \gamma_{1i} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{\gamma,00} & \Sigma_{\gamma,01} \\ \Sigma_{\gamma,01} & \Sigma_{\gamma,11} \end{bmatrix} \right)$$

The β represents fixed effects and γ_k is random effects, along with $\mathbf{X}_k^T$ and $\mathbf{T}_k^T$ as their design matrices.

[14]Using GLMM for binary, count response variable

Results

Complete-case GEE Results and Interpretations

Table 4: Results of Complete-case GEE 1.5

Term	Estimates	Robust SE	Wald	P-value
Working independence				
Intercept	106.9914	5.8823	330.8292	0
AL Score	3.0363	0.9009	11.3583	8e-04
Age	-0.6420	0.0845	57.6703	0
Sex	20.5684	2.0481	100.853	0
Working exchangeable				
Intercept	113.3204	6.1051	344.5386	0
AL Score	2.4295	0.7990	9.2455	0.0024
Age	-0.7133	0.0866	67.9123	0
Sex	20.6172	2.0666	99.5252	0
ρ	0.2780	0.0452		
Working AR-1				
Intercept	110.7835	5.9356	348.3569	0
AL Score	2.2715	0.8202	7.6696	0.0056
Age	-0.6713	0.0845	63.1825	0
Sex	20.2805	2.0635	96.5904	0
ρ	0.3380	0.0489		

We choose to explain the results of complete-case GEE with exchangeable working correlation:

- The estimated average HEE at baseline age for female subjects is 113.3204 kcal/hour.
- For female subjects with 1-unit increase in AL score, the average change of HEE at baseline age is 2.4295 kcal/hour.
- With the age of the female subjects being one year older, the average change in HEE should be -0.7133 kcal/hour.
- When comparing male subject to female subject, the averaged difference of HEE at baseline age is 20.6172 kcal/hour.
- The ρ is the estimated correlation parameter (0.278) for the 'exchangeable' correlation structure, suggesting that there is a weak positive correlation between weights within the same cluster (individual). The standard error for alpha is 0.0452, indicating some precision in this estimate.
- All of them have small p-values and are statistically significant.

Even though we insert different working correlation in GEE (sometimes the working correlation is even incorrect), it provides a consistent point estimate once the mean model is correctly specified. And also, GEE uses robust standard error and would be valid even though the within-cluster correlation structure is incorrectly specified.

Linear Mixed Effects Model Results and Interpretations

Table 5: Results of Random intercepts And Random Slopes Model

Term	Estimate	StdError	DF	t-value	p-value	StdDev	Corr	Value
Fixed Effects								
(Intercept)	116.84576	6.47717	440	18.03964	0			
AL Score	2.36222	0.85408	440	2.7658	0.00592			
Age	-0.75622	0.0859	440	-8.80386	0			
Sex = Male	20.55584	2.05236	219	10.0157	0			
Random Effects								
(Intercept)						34.1740942	(Intr)	
Age						0.3898401	-0.964	
Residual						17.9906908		
Model Statistics								
AIC								5906.10895
BIC								5942.08315
LogLik								-2945.05447

Fixed Effects:

- The estimated average HEE at baseline age for female subjects is 116.84576 kcal/hour.
- For female subjects with 1-unit increase in AL score, the average change of HEE at baseline age is 2.36222 kcal/hour.
- With the age of the female subjects being one year older, the average change in HEE should be -0.75622 kcal/hour.
- When comparing male subject to female subject, the averaged difference of HEE at baseline age is 20.55584 kcal/hour.
- All of them have small p-values and are statistically significant.

Random Effects:

- There is variability in the initial weights of the subjects (as indicated by the intercept) that is not explained by the fixed effects in the model. Each subject has their own intercept, and these intercepts vary with a standard deviation of approximately 34.17 kcal/hour.
- The standard deviation of the random slopes for time is 0.3898401. This suggests the effect of time on HEE is not consistent across all individuals.
- There is a negative correlation (-0.964) between the random intercepts and the random slopes for time. This means that individuals with a higher initial HEE tend to have less increase (or a decrease) in HEE over time compared to individuals with a lower initial HEE.
- The residual standard deviation is 17.9906908, which is the within-group variability in HEE that is not explained by the model.

Conclusion

In our analysis of the Baltimore Longitudinal Study of Aging Dataset, we employed two distinct statistical approaches - the Complete-case Generalized Estimating Equations (CC-GEE) method and the Linear Mixed

Effects Model - to explore the relationship between hourly energy expenditure and allostatic load (AL) levels. Both methodologies yielded congruent findings, highlighting a notable association wherein individuals with higher AL scores, indicative of more abnormal biomarker values, tend to exhibit increased energy expenditure. Specifically, our results suggest that for female participants, the average increase in energy expenditure ranges approximately from 2.36 to 2.4 kilocalories per hour. Moreover, the analysis reveals a differential impact based on gender; male subjects demonstrate a more pronounced change in energy expenditure, exceeding 20 kilocalories, at the baseline age compared to their female counterparts.

This consistency across two methodological frameworks lends robustness to our findings. However, it is important to acknowledge that determining the superiority of one model over the other is not straightforward in this study, given their similar output. The choice between these models should be guided primarily by the specific objectives of the study in question and warrants further deliberation. Our findings contribute to a nuanced understanding of the interplay between allostatic load and energy expenditure, a relationship that is evidently influenced by gender and biomarker variations.

References

1. Blundell, J., Caudwell, P., Gibbons, C., Hopkins, M., Näslund, E., King, N., & Finlayson, G. (2012). Role of resting metabolic rate and energy expenditure in hunger and appetite control: a new formulation. Disease Models & Mechanisms, 5(5), 608–613. https://doi.org/10.1242/dmm.009837

2. Bobba-Alves, N., Sturm, G., Lin, J., Ware, S. A., Karan, K. R., Monzel, A. S., Bris, C., Procaccio, V., Lenaers, G., Higgins-Chen, A., Levine, M. E., Horvath, S., Santhanam, B., Kaufman, B. A., Hirano, M., Epel, E. S., & Picard, M. (2023). Cellular Allostatic Load is linked to Increased Energy Expenditure and Accelerated Biological Aging. Psychoneuroendocrinology, 106322. https://doi.org/10.1016/j.psyneuen.2023.106322

3. Home | National Institute on Aging: Baltimore Longitudinal Study of Aging. (n.d.). https://www.blsa.nih.gov/

4. Miller, M. E., Davis, C. S., & Landis, J. R. (1993). The Analysis of Longitudinal Polytomous Data: Generalized Estimating Equations and Connections with Weighted Least Squares. Biometrics, 49(4), 1033. https://doi.org/10.2307/2532245

5. Top-ranked hospital in the nation. (n.d.). Mayo Clinic. https://www.mayoclinic.org/

6. Score-project by JWiley. (n.d.). https://score-project.org/#collapseOne

7. Wiley, J. F., Gruenewald, T. L., Karlamangla, A. S., & Seeman, T. E. (2016). Modeling multisystem physiological dysregulation. Psychosomatic Medicine, 78(3), 290–301. https://doi.org/10.1097/psy.0000000000000288

8. Wiley, J. F., & Carrington, M. (2016). A metabolic syndrome severity score: A tool to quantify cardiometabolic risk factors. Preventive Medicine, 88, 189–195. https://doi.org/10.1016/j.ypmed.2016.04.006

9. Zunzer, S. C., Von Duvillard, S. P., Tschakert, G., Mangus, B. C., & Hofmann, P. (2013). Energy expenditure and sex differences of golf playing. Journal of Sports Sciences, 31(10), 1045–1053. https://doi.org/10.1080/02640414.2013.764465

Appendix

Code

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
library(tidyverse)
library(readr)
library(readxl)
library(pscore) # Compute the allostatic load
library(table1)
library(knitr)
library(patchwork)
library(kableExtra)
library(geepack) # For marginal model
library(nlme) # Linear mixed effects model
# Data import
blsa = read_csv("./Data/blsa_data_v1.csv")
# Clean the data:
dat = blsa %>%
  ## Convert the negative values as NA
  mutate(across(where(is.numeric), ~if_else(. < 0, NA_real_, .))) %>%
  select(-c(BLSA_Race, educ_years, skin_fold, SF01, contains('DEMO'), contains('SMK'),
            glucose120_adjusted, sex, lean_mass7, CLGLUCOSE, CLURINE24HRCORTISOL,
            CLTOTALT3, CLTOTALT4, CLFREET3, CLFREET4,
            contains("CLIL"), CLSCORTISOL, CLTOTT, RMR02_TF, contains("CLSTNF_R"),
            CLIGF1, CLHOMOCYS, CLESTRAD,
            CLDHEAS, CLDDIMER, CLBIOT, CLAMYLASE, CLGhrelin, CLTNFA, CLRESISTIN,
            CLFIBRINOGEN, CLPAI1, visit, dov, WtKg, HtCm, Waist)) %>%
  ## Convert the variable names based on the variable files
  rename(id = idno, race = RACECD, edu_level = educat, sex = gender,
         hba1c = CLHBA1C, uric_acid = CLURICACID, albumin = CLALBUMIN, creatinine = CLCREAT,
         vb12 = CLVITB12, sodium = CLSODIUM, tsh = CLTSH, cholesterol = CLCHOLESTEROL,
         hdl = CLHDL, ldl = CLLDL, fast_glucose = glucose000_adjusted,
         triglycerides = CLTRIGLYCERIDE, leptin = CLLEPTIN, crp = CLCRP, adiponectin = CLADIPONECTIN) %
  ## Tidy the categories of Race:
  mutate(race = case_when(
    race == 1 ~ 0,
    race == 2 ~ 1,
    race == 3 ~ 2,
    race %in% c(4, 5, 8) ~ 3,
    race %in% c(0, 7) ~ 4)) %>%
  ## Change the variables' type:
  mutate(id = factor(id), race = factor(race), sex = factor(sex), edu_level = factor(edu_level)) %>%
  ## Calculate the response variable 'RMRmean':
  drop_na(contains("RMR")) %>%
  mutate(HEE = (RMRT15EEM*30 + RMRT15EEM*30 + RMRT16EEM*30 + RMRT17EEM*30 + RMRT18EEM*30 + RMRT19EEM*30
  select(-contains("RMRT")) %>%
  ## Reorder the columns:
  relocate(id, age,race, sex, edu_level, HEE, BMI,
           albumin, adiponectin, cholesterol, crp, creatinine, sbp, dbp, fast_glucose,
           hdl, ldl, hba1c, sodium, triglycerides, tsh, leptin, uric_acid, vb12) %>%
  drop_na(everything()) %>%
  ## Delete some cluster who only shows once:
  group_by(id) %>%
  add_count() %>%  ### Adds a count of each 'id'
```

```r
  filter(n %in% c(2, 3)) %>%  ### Keeps only rows where 'id' appears more than once
  select(-n) %>%   ### Removes the count column
  ungroup()
# Create a baseline dataset:
dat_baseline = dat %>%
  group_by(id) %>%
  slice(1)

dat_baseline$sex <-
  factor(dat_baseline$sex, levels = c("M", "F"),
         labels = c("Male", "Female"))
dat_baseline$race <-
  factor(dat_baseline$race, levels = c(0, 1, 2, 3, 4),
         labels = c("White", "Black", "Native India/Alaska", "Asian", "Other races"))
dat_baseline$edu_level <-
  factor(dat_baseline$edu_level, levels = c(0, 1, 2, 3, 4),
         labels = c("No education", "Elementary school", "High school", "Bachelor Degree", "Other Degre

label(dat_baseline$age) = "Age"
label(dat_baseline$edu_level) = "Education level (5 groups)"
label(dat_baseline$HEE) = "Energy Expenditure (kcal/hour)"
label(dat_baseline$race) = "Race (5 groups)"
label(dat_baseline$adiponectin) = "Adiponectin (ng/mL)"
label(dat_baseline$albumin) = "Albumin (g/dL)"
label(dat_baseline$cholesterol) = "Cholesterol (mg/dL)"
label(dat_baseline$crp) = "crp (ug/mL)"
label(dat_baseline$creatinine) = "Serium creatinine (mg/dL)"
label(dat_baseline$sbp) = "Systolic pressure (mmHg)"
label(dat_baseline$dbp) = "Diastolic pressure (mmHg)"
label(dat_baseline$fast_glucose) = "Fasting glucose (mg/dL)"
label(dat_baseline$hdl) = "HDL (mg/dL)"
label(dat_baseline$ldl) = "LDL (mg/dL)"
label(dat_baseline$hba1c) = "HBA1C (%)"
label(dat_baseline$sodium) = "Sodium (mmol/L)"
label(dat_baseline$triglycerides) = "Triglycerides (mg/dL)"
label(dat_baseline$tsh) = "TSH"
label(dat_baseline$leptin) = "leptin (ng/mL)"
label(dat_baseline$uric_acid) = "Uric acid (mg/dL)"
label(dat_baseline$vb12) = "Vitamin B12 (pg/mL)"

foot = "Table 1.1: Summary Statistics of Variables at Baseline Separated by Sex"
foot2 = "Table 1.2: Table 1.1 Cont'd"

# table1(~ HEE + age + race + edu_level + BMI + adiponectin + albumin + cholesterol + crp +
#             creatinine | sex,
#             data = dat_baseline, footnote = foot)
# table1(~ sbp + dbp + fast_glucose + hdl + ldl + hba1c + sodium + triglycerides +
#             tsh + leptin + uric_acid + vb12 | sex,
#             data = dat_baseline, footnote = foot2)
conti_dat =
  dat %>%
  select(-race, -sex, -edu_level) %>%
  group_by(id) %>%
```

```
  summarise(across(everything(), mean, na.rm = TRUE))
## Transfer to long format:
conti_dat_long = conti_dat %>%
  pivot_longer(col = -id)
## Draw the histograms:
conti_hist = ggplot(data = conti_dat_long, aes(x = value)) +
  geom_histogram(aes(fill = name), bins = 30) +
  facet_wrap(~name, scales = "free") +
  labs(x = "Value", y = "Count",
       title = "Figure 1.1: Histograms of Averaged Continuous Variable's Value For Each Individual") +
  theme_bw() +
  theme(legend.position = "none")
conti_hist
cat_dat =
  dat_baseline %>%
  select(id, race, sex, edu_level)
## Transfer to long format:
cat_dat_long = cat_dat %>%
  pivot_longer(col = -id)
# Plot the categorical variable barplots
cat_barplot = ggplot(cat_dat_long, aes(x = value, fill = value)) +
  geom_bar() +
  geom_text(stat = 'count', aes(label = ..count..), vjust = 0.4) +
  facet_wrap(~name, scales = "free") +
  labs(x = "Category", y = "Count", fill = "Category",
       title = "Figure 1.2: Bar Charts of Categorical variable For Each Individual") +
  theme_bw() +
  theme(legend.position = "none") +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1))
cat_barplot
dat_clean = dat

st_5_up <- function(x) {
  mean(x, na.rm = TRUE) + sd(x, na.rm = TRUE)*5
}

st_5_low <- function(x){
  mean(x, na.rm = TRUE) - sd(x, na.rm = TRUE)*5
}

st_5_upper <- lapply(dat_clean[, -c(1, 2, 3, 4, 5, 6)], st_5_up)
st_5_lower <- lapply(dat_clean[, -c(1, 2, 3, 4, 5, 6)], st_5_low)
outlier_name <- colnames(dat_clean[, (-c(1:6))])

for (i in outlier_name) {
  dat_clean[i][dat_clean[i] >= st_5_upper[i]] <- NA
  dat_clean[i][dat_clean[i] <= st_5_lower[i]] <- NA
}
cut_table =   tibble(
  "Biomarker" = factor(c("Adiponectin", "Fasting Glucose", "Albumin", "BMI", "Cholesterol",
                         "Systolic pressure"," Diastolic pressure", "Creatinine", "CRP", "HBA1C (%)",
                         "HDL","LDL", "leptin", "sodium", "Triglycerides","TSH","uric acid","VIT 12")),
  "Male Normal Range" =    c("change based on BMI (ug/mL)", "70 - 100 (mg/dl)", "3.4 - 5.4 (g/dL)",
```

```
                              "18.5 - 30","0 - 200 (mg/dL)","90 - 120 (mmHg)","60 - 80 (mmHg)",
                              "0.74 - 1.35 (mg/dL)","0 - 10 (ug/mL)","4.1 - 5.7 (%)",">= 60 (mg/dL)",
                              "100 - 129 (mg/dL)","change based on BMI (ng/mL)", "135 - 145 (mmol/L)","0
                              "0.3 - 4.2","3.7 - 8 (mg/dL)", "180 - 914 (pg/mL)"),
  "Female Normal Range" =  c("change based on BMI (ug/mL)","70 - 100 (mg/dl)","3.4 - 5.4 (g/dL)","18.5 -
                              "0 - 200 (mg/dL)","90 - 120 (mmHg)","60 - 80 (mmHg)","0.59 - 1.04 (mg/dL)"
                              "4.1 - 5.7 (%)",">= 60 (mg/dL)","100 - 129 (mg/dL)","change based on BMI (n
                              "135 - 145 (mmol/L)", "0 - 150 (mg/dL)","0.3 - 4.2","2.7 - 6.1 (mg/dL)","18
kable(cut_table, booktabs = T, caption = "Mayo Clinic Biomarker Clinical Cutoffs") %>%
  kable_styling(latex_options = c("HOLD_position"), font_size = 6)
# Function for clinical cutoffs
cutoff <- function(df, var_name, var_high = c(Inf, Inf), var_low = c(0, 0)) {
  var_name = enquo(var_name)
  df %>%
    mutate("{{var_name}}_cutoff" := if_else(is.na(!!var_name), 'NA',
                                            if_else(sex == "M" & !!var_name > var_high[1] |
                                                      sex == "M" & !!var_name < var_low[1] |
                                                      sex == "F" & !!var_name > var_high[2] |
                                                      sex == "F" & !!var_name < var_low[2],
                                                    "abnormal",
                                                    "normal")))
}
# Customerized Function for adiponectin and leptin:
adiponectin_cutoff <- function(df, var_name) {
  var_name = enquo(var_name)
  df %>%
    mutate("{{var_name}}_cutoff" := if_else(is.na(!!var_name), 'NA',
                                      if_else(sex == "M" & BMI < 25 & !!var_name > 26 |
                                              sex == "M" & BMI < 25 & !!var_name < 4 |
                                              sex == "F" & BMI < 25 & !!var_name > 37 |
                                              sex == "F" & BMI < 25 & !!var_name < 5 |
                                              sex == "M" & BMI == (25:30) & !!var_name > 20 |
                                              sex == "M" & BMI == (25:30) & !!var_name < 4 |
                                              sex == "F" & BMI == (25:30) & !!var_name > 28 |
                                              sex == "F" & BMI == (25:30) & !!var_name < 5 |
                                              sex == "M" & BMI > 30 & !!var_name > 20 |
                                              sex == "M" & BMI > 30 & !!var_name < 2 |
                                              sex == "F" & BMI > 30 & !!var_name > 22 |
                                              sex == "F" & BMI > 30 & !!var_name < 4,
                                              "abnormal", "normal")))
}
leptin_cutoff <- function(df, var_name) {
  var_name = enquo(var_name) # Makes it so I can work with the var_name more easily
  df %>%
    mutate("{{var_name}}_cutoff" := if_else(is.na(!!var_name), 'NA',
                                      if_else(sex == "M" & BMI == 11 & !!var_name > 0.4 |
                                              sex == "M" & BMI == 11 & !!var_name < 0.1 |
                                              sex == "F" & BMI == 11 & !!var_name > 3.6 |
                                              sex == "F" & BMI == 11 & !!var_name < 0.7 |
                                              sex == "M" & BMI == 12 & !!var_name > 0.6 |
                                              sex == "M" & BMI == 12 & !!var_name < 0.1 |
                                              sex == "F" & BMI == 12 & !!var_name > 4.2 |
                                              sex == "F" & BMI == 12 & !!var_name < 0.8 |
```

```
                                    sex == "M" & BMI == 13 & !!var_name > 0.7 |
                                    sex == "M" & BMI == 13 & !!var_name < 0.1 |
                                    sex == "F" & BMI == 13 & !!var_name > 4.8 |
                                    sex == "F" & BMI == 13 & !!var_name < 0.9 |
                                    sex == "M" & BMI == 14 & !!var_name > 0.9 |
                                    sex == "M" & BMI == 14 & !!var_name < 0.1 |
                                    sex == "F" & BMI == 14 & !!var_name > 5.6 |
                                    sex == "F" & BMI == 14 & !!var_name < 1.0 |
                                    sex == "M" & BMI == 15 & !!var_name > 1.1 |
                                    sex == "M" & BMI == 15 & !!var_name < 0.1 |
                                    sex == "F" & BMI == 15 & !!var_name > 6.5 |
                                    sex == "F" & BMI == 15 & !!var_name < 1.2 |
                                    sex == "M" & BMI == 16 & !!var_name > 1.3 |
                                    sex == "M" & BMI == 16 & !!var_name < 0.2 |
                                    sex == "F" & BMI == 16 & !!var_name > 7.5 |
                                    sex == "F" & BMI == 16 & !!var_name < 1.4 |
                                    sex == "M" & BMI == 17 & !!var_name > 1.7 |
                                    sex == "M" & BMI == 17 & !!var_name < 0.2 |
                                    sex == "F" & BMI == 17 & !!var_name > 8.7 |
                                    sex == "F" & BMI == 17 & !!var_name < 0.6 |
                                    sex == "M" & BMI == 18 & !!var_name > 2.1 |
                                    sex == "M" & BMI == 18 & !!var_name < 0.2 |
                                    sex == "F" & BMI == 18 & !!var_name > 10 |
                                    sex == "F" & BMI == 18 & !!var_name < 1.8 |
                                    sex == "M" & BMI == 19 & !!var_name > 2.6 |
                                    sex == "M" & BMI == 19 & !!var_name < 0.3 |
                                    sex == "F" & BMI == 19 & !!var_name > 11.6 |
                                    sex == "F" & BMI == 19 & !!var_name < 2.1 |
                                    sex == "M" & BMI == 20 & !!var_name > 3.2 |
                                    sex == "M" & BMI == 20 & !!var_name < 0.4 |
                                    sex == "F" & BMI == 20 & !!var_name > 13.4 |
                                    sex == "F" & BMI == 20 & !!var_name < 2.4 |
                                    sex == "M" & BMI == 21 & !!var_name > 4 |
                                    sex == "M" & BMI == 21 & !!var_name < 0.4 |
                                    sex == "F" & BMI == 21 & !!var_name > 15.6 |
                                    sex == "F" & BMI == 21 & !!var_name < 2.8 |
                                    sex == "M" & BMI == 22 & !!var_name > 5.0 |
                                    sex == "M" & BMI == 22 & !!var_name < 0.5 |
                                    sex == "F" & BMI == 22 & !!var_name > 18 |
                                    sex == "F" & BMI == 22 & !!var_name < 3.3 |
                                    sex == "M" & BMI == 23 & !!var_name > 6.2 |
                                    sex == "M" & BMI == 23 & !!var_name < 0.8 |
                                    sex == "F" & BMI == 23 & !!var_name > 20.9 |
                                    sex == "F" & BMI == 23 & !!var_name < 3.8 |
                                    sex == "M" & BMI == 24 & !!var_name > 7.7 |
                                    sex == "M" & BMI == 24 & !!var_name < 0.9 |
                                    sex == "F" & BMI == 24 & !!var_name > 24.2 |
                                    sex == "F" & BMI == 24 & !!var_name < 4.4 |
                                    sex == "M" & BMI == 25 & !!var_name > 9.6 |
                                    sex == "M" & BMI == 25 & !!var_name < 1.1 |
                                    sex == "F" & BMI == 25 & !!var_name > 28 |
                                    sex == "F" & BMI == 25 & !!var_name < 5.1 |
                                    sex == "M" & BMI == 26 & !!var_name > 12 |
```

```
                                        sex == "M" & BMI == 26 & !!var_name < 1.3 |
                                        sex == "F" & BMI == 26 & !!var_name > 32.4 |
                                        sex == "F" & BMI == 26 & !!var_name < 5.9 |
                                        sex == "M" & BMI == 27 & !!var_name > 14.9 |
                                        sex == "M" & BMI == 27 & !!var_name < 1.6 |
                                        sex == "F" & BMI == 27 & !!var_name > 37.5 |
                                        sex == "F" & BMI == 27 & !!var_name < 6.8 |
                                        sex == "M" & BMI == 28 & !!var_name > 18.6 |
                                        sex == "M" & BMI == 28 & !!var_name < 2 |
                                        sex == "F" & BMI == 28 & !!var_name > 43.5 |
                                        sex == "F" & BMI == 28 & !!var_name < 7.9 |
                                        sex == "M" & BMI == 29 & !!var_name > 23.2 |
                                        sex == "M" & BMI == 29 & !!var_name < 2.5 |
                                        sex == "F" & BMI == 29 & !!var_name > 50.4 |
                                        sex == "F" & BMI == 29 & !!var_name < 9.1 |
                                        sex == "M" & BMI == 30 & !!var_name > 28.9 |
                                        sex == "M" & BMI == 30 & !!var_name < 3.2 |
                                        sex == "F" & BMI == 30 & !!var_name > 58.3 |
                                        sex == "F" & BMI == 30 & !!var_name < 10.6 |
                                        sex == "M" & BMI == 31 & !!var_name > 36 |
                                        sex == "M" & BMI == 31 & !!var_name < 3.9 |
                                        sex == "F" & BMI == 31 & !!var_name > 67.5 |
                                        sex == "F" & BMI == 31 & !!var_name < 12.2 |
                                        sex == "M" & BMI == 32 & !!var_name > 44.9 |
                                        sex == "M" & BMI == 32 & !!var_name < 4.9 |
                                        sex == "F" & BMI == 32 & !!var_name > 78.2 |
                                        sex == "F" & BMI == 32 & !!var_name < 14.1 |
                                        sex == "M" & BMI == 33 & !!var_name > 55.8 |
                                        sex == "M" & BMI == 33 & !!var_name < 6.1 |
                                        sex == "F" & BMI == 33 & !!var_name > 90.5 |
                                        sex == "F" & BMI == 33 & !!var_name < 16.4 |
                                        sex == "M" & BMI == 34 & !!var_name > 69.6 |
                                        sex == "M" & BMI == 34 & !!var_name < 7.6 |
                                        sex == "F" & BMI == 34 & !!var_name > 105 |
                                        sex == "F" & BMI == 34 & !!var_name < 19 |
                                        sex == "M" & BMI == 35 & !!var_name > 86.7 |
                                        sex == "M" & BMI == 35 & !!var_name < 9.5 |
                                        sex == "F" & BMI == 35 & !!var_name > 121 |
                                        sex == "F" & BMI == 35 & !!var_name < 22 |
                                        sex == "M" & BMI == 36 & !!var_name > 108 |
                                        sex == "M" & BMI == 36 & !!var_name < 11.8 |
                                        sex == "F" & BMI == 36 & !!var_name > 141 |
                                        sex == "F" & BMI == 36 & !!var_name < 25.4 |
                                        sex == "M" & BMI == 37 & !!var_name > 135 |
                                        sex == "M" & BMI == 37 & !!var_name < 14.8 |
                                        sex == "F" & BMI == 37 & !!var_name > Inf |
                                        sex == "F" & BMI == 37 & !!var_name < 0 ,
                                        "abnormal", "normal")))
}
dat_clean2 = dat_clean
## Adiponectin (ug/mL):
dat_clean2 = adiponectin_cutoff(dat_clean2, `adiponectin`)
## Fasting Glucose value (mg/dl)
```

```r
dat_clean2 = cutoff(dat_clean2, `fast_glucose`, var_high = c(100, 100), var_low = c(70, 70))
## ALBUMIN (g/dL)
dat_clean2 = cutoff(dat_clean2, `albumin`, var_high = c(5.4, 5.4), var_low = c(3.4, 3.4))
## BMI
dat_clean2 = cutoff(dat_clean2, `BMI`, var_high = c(30, 30), var_low = c(18.5, 18.5))
## CHOLESTEROL (mg/dL)
dat_clean2 = cutoff(dat_clean2, `cholesterol`, var_high = c(200, 200), var_low = c(0, 0))
## Composed systolic pressure (mmHg)
dat_clean2 = cutoff(dat_clean2, `sbp`, var_high = c(120, 120), var_low = c(90, 90))
## Composed Diastolic  pressure (mmHg)
dat_clean2 = cutoff(dat_clean2, `dbp`, var_high = c(80, 80), var_low = c(60, 60))
## CREATININE (mg/dL)
dat_clean2 = cutoff(dat_clean2, `creatinine`, var_high = c(1.35, 1.04), var_low = c(0.74, 0.59))
## CRP (ug/mL)
dat_clean2 = cutoff(dat_clean2, `crp`, var_high = c(10, 10), var_low = c(0, 0))
## HBA1C (%)
dat_clean2 = cutoff(dat_clean2, `hba1c`, var_high = c(5.7, 5.7), var_low = c(4.1, 4.1))
## HDL (mg/dL)
dat_clean2 = cutoff(dat_clean2, `hdl`, var_high = c(Inf, Inf), var_low = c(60, 60))
## LDL (mg/dL)
dat_clean2 = cutoff(dat_clean2, `ldl`, var_high = c(129, 129), var_low = c(100, 100))
## LEPTIN (ng/mL) based on the bmi
dat_clean2 = leptin_cutoff(dat_clean2, `leptin`)
## SODIUM (mmol/L)
dat_clean2 = cutoff(dat_clean2, `sodium`, var_high = c(145, 145), var_low = c(135, 135))
## TRIGLYCERIDES (mg/dL)
dat_clean2 = cutoff(dat_clean2, `triglycerides`, var_high = c(150, 150), var_low = c(0, 0))
## TSH
dat_clean2 = cutoff(dat_clean2, `tsh`, var_high = c(4.2, 4.2), var_low = c(0.3, 0.3))
## URIC ACID (mg/dL)
dat_clean2 = cutoff(dat_clean2, `uric_acid`, var_high = c(8, 6.1), var_low = c(3.7, 2.7))
## VIT B12 (pg/mL)
dat_clean2 = cutoff(dat_clean2, `vb12`, var_high = c(914, 914), var_low = c(180, 180))
dat_clean3 = dat_clean2 %>%
  rowwise() %>%
  mutate(abnormal_count = sum(c_across(contains('cutoff')) == "abnormal", na.rm = TRUE)) %>%
  mutate(nonmissing_count = sum(!is.na(c_across(BMI:vb12)))) %>%
  mutate(ALscore = (abnormal_count/nonmissing_count)*10) %>%
  select(id, age, race, sex, edu_level, HEE, BMI, ALscore)
dat_n = dat_clean3 %>%
  group_by(id) %>% summarize(n = n()) %>% filter(n < 6)
dat_final = inner_join(dat_clean3, dat_n, by = "id") %>%
  mutate(ALscore = round(ALscore, 2)) %>%
  filter(n == 3) %>%
  select(-n) %>%
  filter(id != 4850)
head(dat_final, 6)
al_dat =
  dat_final %>%
  select(ALscore, id) %>%
  group_by(id) %>%
  summarise(ALscore = mean(ALscore))
al_hist = al_dat %>%
```

```
  ggplot(aes(x = ALscore)) +
  geom_histogram(bins = 30, fill = "orange")
al_hist
sp = ggplot(data = dat_final, aes(x = ALscore, y = HEE, group = id)) +
       geom_line(aes(color = sex)) +
       geom_point(aes(color = sex), size = 0.5) +
       labs(x = "AL Score", y = "Hourly Energy Expenditure") +
       theme(legend.position = "none")
fl = ggplot(data = dat_final) +
     geom_smooth(aes(group = id, x = ALscore, y = HEE, color = sex),
                 method = "lm", se = FALSE, linetype = 2, size = 0.5) +
     geom_smooth(aes(x = ALscore, y = HEE),
                 method = 'lm', se = FALSE, color = 'black', size = 0.6) +
     labs(x = "AL Score", y = "Hourly Energy Expenditure", color = "Sex")
sp + fl +
  plot_annotation(title = "Individual Trajectory Plots Between AL Score and HEE",
  theme = theme(plot.title = element_text(size = 11)))
sp1 = ggplot(data = dat_final, aes(x = age, y = HEE, group = id)) +
       geom_line(aes(color = sex)) +
       geom_point(aes(color = sex), size = 0.5) +
       labs(x = "Age", y = "Hourly Energy Expenditure") +
       theme(legend.position = "none")
fl1 = ggplot(data = dat_final) +
     geom_smooth(aes(group = id, x = age, y = HEE, color = sex),
                 method = "lm", se = FALSE, linetype = 2, size = 0.5) +
     geom_smooth(aes(x = age, y = HEE),
                 method = 'lm', se = FALSE, color = 'black', size = 0.6) +
     labs(x = "Age", y = "Hourly Energy Expenditure", color = "Sex")
sp1 + fl1 +
  plot_annotation(title = "Individual Trajectory Plots Between Age and HEE",
  theme = theme(plot.title = element_text(size = 11)))
# Independent, homoskedastic errors
model1 <- glm(HEE ~ ALscore + age + sex,
              data = dat_final,
              family = gaussian)
# random intercepts + independent, homoskedastic errors
model2 <- lme(HEE ~ ALscore + age + sex,
              data = dat_final,
              random = reStruct(~1|id),
              method = "ML")
# random intercepts/slopes + independent, homoskedastic errors
model3 <- lme(HEE ~ ALscore + age + sex,
              data = dat_final,
              random = reStruct(~age|id),
              method = "ML")
# random intercepts + auto-regressive errors
model4 <- lme(HEE ~ ALscore + age  + sex,
              data = dat_final,
              random = reStruct(~1|id),
              correlation = corAR1(form = ~age|id),
              method = "ML")
# random intercepts + exponential spatial errors
model5 <- lme(HEE ~ ALscore + age + sex,
```

```
            data = dat_final,
            random = reStruct(~1|id),
            correlation = corExp(form = ~age|id),
            method = "ML")
# random intercepts + exponential spatial errors and independent, homoskedastic errors
model6 <- lme(HEE ~ ALscore + age + sex,
            data = dat_final,
            random = reStruct(~1|id),
            correlation = corExp(form = ~age|id, nugget = TRUE),
            method = "ML")
# random intercepts/slope + exponential spatial errors
model7 <- lme(HEE ~ ALscore + age + sex,
            data = dat_final,
            random = reStruct(~age|id),
            correlation = corExp(form = ~age|id),
            method = "ML")
# random intercepts/slope + exponential spatial errors and independent, homoskedastic errors
model8 <- lme(HEE ~ ALscore + age + sex,
            data = dat_final,
            random = reStruct(~age|id),
            correlation = corExp(form = ~age|id, nugget = TRUE),
            method = "ML")
model_list <- list(
  "Independent, homoskedastic errors" = model1,
  "Random intercepts + independent, homoskedastic errors" = model2,
  "Random intercepts/slopes + independent, homoskedastic errors" = model3,
  "Random intercepts + auto-regressive errors" = model4,
  "Random intercepts + exponential spatial errors" = model5,
  "Random intercepts + exponential spatial and independent, homoskedastic errors" = model6,
  "Random intercepts/slope + exponential spatial errors" = model7,
  "Random intercepts/slope + exponential spatial and independent, homoskedastic errors" = model8
)

model_summary <- data.frame(Model = character(), AIC = numeric(), 'Log-likelihood' = numeric())

for (i in 1:8) {
  model_name <- names(model_list)[i]
  aic <- AIC(model_list[[i]])
  log_likelihood <- logLik(model_list[[i]])[1]
  model_summary <- rbind(model_summary, data.frame(Model = model_name, AIC = aic, 'Log-likelihood' = log
}

kable(model_summary, format = "latex", caption = "LMM Model Comparison", booktabs = TRUE, digits = 4, es
  kable_paper("striped", full_width = F, latex_options = c("HOLD_position"), font_size = 8)
gee.i = geeglm(HEE ~ ALscore + age + sex, data = dat_final, id = id, family = gaussian, corstr = "indepe
gee.e = geeglm(HEE ~ ALscore + age + sex, data = dat_final, id = id, family = gaussian, corstr = "exchan
gee.ar1 = geeglm(HEE ~ ALscore + age + sex, data = dat_final, id = id, family = gaussian, corstr = "ar1
isum = summary(gee.i)
esum = summary(gee.e)
asum = summary(gee.ar1)

gee_table = tibble(
  Term = c("Intercept", "AL Score", "Age", "Sex",
```

```
           "Intercept", "AL Score", "Age", "Sex", "$\\rho$",
           "Intercept", "AL Score", "Age", "Sex", "$\\rho$"),
  Estimates = c(isum$coef$Estimate[1], isum$coef$Estimate[2], isum$coef$Estimate[3],
                isum$coef$Estimate[4],
                esum$coef$Estimate[1], esum$coef$Estimate[2], esum$coef$Estimate[3],
                esum$coef$Estimate[4], 0.278,
                asum$coef$Estimate[1], asum$coef$Estimate[2], asum$coef$Estimate[3],
                asum$coef$Estimate[4], 0.338),
  "Robust SE" = c(isum$coef$Std.err[1], isum$coef$Std.err[2], isum$coef$Std.err[3],
                  isum$coef$Std.err[4],
                  esum$coef$Std.err[1], esum$coef$Std.err[2], esum$coef$Std.err[3],
                  esum$coef$Std.err[4], 0.0452,
                  asum$coef$Std.err[1], asum$coef$Std.err[2], asum$coef$Std.err[3],
                  asum$coef$Std.err[4], 0.0489),
  Wald = c(isum$coef$Wald[1], isum$coef$Wald[2], isum$coef$Wald[3],
           isum$coef$Wald[4],
           esum$coef$Wald[1], esum$coef$Wald[2], esum$coef$Wald[3],
           esum$coef$Wald[4], NA,
           asum$coef$Wald[1], asum$coef$Wald[2], asum$coef$Wald[3],
           asum$coef$Wald[4], NA),
  "P-value" = c(isum$coef$"Pr(>|W|)"[1], isum$coef$"Pr(>|W|)"[2], isum$coef$"Pr(>|W|)"[3],
                isum$coef$"Pr(>|W|)"[4],
                esum$coef$"Pr(>|W|)"[1], esum$coef$"Pr(>|W|)"[2], esum$coef$"Pr(>|W|)"[3],
                esum$coef$"Pr(>|W|)"[4], NA,
                asum$coef$"Pr(>|W|)"[1], asum$coef$"Pr(>|W|)"[2], asum$coef$"Pr(>|W|)"[3],
                asum$coef$"Pr(>|W|)"[4], NA)
) %>%
  mutate(Wald = round(as.numeric(Wald), 4),
         `P-value` = round(as.numeric(`P-value`), 4)) %>%
  mutate_all(~ifelse(is.na(.), "", .))

kable(gee_table, format = "latex", caption = "Results of Complete-case GEE 1.5", booktabs = TRUE, digits
  kable_paper("striped", full_width = F, latex_options = c("HOLD_position")) %>%
  pack_rows("Working independence", 1, 4) %>%
  pack_rows("Working exchangeable", 5, 9) %>%
  pack_rows("Working AR-1", 10, 14)
# Extracting fixed effects
fixed_effects2 <- summary(model3)$tTable %>% round(5)

# Extracting random effects
random_effects2 <- VarCorr(model3) %>% as.data.frame()
random_effects2.2 = random_effects2[1:3,]

# Extracting lme1 statistics like AIC, BIC, log-likelihood
model_stats2 <- c(AIC = AIC(model3), BIC = BIC(model3), LogLik = logLik(model3)) %>% round(5)

table2 <- bind_rows(
  tibble(Term = rownames(fixed_effects2), Estimate = fixed_effects2[, "Value"],
         StdError = fixed_effects2[, "Std.Error"], DF = fixed_effects2[, "DF"],
         "t-value" = fixed_effects2[, "t-value"], "p-value" = fixed_effects2[, "p-value"]),
  tibble(Term = rownames(random_effects2.2), StdDev = random_effects2.2[, "StdDev"], Corr = random_effe
  tibble(Term = names(model_stats2), Value = model_stats2)
) %>%
```

```
  mutate_all(~ifelse(is.na(.), "", .)) %>%
  mutate(Term = ifelse(Term == "ALscore", "AL Score", Term)) %>%
  mutate(Term = ifelse(Term == 'age', "Age", Term)) %>%
  mutate(Term = ifelse(Term == 'sexM', "Sex = Male", Term))
# Kable our table
kable(table2, format = "latex", caption = "Results of Random intercepts And Random Slopes Model", bookta
  kable_paper("striped", full_width = F, latex_options = c("HOLD_position")) %>%
  pack_rows("Fixed Effects", 1, 4) %>%
  pack_rows("Random Effects", 5, 7) %>%
  pack_rows("Model Statistics", 8, 10)
table1(~ HEE + age + race + edu_level + BMI + adiponectin + albumin + cholesterol + crp +
            creatinine | sex,
            data = dat_baseline, footnote = foot)
table1(~ sbp + dbp + fast_glucose + hdl + ldl + hba1c + sodium + triglycerides +
            tsh + leptin + uric_acid + vb12 | sex,
            data = dat_baseline, footnote = foot2)
```

Table

	Male	Female	Overall
	(N=232)	(N=244)	(N=476)
Energy Expenditure (kcal/hour)			
Mean (SD)	97.0 (28.0)	75.3 (23.3)	85.9 (27.9)
Median [Min, Max]	90.1 [48.5, 195]	71.3 [35.3, 185]	80.0 [35.3, 195]
Age			
Mean (SD)	66.1 (15.1)	65.9 (13.7)	66.0 (14.4)
Median [Min, Max]	70.0 [26.0, 95.0]	67.0 [31.0, 95.0]	68.0 [26.0, 95.0]
Race (5 groups)			
White	162 (69.8%)	154 (63.1%)	316 (66.4%)
Black	53 (22.8%)	74 (30.3%)	127 (26.7%)
Native India/Alaska	2 (0.9%)	0 (0%)	2 (0.4%)
Asian	2 (0.9%)	8 (3.3%)	10 (2.1%)
Other races	13 (5.6%)	8 (3.3%)	21 (4.4%)
Education level (5 groups)			
No education	2 (0.9%)	0 (0%)	2 (0.4%)
Elementary school	6 (2.6%)	16 (6.6%)	22 (4.6%)
High school	25 (10.8%)	29 (11.9%)	54 (11.3%)
Bachelor Degree	45 (19.4%)	52 (21.3%)	97 (20.4%)
Other Degree	154 (66.4%)	147 (60.2%)	301 (63.2%)
BMI			
Mean (SD)	27.3 (3.93)	26.7 (4.91)	27.0 (4.47)
Median [Min, Max]	26.9 [17.9, 39.7]	25.7 [18.5, 49.3]	26.4 [17.9, 49.3]
Adiponectin (ng/mL)			
Mean (SD)	11.7 (9.10)	16.5 (12.0)	14.2 (10.9)
Median [Min, Max]	8.86 [1.17, 58.3]	13.2 [0.780, 65.6]	11.3 [0.780, 65.6]
Albumin (g/dL)			
Mean (SD)	4.13 (0.361)	4.00 (0.408)	4.07 (0.391)
Median [Min, Max]	4.20 [2.90, 5.00]	4.00 [3.00, 5.50]	4.10 [2.90, 5.50]
Cholesterol (mg/dL)			
Mean (SD)	179 (39.3)	199 (34.7)	189 (38.3)
Median [Min, Max]	174 [83.0, 330]	195 [126, 341]	186 [83.0, 341]
crp (ug/mL)			
Mean (SD)	2.35 (3.88)	3.35 (6.43)	2.86 (5.36)
Median [Min, Max]	1.15 [0, 27.3]	1.37 [0, 64.4]	1.23 [0, 64.4]
Serium creatinine (mg/dL)			
Mean (SD)	1.06 (0.210)	0.846 (0.228)	0.951 (0.244)
Median [Min, Max]	1.01 [0.660, 2.10]	0.800 [0.500, 3.40]	0.900 [0.500, 3.40]

Table 1.1: Summary Statistics of Variables at Baseline Separated by Sex

	Male	Female	Overall
	(N=232)	(N=244)	(N=476)
Systolic pressure (mmHg)			
Mean (SD)	118 (13.7)	113 (14.2)	115 (14.1)
Median [Min, Max]	117 [81.0, 180]	113 [80.0, 177]	115 [80.0, 180]
Diastolic pressure (mmHg)			
Mean (SD)	67.5 (8.77)	64.5 (8.83)	66.0 (8.92)
Median [Min, Max]	68.0 [43.0, 97.0]	63.0 [40.0, 89.0]	65.0 [40.0, 97.0]
Fasting glucose (mg/dL)			
Mean (SD)	98.2 (14.7)	95.0 (12.6)	96.5 (13.8)
Median [Min, Max]	96.0 [58.0, 228]	95.0 [42.0, 168]	95.0 [42.0, 228]
HDL (mg/dL)			
Mean (SD)	53.7 (14.1)	66.0 (16.7)	60.0 (16.7)
Median [Min, Max]	52.0 [25.0, 108]	64.0 [33.0, 132]	58.0 [25.0, 132]
LDL (mg/dL)			
Mean (SD)	106 (34.5)	114 (32.7)	110 (33.8)
Median [Min, Max]	103 [30.0, 232]	109 [38.0, 245]	107 [30.0, 245]
HBA1C (%)			
Mean (SD)	5.81 (0.674)	5.78 (0.441)	5.79 (0.566)
Median [Min, Max]	5.70 [4.30, 10.1]	5.70 [4.80, 8.10]	5.70 [4.30, 10.1]
Sodium (mmol/L)			
Mean (SD)	141 (2.40)	141 (2.50)	141 (2.45)
Median [Min, Max]	141 [133, 148]	141 [122, 146]	141 [122, 148]
Triglycerides (mg/dL)			
Mean (SD)	99.0 (53.6)	96.8 (43.5)	97.9 (48.7)
Median [Min, Max]	85.0 [20.0, 427]	87.5 [32.0, 267]	86.0 [20.0, 427]
TSH			
Mean (SD)	2.36 (1.17)	2.66 (2.47)	2.52 (1.96)
Median [Min, Max]	2.14 [0.530, 6.92]	2.29 [0.0560, 34.3]	2.23 [0.0560, 34.3]
leptin (ng/mL)			
Mean (SD)	11.6 (12.0)	26.3 (21.5)	19.1 (19.0)
Median [Min, Max]	7.50 [0.500, 105]	19.5 [1.20, 130]	13.2 [0.500, 130]
Uric acid (mg/dL)			
Mean (SD)	5.54 (1.15)	4.52 (1.11)	5.02 (1.24)
Median [Min, Max]	5.60 [1.50, 9.70]	4.40 [1.70, 7.90]	5.00 [1.50, 9.70]
Vitamin B12 (pg/mL)			
Mean (SD)	579 (304)	691 (447)	637 (388)
Median [Min, Max]	496 [107, 2210]	577 [190, 4420]	548 [107, 4420]

Table 1.2: Table 1.1 Cont'd