Application of novel technologies to cardiovascular prevention

HARVARD UNIVERSITY

Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Committee on Higher Degrees in Population Health Sciences,
have examined a dissertation, entitled,

"Application of Novel Methodologies to Cardiovascular Prevention Research"

presented by

RIKUTA HAMAYA

candidate for the degree of Doctor of Philosophy
and hereby certify that it is worthy of acceptance.

Eric Rimm (Dec 28, 2023 10:44 EST)

Dr. Eric B. Rimm, Sc.D., Committee Chair,
Harvard Medical School, Harvard T.H. Chan School of Public Health

Nancy R. Cook
Nancy R. Cook (Dec 28, 2023 11:15 PST)

Dr. Nancy R. Cook, Sc.D.,
Harvard Medical School, Harvard T.H. Chan School of Public Health

JoAnn Manson (Jan 1, 2024 17:22 EST)

Dr. JoAnn E. Manson, M.D., Dr.P.H.,
Harvard Medical School, Harvard T.H. Chan School of Public Health

Date: 08 December 2023

Application of novel technologies to cardiovascular prevention

A dissertation presented

by

Rikuta Hamaya

to

The Department of Population Health Sciences

In the Epidemiology Area of Study

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Population Health Sciences

Harvard University

Cambridge, Massachusetts

December 2023

Dissertation Advisor:
Eric B. Rimm

Author:
Rikuta Hamaya

Application of novel technologies to cardiovascular prevention

ABSTRACT

This dissertation aimed to employ advanced methodologies in machine learning (ML), causal inference, and metabolomics in epidemiological research to contribute to the field of CVD prevention. First, quantifying sodium intake has been a persistent challenge in epidemiological research due to the measurement errors, limiting the explorations of sodium exposure in relation to disease outcomes, such as early-onset hypertension. Chapter 1 describes the development of ML algorithms that predict dietary sodium intake—measured via multiple 24-hour urinary sodium excretion—based on questionnaire data in the subcohorts of the Nurses' Health Study (NHS), NHS II, and Health Professional Follow-up Study (HPFS). These algorithms well predicted absolute sodium intake, albeit with a modest improvement in mitigating measurement error bias. In Chapter 2, using the developed algorithms, we evaluated the population-attributable risks (PARs) of modifiable lifestyle factors concerning the incidence of overall and early-onset hypertension across the full cohorts of NHS, NHS II, and HPFS. The findings suggest that maintaining healthy weight may substantially lower the risk of hypertension, and that adopting a combination of lifestyle modifications could further reduce the risk across all ages, particularly stronger in younger populations. Second, studies exploring heterogeneous treatment effect (HTE) gain traction in epidemiology, potentially offering advancements in precision medicine. Chapter

3 delineates the methodological frameworks for HTE research using a single randomized controlled trial. In an application example using the Preventing Overweight Using Novel Strategies (POUNDS LOST) trial, we developed and validated an ITR to optimize the efficacy of high- and low-fat diet interventions for weight loss over two years. Third, Chapter 4 delves into the development of a novel biomarker identified through metabolomics technology. We investigated the associations between circulating branched-chain amino acids (BCAAs)—metabolites indicative of diabetes risk—and established cardiometabolic biomarkers in the Women's Health Study, thus characterizing BCAAs as biomarkers representing CVD risk independent of glucose metabolism.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

To my beloved wife, Momoko

and my cherished children, Yu, Syuta, and Riko

whose presence has filled my life with immeasurable joy and happiness

ACKNOWLEDGEMENTS

I am profoundly thankful to my advisors Drs. Eric Rimm, JoAnn Manson, and Nancy Cook, for their exceptional support and guidance, which extended far beyond the scope of this dissertation. Dr. Rimm has been not only a primary mentor but also a valued friend, offering genuine care and encouragement that helped me gain confidence and discover my strengths. Dr. Manson introduced me to the fascinating world of cardiovascular epidemiology, fostering a nurturing and supportive environment for my learning and growth. Dr. Cook has been indispensable for her expertise in statistics, playing a pivotal role in both this dissertation and other projects. During a particularly challenging period when my daughter was battling cancer, their collective support was a beacon of strength, helping me navigate through those difficult times. I am eternally grateful for their mentorship, inspiration, and the myriad opportunities they have provided.

I would also like to express my deep appreciation to Drs. Molin Wang, Walter Willett, Stefania Papatheodorou, Yuan Ma, Deirdre Tobias, and I-Min Lee for their dedication, interest, and unwavering support in a multitude of projects. Their mentorship has been a cornerstone of my PhD journey. They have not only contributed significantly to the depth and breadth of my learning but also made the entire process immensely enriching and enjoyable. Their wisdom and encouragement have been pivotal in shaping both my professional development and personal growth during this pivotal phase of my academic career.

My gratitude extends to the program administrators Eric DiGiovanni, Bruce Villineau, and Matthew Boccuzzi, and to my friends and colleagues at Harvard, who have been instrumental in guiding me through the past three years of the program.

Lastly, my heartfelt thanks go to my dearest family. The unwavering support and encouragement from my beloved wife and children have made my life. To my parents – thank you for everything. Your love and guidance have shaped who I am today. I am also thankful to my younger brother, cousins, and grandparents, who have always stood by my side throughout my life. I love you all.

Introduction

Cardiovascular disease (CVD) remains a paramount global health concern issue.[1] Advancing the primary prevention strategies is challenging due to the complexity of its determinants and inconsistent efficacy of preventive interventions across diverse populations. Concurrently, advancements in methods and technologies for CVD prevention are creating opportunities for novel scientific insights that could potentially develop and optimize preventive strategies. The application of these novel technologies necessitate a multidisciplinary synthesis of knowledge from fields including epidemiology, nutrition, computer science, economics, and biology. This dissertation represents collaborative efforts by specialists across these domains to contribute to CVD prevention through the application of state-of-the-art machine learning (ML), causal inference, and metabolomics methodologies. It consists of three primary objectives:

1) *Application of ML algorithms to address unresolved issues in epidemiological research*

The accurate quantification of dietary sodium intake has been a persistent challenge in epidemiological research due to the measurement error that hinder the assessment of sodium's impact on disease outcomes.[2,3] ML algorithms have potential to capture complex interactions between predictors to improve the prediction accuracy. The first project aimed to employ ML algorithms to enhance the prediction of dietary sodium intake, as measured by multiple 24-hour

urinary sodium excretion, using questionnaire data in the subcohorts of the Nurses' Health Study (NHS), NHS II, and Health Professional Follow-up Study (HPFS) (**Chapter 1**).

Evidence is lacking on the relative contributions of specific lifestyle factors on the prevention of hypertension, in particular early-onset hypertension.[4,5] While early-onset hypertension is often deemed highly heritable, less has been characterized with respect to the role of lifestyle modification.[5,6] Therefore, the second project aimed to determine the population-attributable risks of modifiable lifestyle factors in relation to incident overall and early-onset hypertension, within three large US cohorts spanning ages 27 to 75, across 27-31 years of follow-ups (**Chapter 2**). We leveraged ML algorithms to refine dietary sodium intake estimates to reduce the measurement error.

2) Application of novel causal inference methods in randomized controlled trials (RCTs)

Moving beyond the conventional 'one-fits-for-all' approaches, the exploration of heterogeneous treatment effect (HTE) in medical and epidemiological studies is gaining momentum.[7–9] Investigating HTE may provide pivotal evidence for precision medicine and individualized prevention strategies. However, the focuses of current application studies are often not clear. The third project sought to delineate the research concepts and methodological approaches (**Chapter 3**). Specifically, we defined two major study aims of HTE studies: the exploration of sources of effect heterogeneity and the development and validation of individualized treatment rules (ITRs).[10,11] The study further illustrates the application of ITRs within the Preventing Overweight Using Novel Strategies (POUNDS LOST) trial.[12]

3) Exploration of metabolomics data for novel biomarker identification and establishment

Metabolomics technologies have led to a discovery of various new biomarkers that can be potentially useful in clinical settings.[13,14] Among these, circulating branched-chain amino acids (BCAAs) have emerged as predictive biomarkers for the future risk of type 2 diabetes and CVD.[15,16] The forth project sought to elaborate on the role of BCAAs as the risk of CVD, independently of glucose metabolism investigating their relationship with established cardiometabolic biomarkers in the Women's Health Study (**Chapter 4**).

Each aim of this dissertation may not only contribute uniquely to the corpus of knowledge on CVD prevention but also promote the integration of innovative technologies and analytical methods into future medical and epidemiological studies. I hope this work will contribute incrementally towards ameliorating the global burden of CVD.

Chapter 1

Prediction of 24-hour urinary sodium excretion using machine-learning algorithms

Author: Rikuta Hamaya, MD, MSc[1,2], Molin Wang, PhD[1,3,4], Stephen P. Juraschek, MD, PhD[5], Kenneth J. Mukamal, MD, MPH[5], JoAnn E. Manson, MD, DrPH[1,2,6], Deirdre K. Tobias, ScD[2,7], Qi Sun, MD, ScD[7,8], Gary C. Curhan, MD, ScD[4,9], Walter C. Willett, MD, DrPH[1, 4,7], Eric B. Rimm, ScD[1,4,7*], and Nancy R. Cook, ScD[1,2*].

[*]These two last authors contributed equally

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[2] Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[3] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

[4] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[5] Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

[6] Mary Horrigan Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[7] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

[8] Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA

[9] Renal Division, Department of Medicine, Brigham and Women's Hospital, Boston, MA

1.1 Abstract

Background: Accurate quantification of sodium intake based on self-reported dietary assessments has been a persistent challenge. We aimed to apply and validate machine-learning (ML) algorithms to predict 24-hour urinary sodium excretion from self-reported questionnaire information.

Methods: We analyzed 3,454 participants from the Nurses' Health Study (NHS), the Nurses' Health Study II (NHS-II), and the Health Professionals Follow-up Study (HPFS), with repeated measures of 24-hour urinary sodium excretion over one year. We used an ensemble approach to predict averaged 24-hour urinary sodium excretion using 36 characteristics. The agreement and calibration of the ML-predicted sodium were tested internally and externally using the Trial of Hypertension Prevention I (TOHP-I). In addition, doubly-labelled water (DLW)-derived energy expenditure-adjusted 24-hour sodium excretion was predicted in 1,085 participants using the same ML approaches. The final ML algorithms were applied to 167,920 non-hypertensive adults with repeated assessments of diet over 30 years of follow-up. Confounder-adjusted Cox proportional hazard models were fit to estimate the hazard ratio [HR] of incident hypertension for predicted sodium in quintiles.

Results: Averaged 24-hour urinary sodium excretion was better predicted and calibrated with ML compared with FFQ, and the findings were externally validated (Spearman correlation coefficient [95% CI]: 0.51 [0.49, 0.54] in ML; 0.19 [0.16, 0.23] in FFQ; 0.46 [0.42, 0.50] in TOHP-I). However, prediction of DLW-based energy-adjusted 24-hour sodium excretion was only modestly better with ML (0.37 [0.31, 0.42] in ML; 0.32 [0.26, 0.37] in FFQ). ML-predicted sodium excretion was modestly more strongly associated than FFQ-based sodium intake in the

NHS-II (HR [95% CI] comparing Q5 vs. Q1: 1.48 [1.40, 1.56] in ML; 1.04 [0.99, 1.08] in FFQ) but no material differences were observed in NHS or HPFS.

Conclusions: The present ML algorithm improved prediction of participants' averaged absolute 24-hour urinary sodium excretion. However, it heavily depended on body size, and prediction of energy-adjusted 24-hour sodium excretion was only modestly better using ML compared with the FFQ. The present algorithms may prove to be a generalizable approach for predicting absolute sodium intake but do not yet substantially reduce the bias stemming from measurement error in disease associations.

1.2 Introduction

Self-reported dietary intake is essential for exposure measurement in nutritional epidemiological research, yet error remains a methodological concern[1]. In particular, dietary sodium varies substantially between foods and can be difficult to quantify from traditional dietary assessment instruments, especially due to its use in food processing or at the table. The influence of random and systematic measurement error in quantitating sodium intake is often high and can interfere with the ability to assess sodium as an exposure in epidemiological studies of chronic disease outcomes.[2,20,21] Although evidence from randomized controlled trials (RCTs) demonstrates sodium intake plays a causal role in development of hypertension, observational studies with longer follow-up periods can provide additional insights, including contributions of other modifiable lifestyles factors to incident hypertension[22,23] or other chronic disease endpoints. Given the high prevalence of inadequately controlled blood pressure (BP) globally, accurately measuring dietary sodium intake is of particular importance in reducing the global

burden of cardiovascular disease (CVD).[24,25] Sodium excretion based on multiple 24-hour urine samples is considered the gold standard method for assessing sodium intake[26], but is prohibitively expensive in large epidemiological cohorts and burdensome for participants, highlighting the need for a more cost-effective strategy for accurate dietary sodium assessment.

Regression calibration methods have been proposed to address the issue of measurement error. Widely recognized techniques include Rosner's method[27] and Carroll's method[3,17], which are based on a simple linear or logistic regression in an accompanying validation study. Huang et al. proposed a use of linear regression with subject characteristics data for the prediction of exposures with measurement error[2], but non-linear associations and complex interactions of covariates are not captured by simple linear regression, and their application was prone to overfitting with many covariates. In addition, this initial study was limited by sample size, a homogeneous population of post-menopausal women, and a lack of internal and external validation.[2] These limitations hinder the wider application of the previously proposed calibration or prediction methods for measurement error-correction purpose in epidemiological research.

Therefore, we sought to provide a more reliable and generalizable method for accurately predicting sodium intake. We applied machine-learning (ML) algorithms to various questionnaire-based items including food frequency questionnaire (FFQ)-based dietary intake, demographics, and behavioral patterns, examining whether the ML-predicted true sodium exposure approximates the expected value of the true exposure given a large number of covariates in the prediction model. We then examined the associations of FFQ-based sodium intake and the ML-based predicted sodium intake with incident hypertension. We hypothesized

that the ML algorithm would more effectively capture the between-person variation of 24-hour urinary sodium excretion and reduce the measurement error bias in the disease associations.

1.3 Methods

Population

We analyzed participants from the Nurses' Health Study (NHS), the Nurses' Health Study 2 (NHS-II), and the Health Professionals Follow-up Study (HPFS) (original cohort sizes are 121,700, 116,429, and 51,529, respectively).[28] We included a subgroup of 2505, 365, and 686 participants with previously assessed two, three, and four 24-hour urinary sodium excretion data collected within one year[26], respectively. For each study, baseline was defined as the period during which the pre-specified numbers of 24-hour urine samples were obtained. Written or oral informed consent was obtained from all participants. The Trial of Hypertension Prevention (TOHP) I was used to assess external validation of the present approach.[29] TOHP-I was designed to test the short-term feasibility and efficacy of seven nonpharmacologic interventions in persons with high normal blood pressure. From the trial, a total of 1,423 participants with two available 24-hour urinary samples at baseline were included in the analysis. This study was approved by the Institutional Review Board of Brigham and Women's Hospital and the Harvard T.H. Chan School of Public Health.

24-Hour urinary sodium excretion

We assessed urinary sodium excretion by averaging all available 24-hour urine samples for each participant. To reduce measurement error that might arise from the under-collection or

overcollection of 24-hour urine samples, we excluded samples that met at least one of the following criteria: the volume of the 24-hour urine sample was less than 500 ml or more than 5000 ml; the start and end times of urine collection were unavailable; the duration of urine collection was less than 20 hours or was more than 28 hours; or the volume loss was more than 100 ml if the estimated volume lost was reported.

Among the cohorts, two 24-hour urine samples were collected within an average of one week in 2,651 participants between 2003 and 2007 to examine risk factors for kidney stones.[30] In second cycle, four 24-hour urine samples were collected, in periods evenly spanned over four seasons within one year, among 1,232 participants between 2010 and 2013, as part of the Women's Lifestyle Validation Study and Men's Lifestyle Validation Study.[31] We excluded participants with missing age, height, or BMI, and those without available at least two 24-hour urine samples based on the aforementioned criteria, leaving 3,454 participants in our main analysis. For developing the ML prediction algorithm, the averaged 24-hour urinary sodium excretion over all available samples was used as the prediction target.

In secondary analysis predicting averaged 24-hour urinary sodium excretion adjusted for doubly-labelled water (DLW)-based total energy expenditure, 1,085 participants with available DLW data were included.

Questionnaires

In NHS, NHS-II and HPFS, participants returned a biennial questionnaire to ascertain lifestyle information and new onset disease diagnoses with follow-up rates greater than 90% of eligible person-time. Participants also answered semi-quantitative FFQs every four years, reporting intake of more than 130 foods and beverages. For the FFQ, respondents were asked

how often, on average, they consumed the specified amount of each food or beverage during the preceding year; 9 possible frequency categories ranged from never/almost never to ≥ 6 times per day. Nutrient intake, including FFQ-based sodium intake, was calculated by multiplying the frequency of intake by the nutrient composition for the portion size specified for each food or vitamin supplement using the Harvard University nutrient database. The database is derived from the USDA food composition manufacturers information, derived information from food ingredient lists, and direct analyses of foods. Frequency of using added table salt was asked with the same scale. Predictors were defined from demographics, behavioral and dietary information based on the baseline questionnaire, defined as the questionnaire that covered the period during which the prespecified numbers of 24-hour urine samples were obtained.

Predictors

We used the following covariates to develop the ML prediction model aiming for external usage: FFQ-based sodium intake, FFQ-based potassium intake, sodium-to-potassium ratio, age, sex, body weight, height, BMI, hours per week of moderate-to-vigorous physical activity (PA hours), history of hypertension, use of antihypertensives/diuretics, menopausal status, FFQ-based total calorie intake, DASH score without a sodium component, and FFQ-based intake of red meat, vegetables, fruits, tea or coffee, alcohol, and energy-adjusted sodium and potassium (FFQ-based sodium or potassium divided by FFQ-based total calorie intake). These predictors (selected list of predictors) were selected considering that this prediction algorithm would have external usage and most epidemiologic studies of chronic disease would collect this information. For internal usage only (i.e. in NHS/NHS-II/HPFS), another ML model was developed using the following predictors in addition to the list above (comprehensive list of

predictors): cohorts, ever use of oral contraceptives, living alone, marital status, frequency of fried foods at home/ away from home, family history of hypertension, frequency of using added table salt, menopausal hormone therapy (past and current), parity, and FFQ-based intake of processed meat, hot dogs, hamburgers, and cheese. Sex was omitted in the comprehensive list of predictors when cohort was used for a predictor since the cohorts were exclusively male or female.

Machine-learning

Figure 1.1 illustrates the present ML prediction approach. Missing values of each predictor were imputed using single imputation based on the Multiple Imputation by Chained Equations (MICE) algorithm.[32] Continuous variables were centered and scaled. Dummy variables were created for categorical variables.

After preprocessing, we used 36 predictors for internal usage and 22 for external usage. The outcome was averaged 24-hour urinary sodium excretion over all available 24-hour urine samples per participant. A 5-fold cross-fitting procedure was applied, in which data was randomly split into 5 subsets to calculate the predicted sodium intake in every participant[33]. Each subset was used in turn as the validation data while the remaining four subsets served as the corresponding training data for building a prediction algorithm. For prediction, three fully tuned distinct ML algorithms were used: elastic net, eXtreme Gradient Boosting (XGBoost), and Bayesian regularization for feed-forward neural net (BRNN). Further description of these algorithms is provided in the **Appendix 1.1**. Five-fold cross-validation (CV) was conducted for the hyperparameter tuning within each training set. Finally, the predicted sodium intake from the

three ML algorithms were simply averaged (ensemble method). The output was used to assess the prediction accuracy with averaged 24-hour urinary sodium excretion.

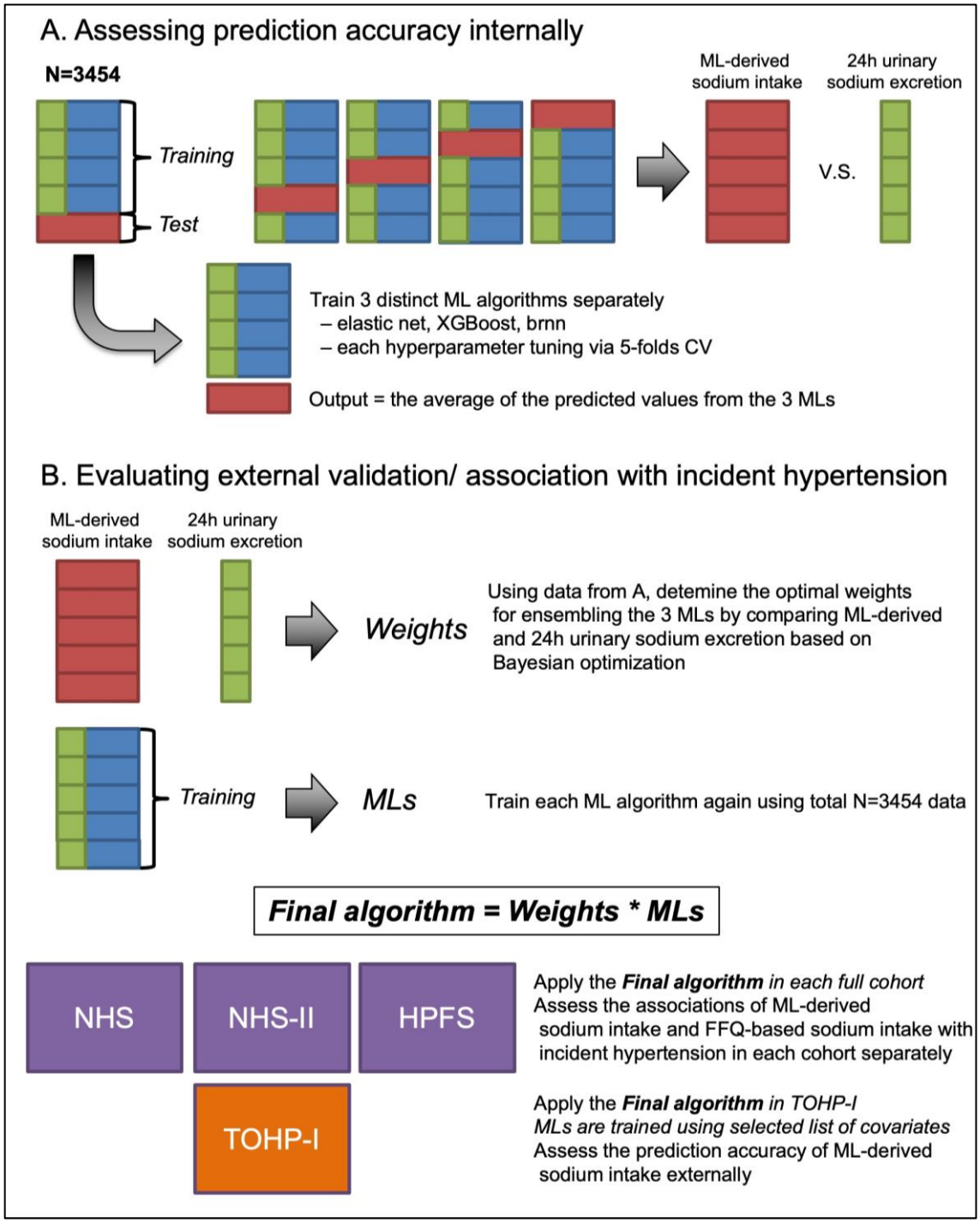


Figure 1.1: Summary of the present study methods

Figure 1.1 (continued)

This study aimed to assess the prediction accuracy of the machine-learning (ML)-derived sodium intake against 24-hour urinary sodium excretion (**A**), and evaluate the association of the ML- and food frequency questionnaire (FFQ)-derived sodium intake with incident hypertension (**B**). **A**) Using derivation dataset (N=3,454) in which data on multiple 24-hour urinary sodium excretion was available, we computed ML-derived sodium intake in all the participants by the cross-fitting approach (out-of-fold prediction). After training/test data splitting, each fold was used as the validation data while the remaining folds will serve as the corresponding training data for building a prediction algorithm. For prediction, an ensemble algorithm over fully tuned four distinct MLs as was used: elastic net, eXtreme Gradient Boosting (XGBoost), and Bayesian regularization for feed-forward neural net (brnn). Five-fold cross-validations (CV) were conducted for the hyperparameter tuning within each training set. Finally, the prediction accuracy of the ML-derived sodium intake was assessed against the corresponding 24-hour urinary sodium excretion. **B**) We then finalized the prediction algorithm based on fully tuned MLs and optimized ensemble weights for each algorithm. Each ML algorithms were trained based on the whole derivation dataset, and the weights were calculated via Bayesian optimization using the out-of-fold prediction in **A**. Using the final algorithm, the ML-derived sodium intake was calculated in the whole participants in NHS, NHS-II, and HPFS for every time point. Cox proportional hazard models were used to evaluate the associations between ML- and FFQ-derived sodium intake and incident hypertension in the three cohorts separately.

Prediction of energy-adjusted sodium

In a secondary analysis, we set DLW-based energy-adjusted averaged 24-hour urinary sodium excretion, based on the density method[34], as our prediction target in the subset with available measures of DLW (N=1,085). The aim of this analysis was to account for strong confounding and extraneous variation due to body size.[34] For interpretability, we multiplied the median of the DLW-based energy expenditure in the derivation data to the predicted values by our ML algorithm, which was the final output of the prediction algorithm (ML-energy-adjusted sodium).

We conducted another analysis, in which averaged 24-hour urinary sodium excretion and DLW-based total energy expenditure were predicted separately (in N=3,454 and N=1,085,

respectively). Then the predicted energy-adjusted sodium was evaluated as the ratio of ML-predicted sodium and ML-predicted energy (ML-energy-adjusted ML-sodium).

Association with incident hypertension

For a use of the prediction algorithm to examine the association with hypertension in the total cohorts of NHS/NHS-II/HPFS, we refit the ML models using the total sample and the hyperparameters were tuned by 5-fold CV. Then we applied Bayesian optimization to calculate the optimal weights for each ML algorithm for the ensemble method[35]. The ML algorithm was applied to the NHS, NHS-II, and HPFS cohorts for every 4-year time period, separately, to estimate the time-varying sodium intake for all the participants. We then compared the association of FFQ-derived and ML-predicted sodium intake with incident hypertension among those without a history of hypertension. For this analysis, participants were excluded if they reported a diagnosis of hypertension, cancer, or stroke at the baseline questionnaire (1986 for NHS and HPFS, 1991 for NHS-II).

We examined the following exposures: FFQ-derived sodium intake (FFQ-sodium), FFQ-derived sodium intake divided by FFQ-derived calorie intake (FFQ-energy-adjusted sodium), ML-predicted 24-hour sodium excretion (ML-sodium), ML-predicted DLW-based energy expenditure-adjusted 24-hour sodium excretion (ML-energy-adjusted sodium), and ML-predicted 24-hour sodium excretion divided by ML-predicted DLW-based energy expenditure (ML-energy-adjusted ML-sodium).

Physician diagnosis of hypertension was self-reported on the baseline and biennial questionnaires. This method of reporting a diagnosis of hypertension among female and male health professionals was shown to be valid in these three cohorts when compared to ascertained

medical records in a subsample.[36–38] Participants were determined to be cases if they reported a diagnosis of hypertension and year of diagnosis after the baseline questionnaire as previously described.[36,37]

Statistical analysis

Overview of the present analyses is in **Figure 1.2**. For the assessment of prediction accuracy in the derivation cohorts (i.e. NHS/NHS-II/HPFS) and in external dataset (TOHP-I), we computed Spearman correlation coefficients between FFQ-based sodium intake, ML-derived sodium intake (ML-sodium), and average 24-hour urinary sodium excretion. Calibration plots, Bland-Altman plots, and calibration intercepts and slopes were also computed. Calibration intercepts close to 0 and calibration slopes close to 1 were considered "good" calibration in general.[39]

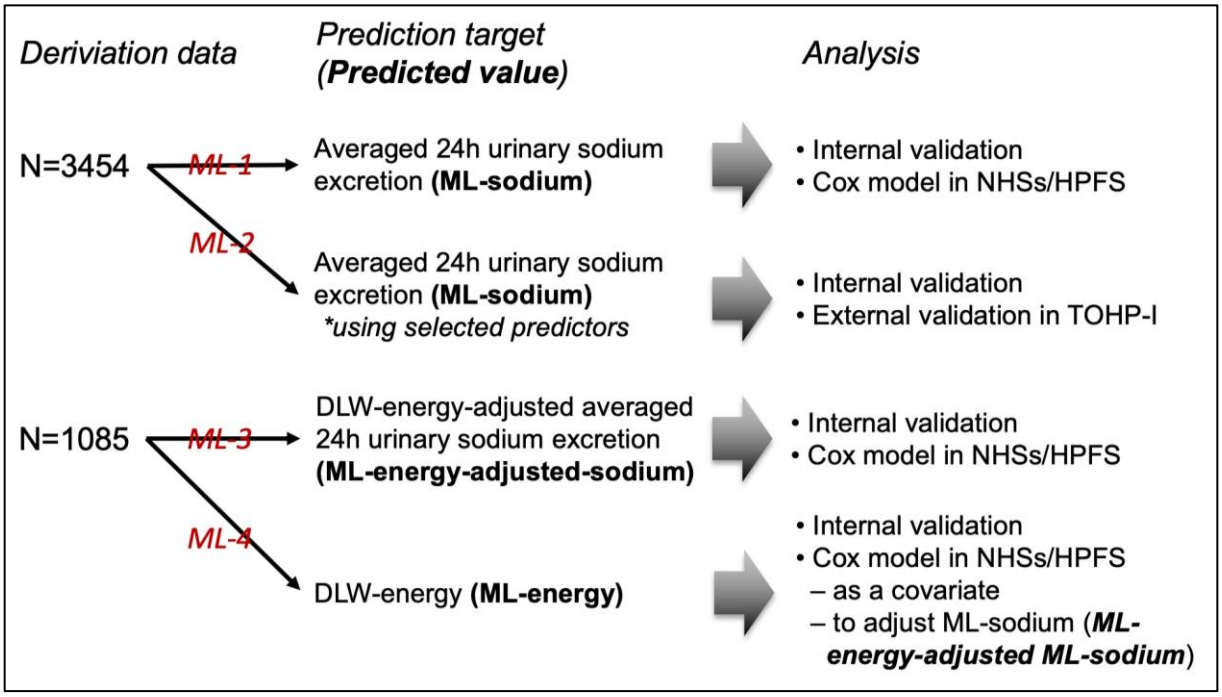


Figure 1.2: Overview of the present analyses

Figure 1.2 (continued)
Using N=3,454 dataset, averaged 24-hour urinary sodium excretion was predicted (ML-1), which was internally validated and examined with respect to incident hypertension in NHS and NHS-II (NHSs) and HPFS. An ML model for 24-hour sodium was developed using selected predictors (ML-2) and the model was internally and externally validated. Using N=1,085 dataset, DLW-energy-adjusted averaged 24-hour urinary sodium excretion was predicted (ML-3), which was also internally validated and examined with respect to incident hypertension in NHSs and HPFS. Finally DLW-based energy expenditure was predicted in the same dataset (ML-4), which was used either as a covariate in Cox models or to calculate ML-energy-adjusted ML-sodium.

We fit Cox proportional models in the three cohorts to assess the association of FFQ- and ML-derived sodium measures with incident hypertension. The exposures were categorized into quartiles to compute the hazard ratio (HR) and 95% confidence intervals (CIs) and treated as time-varying. Age was used as the time axis for the models. The models were adjusted for race (white; non-white), family history of hypertension (yes; no), smoking status (never; past; current), BMI (continuous, kg/m^2), ML-energy (continuous, kcal/day), alcohol intake (0; 0.1 to <5; 5 to <15; 15 to <30; ≥30 g/day), FFQ-based potassium intake (continuous, mg/day), DASH score without the sodium component (continuous, unit), moderate-to-vigorous physical activity (continuous, hours/week), and use of aspirin, acetaminophen, and non-steroidal anti-inflammatory drugs (yes/no). In the NHS and NHS-II, menopausal status (pre-; post-menopausal), use of menopausal hormone therapy (never/past/current), and parity (none; 1 or 2; 3 or 4; ≥5 kids) were additionally adjusted. In the NHS-II only, use of oral contraceptive (never; past; current) was further adjusted. All of these covariates were included in the ML prediction model. Estimates and 95% CIs were calculated in each cohort separately.

The statistical analyses were performed using R 4.2.0 (R Foundation for Statistical Computing, Vienna, Austria.), Python 3.9.13, and SAS 9.4 (SAS Institute Inc., Cary, NC, USA).

1.4 Results

Baseline characteristics are summarized in **Table 1.1.** A total of 3,454 participants with at least two 24-hour urinary sodium excretion data were included from the three cohorts. Higher urinary sodium excretion was characterized by younger age, greater proportion of men or pre-menopausal women, non-White race, larger body size, higher FFQ-based sodium intake, history of hypertension and cancer, and ever usage of oral contraceptives. **Supplemental Table 1.1** contrasts baseline characteristics according to the quartile of energy-adjusted 24-hour sodium excretion. Higher energy-adjusted sodium excretion was associated with smaller DLW-based energy expenditure and higher 24-hour sodium excretion, and the correlation with body size was not strong.

	Quartile 1 N = 864	Quartile 2 N = 864	Quartile 3 N = 863	Quartile 4 N = 863
Cohort				
1: HPFS (Men only)	151 (17.5)	200 (23.1)	290 (33.6)	387 (44.8)
2: NHS (Women only)	417 (48.3)	295 (34.1)	243 (28.2)	164 (19.0)
3: NHS-II (Women only)	296 (34.3)	369 (42.7)	330 (38.2)	312 (36.2)
Age, years	63.2 (10.4)	61.2 (10.5)	61.4 (10.3)	59.9 (9.7)
Body mass index, kg/m^2	24.6 (4.5)	25.7 (4.4)	26.7 (5.0)	28.9 (6.0)
Non-White race, %	36 (4.2)	41 (4.7)	45 (5.2)	54 (6.3)
History of hypertension, %	306 (35.4)	299 (34.6)	318 (36.8)	352 (40.8)
History of cancer, %	100 (11.6)	102 (11.8)	122 (14.1)	156 (18.1)
Use of antihypertensives, %	282 (32.6)	287 (33.2)	287 (33.3)	327 (37.9)
Family history of hypertension, %	437 (50.6)	426 (49.3)	380 (44.0)	413 (47.9)
Family history of CVD, %	197 (22.8)	175 (20.3)	184 (21.3)	159 (18.4)
Moderate to vigorous PA, hours/w	3.9 (5.0)	4.1 (5.7)	4.2 (4.9)	4.2 (5.1)
Smoking status				
1: Never	506 (58.6)	530 (61.3)	520 (60.3)	517 (59.9)
2: Past	314 (36.3)	294 (34.0)	302 (35.0)	313 (36.3)
3: Current	44 (5.1)	40 (4.6)	41 (4.8)	33 (3.8)
Alcohol, gram/day	8.8 (13.2)	8.1 (13.0)	8.4 (13.2)	7.9 (11.6)
DASH score without sodium component	21.2 (4.3)	21.1 (4.1)	20.6 (4.3)	20.2 (4.1)
FFQ-based sodium, mg/day	2000 (729)	2159 (745)	2263 (774)	2411 (864)
FFQ-based pottasium, mg/day	3271 (1013)	3314 (1069)	3322 (1109)	3397 (1113)

FFQ-based total calorie, kcal/day	1830 (557)	1871 (592)	1900 (601)	1941 (636)
Averaged 24-hour urinary sodium excretion, mg/day	2074 (372)	2889 (185)	3591 (237)	5014 (906)

Table 1.1: Baseline characteristics according to the quartile of the averaged 24-hour urinary sodium excretion in N=3,454 participants of NHS, NHS-II , and HPFS

Prediction of averaged 24-hour urinary sodium excretion

Averaged 24-hour urinary sodium excretion was weakly correlated with FFQ-based sodium intake (Spearman correlation coefficient [95% CI]: 0.19 [0.16, 0.23]) and moderately correlated with ML-based predicted value (0.51 [0.49, 0.54]; **Figure 1.3)**. Elastic net, XGBoost, and BRNN algorithms predicted the target with Spearman correlation coefficient [95% CI] of 0.51 [0.49, 0.54], 0.50 [0.48, 0.53], and 0.51 [0.48, 0.53], respectively (**Supplemental Figure 1.1**). The calibration of the ML-predicted sodium was good (calibration intercept [SE]: -92 mg/day [97]; calibration slope: 1.03 [0.03]). However, the ML output did not cover 24-hour urinary sodium excretion values below 1944 mg/day or above 5682 mg/day, partly reflecting the limited number of participants with such sodium intake values (N=265 and 163, respectively). The mean difference between ML-predicted sodium intake and averaged 24-hour urinary sodium excretion was -0 mg/day (95% CI: -1986, 1987), and higher actual intake was associated with systematically underestimated ML estimate (**Supplemental Figure 1.2**). ML models based on a selected list of covariates yielded similar prediction accuracy (**Supplemental Figure 1.3**).

Body size had the largest contribution to the prediction algorithm of absolute sodium excretion, followed by age and sex (**Table 1.2**). In the elastic net algorithm, weight, FFQ-sodium measures and use of table salt were strongly and positively predictive, while age and membership in NHS or NHS-II (i.e. women) were negatively associated. In the XGBoost algorithm, weight, FFQ-based calorie-adjusted sodium intake, BMI, and age were most highly

predictive. The strong predictors were similar when the subset of selected covariates was used (**Supplemental Table 1.2**).

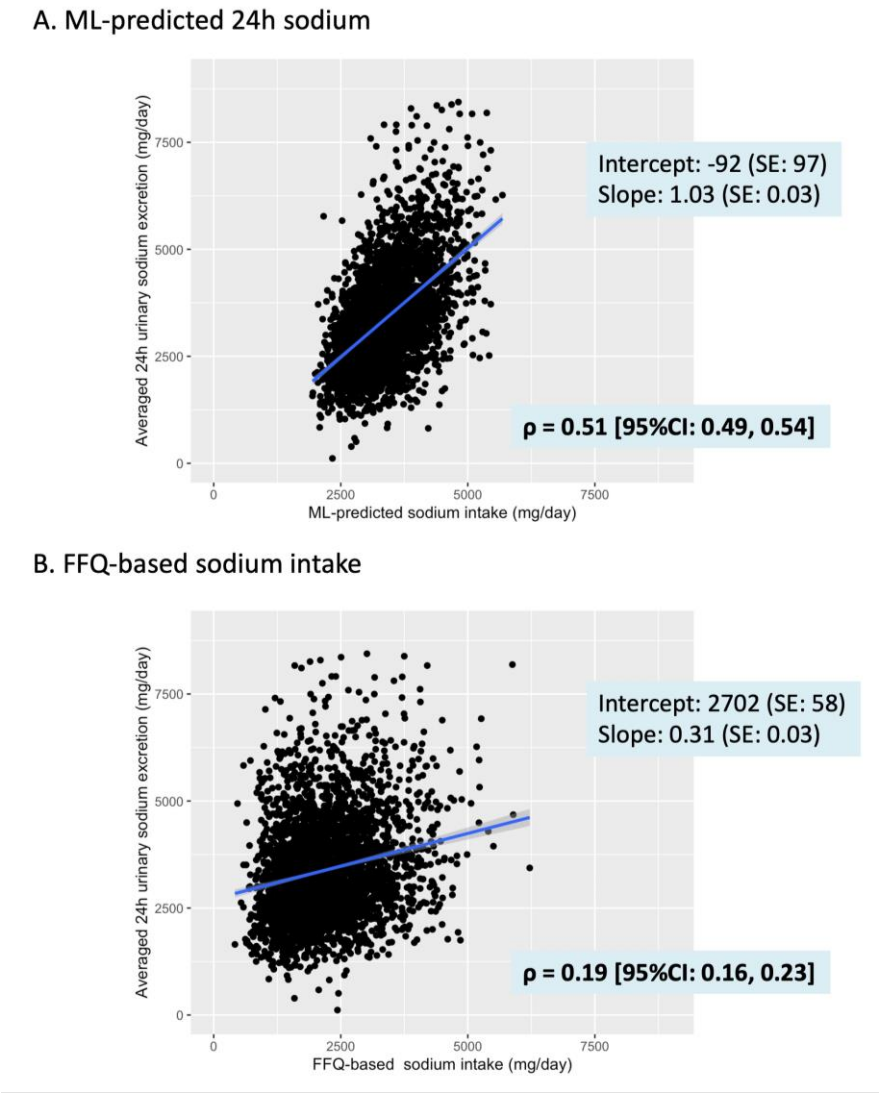


Figure 1.3: Comparison of ML-predicted and FFQ-based sodium intake with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS
Scatter plots describing the association between ML-predicted sodium intake (A) and FFQ-based sodium intake (B) with respect to averaged 24-hour urinary sodium excretion in N=3,454 participants. ML-predicted sodium intake was based on out-of-fold prediction based on 36 predictors using ensemble over three ML algorithms. Each dot represents each participant. Blue straight lines are the calibration curves, and the values of calibration intercept and calibration slope are provided. ρ represents Spearman correlation coefficient.

A. Coefficients of elastic net

Positive association		***Shrunk to null***		***Negative association***	
	Coef		*Coef*		*Coef*
Weight	371	FFQ-sodium	0	Age	-258
FFQ-sodium/FFQ-calorie	191	BMI	0	NHS	-218
Use of table salt	81	FFQ-sodium/ FFQ-potassium	0	NHS-II	-187
Height	75	FFQ-potassium	0	Use of oral contraceptives	-113
Fried food away from home	33	Fruit	0	FFQ-calorie	-31
Red meat	33	PA hours	0	Past use of HRT	-31
Processed meat	33	Post-menopausal	0	Alcohol	-19
Use of diuretics	32	FH of CVD	0	DASH score	-16
Fried food at home	31	Hotdog	0	FH of hypertension	-2
FFQ-potassium/FFQ-calorie	29	Married	0		
Living alone	19	Current use of HRT	0		
Use of antihypertensives	13	Parity	0		
Hamburger	11				
Cheese	6				
Vegetable	6				
Tea or coffee	3				
History of hypertension	2				

B. Top 10 variable importance in XGBoost

	Variable importance
Weight	100
FFQ-sodium/FFQ-calorie	30
BMI	24
Age	22
Height	20
FFQ-sodium/ FFQ-potassium	14
Post menopausal	11
FFQ-sodium	9
FFQ-calorie	8
Fruit	7

Table 1.2: Interpretation of ML models based on full covariate list for predicting averaged 24-hour urinary sodium excretion in N=3,454 participants

A) Values are coefficients per category for categorical variables and per SD for continuous variables

B) Variable importance measures are scaled into 0 to 100.

The averaged 24-hour urinary sodium excretion in TOHP-I was moderately well-predicted by the ML based on selected list of covariates (Spearman correlation coefficient [95% CI]: 0.46 [0.42, 0.50]; **Figure 1.4**). The calibration intercept [SE] was -269 mg/day [209] and calibration slope was 0.83 [0.044]), systematically over-estimating sodium intake by ML. Nevertheless, it substantially outperformed prediction by FFQ-based sodium (Spearman correlation coefficient: 0.22 [0.17, 0.27]). (calibration intercept [SE]: 2852 [105]; calibration slope: 0.21 [0.027]). The mean difference between 24-hour urinary sodium excretion and ML-predicted sodium intake was -1076 mg/day [95%CI: -3638, 1497] (**Supplemental Figure 1.4**).

Prediction of DLW-based energy and energy-adjusted averaged 24-hour urinary sodium excretion

Among 1,085 participants whose DLW-information was available, DLW-based energy expenditure was well predicted and calibrated by ML approach compared with FFQ-derived total calorie intake (Spearman correlation coefficient [95% CI]: 0.77 [0.75, 0.80] in ML, 0.29 [0.24, 0.35] in FFQ; **Supplemental Figure 1.5-1.6**), with weight being the strongest predictor (**Supplemental Table 1.3**). However, DLW-based energy-adjusted averaged 24-hour urinary sodium excretion (i.e. sodium excretion / DLW-energy) was only modestly better predicted in ML approaches compared with FFQ-based calorie-adjusted FFQ-based sodium intake (Spearman correlation coefficient: 0.37 [0.31, 0.42] in ML outputs predicting '24-sodium / DLW-energy', 0.32 [0.27, 0.38] in ML outputs predicting 24-hour sodium divided by ML outputs predicting DLW-energy, and 0.32 [0.26, 0.37] in FFQ-calorie-adjusted FFQ-sodium; **Figure 1.5**). Calibration was better and mean differences were smaller in ML approaches compared with a FFQ measure (**Figure 1.5** and **Supplemental Figure 1.7**). In ML models predicting DLW-

energy-adjusted sodium, FFQ-based nutritional or food intakes related to sodium and potassium were most predictive, and the contributions of body size were smaller.

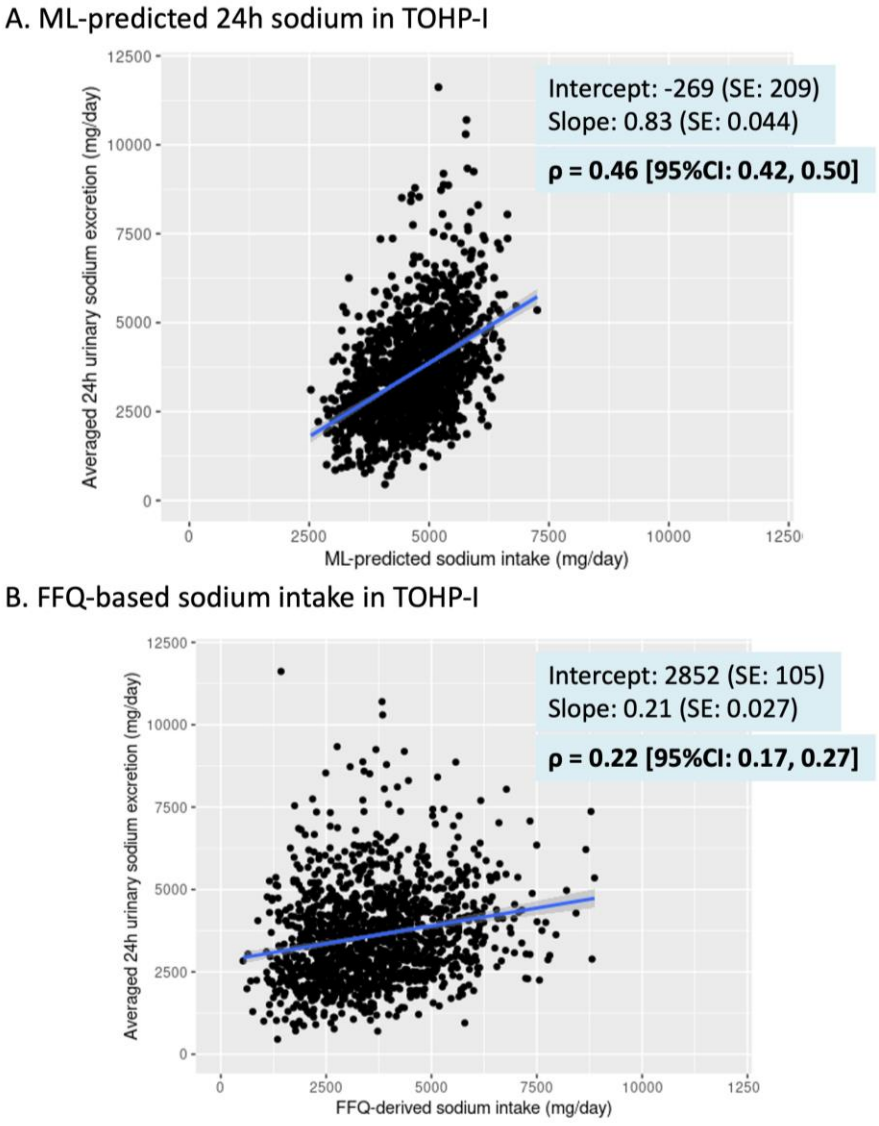


Figure 1.4: Comparison of ML-predicted and FFQ-based sodium intake with averaged 24-hour urinary sodium excretion in N=1,423 participants in the TOHP-I

Scatter plots describing the association between ML-predicted sodium intake (A) and FFQ-based sodium intake (B) with respect to averaged 24-hour urinary sodium excretion in N=1,423 participants in the TOHP-I. The ML prediction algorithm was externally developed based on 22 predictors in the NHS, NHS-II, and HPFS. Each dot represents each participant. Blue straight lines are the calibration curves, and the values of calibration intercept and calibration slope are provided. ρ represents Spearman correlation coefficient.

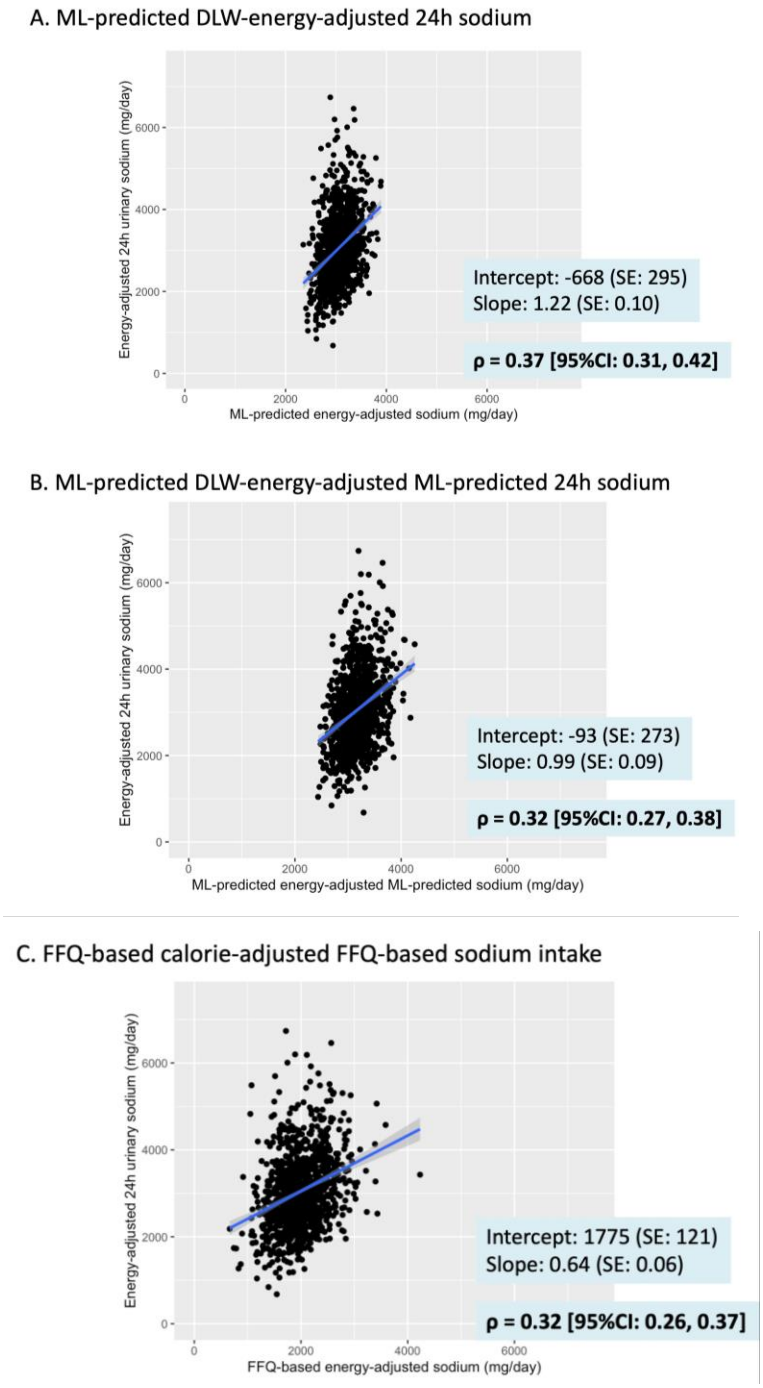


Figure 1.5: Comparison of ML-derived and FFQ-based energy-adjusted sodium intake with DLW-based energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants from NHS, NHS-II, and HPFS

Scatter plots describing the association between ML-predicted DLW-energy-adjusted 24-hour urinary sodium excretion (A), ML-predicted DLW-energy-adjusted ML-predicted 24-hour urinary sodium excretion (B), and FFQ-based calorie-adjusted FFQ-based sodium intake (C) with respect to DLW-energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants. ML-predicted sodium measures was based on out-of-fold prediction based on 36 predictors using ensemble over three ML algorithms. Each dot represents each participant.

Figure 1.5 (continued)
Blue straight lines are the calibration curves, and the values of calibration intercept and calibration slope are provided. ρ represents Spearman correlation coefficient.

Association between FFQ- or ML-based sodium measures and incident hypertension

We applied the ML algorithm to the NHS (N=52,780), NHS-II (N=83,871), and HPFS (N=31,269) cohorts for each repeated FFQ in every 4-year period to estimate the time-varying intake of sodium intake (**Table 1.3**). In NHS, FFQ-energy-adjusted sodium was similarly associated with incident hypertension than ML-based measures (Q5 vs. Q1: HR = 1.00 for FFQ-energy-adjusted sodium; HR = 0.99-1.01 for ML-based measures). In NHS-II, the associations with ML-based measures were stronger than FFQ-based ones (Q5 vs. Q1: HR = 1.05 for FFQ-energy-adjusted sodium; HR = 1.13-1.48 for ML-based measures). In HPFS, similar inverse associations were observed for FFQ- and ML-based measures ones (Q5 vs. Q1: HR = 0.92 for FFQ-energy-adjusted sodium; HR = 0.93-0.97 for ML-based measures).

NHS

	Q1	Q2	Q3	Q4	Q5
FFQ-sodium	–	0.99 (0.96, 1.03)	0.99 (0.96, 1.03)	0.97 (0.93, 1.00)	0.99 (0.95, 1.03)
FFQ-energy-adjusted sodium	–	1.00 (0.97, 1.04)	1.03 (1.00, 1.07)	1.00 (0.97, 1.04)	1.00 (0.96, 1.04)
ML-sodium	–	0.97 (0.93, 1.00)	0.96 (0.92, 1.00)	0.99 (0.94, 1.03)	1.01 (0.95, 1.07)
ML-energy-adjusted sodium	–	1.00 (0.97, 1.04)	1.01 (0.97, 1.05)	1.00 (0.97, 1.04)	0.99 (0.96, 1.03)
ML-energy-adjusted ML-sodium	–	1.00 (0.96, 1.04)	1.03 (0.99, 1.07)	1.01 (0.97, 1.05)	1.00 (0.96, 1.05)

NHS-II

	Q1	Q2	Q3	Q4	Q5
FFQ-sodium	–	0.99 (0.95, 1.02)	1.04 (1.00, 1.08)	1.03 (0.99, 1.07)	1.04 (0.99, 1.08)
FFQ-energy-adjusted sodium	–	1.03 (0.99, 1.07)	1.04 (1.01, 1.08)	1.02 (0.98, 1.06)	1.05 (1.02, 1.09)
ML-sodium	–	1.11 (1.07, 1.16)	1.23 (1.18, 1.29)	1.34 (1.28, 1.40)	1.48 (1.40, 1.56)

ML-energy-adjusted sodium	–	1.07 (1.02, 1.11)	1.09 (1.05, 1.14)	1.12 (1.07, 1.16)	1.13 (1.08, 1.17)
ML-energy-adjusted ML-sodium	–	1.07 (1.02, 1.11)	1.14 (1.09, 1.19)	1.19 (1.14, 1.24)	1.19 (1.14, 1.24)
HPFS					
	Q1	Q2	Q3	Q4	Q5
FFQ-sodium	–	0.98 (0.93, 1.03)	0.95 (0.90, 1.00)	0.96 (0.91, 1.01)	0.90 (0.84, 0.95)
FFQ-energy-adjusted sodium	–	0.99 (0.94, 1.03)	0.98 (0.93, 1.02)	0.96 (0.91, 1.01)	0.92 (0.87, 0.98)
ML-sodium	–	0.97 (0.92, 1.02)	0.94 (0.88, 1.00)	0.92 (0.86, 0.99)	0.97 (0.89, 1.06)
ML-energy-adjusted sodium	–	1.00 (0.95, 1.05)	1.00 (0.94, 1.05)	0.98 (0.93, 1.03)	0.93 (0.88, 0.98)
ML-energy-adjusted ML-sodium	–	1.01 (0.96, 1.06)	1.01 (0.95, 1.06)	0.99 (0.93, 1.05)	0.96 (0.90, 1.02)

Table 1.3: Multivariable-adjusted Hazard Ratios (95% CI) between FFQ-based energy-adjusted ML-predicted or FFQ-derived sodium measures and incident hypertension among 52,780 women in the in the NHS, 83,871 women in the NHS-II , and 31,269 men in the HPFS

Numbers are case numbers or hazard ratios (95% confidence intervals).

The exposures are the quintiles of FFQ-derived sodium intake (FFQ-sodium), FFQ-derived sodium intake divided by FFQ-derived calorie intake (FFQ-energy-adjusted sodium), ML-predicted 24-hour sodium excretion (ML-sodium), ML-predicted DLW-based energy expenditure-adjusted 24-hour sodium excretion (ML-energy-adjusted sodium), and ML-predicted 24-hour sodium excretion divided by ML-predicted DLW-based energy expenditure (ML-energy-adjusted ML-sodium).

The outcome is incident self-reported hypertension.

The multivariable-adjusted models are adjusted for age, race, family history of hypertension, smoking status, body mass index, DASH score without sodium component, moderate-to-vigorous physical activity, alcohol intake, potassium intake, and total calorie in HPFS; plus parity, menopausal status, and hormone replacement therapy in NHS; and plus contraceptive use in NHS-II .

The 95% CIs did not take into account the variation due to fitting the prediction model.

Abbreviations: NHS, Nurses' Health Study; HPFS, Health Professional Follow-up Study; DASH, Dietary Approaches to Stop Hypertension score; P-Y, person-years; HR, hazard ratio.

1.5 Discussion

In this study, using data from 3,454 participants from the NHS, NHS-II, and HPFS, we applied ML algorithms to predict average 24-hour urinary sodium excretion over multiple samples. Our algorithms yielded a more accurate prediction and calibration of sodium intake compared with that calculated based on a single FFQ. The prediction accuracy was externally validated in TOHP-I but the calibration was poor. However, the improvement in prediction was

largely due to the incorporation of body size, and prediction of DLW-based energy-adjusted 24-hour sodium excretion was only modestly better in ML compared with FFQ-calorie-adjusted FFQ-sodium. In the NHS-II consisting of younger women (mean age of 36y), more robust associations with incident hypertension were observed in ML-based sodium measures than FFQ-based measures. However, in NHS (mean 52y) and HPFS (mean 53y), no apparent associations were observed between ML-based sodium measures and incident hypertension. Therefore, the proposed algorithms may serve as a novel and more generalizable method for predicting more accurate absolute sodium intake compared with a FFQ, but given the widely-observed association of sodium with hypertension from both trials and observational studies, ML-predicted sodium could not greatly attenuate the apparent bias stemming from measurement error in disease association.

Measurement error in estimating sodium intake based on dietary assessment has been a persistent challenge in epidemiological research due to the great variability of sodium in processed foods and food prepared away from home.[2,3,21] This study represents a pioneering effort to incorporate all available information from cohort questionnaires using ML algorithms to improve the calibration of sodium exposure. Our algorithms, which referenced the averaged 24-hour urinary sodium excretion across multiple samples, provided substantially more accurate estimates of sodium intake compared with those based solely on FFQ. We leveraged three distinct algorithms (linear, tree-based, and neural network algorithms) to capture multiple aspects of the contributions of the predictors, and the ensemble approach resulted in the best performance. The algorithm performed fairly well in TOHP-I, despite distinct differences in participant characteristics from those in the NHS, NHS-II, and HPFS. Therefore, our approach

may have broad applicability and provide a more robust prediction of absolute sodium intake. In addition, DLW-based energy expenditure was well predicted using the same ML approach despite the limited sample size, also suggestive of its potential utility in future investigations.

However, the better prediction did not necessary lead to a greater reduction of measurement error bias in the disease association. The variable importance analyses indicated that body size, age, and sex, were the strongest predictors of sodium intake. This reflects the fact that larger and younger participants eat more food, including salt, and the contributions of these variables are of no interest in evaluating the association between sodium and disease outcome – i.e. these variables are confounders included in the disease model. Therefore, a primary nutritional exposure in disease association is energy-adjusted intake to minimize such confounding, inferring the effect of changing the composition of diet on disease outcomes. The present MLs only modestly improved the prediction accuracy for 'gold-standard' DLW-energy-adjusted 24-hour sodium excretion compared with FFQ-based calorie-adjusted FFQ-sodium, although the predictions for either 24-hour sodium excretion or DLW-based energy expenditure were far more accurate than FFQ-based measures. The improvement in ML prediction was largely due to FFQ-based dietary information related to sodium or potassium. Consequently, we observed only slight improvement in removing measurement error bias using the ML in the association with incident hypertension in the younger female cohort and no improvement in the other cohorts.

Accurate calculation of salt intake from a generic list of foods in the FFQ is challenging, whereas signals from other questionnaire-based information were meaningful for the prediction

by reflecting salt-related eating behaviors. This aligns with prior epidemiological studies in which salt preference was used as a proxy for salt intake.[40,41] Our prediction algorithms were robust in that they accounted for the flexible contributions of multiple characteristics and have the potential to be applied in a wide range of settings. For example, quantifying sodium intake purely based on questionnaire information would be informative as well as cost saving compared with 24-hour urinary sampling, having the potential to be widely used in the clinical setting for counselling/lifestyle interventions, given that dietary guidelines define cutoffs for absolute sodium intake. However, the use of absolute sodium excretion has dubious value in clinical applications because of its strong dependence on age, sex, and body size, whereas the intervention would be on diet quality or adiposity if appropriate. Further work to develop sodium screeners targeting diet quality and incorporating additional behavioral questions would be desirable.

A limited number of longer-term RCTs have reported the effect of salt reduction interventions on incident hypertension. In the meta-analysis conducted within the Dietary Reference Intakes for Sodium and Potassium, the relative risk estimate integrating TOHP-I, TOHP-II, and Hypertension Prevention Trial was 0.79 (95%CI: 0.67, 0.93).[42] Although our estimates may not be directly compared to these trial results as we did not emulate the target trial of the comparable intervention, we only observed 1.2 to 1.4-fold higher risk in the highest intake group versus the lowest one in the NHS-II and null results in NHS and HPFS. The notable differences across NHS, NHS-II, and HPFS suggest effect modification by age; NHS and HPFS consisted of older populations, and the participants most susceptible to hypertension may have been already excluded at baseline. Effect modification by sex may also be present, as studies

reporting women are more salt-sensitive compared with men.[43–45] Further investigation is warranted to clarify the relationship between predicted sodium intake and hypertension to identify specific targets for sodium reduction interventions.

Strengths and limitations

Strengths of this study include the use of multiple 24-hour urinary samples and DLW information, its large sample size, rigorous ML algorithms, external validation, and application examples for the actual use of the predicted sodium measures in outcome analyses targeting incident hypertension. However, there are several limitations that warrant consideration. Although averaged 24-hour urinary sodium excretion is considered the gold standard approach to estimate sodium intake, the measure still suffers from measurement error due to true variability from day-to-day sodium exposure. There is also likely to be measurement error in the DLW-based energy expenditure used for calculating energy-adjusted sodium measures. Although we externally validated the present algorithm in TOHP-I, which had sizeable minority participation, further evaluation is warranted in other racial/ethnic groups and across different cultures. Urine and DLW samples were based on smaller number of participants, and thus the generalizability of the ML algorithm might be limited. In addition, DLW information was not available in TOHP-I and external validation could not be assessed. The 95% CIs in the Cox models did not take into account the variation due to fitting the calibration model, whereas a simulation study showed such CI estimates would work when the predicted value is used as a categorical variable.[46] Finally, our prediction of incident hypertension was in three large prospective observational studies with all self-reported lifestyle and demographic information, which in themselves may have modest measurement error, although the reliability of self-reported hypertension has been

validated [36–38]. This error is likely mostly random with respect to our predicted sodium intake and thus our risk estimates may be underestimated with respect to predicted incidence of hypertension.

Conclusion

ML algorithms provided improved prediction of average 24-hour urinary sodium excretion and DLW-based energy expenditure compared with the FFQ-based estimates. The prediction accuracy was externally valid. However, the ML algorithms heavily depended on body size, age, and sex, which were confounders in disease associations and included in the disease prediction models. Therefore, the prediction of energy-adjusted 24-hour sodium excretion was only modestly better using ML compared with FFQ. The ML-predicted sodium intake was more strongly associated with incident hypertension than FFQ-based ones in NHS-II, whereas there were no appreciable differences in NHS or HPFS. The present algorithms may thus prove to be a more effective and generalizable approach for predicting absolute sodium intake but did not materially reduce the measurement error bias in disease associations.

Chapter 2

Modifiable lifestyle factors in the primordial prevention of hypertension in three US cohorts

Author: Rikuta Hamaya, MD, PhD, MS[1,2], Molin Wang, PhD[1,3,7], Ellen Hertzmark, MA[4], Nancy R Cook, ScD[1,2], JoAnn E Manson, MD, DrPH[1,2,5], Qi Sun, MD, ScD[6,7], Meir J Stampfer, MD, DrPH[1,6,7], Kenneth J. Mukamal, MD, MPH[8], Gary Curhan, MD, ScD[1,7,9], Eric B Rimm, ScD[1,6,7]

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[2] Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[3] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

[4] Department of Global Health and Population, Harvard TH Chan School of Public Health, Boston, MA

[5] Mary Horrigan Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[6] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

[7] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[8] Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA

[9] Renal Division, Department of Medicine, Brigham and Womens' Hospital, Boston, MA

2.1 Abstract

Importance: Primordial prevention of hypertension by lifestyle modification is a high priority. Evidence is lacking on the relative contributions of specific lifestyle factors and their overall contribution to prevention of hypertension, in particular early-onset hypertension.

Objective: To investigate the population attributable risks (PARs) and restricted mean survival time (RMST) differences of lifestyle factors in incident hypertension and early-onset hypertension in three large US cohorts.

Design: Prospective cohort study.

Setting, and Participants: We included participants of the Nurses' Health Study (NHS, N=52,780 women, aged 40–67 in 1986), the Nurses' Health Study II (NHS II, N=83,871 women, aged 27–46 in 1991), and the Health Professionals Follow-up Study (HPFS, N =31,269 men, aged 40–75 in 1986), who were free from hypertension, cardiovascular disease and cancer at baseline.

Exposures: Four modifiable lifestyles based on the 2020 International Society of Hypertension Global Hypertension Practice Guidelines; body mass index (BMI), moderate-to-vigorous physical activity (MVPA), Dietary Approaches to Stop Hypertension (DASH) score, and alcohol intake, updated every 2 to 4 years.

Main Outcomes and Measures: Primary outcome was incident self-reported diagnosis of hypertension with 27-31 years of follow-up. Diagnosed hypertension treated with antihypertensive medication was assessed in sensitivity analyses. We calculated hazard ratios, PARs, and RMST differences that accounted for time-varying exposures/covariates to infer the contributions of each exposure.

Results: Each lifestyle factor was associated with incident hypertension in dose-dependent manners across the cohorts, with BMI having the strongest associations with HRs [95% CIs] comparing BMI ≥35 kg/m^2 versus 18.5 to <25 kg/m^2 of 2.13 [95% CI 1.99, 2.28], 3.29 [3.12, 3.48], and 2.19 [1.93, 2.48] in NHS, NHS II and HPFS, respectively. On average, adhering to BMI <25 kg/m^2 was associated with 20.3 [18.5, 22.0], 25.0 [23.2, 26.8], and 18.6 [16.7, 20.7] months longer periods free from hypertension during 25-year follow-up in NHS, NHS II and HPFS, respectively. BMI accounted for approximately 20% of incident hypertension in NHS and HPFS, and 35% of early-onset hypertension (age < 55y). MVPA and diet accounted for 10-15% of incident hypertension in women, and the contributions were greater for early-onset hypertension. Similar results were observed for incident hypertension treated with anti-hypertensive medication.

Conclusions and Relevance: Maintaining healthy weight during adulthood may provide the greatest reduction in the burden of hypertension, but several other lifestyle choices in combination could further reduce risk across all ages, and delay onset of hypertension, with stronger associations in younger individuals.

2.2 Introduction

Hypertension is a leading cause of cardiovascular diseases (CVD). The prevalence has doubled from 1990 to 2019[47] and now impacts almost half of the US adult population[48], and the condition accounts for a major proportion of CVD incidence[49] and burden of death and disability.[50] Hypertension also leads to debilitating, financially costly chronic conditions including kidney disease, stroke, and heart failure. Although numerous therapeutic options are

available, this condition is poorly controlled at the population level, as a recent study from the National Health and Nutrition Examination Survey data reported only 43.7% of the patients in the US in 2017 and 2018 had systolic/diastolic blood pressure (BP) below 140/90 mmHg.[24] Furthermore, there are increasing global disparities in the prevalence, awareness, treatment and control of hypertension.[25] As such, prevention of hypertension is an increasingly important public health aim.

The adoption of healthier lifestyle habits can prevent or delay the development of the condition, lower BP, and enhance the efficacy of antihypertensive medications.[4,51] The 2020 International Society of Hypertension Global Hypertension Practice Guidelines listed 10 lifestyle modifications that should be included in the management of hypertension; in particular, the totality of evidence suggests consistent roles of weight reduction[52], sodium reduction[22,53], adherence to an overall healthy diet pattern such as Dietary Approaches to Stop Hypertension (DASH)[54], moderation of alcohol consumption[55], and enhanced regular, aerobic physical activity[56], in reducing BP. The guidelines also relate healthy drinks such as coffee or tea, stress reduction, complementary alternative medicines, and reduction of exposures to air pollution or cold temperature as other factors for hypertension management/prevention.

Given the high prevalence of uncontrolled BP, to properly guide prevention and interventional strategies at the population level, clarifying the attributable fraction of specific lifestyle factors on incidence hypertension is a high priority. Previously, more than a decade ago, we investigated the population attributable risks (PAR) of hypertension among younger women in the Nurses' Health Study II[57], but such large studies among men or in older populations of women have not been conducted. Previous studies have shown important age- and sex-specific differences in incident hypertension.[58,59] Furthermore, the guidelines for prevention and

diagnosis of hypertension has substantially changed with new evidence[8] and an emphasis on early-onset hypertension.[6,60] Many consider early onset hypertension a highly heritable trait, but less has been characterized with respect to the role of lifestyle modification.

In this study, we aimed to comprehensively investigate the contributions of modifiable lifestyle factors on incident hypertension in three large US cohorts with an age range from 27-75 years, with 27-31 years of follow-up. We evaluated PARs to estimate the proportion of hypertension cases attributable to poor adherence to a healthy lifestyle and computed restricted mean survival time (RMST) differences[61] to assess the estimated gain in disease-free months if we had intervened and each participant had shifted to the healthy category for each lifestyle factor.

2.3 Methods

Population

Participants consisted of the Nurses' Health Study (NHS), the Nurses' Health Study II (NHS II), and the Health Professionals Follow-up Study (HPFS). We excluded participants who reported a diagnosis of hypertension, cancer, heart failure, or stroke at the baseline questionnaire (1986 in NHS, 1991 in NHS II, and 1986 in HPFS). The resulting study population consisted of 52,780 women (aged 40–67 in 1986) from NHS, 83,871 women (aged 27–46 in 1991) from NHS II, and 31,269 men (aged 40–75 in 1986) from HPFS. Participants returned a biennial questionnaire every two years to ascertain lifestyle information and new onset disease diagnoses with follow-up rates greater than 90% of eligible person-time. Participants also answered semi-quantitative food frequency questionnaires (FFQs) every four years, reporting intake of more

than 130 foods and beverages. This study was approved by the Institutional Review Board of Brigham and Women's Hospital and the Harvard T.H. Chan School of Public Health.

Assessment of modifiable lifestyle factors

On baseline and follow-up questionnaires, body weight and time spent on recreational moderate-to-vigorous physical activity (MVPA) were ascertained. Components of physical activity included brisk walking with ≥3 mph, jogging, running, swimming, racquet sports, bicycling, weight training, or other aerobic activity. Waist was assessed similarly only in NHS II.

For the FFQ, respondents were asked how often, on average, they consumed the specified amount of each food or beverage during the preceding year; 9 possible frequency categories ranged from never/almost never to ≥ 6 times per day. Open-ended questions were used for usual brand and type of margarine, cooking oil, cold breakfast cereal, and multivitamins. We also collected detailed information regarding the type of fat used at the table and in food preparation. Nutrient intake was calculated by multiplying the frequency of intake by the nutrient composition for the portion size specified for each food or vitamin supplement using the Harvard University nutrient database. The database is derived from the USDA food composition database[21] manufacturers information, derived information from food ingredient lists, and direct analyses of foods. The FFQ has been extensively validated in these populations against 2X7 day diet records and biomarkers of intake.[18,19,62] Based on each year's FFQs, We calculated each participant's DASH score[63,64], consisting of fruits, vegetables, whole grains, nuts and legumes, low-fat dairy, red and processed meats, sweetened beverages, and sodium. Since FFQs are known to poorly estimate sodium intake on their own, we used ML to derive estimated sodium intake from other variables based upon a subsample of 3,454 participants who underwent

measured 24-hour sodium excretion. We also developed an algorithm to predict doubly labelled water (DLW)-based energy expenditure using ML. As the sodium component of the DASH score, we used predicted energy-adjusted sodium levels (predicted sodium excretion divided by predicted DLW-based energy expenditure).

Ascertainment of Hypertension

Physician diagnosis of hypertension was self-reported on the baseline and biennial questionnaires. This method of reporting a diagnosis of hypertension among female and male health professionals was shown to be valid in these three cohorts when compared to ascertained medical records in a subsample.[36–38] Participants were determined to be cases if they reported a diagnosis of hypertension and year of diagnosis after the baseline questionnaire. Early-onset hypertension was defined as incident hypertension at age less than 55 years.[6,60,65,66]

Covariates

Age, smoking status, use of aspirin, acetaminophen, and non-steroidal anti-inflammatory drugs (NSAIDs), and for women, parity and current use of oral contraceptives and postmenopausal hormone therapy were assessed on biennial questionnaires, and potassium intake was evaluated on FFQs. Predicted DLW-based energy expenditure was also used as a covariate. In addition, participants reported race, ethnicity, and family history of hypertension. All covariates except race and family history of hypertension were used as time-varying covariates.

Statistical analysis

The person-time for each participant was calculated from the date of the return of the questionnaire in 1986 in NHS, 1991 in NHS II, and 1986 in HPFS, to the date at which hypertension was first diagnosed, death, or June 2017 in NHS and HPFS, and June 2018 in NHS II, whichever came first. Cumulative incidence curves that accounted for the time-varying associations were drawn for each exposure with age as the time scale. Confounder-adjusted associations between modifiable lifestyles factors and incident hypertension and early-onset hypertension were evaluated by three distinct approaches: hazard ratios, PAR %, and RMST differences.

Cox proportional hazards regression, with age as the time scale, was used to compute the hazard ratio (HR) and 95% confidence intervals (CIs) for categories of each factor. The counting process data structure was used to handle time-varying covariates and the left truncation issue. Models were stratified by age and calendar year of the current questionnaire cycle. Four modifiable lifestyle factors were categorized in accordance with current guidelines as follows[4]: BMI (<18.5 kg/m^2, 18.5 to 25 kg/m^2, 25 to <30 kg/m^2, 30 to <35 kg/m^2, and ≥35 kg/m^2); MVPA (<0.2 hours/week, 0.2 to <1 hours/week, 1 to <2.5 hours/week, 2.5 to <5 hours/week, and ≥5 hours/week); DASH score (quintiles); and alcohol (none, <5 g/day, 5 to <15 g/day, 15 to <30 g/day, and ≥30 g/day).

The PAR %[67], an estimate of the percentage of new hypertension cases occurring in this population that hypothetically could have been prevented if all participants had been in the guideline-recommended group, assuming causal relationships of the modifiable lifestyle factors and incident hypertension, was calculated for each of the dichotomized lifestyle factors. For the analyses, each lifestyle factor was dichotomized into adherence with guideline vs non–adherence with guideline categories as follows; BMI (<25 kg/m^2 vs. ≥25 kg/m^2); MVPA (≥2.5 hours/week

vs. <2.5 hours/week); DASH score (40th percentile or higher vs. lower than 40th percentile); and alcohol intake (men: ≤20 gram/day vs. >20 gram/day; women: ≤15 gram/day vs. >15 gram/day).

Confounders-adjusted RMST differences between the guideline-adhered vs non–adhered groups were evaluated. RMST difference in this context is interpreted as months free from hypertension in guideline-adherence group compared with non-adherence group, holding the covariates constant for each exposure.[68] 95% confidence intervals [CIs] for RMST differences were obtained based on N=500 bootstrapping. The cutoffs of the maximum follow-ups for the RMST calculation were defined as 25 years for hypertension outcome and 15 years for early-onset hypertension outcome. Dichotomized exposures were used for the analyses, which were same as the PAR models.

Models were adjusted mutually for all four modifiable risk factors and the following covariates: age at start of follow-up for the questionnaire cycle (years), calendar year, race (Caucasian or not), family history of hypertension (yes/no), smoking status (never/past/current), potassium intake (mg), use of aspirin (no/once per week/ more than once per week), acetaminophen (yes/no), non-steroidal anti-inflammatory drugs (yes/no)[69], and predicted total energy expenditure (kcal) in all cohorts; plus parity (none, 1, 2, or ≥3), menopausal status (premenopausal, postmenopausal, or unsure), and postmenopausal hormone therapy (never/ever/current) in NHS and NHS II; and plus contraceptive use (never/past/current) in NHS II. The four modifiable lifestyle factors and covariates excluding race and family history of hypertension were updated as time-dependent variables with each questionnaire cycle to reflect the most recent information. For PAR and RMST models, all covariates were categorized (age by 5 years; potassium intake and predicted total energy expenditure in quintiles). For RMST models, Covariate adjustment was done by setting the target values; median for categories

stemming from continuous covariates (e.g. quintile of potassium) and most frequent category for categorical covariates (e.g. smoking).

We conducted sensitivity analyses by use of self-reported hypertension combined with the initiation of antihypertensives as the outcome, by excluding those using any antihypertensives from the source populations, and using DASH score based on predicted absolute (i.e. not energy-adjusted) sodium levels. We also examined the contribution of smoking on incident hypertension by including the time-varying exposure (a total of five exposures). In NHS II, contribution of optimal weight, defined as BMI<25 kg/m^2 as well as waist<80cm, was additionally evaluated.

All statistical analyses were performed using SAS statistical software version 9.4 (SAS Institute Inc, Cary, North Carolina).

2.4. Results

Baseline characteristics are summarized in **Table 2.1**. At baseline, the mean (SD) age was 52.1 (7.1) years in NHS, 36.4 (4.6) years in NHS II, and 52.7 (9.5) years in HPFS. The mean BMI was 24.8 (4.5) kg/m^2 in NHS, 24.3 (5.0) kg/m^2 in NHS II, and 25.2 (3.1) kg/m^2 in HPFS. Throughout the follow-up periods of 31, 27, and 31 years, incident hypertension was reported among 34,537 (65.4%), 32,520 (38.8%), and 15,937 (51.0%) participants in NHS, NHS II, and HPFS, respectively. **Figures 2.1-2.3** depict age-adjusted time-to-event associations between each of the four lifestyle factor and incident hypertension.

	NHS N=52,780	NHS II N=83,871	HPFS N=31,269
Age, years	52.1 (7.1)	36.4 (4.6)	52.7 (9.5)
Caucasian	50,074 (94.9%)	81,061 (96.7%)	28,543 (91.3%)
Family history of hypertension	22,935 (43.5%)	42,109 (50.2%)	9,820 (31.4%)
BMI, kg/m^2	24.7 (4.4)	24.3 (5.0)	25.2 (3.1)
Smoking			
Never	23,430 (44.4%)	55,263 (65.9%)	16,064 (51.4%)
Past	18,082 (34.3%)	18,531 (22.1%)	12,241 (39.2%)
Current	11,268 (21.4%)	10,077 (12.0%)	2,964 (9.5%)
Moderate-to-vigorous PA, hours/week	0.67 (0, 2.5)	1.7 (0.4, 4.1)	1.3 (0.2, 3.8)
Sodium, gram/day	3.8 (3.5, 4.2)	3.4 (3.2, 3.6)	3.7 (3.4, 3.7)
Potassium, gram/day	3.2 (2.5, 3.9)	2.8 (2.2, 3.5)	3.3 (2.6, 4.0)
Alcohol, gram/day	1.8 (0, 7.6)	0.9 (0, 3.5)	5.5 (0.9, 14.6)
DASH score, unit	22 (19, 26)	23 (19, 27)	23 (19, 27)

Table 2.1: Baseline characteristics of the study populations of Nurses' Health Study (NHS), NHS II, and Health Professional Follow-up Study (HPFS)

Values are frequency (%) for categorical variables and mean (SD) or median (IQR) for continuous variables. Sodium was predicted based on ML algorithms, and potassium and alcohol were based on FFQ.

Hazard ratios

Table 2.2A summarizes the association of modifiable lifestyles and incident hypertension using Cox proportional hazard models. After mutual adjustment of exposures as well as for other confounders, BMI, MVPA, DASH score, and alcohol were robustly associated with incident hypertension in dose-dependent manners across three cohorts. The associations were strongest with BMI, with HRs [95% CIs] of 2.13 [1.99, 2.27], 3.29 [3.12, 3.48], and 2.19 [1.93, 2.48] in BMI ≥35 kg/m^2 compared with 18.5 to <25 kg/m^2 in NHS, NHS II and HPFS, respectively. Alcohol intake ≥30 g/day was more strongly associated with incident hypertension in NHS II compared with other cohorts (HRs referenced of never-drinker: 1.28 [1.21, 1.35] in NHS, 1.73 [1.62, 1.83] in NHS II, and 1.32 [1.24, 1.40] in HPFS).

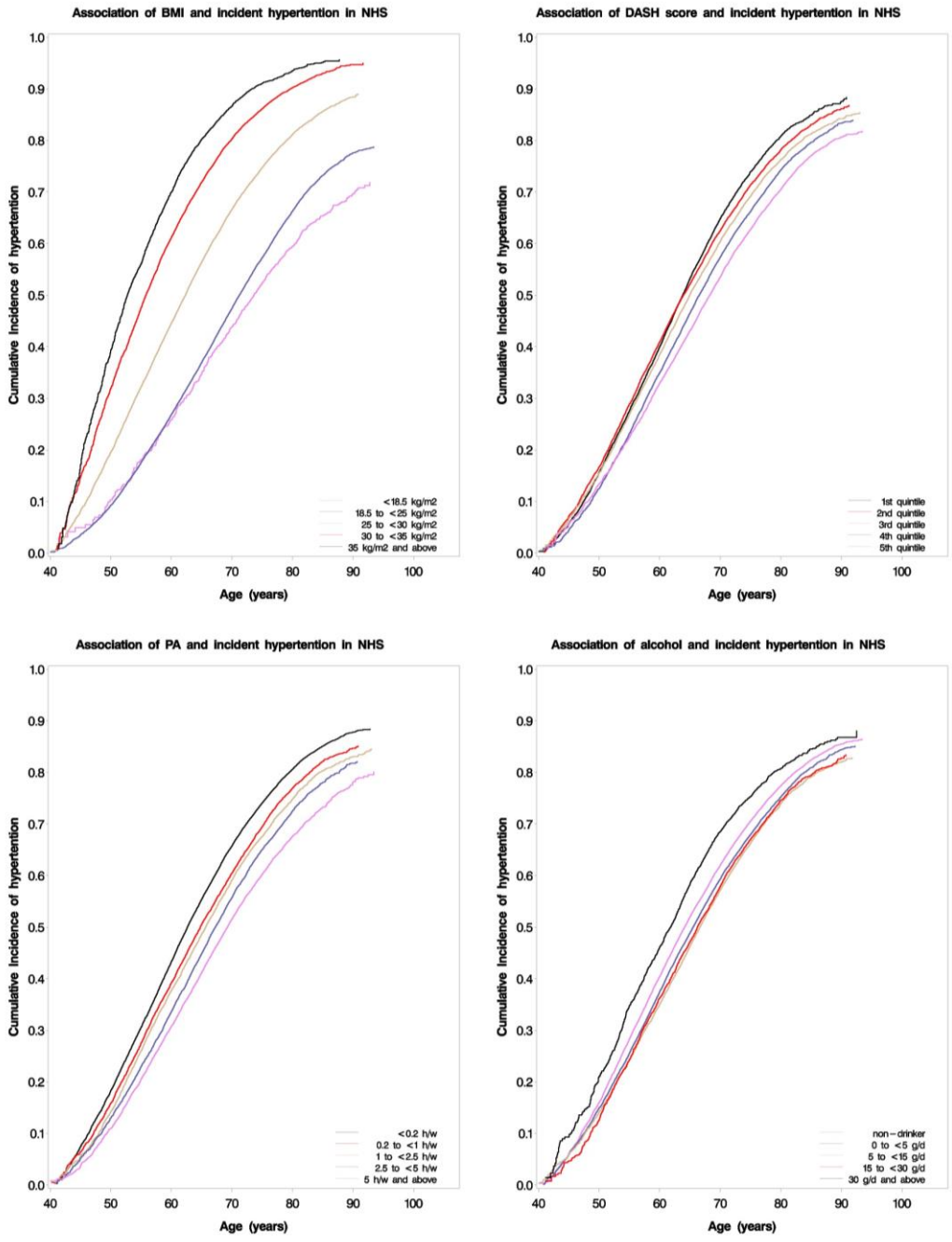


Figure 2.1: Cumulative incidence curves of hypertension according to each lifestyle factor in NHS

Cumulative incidence curves of hypertension by age (x-axis) in NHS by each lifestyle factor: BMI (<18.5 kg/m^2, 18.5 to 25 kg/m^2, 25 to 30 kg/m^2, 30 to 35 kg/m^2, and ≥35 kg/m^2); moderate-to-vigorous PA (<0.2 hours/week, 0.2 to <1 hours/week, 1 to <2.5 hours/week, 2.5 to <5 hours/week, and ≥5 hours/week); DASH score (quintiles); and alcohol (none, <5 g/day, 5 to <15 g/day, 15 to <30 g/day, and ≥30 g/day). Adherence to guideline-recommended lifestyles were represented as black, red, brown, blue, purple lines orderly from the best to worst in each exposure.

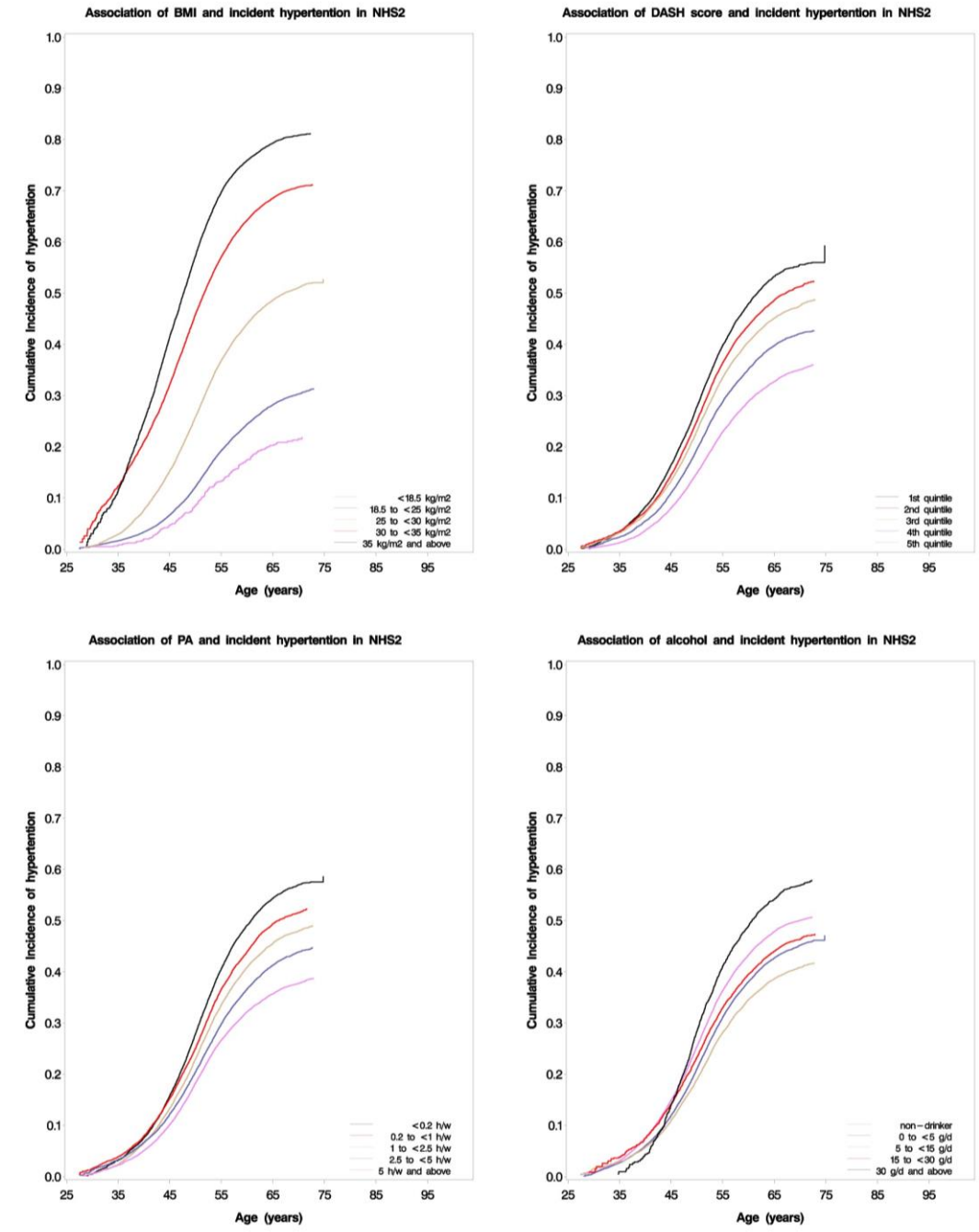


Figure 2.2: Cumulative incidence curves of hypertension according to each lifestyle factor in NHS II

Cumulative incidence curves of hypertension by age (x-axis) in NHS II by each lifestyle factor: BMI (<18.5 kg/m^2, 18.5 to 25 kg/m^2, 25 to 30 kg/m^2, 30 to 35 kg/m^2, and ≥35 kg/m^2); moderate-to-vigorous PA (<0.2 hours/week, 0.2 to <1 hours/week, 1 to <2.5 hours/week, 2.5 to <5 hours/week, and ≥5 hours/week); DASH score (quintiles); and alcohol (none, <5 g/day, 5 to <15 g/day, 15 to <30 g/day, and ≥30 g/day). Adherence to guideline-recommended lifestyles were represented as black, red, brown, blue, purple lines orderly from the best to worst in each exposure.

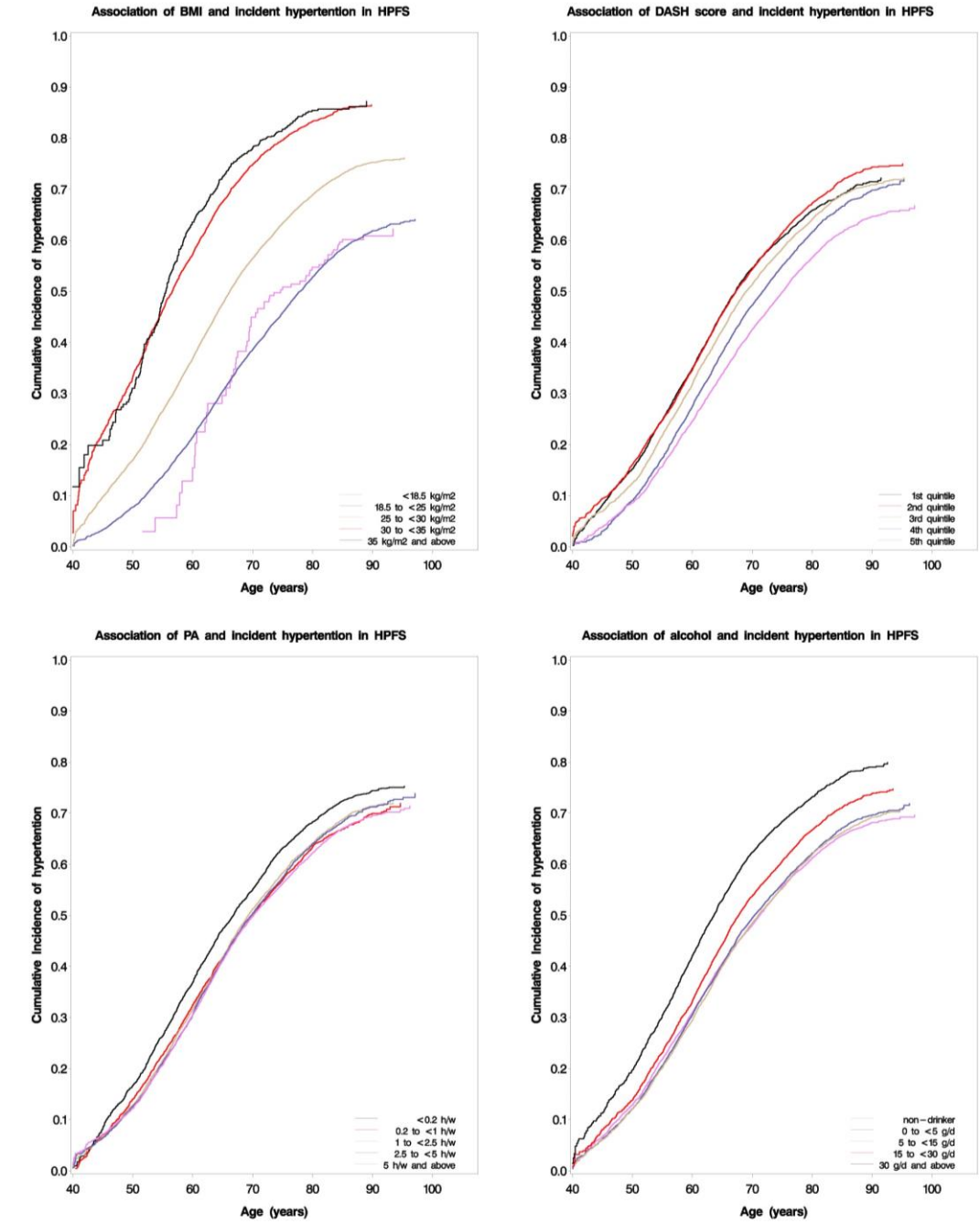


Figure 2.3: Cumulative incidence curves of hypertension according to each lifestyle factor in HPFS

Cumulative incidence curves of hypertension by age (x-axis) in HPFS by each lifestyle factor: BMI (<18.5 kg/m^2, 18.5 to 25 kg/m^2, 25 to 30 kg/m^2, 30 to 35 kg/m^2, and ≥35 kg/m^2); moderate-to-vigorous PA (<0.2 hours/week, 0.2 to <1 hours/week, 1 to <2.5 hours/week, 2.5 to <5 hours/week, and ≥5 hours/week); DASH score (quintiles); and alcohol (none, <5 g/day, 5 to <15 g/day, 15 to <30 g/day, and ≥30 g/day). Adherence to guideline-recommended lifestyles were represented as black, red, brown, blue, purple lines orderly from the best to worst in each exposure.

A. Incident hypertension

	NHS	NHS II	HPFS
BMI			
<18.5 kg/m2	0.90 (0.82, 0.98)	0.76 (0.66, 0.88)	1.02 (0.78, 1.34)
18.5 to <25 kg/m2	Reference	Reference	Reference
25 to <30 kg/m2	1.42 (1.38, 1.46)	1.81 (1.76, 1.87)	1.42 (1.36, 1.47)
30 to <35 kg/m2	1.85 (1.77, 1.93)	2.63 (2.52, 2.73)	2.00 (1.88, 2.14)
≥35 kg/m2	2.13 (1.99, 2.27)	3.29 (3.12, 3.48)	2.19 (1.93, 2.48)
Moderate-to-vigorous PA			
<0.2 hours/week	Reference	Reference	Reference
0.2 to <1 hours/week	0.94 (0.91, 0.97)	0.97 (0.94, 1.01)	0.93 (0.87, 0.99)
1 to <2.5 hours/week	0.96 (0.93, 1.00)	0.98 (0.94, 1.01)	0.98 (0.93, 1.04)
2.5 to <5 hours/week	0.94 (0.91, 0.98)	0.93 (0.90, 0.96)	0.98 (0.92, 1.03)
≥5 hours/week	0.91 (0.88, 0.94)	0.91 (0.88, 0.94)	0.95 (0.90, 1.00)
DASH score			
≤20th percentile	Reference	Reference	Reference
>20th to 40th percentile	0.96 (0.93, 0.99)	0.96 (0.92, 0.99)	0.98 (0.93, 1.03)
>40th to 60th percentile	0.92 (0.89, 0.95)	0.89 (0.86, 0.92)	0.94 (0.89, 0.98)
>60th to 80th percentile	0.91 (0.87, 0.94)	0.81 (0.78, 0.84)	0.93 (0.88, 0.98)
>80th percentile	0.88 (0.85, 0.92)	0.72 (0.69, 0.76)	0.87 (0.81, 0.92)
Alcohol			
Never	Reference	Reference	Reference
<5 g/day	0.98 (0.96, 1.01)	0.95 (0.92, 0.97)	1.03 (0.98, 1.08)
5 to <15 g/day	1.01 (0.98, 1.04)	0.98 (0.95, 1.01)	1.00 (0.95, 1.05)
15 to <30 g/day	1.10 (1.05, 1.15)	1.22 (1.16, 1.28)	1.12 (1.06, 1.18)
≥30 g/day	1.28 (1.21, 1.35)	1.73 (1.62, 1.83)	1.32 (1.24, 1.40)

B. Incident early-onset hypertension (incident hypertension in age < 55 years)

	NHS	NHS II	HPFS
BMI			
<18.5 kg/m2	1.01 (0.76, 1.34)	0.75 (0.63, 0.88)	0.45 (0.11, 1.79)
18.5 to <25 kg/m2	Reference	Reference	Reference
25 to <30 kg/m2	1.70 (1.59, 1.81)	1.90 (1.83, 1.97)	1.70 (1.55, 1.86)
30 to <35 kg/m2	2.50 (2.28, 2.74)	2.81 (2.69, 2.94)	2.86 (2.47, 3.31)
≥35 kg/m2	2.93 (2.55, 3.36)	3.58 (3.37, 3.81)	3.47 (2.71, 4.46)
Moderate-to-vigorous PA			
<0.2 hours/week	Reference	Reference	Reference
0.2 to <1 hours/week	0.97 (0.9, 1.04)	0.96 (0.93, 1.00)	0.93 (0.81, 1.06)
1 to <2.5 hours/week	1.01 (0.94, 1.09)	0.97 (0.93, 1.00)	0.89 (0.78, 1.00)
2.5 to <5 hours/week	0.91 (0.83, 0.98)	0.91 (0.87, 0.95)	0.88 (0.77, 0.99)
≥5 hours/week	0.88 (0.81, 0.96)	0.91 (0.87, 0.94)	0.83 (0.74, 0.94)
DASH score			
≤20th percentile	Reference	Reference	Reference
>20th to 40th percentile	1.06 (0.99, 1.13)	0.96 (0.93, 1.00)	0.92 (0.83, 1.02)
>40th to 60th percentile	0.98 (0.91, 1.06)	0.90 (0.87, 0.94)	0.93 (0.84, 1.04)
>60th to 80th percentile	0.94 (0.86, 1.02)	0.81 (0.78, 0.85)	0.90 (0.79, 1.02)
>80th percentile	0.91 (0.82, 1.01)	0.72 (0.68, 0.75)	0.86 (0.74, 0.99)

Table 2.2: Hazard ratios (95% CI) for modifiable lifestyles and incident hypertension and early-onset hypertension in Cox proportional hazard models among women and men in the Nurses' Health Studies and the Health Professionals Follow-up Study

Table 2 (continued)

Alcohol			
Never	Reference	Reference	Reference
<5 g/day	0.95 (0.90, 1.01)	0.93 (0.90, 0.96)	0.95 (0.86, 1.06)
5 to <15 g/day	0.99 (0.92, 1.07)	0.96 (0.92, 1.00)	0.94 (0.84, 1.04)
15 to <30 g/day	1.11 (0.99, 1.24)	1.20 (1.13, 1.28)	1.05 (0.93, 1.19)
≥30 g/day	1.49 (1.32, 1.70)	1.72 (1.59, 1.86)	1.36 (1.19, 1.55)

Numbers were hazard ratios (95% confidence intervals).
The multivariable-adjusted models were mutually adjusted for four modifiable risk factors and the following covariates: age, race , family history of hypertension, smoking status, potassium intake, use of aspirin, acetaminophen, and NSAIDs, and energy expenditure in all cohorts; plus parity, menopausal status, and hormone replacement therapy in NHS and NHS II; and plus contraceptive use in NHS II.

Associations of lifestyles with incident early-onset hypertension (incident hypertension in age<55 years) was generally similar to the overall except for BMI in hazard ratio scale (**Table 2.2B**). BMI was more strongly associated with early-onset hypertension than with overall hypertension, with HRs of 2.93 [2.55, 3.36], 3.58 [3.37, 3.81], and 3.47 [2.71, 4.46] in BMI ≥35 kg/m^2 compared with 18.5 to <25 kg/m^2 in NHS, NHS II and HPFS, respectively.

Population attributable risks

Results of analyses on PARs are summarized in **Table 2.3A**. The highest partial PAR values for each cohort were for BMI, with 19.6% [18.1, 21.1] in NHS, 35.5% [33.7, 37.2] in NHS II, and 21.7% [19.4, 23.9] in HPFS. MVPA accounted for approximately 10% of incident hypertension in NHS and NHS II, while the contribution was smaller in HPFS (12.3% [10.7, 14.0] in NHS, 10.8% [9.5, 12.1] in NHS II, and 2.5% [1.1, 3.9] in HPFS). Contribution of DASH diet was higher in NHS II (14.0% [12.4, 15.6]) compared with NHS (4.0% [2.5, 5.5]) and HPFS (3.4% [1.1, 5.7]).

The PARs of lifestyles for incident early-onset hypertension were generally stronger than those for incident hypertension in older individuals (**Table 2.3B**). Notably, the contributions of BMI were 30.6% [27.4, 33.7] in NHS and 33.6% [28.9, 38.0] in HPFS for early-onset hypertension, which was approximately 50% higher than those for incident hypertension overall. The PARs of MVPA and the DASH diet were more than two times higher in NHS and HPFS for early-onset hypertension compared with incident hypertension overall.

A. Incident hypertension

	NHS	NHS II	HPFS
BMI	19.6 (18.1, 21.1)	35.5 (33.7, 37.2)	21.7 (19.4, 23.9)
Moderate-to-vigorous PA	12.3 (10.7, 14.0)	10.8 (9.5, 12.1)	2.5 (1.1, 3.9)
DASH score	4.0 (2.5, 5.5)	14.0 (12.4, 15.6)	3.4 (1.1, 5.7)
Alcohol	1.2 (0.8, 1.6)	2.6 (2.2, 2.9)	3.5 (2.6, 4.3)

B. Incident early-onset hypertension (incident hypertension in age < 55 years)

	NHS	NHS II	HPFS
BMI	30.6 (27.4, 33.7)	37.5 (35.6, 39.5)	33.6 (28.9, 38.0)
Moderate-to-vigorous PA	13.8 (9.9, 17.7)	11.2 (9.7, 12.8)	5.5 (1.9, 9.0)
DASH score	9.6 (5.1, 14.1)	15.1 (13.1, 17.0)	7.5 (11.0, 13.9)
Alcohol	2.0 (1.2, 2.9)	2.1 (1.8, 2.5)	4.4 (2.6, 6.3)

Table 2.3: Association of modifiable lifestyles and incident hypertension and early-onset hypertension in population attributable risk (PAR) % among women and men in the Nurses' Health Studies and the Health Professionals Follow-up Study

Numbers indicate PAR % (95% CI) computed for each dichotomized lifestyle factor; BMI (<25 kg/m2 vs. ≥25 kg/m2); moderate-to-vigorous physical activity (≥2.5 hours/week vs. <2.5 hours/week); DASH score (40th percentile or higher vs. lower than 40th percentile); and alcohol intake (≤20 gram/day vs. >20 gram/day in male; ≤15 gram/day vs. >15 gram/day in female).

Models were mutually adjusted for four modifiable risk factors and the following covariates: age, race, family history of hypertension, smoking status, potassium intake, use of aspirin, acetaminophen, and NSAIDs, and energy expenditure in all cohorts; plus parity, menopausal status, and hormone replacement therapy in NHS and NHS II; and plus contraceptive use in NHS II.

Periods free from hypertension

RMST differences until the cutoff years of the follow-ups are summarized in **Table 2.4**. Periods free from incident hypertension [95% CI] were on average longer by 20.3 [18.5, 22.0] months in NHS, 25.0 [23.2, 26.8] months in NHS II, and 18.6 [16.7, 20.7] months in HPFS, among those with BMI ≤25 kg/m^2 compared with >25 kg/m^2, in 25-year follow-up. RMST differences for early-onset hypertension were similarly greater by adhering optimal weight and alcohol intake compared with other lifestyles.

A. Incident hypertension by a cutoff of 25 years

	NHS	NHS II	HPFS
BMI	20.3 (18.5, 22.0)	25.0 (23.2, 26.8)	18.6 (16.7, 20.7)
Moderate-to-vigorous PA	3.6 (2.5, 4.8)	1.9 (1.2, 2.5)	0.3 (-1.1, 1.9)
DASH score	2.1 (0.9, 3.3)	3.2 (2.6, 4.1)	2.4 (1.0, 4.2)
Alcohol	7.1 (5.3, 9.0)	9.9 (8.5, 11.4)	8.9 (7.1, 10.8)

B. Incident early-onset hypertension (incident hypertension in age < 55 years) by a cutoff of 15 years

	NHS	NHS II	HPFS
BMI	6.0 (4.6, 7.8)	3.9 (3.5, 4.2)	6.7 (5.2, 8.4)
Moderate-to-vigorous PA	1.0 (0.5, 1.5)	0.3 (0.2, 0.4)	0.7 (0.0, 1.5)
DASH score	0.7 (0.2, 1.2)	0.5 (0.4, 0.6)	0.5 (-0.1, 1.1)
Alcohol	2.3 (1.4, 3.5)	1.5 (1.3, 1.8)	2.7 (1.7, 4.0)

Table 2.4: Months free from hypertension and early-onset hypertension by adhering optimal lifestyles among women and men in the Nurses' Health Studies and the Health Professionals Follow-up Study

Numbers indicated months (95% CI) free from hypertension or early-onset hypertension by adhering optimal lifestyles, with a cutoff of 25 years for incident hypertension and 15 years for early-onset hypertension.
RMST differences were computed for each dichotomized lifestyle factor; BMI (<25 kg/m2 vs. ≥25 kg/m2); moderate-to-vigorous physical activity (≥2.5 hours/week vs. <2.5 hours/week); DASH score (40th percentile or higher vs. lower than 40th percentile); and alcohol intake (≤20 gram/day vs. >20 gram/day in male; ≤15 gram/day vs. >15 gram/day in female).
Models were mutually adjusted for four modifiable risk factors and the following covariates: age, race , family history of hypertension, smoking status, potassium intake, use of aspirin, acetaminophen, and NSAIDs, and energy expenditure in all cohorts; plus parity, menopausal status, and hormone replacement therapy in NHS and NHS II; and plus contraceptive use in NHS II. Covariate adjustment was done by setting the target values; median for categories stemming from continuous covariates (e.g. quintile of potassium) and most frequent category for categorical covariates (e.g. smoking).

Sensitivity analyses

We conducted sensitivity analyses using various PAR models (**Supplemental Table 2.1**). The patterns of the associations between lifestyle factors and incident hypertension and early-onset hypertension were generally similar to those of the main analyses, with the largest contributions by BMI in particular for early-onset hypertension. In HPFS, the contributions of BMI, MVPA, and DASH diet were smaller when the hypertension outcome was defined as self-reporting plus initiating anti-hypertensives; in contrast, for early-onset hypertension, the stricter outcome definition was associated with higher PAR of MVPA, while there was much uncertainty due to limited sample size. An inclusion of smoking as a 5th exposure did not influence the contributions of other lifestyle factors, with little contribution of smoking to the outcomes. Using predicted absolute sodium instead of energy-adjusted sodium in DASH score led to decreased PARs of diet. Finally, in NHS II, adhering to both optimal BMI and waist had even larger PARs (38.6% for overall hypertension and 40.0% for early-onset hypertension).

2.5 Discussion

In three large cohorts of US adults, adhering to a healthy lifestyle was associated with lower risk of incident hypertension in approximately 30 years of follow-up with stronger associations at younger ages among men and women. Higher BMI was most consistently and strongly associated with risk of incident hypertension across all three cohorts, accounting for approximately 20% of hypertension overall and 35% of early-onset hypertension. The RMST analysis showed that there would be an additional 20-30 months free from hypertension on average by maintaining a healthy BMI throughout 25-year follow-up. MVPA and diet accounted

for 10-15% of early-onset hypertension in women. The present observations highlight not only the potential impact of primordial prevention of hypertension, but the significance of prevention strategies starting at younger ages, focusing most importantly on body weight.

In general, unhealthy lifestyle choices are known major causes of middle-to-late-onset hypertension, whereas early-onset hypertension is usually considered a highly heritable trait.[6,60] The present study, in contrast, found that lifestyle modification at younger ages could contribute to preventing up to half of early-onset hypertension. Since early-onset hypertension is associated with potentially more devastating consequences such as increased CVD incidence and mortality[58,70], our observation is particularly important and should lead to enhanced efforts by medical professionals to screen and provide guidance for primordial prevention via lifestyle modification in younger populations. We observed relatively modest contributions of lifestyle factors in older cohorts with mean age of 52-53 years, which likely is due to depletion of the most susceptible from the at risk pool and because with many decades of exposure to dozens of small risk factors there is a dilution of effect of the five we identified. However, lifestyle choices still accounted for approximately 30% of the incident hypertension in older populations, highlighting the importance of choosing a healthy lifestyle across the adult lifespan. Other lifestyle scores have shown a similar pattern for primordial prevention of CVD based on risk exposure from mid to later life[71].

Our findings indicate that a BMI below 25kg/m^2 may be the cornerstone to preventing incident hypertension, concordant with multiple RCTs[52,72] that documented blood pressure lowering effects by weight loss interventions. Higher MVPA and better adherence to the DASH

diet also importantly contributed to lower hypertension risk (~10% depending on the population). The lower effect estimate for physical activity could partly be attributable to the importance of BMI on mediating the effects of PA on incident hypertension. The effect size of alcohol abstinence or excessive consumption was modest and might be driven by the non-linear effects of alcohol as well as on the importance of drinking patterns rather than just average alcohol consumption on incident hypertension[73]. We did not separate the contribution of sodium from other diet because of the complexity of sodium assessment in the diet, where it is spread across most foods with tremendous variability (due to processing, cooking methods, and individual preference for adding salt to meals), leading to measurement errors, which were not addressed by our ML algorithm. There may be other contributions to the heterogeneity of the causal role of sodium in raising BP due to sodium sensitivities according to age, genotypes, and other factors.[4,74–76] The present study does not undermine the roles of PA or diet in reducing the risk of hypertension but emphasizes that weight loss may play an essential role regardless of other lifestyle choices. This is further supported by a recent trial showing that weight loss due to glucagon-like peptide-1 receptor agonists resulted in a pronounced reduction in BP.[72]

Strengths and limitations

Strengths of our study include the prospective cohort designs, large sample size, very long follow-up periods, large number of events, detailed information on repeatedly measured lifestyle factors and the covariates over follow-up periods, a use of newly developed ML algorithms predicting sodium intake, and rigorous sensitivity analyses.

Our study also has several limitations. First, the interpretation of PAR depends on assumption that the association is causal, which might be too strong given the observational

study design. However, many of the lifestyle factors we chose have been shown to be causally related to BP lowering. Second, the populations were predominantly white health professionals and thus our findings might not be generalizable to other populations or cultures. However, the homogeneity of the study populations and comprehensive lifestyle assessment minimized potential confounding by a myriad of factors associated with lower socio-economic status. Third, self-reported lifestyles and hypertension status could have led to some modest misclassification of exposure or to under screened populations, although we have previously validated self-report of PA[77], diet[18], weight[78], alcohol[18,79], and hypertensive status.[36–38] In addition, we performed several sensitivity analyses, which provided generally consistent results. Still, sodium was subject to measurement error despite using ML prediction, and the impact of measurement error could particularly impact PAR results due to misclassification-caused biases in the prevalence.[80] Finally, reverse causation was possible. With BP within a prehypertension range, the participants might have been advised to changes lifestyles, such as reducing weight. Such biases would generally underestimate the contributions of lifestyle factors to incident hypertension.

Conclusion

Our data provide a strong public health message that avoiding overweight and obesity will substantially reduce the burden of hypertension among all adults, with the strongest contribution among populations under 55 years of age. Primordial prevention through weight control combined with a healthy DASH diet, avoidance of excessive alcohol consumption, and moderate or vigorous aerobic activity could substantially reduce hypertension risk, in particular early-onset hypertension in the United States.

Chapter 3

Methodological approaches to studies related to heterogeneous treatment effect and individualized treatment rule based on randomized controlled trials

Author: Rikuta Hamaya, MD, PhD, MS[1,2], Konan Hara, MD, PhD[3], JoAnn E Manson, MD, DrPH[1,2,4], Eric B Rimm, ScD[1,5,6], Frank M. Sacks, MD[5], Nancy R Cook, ScD[1,2]

[1] Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[2] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[3] Department of Econometrics, University of Arizona, Tucson, AZ

[4] Mary Horrigan Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[5] Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

[6] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

3.1 Abstract

Estimating conditional average treatment effect (CATE) and developing an individualized treatment rule (ITR) using a well-powered randomized controlled trial (RCT) may directly inform practical evidence to advance precision medicine, while there are some pitfalls in the application methodology. We overviewed methodological approaches of heterogeneous treatment effect (HTE)-related studies, which differ according to the study aims: 1) to investigate the sources of effect heterogeneity; 2a) to derive and validate a new ITR; and 2b) to validate existing ITRs in a new trial dataset. Meta-learners and causal forest are popular and easy methodologies to predict CATE, while double machine-learning including R-learner and doubly-robust learner are more efficient in high-dimensional settings. ITR evaluation approaches include population average value (PAV), population average prescription effect (PAPE), area under the prescriptive effect curve (AUPEC), and the sorted grouped average treatment effect (GATE). In the application with the POUNDS Lost trial, we demonstrate the development and validation of an effective ITR of fat diet intervention on 2-year weight change.

3.2 Introduction

Evidence-based medicine, prevention and epidemiology, as well as evidence-based policy-making (EBPM), have largely evolved from the evaluation of average treatment effects (ATE) in target populations, primarily relying on data from randomized controlled trials (RCTs). However, treatment effects are often heterogeneous depending on participants' characteristics, leading to the presence of heterogeneous treatment effects (HTEs).[9,81] Recent advancements in

machine learning (ML) approaches have led to significant developments in applications related to HTE, in particular in high-dimensional settings.[7,8,33,82–85] HTE exploration could lead to a development of individualized treatment rules (ITRs) – algorithms that optimize intervention delivery to maximize population-level treatment effects – that clinicians and policymakers ultimately need.[10,11] Rigorous medical and epidemiological studies, combined with valid statistical techniques for development and evaluation of ITRs, will usher in a new era of evidence-based personalized medicine, precision nutrition, and individualized prevention.

HTEs have been thoroughly investigated in medical RCTs, typically through analysis of effect measure modification (EMM) by a single (often categorical) variable. Generally, the variables of interest, known as effect modifiers, are pre-specified in an RCT analysis plan.[81] Regression models incorporating interaction terms between the intervention variable and each effect modifier are employed to evaluate EMM, with assessments often based on p-values of the coefficient of the interaction term. Although this conventional approach is generally pre-specified and boosts interpretability, it suffers from multiple testing issues and cannot capture nonlinear contributions from continuous effect modifiers or complex interactions between effect modifiers. In addition, developing an ITR may better incorporate information from multiple covariates. This is where ML-based, data-driven approaches for exploring HTE have an advantage.

There are distinct study aims related to HTEs, and the methodological approaches can greatly differ accordingly – however, the focuses of current application studies are often not clear, partly because similar methods can be used for different aims. Specifically, the major aims

of HTE-related studies can be 1) to investigate the sources of effect heterogeneity and 2) to derive and validate a new ITR. In the latter situation, methodology can differ if researchers validate the ITR in an external dataset or not. In this paper, we sought to outline research concepts and approaches according to HTE-related study aims. Although a thorough understanding of causal inference language and relevant statistics is necessary to grasp these concepts formally, we strive to explain them practically without resorting to technical jargon. We focus on the setting of a well-powered RCTs in which identifiability assumptions are met.[86] For simplicity, we consider the intention-to-treat basis and do not account for post-randomization selection bias.[87,88] We do not cover the methodological extension to survival data.

In Part A, we provide an overview of different methodological approaches according to the study aims. In Part B, we introduce estimation methods for conditional average treatment effect (CATE) and the cross-fitting approach, which are the basis of HTE-related research. In Part C, we present different methodological approaches for studies aiming at interpreting sources of HTE. In Part D, we explain various evaluation approaches for studies aiming at developing an ITR. Lastly, in Part E, we present an applied example using the Preventing Overweight Using Novel Strategies (POUNDS Lost) trial - a 2x2 factorial RCT of high vs. low fat and high vs. low protein diets.[12] In this application, we created and evaluated ITRs of distinct macronutrient compositions on two-year weight loss outcomes according to baseline characteristics. To enhance the readability of this paper, **Table 3.1** summarizes the terminology specific to HTE-related research.

Types of treatment effects	
ATE = average treatment effect	As it stands, the average effect of an intervention on an outcome. Usually the primary target to estimate in RCTs

CATE = conditional average treatment effect	Treatment effect conditional on the participants' characteristics (≒individualized treatment effect [ITE])
Concepts related to heterogeneous treatment effect	
HTE = heterogeneous treatment effect	Variations in the treatment effect across individuals or subgroups within a study population.
ITR = individualized treatment rule	A rule that assigns a treatment to an individual based on their observable characteristics with the goal of optimizing an outcome
High-dimensional setting	In a situation where many covariates need to be accounted for with a limited sample size. This is almost always the case when aiming to estimate CATE in medical and epidemiological RCTs.
CATE estimation approaches	
Causal Forest	An ML algorithm adapting random forest specifically designed to estimate CATE. Causal forest is essentially a DML.
Meta-learner	Overarching algorithms that learn from different models, aiming to estimate CATE. Examples include S-, T-, X-, DR-, and R-learner.
S-learner	A type of meta-learner based on a single model that includes treatment indicators as predictors.
T-learner	A type of meta-learner training separate models for the treated and control groups.
DR-learner	A type of meta-learner leveraging doubly robust estimator. One of the DML approaches.
R-learner	A type of meta-learner leveraging residuals of outcome and propensity score models. One of the DML approaches.
Cross-fitting	Similar to cross-validation, but it is used to calculate CATE in all samples in the same dataset in which model training is conducted.
Concepts related to DML	
DML = double machine learning	A statistical approach to estimate ATE or CATE in which scores obeying Neyman orthogonality conditions are used, including DR-learner and R-learner.
Nuisance model	Models that are not of direct interest but that must be accounted for in the analysis, including the outcome and propensity score models in DR-learner or R-learner.
Doubly robust estimator	An estimator in which correct specification of either the outcome model or the propensity score model is necessary to get unbiased estimates, but not necessarily both. DR-learner is based on this estimator.
ITR evaluation approaches	
PAV = population average value	The expected average outcome under a particular treatment rule over the entire population.
PAPE = population average prescription effect	Difference between the overall average outcome in the population under an ITR versus under random assignment of the intervention with a fixed proportion of the intervention assignment as in the ITR.
AUPEC = area under the prescriptive effect curve	Area between the curve indicating PAV and the straight line indicating the average outcome under random assignment of the intervention, with varying proportion of intervention
GATE = the sorted grouped average treatment effect	An estimate of the average treatment effect within pre-specified subgroups formed by sorting individuals according to their estimated CATEs.
TOC = targeting operator characteristic	The difference between an average effect among those with ranks of the priority score higher than a cutoff and ATE.

AUTOC = are under the TOC curve	Area between the curve indicating TOC and the straight line indicating ATE with varying the quantile cutoff for TOC.
RATE = the rank-weighted average treatment effect	Weighted average of TOC. An example is AUTOC. RATE metrics only depend on the ranks of priority score (e.g. estimated CATE).

Table 3.1: Terminologies of methods related heterogeneous treatment effects

3.3 Part A: An overview of the methodological approaches according to the study aims

Clarifying the objective of an applied HTE research is pivotal to conducting high-quality applied research. There are two major distinct aims – 1) to explore the sources of effect heterogeneity in interpretable ways; and 2) to develop and validate an ITR. Here we focus on these two most clinically relevant aims, named Study type 1 and type 2, to explain the conceptual and methodological differences – although we acknowledge that there can be other aims, such as developing a recruitment algorithm to an RCT to maximally detect HTE. Note, for ITR evaluation, the methodological approaches can differ according to whether the evaluation is conducted in an internal (or derivation) dataset or an external one, which are named Study type 2a and 2b in the following.

In either study aim, estimating the CATE, referring to a potentially different treatment effect conditional on a combination of specific participant characteristics[83], can be a major milestone. CATE largely corresponds to the concept of "individualized treatment effect (ITE)", and these have used these terms interchangeably in applied research. Strictly speaking, CATE is not necessarily a treatment effect for an individual but an effect conditional on combinations of

multiple covariates,[89] whereas such distinction may not be practically important. In general, there are numerous combinations of the covariates, and the estimation of CATE can be challenging even in a large-scale RCT, i.e. a high-dimensional setting (limited sample size relative to the number of covariates). This is because the primary aim of any medical/epidemiological RCTs is to estimate the ATE, treatment effect in a target population, and the trials are typically only powered for ATE estimation.[90] CATE can be estimated for all participants in an RCT with use of a cross-fitting approach. A detailed description of estimation approaches to CATE is provided in **Part B**.

In this section, we outline study aim-specific methodological approaches using CATE in an RCT. An overview of the distinct aims and approaches of HTE-related studies is summarized in **Figure 3.1**.

Study type 1: To explore the source of effect heterogeneity in interpretable ways

In epidemiology and medicine, practitioners almost always need a 'rational' for employing an intervention. Documented and statistically tested ATE in a well-conducted RCT is typically considered strong evidence supporting such decisions. At the same time, potentially different treatment effects according to population characteristics are often taken into account, in particular when the intervention is costly and reasonable EMM is observed. Therefore, interpreting the sources of effect heterogeneity align with current use of health interventions and are essential to personalizing existing interventions.

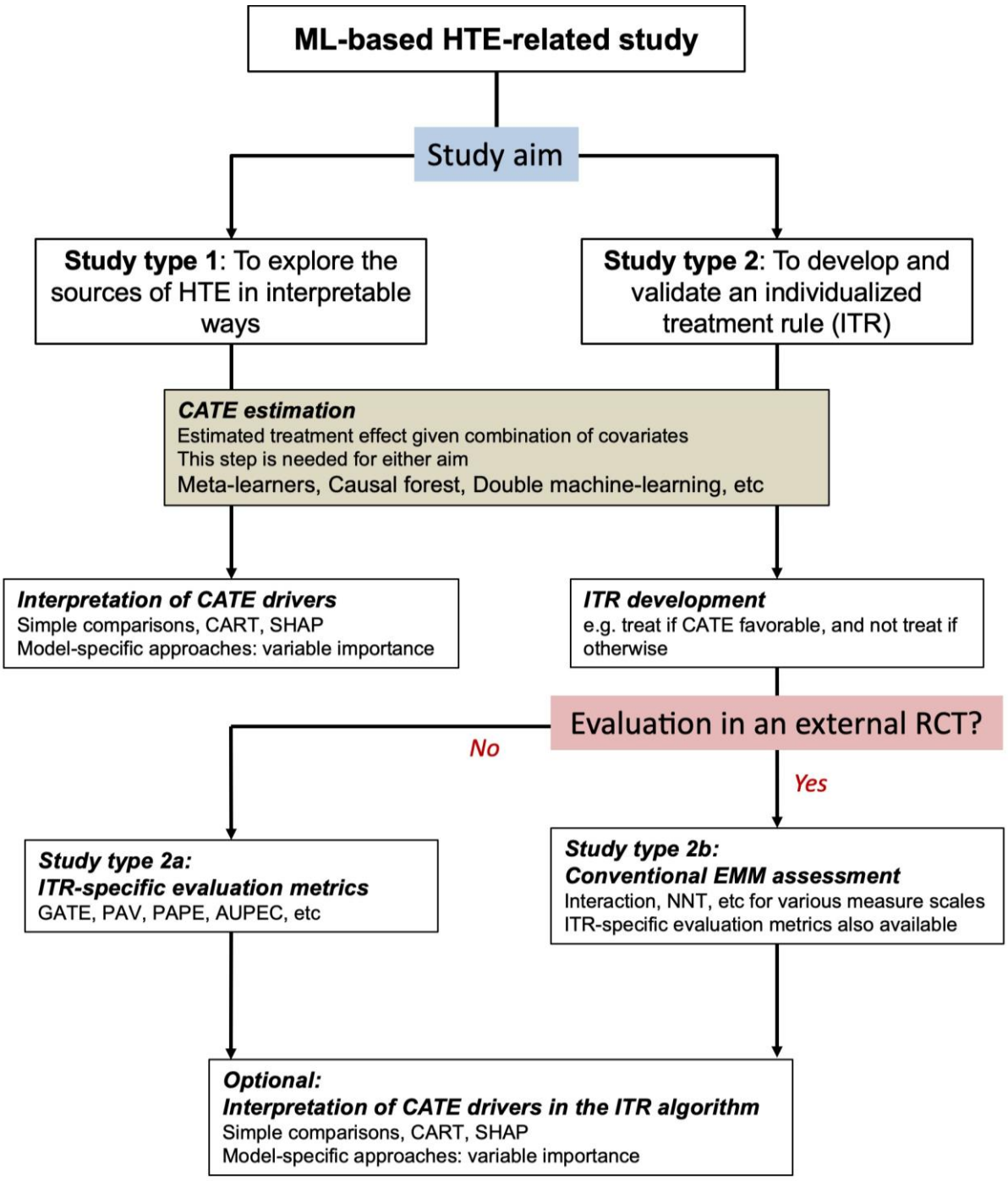


Figure 3.1: Types of studies related to heterogeneous treatment effect

There are major two study types: to explore the sources of heterogeneous treatment effect (HTE) (type 1) and to develop and validate an individualized treatment rule (ITR). In each aim, the first step is to estimate conditional average treatment effect (CATE), for example using meta-learners, causal forest, and double machine learning. For study aim 1, the sources of HTE is explored by interpreting the drivers of estimated CATE. For study aim 2, an ITR is developed based on estimated CATE – and the evaluation may be according to whether the evaluation is internal (type 2a) or external (type 2b). Optionally, as is in study aim 1, the drivers of estimated CATE may be explored in study aim 2.

When the study aim is to explore sources of effect heterogeneity, a simple, very effective approach is to assess the EMM by single covariates one at a time, as is almost always implemented in recent RCTs. Candidate effect modifiers are best if they are a priori specified to avoid "p-hacking", and this simple approach boosts interpretability with little risk of model misspecification. Using ML-based HTE methods may shed light on other aspects, for example accounting for complex interactions and avoiding multiple testing issues, while these complex approaches may or may not complement the simple EMM evaluation.

Generic approaches to interpreting CATE

Given that CATE is validly estimated for all the individuals in an RCT, there are several 'generic' methodologies to explore the source of effect heterogeneity for any approaches to CATE estimation; 'generic' in terms of what can be done irrespective of how CATE is estimated. Three approaches are given below that can be easily implemented while informative:

1) One simple approach is to compare covariates one by one according to quartile (or quintile etc) of the estimated CATE, similar to Table 1 in any epidemiological research. Standardized mean difference (SMD) may be used to quantify the difference across the CATE quartile. This allows an interpretation of which variables are correlated with the estimated CATE.

2) Another way is to show characteristics of the participants with top and bottom estimated CATE.[82] For example, by comparing individuals with top 10 and bottom 10 CATE, one can qualitatively understand which characteristics may drive the CATE estimates.

3) A Classification and Regression Tree (CART) can be employed to visualize the contributions of covariates to the estimated CATE.[91] Fitting CART using the estimated CATE as the outcome variable, a decision tree can be visualized showing the interactions between covariates to determine CATE. Variables that appear in the decision tree can be considered as those driving the HTE.

Another more advanced approach includes a use of Shapley Additive Explanations (SHAP)[92,93], one of popular approaches to explain black-box ML algorithms. SHAP is rooted in cooperative game theory and quantifies the importance of each covariates' impact on a model's prediction. Briefly, SHAP determines how much every feature sways the prediction from a default baseline. By giving a detailed breakdown, it offers a clear understanding of the factors driving CATE values.

Model-specific approaches to interpret CATE

In recent application studies, the variable importance (VI) measure in the Causal forest algorithm (see **Part B–CATE estimation**) is often used to interpret CATE. Similar to random forest algorithm, the VI measure in the Causal forest indicates the relative importance of a variable in estimating the treatment effect. A variable is deemed to be important if splitting on that variable greatly improves the accuracy of predicting the outcome difference between intervention and control groups. When using doubly robust-learner or R-learner (also see **Part B –CATE estimation**), we can get direct estimates of the contribution to CATE of all covariates. In those two-stage approaches, we can get interpretable values for each covariates especially when resorting to simple linear regression in the final (i.e. 2nd stage) models.

Of note, caution is needed that causality of the EMM may not be inferred by using any of the above approaches to interpret CATE. Differentiating causal effect modifiers and surrogate effect modifiers is an issue of causal structure.[86] Thus, even if one variable has a high influence in differentiating HTE, one may not know how much the variable performs in another dataset.

Study type 2a: To develop and validate an ITR in a single RCT

Another major study aim related to HTE exploration is to develop an ITR, a tailored treatment rule according to participants' characteristics, maximizing the benefit of the intervention. An ITR is a treatment rule based on a scoring rule, that is, a single covariate. CATE is an intuitive scoring rule since it is rationale to treat someone only if the estimated CATE indicates a treatment benefit. Although the scoring rule may be based on another metric, here we focus on the case with use of estimated CATE.

We can estimate CATE in all the RCT participants using cross-fitting (see Section 2–Cross fitting). The next step is ITR evaluation. For this purpose, we would like to observe two potential outcomes[86,94] per participant, either treated or not treated, in order to directly evaluate the ITR performance. However, this is impossible, and thus we need evaluation metrics specifically designed for ITR evaluation. Such metrics should reflect the performance of an ITR with respect to the overall treatment effect in the target population. In addition, the evaluation methods should be generalizable to any ITR development approach in practice. Given that the CATE estimation approach is data-driven, traditional epidemiological approaches may not be

valid for an ITR evaluation, such as using it in a simple interaction in a Cox proportional hazard model and referring the p-value of the coefficient of the interaction term. Statistically valid metrics include the population average value (PAV)[10,11], the population average prescription effect (PAPE)[10], the area under the prescriptive effect curve (AUPEC), and the sorted group average treatment effect (GATE).[85,95] Details of these metrics are explained in **Part C**.

An ITR can be a 'black-box' algorithm suggesting who to treat or not, while an interpretation of the black-box is generally informative in thinking of an actual application. In essence, interpretation of an ITR is same as interpretation of the CATE that shapes the ITR. Therefore, for this purpose, the same approach to Study type 1 can be applied, such as simple comparison across CATE quartile or a use of CART using estimated CATE as the outcome variable. However, this endeavor is only to supplement the understanding of an ITR, different from the aim of Study type 1, although the same methodologies can be used.

Study type 2b: Validation of an externally developed ITR

Just like studies aiming at developing a prediction model, it is important to evaluate a performance of an ITR in external settings. In cases of ITR, the external dataset should be an RCT with the same (or similar) intervention as that of the derivation RCT. Although similar to Study type 2a, external evaluation approaches can be more flexible.

If an ITR is developed externally, the resulting dichotomous variable indicating who to treat can be used safely with usual epidemiological methods, such as interaction analysis for

EMM evaluation. The analytic approach is same as a "risk score" based on several variables.[81] For the assessment of EMM, it is advisable to evaluate several scales, including risk difference (e.g. in linear regression) and risk ratio (e.g. in pooled logistic regression) scales, since the extent of effect modification depends on the outcome measure.[86] Contrasting number needed to treat (NNT) metrics between participants with distinct ITR value is also informative. Using PAV, PAPE, AUPEC, and GATEs is also valid and informative when contrasting the performance of an ITR in internal (i.e. derivation) versus external dataset.

3.4 Part B: CATE estimation and cross-fitting

As stated in a previous section, estimating CATE, individualized treatment effect conditional on a combination of specific participant characteristic, is necessary in each study aim. Here we describe possible approaches to estimate CATE based on RCTs, including meta-learners, double machine learning (DML), and causal forest. Then, an explanation of cross-fitting, an important technique of training/test data splitting, follows. Throughout this section, the aim is set to estimate CATE for every participant in an RCT while avoiding bias.

CATE estimation: naïve but valid approaches

Meta-learners comprise a family of ML methods that integrate existing supervised ML algorithms to estimate CATE through a two-step estimation process: model training and CATE estimation.[8,96] Meta-learners can be classified into several types, including T-learner, S-learner, X-learner, R-learner, and doubly-robust (DR)-learner. Among them, we discuss the two simplest

approaches: T-learner and S-learner. More efficient approaches especially in high-dimensional settings are explained in the next section.

The T-learner method, where "T" signifies "two," is based on two prediction models **(Figure 3.2A)**.[8,11,97] In the first stage, using the training dataset, outcome prediction models are developed separately for the intervention and control arms. These prediction models, called base-learners, can be any prediction algorithm, such as simple linear or logistic regression, random forest, penalized regression models, support vector machines, and ensembles over multiple MLs. The input of the prediction algorithms includes covariates that are selected a priori. In the second stage, two prediction models are fitted to the covariate test data, separately, regardless of the actual intervention received. The predicted values from the prediction model based on the intervention arm represent the estimates of the outcome had the participants been in the intervention arm, while the predicted values from the control arm-derived prediction model are the estimates of the outcome had they been in the control arm. Finally, the difference of the two predicted values (i.e. predicted value if in the treatment arm minus that if in the control arm) is the estimate of CATE.

In contrast, the S-learner method, where "S" denotes "single," relies on a single prediction model **(Figure 3.2B)**.[8,82,98,99] In the first stage, an outcome prediction model based on any algorithm is built using both intervention status and covariate information. This base learner is more effective in capturing HTE if it better captures the interaction between intervention status and covariates, for example making Bayesian additive regression trees (BART) a suitable choice. In the second stage, the prediction algorithm is fitted to two test datasets in which the

intervention status is imputed as 0 and 1 for every participant. The predicted values from the intervention status = 0 represent the estimate if participants had been in the control arm, and those from intervention status = 1 is the estimate if they had been in the control arm. The differences between those two predicted values provide the estimates of CATE.

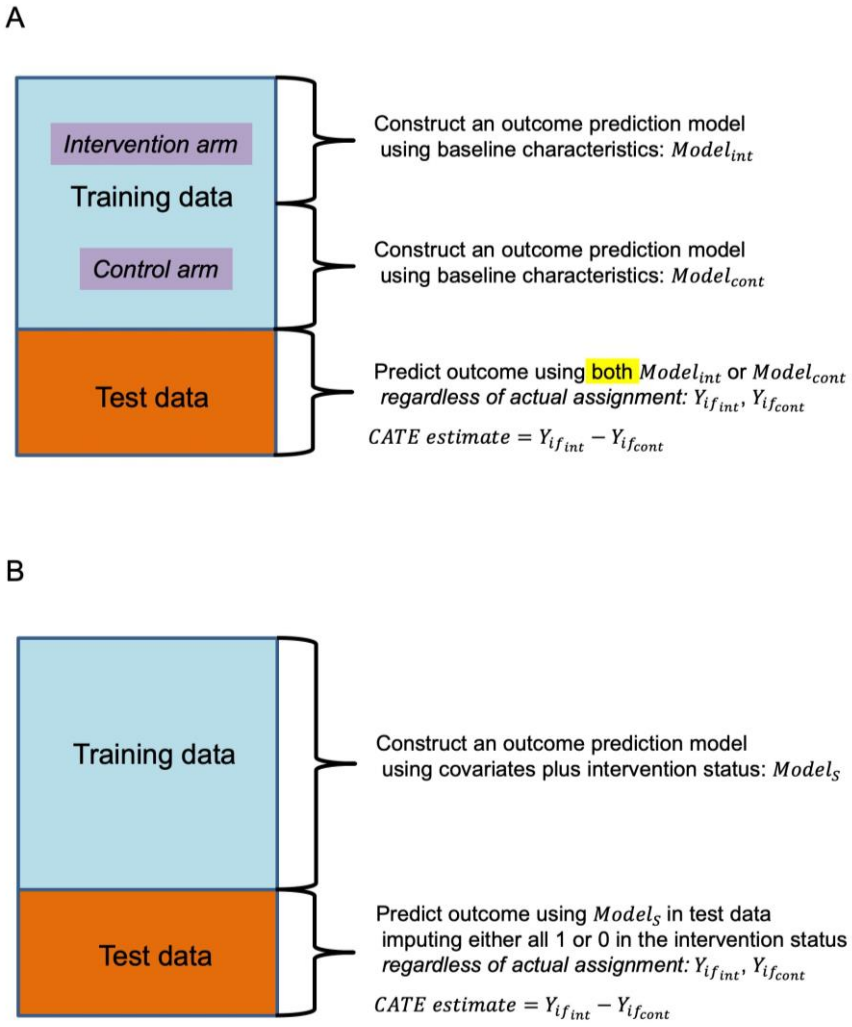


Figure 3.2: Graphical explanation of simple meta-learner approaches
T-learner (A) and S-learner (B) approaches to estimate CATE are explained. In either approach, first split dataset into training data and test data. For T-learner, two outcome prediction models are developed in either intervention arm or control arm only ($Model_{int}$ and $Model_{cont}$, respectively). CATE is estimated as the differences in outputs of the two models in test dataset. For S-learner, one outcome prediction model is developed based on covariates and intervention status. CATE estimate is estimated as the differences in outputs of the model in test data inputting either all 1 or 0 in the intervention status.

In any meta-learner approaches, the base learner(s) can be any ML prediction algorithms, such as random forests, regularized regression, deep neural nets, boosted trees, and ensembles of these methods.[100] In selecting the algorithm, understanding the concept of each approach would be helpful; for instance, a simple linear or logistic regression without interaction terms does not work in the S-learner, since it generates a constant treatment effect for every individual. In addition, although the T-learner effectively accounts for the covariate interactions by the intervention status through stratification, base-learners with weak predictive accuracy may not perform well, since the last procedure, subtraction of the absolute predicted values from the two algorithms, relies on the assumption that each absolute predicted value is reliable.

In T- and S-learner, the base-learners are trained to minimize the prediction error, not the error of the treatment effect, which may not be efficient in exploring HTE (i.e. the CATE estimator is consistent but the variance is high).[8,100,101] Therefore, in practice, a use of more efficient approaches will be required to explore HTE or develop an ITR in medical/epidemiological RCTs.

CATE estimation: efficient approaches

Double/Debiased machine-learning (DML) is a more involving but efficient approach. DML refers to a statistical approach in which scores obeying Neyman orthogonality conditions are used[100] (details in **Appendix 3.1**), which includes DR-learner[100,102] and R-learner[96,100]. DML approaches need additional data splitting within the training dataset, and thus a use of multiple cross-validation is necessary to gain statistical efficiency, as described in the next section.

In DR-learner, a pseudo-outcome is created using a doubly robust estimator, which is an unbiased estimator for CATE, and the pseudo-outcomes are regressed on covariates, thereby minimizing the mean squared error for estimated CATE **(Figure 3.3)**.

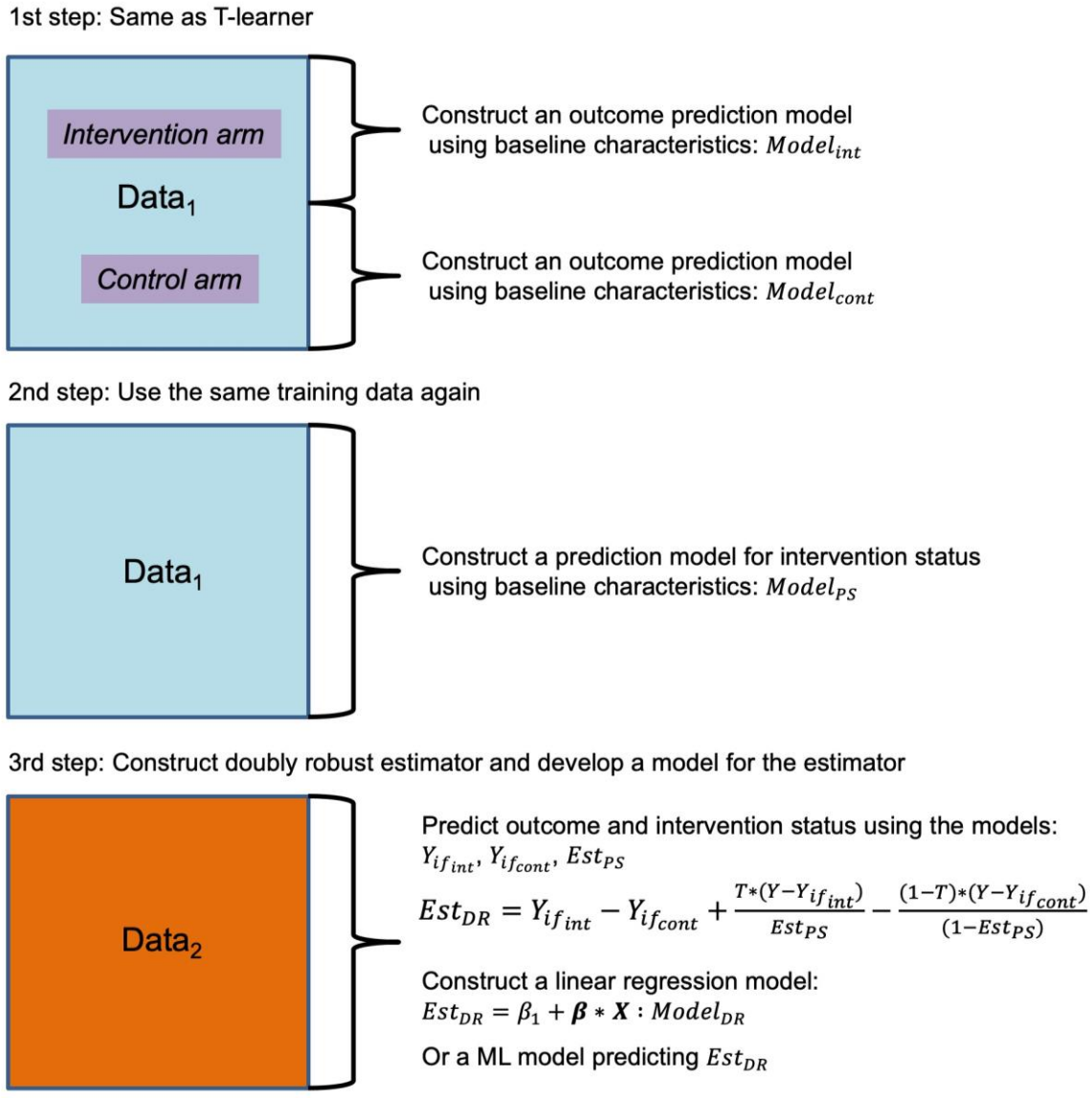


Figure 3.3: Graphical explanation of doubly robust-learner (DR-learner) approach
First split the data into $Data_1$ and $Data_2$ (note either data can be considered as training datasets). Using $Data_1$, three prediction models are developed: two models predicting outcome in either intervention arm or control arm only ($Model_{int}$ and $Model_{cont}$, respectively), and one predicting intervention status ($Model_{PS}$). Using $Data_2$, construct doubly robust estimator (EST_{DR}) using outputs in the three prediction models. Lastly, construct a final model to predict EST_{DR} using linear regression or ML models. To apply the final model to estimate CATE, another test dataset may be needed.

Specifically, first randomly split the dataset into two-folds, for example by half, denoting $Data_1$ and $Data_2$ (note, these are not training/test). This data splitting prevents overfitting and is necessary to obtain valid CATE estimates. In $Data_1$, base-learners for S- or T-learner and propensity score (PS) model are each constructed based on covariates using the same data partition. The PS model may better be fit even in an RCT setting, in particular with relatively small sample size, although the true propensity score is constant (e.g. 0.5)[103,104]. Of note, using ML algorithms is recommended for the PS model as well as the base-learners, although logistic regression has been typically used for PS models in epidemiological studies. The models developed in $Data_1$ are called nuisance models or functions, since they are not directly of interest but are needed in order to produce an unbiased and consistent CATE estimator. In $Data_2$, the doubly robust estimator of treatment effect is calculated based on the predicted values from these models. The estimator is called "doubly robust" since correct specification of either the outcome model or the PS model is necessary to get unbiased estimates, but not necessarily both. Finally, a covariate-based linear regression is built using the doubly robust estimator as the independent variable. The use of the final model is basically intended for causal inference; i.e. to determine which variables contribute to CATE. The final model can be based on an ML algorithm. The DR-learner works well in high-dimensional settings, with narrower variance of the estimator (i.e. more efficient) compared with those from S- or T-learner.

The R-learner relies on calculating residuals for the CATE estimation.[100] Estimation steps are similar to those of DR-learner, in which prediction algorithms are trained in two-stages (**Figure 3.4**). First, split the dataset into two-folds ($Data_1$ and $Data_2$). In $Data_1$, two prediction algorithms are developed for outcome and intervention status based on baseline covariates;

different from S-learner, the outcome prediction model does not include intervention status. Then, in $Data_2$, residuals for outcome and intervention status are calculated separately, as the differences between observed minus predicted values based on the prediction algorithms. Here, the residuals for intervention status should be close to the true proportion of intervention (e.g. 0.5). There are several approaches afterwards. A linear regression model may be used incorporating interaction terms between residual for intervention status and covariates. CATE can be estimated as subtracting the predicted values from the linear model imputing either all 1 or 0 in the intervention status in $Data_2$, similarly in case of S-learner. Also the coefficients for each interaction term may be interpreted as the contribution of each covariate in CATE. An ML algorithm may also be used as the final model, whereas the process differs a bit. The final ML model is developed to predict the ratio of two residuals, using squared value of residual for intervention status as the weight (details in **Appendix 3.1)**. The predicted value from the last prediction model is the CATE estimate.

Of note, although R-learner is one of the DML approaches, it is sometimes called as parametric (e.g. when linear regression is used in the final step) or non-parametric (e.g. when an ML is used) DML. Similar to DR-learner, R-learner is trained so as to optimize HTE, processing doubly robust property and performing well in a setting of high dimensionality.

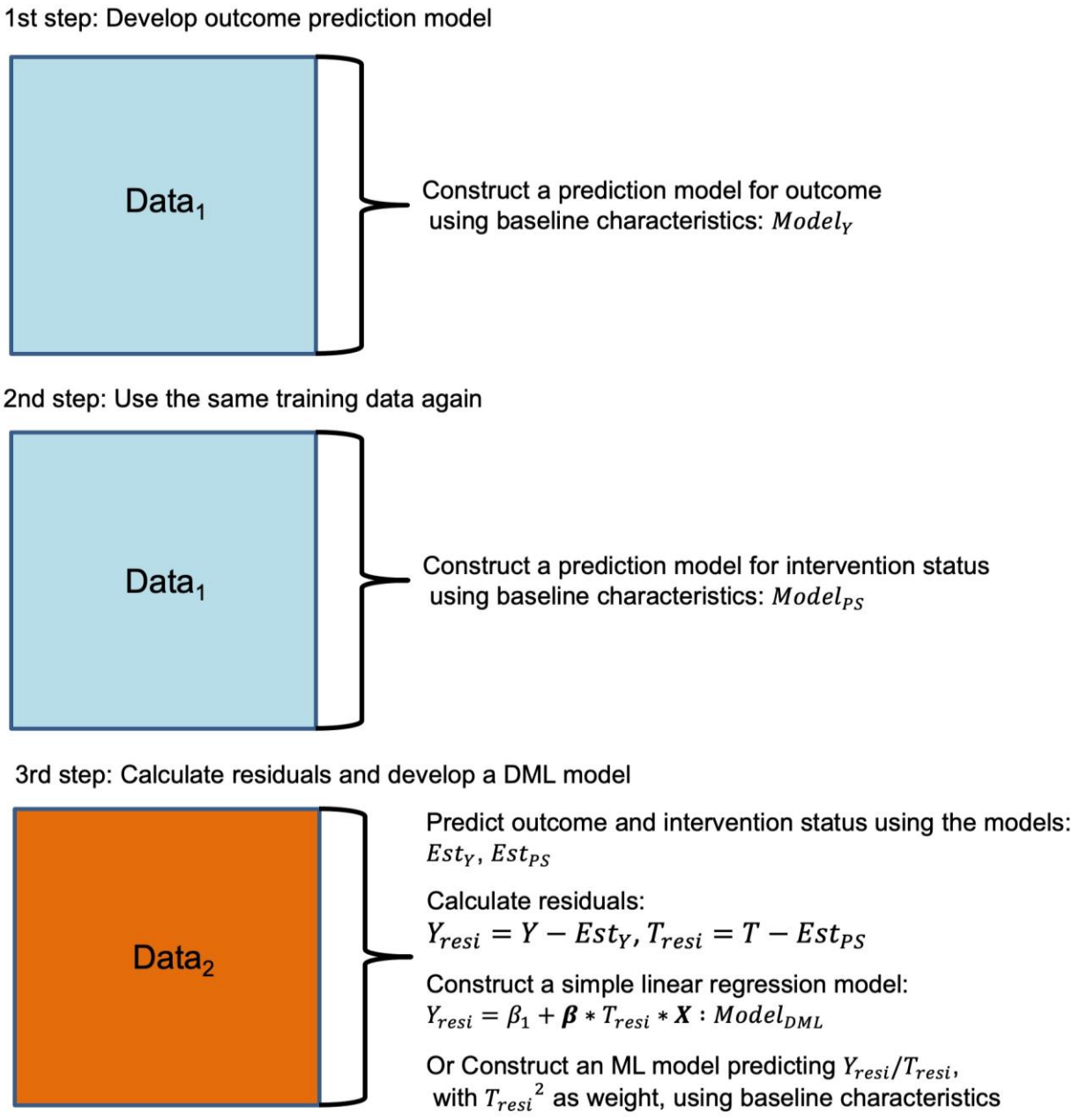


Figure 3.4: Graphical explanation of R-learner approach

First split the data into $Data_1$ and $Data_2$ (note either data can be considered as training datasets). Using $Data_1$, two prediction models are developed: one model predicting outcome ($Model_Y$), and one predicting intervention status ($Model_{PS}$). Using $Data_2$, calculate residuals for outcome (Y_{resi}) and intervention status (T_{resi}) using the two prediction models. Lastly, construct a final model to predict Y_{resi}/T_{resi} with ${Y_{resi}}^2$ as weight, using linear regression or ML models. Or, a final model may be a linear regression with interactions between covariates and T_{resi}. To apply the final model to estimate CATE, another test dataset may be needed.

CATE estimation: Causal Forest

The generalized random forest algorithm, also known as the Causal Forest, is an adaptation of random forest, specifically designed to estimate CATE.[7,105] Causal Forest employs recursive partitioning to identify neighborhoods in covariate space. These forests are built from causal trees in which the splitting criterion is optimized to represent treatment effect heterogeneity. By assuming a constant CATE within neighborhoods, the algorithm aims to locate leaves with constant but unique treatment effects across the leaves. The Causal Forest averages multiple causal trees from bootstrap samples, with variations arising from subsampling. Treatment effect predictions rely on the difference in average outcomes between intervention and control arms in terminal leaves.

A key of Causal Forest is the honesty condition, which is essentially a DML (i.e. meets Neyman orthogonality conditions). Half of the training data is utilized to estimate the tree structure (splitting subsample), while the remaining half is used to estimate treatment effects in each leaf (estimating subsample), thus preventing overfitting. Estimates of Causal Forest guarantee valid confidence intervals with asymptotic variance estimators based on the honesty condition.

Cross-fitting approach

Cross-fitting, a form of cross-validation, is a technique employed to develop a CATE algorithm while estimating CATE in the same samples. This is particularly important in high-dimensional settings where the number of covariates is large relative to the sample size, as is generally the case in medical and epidemiological trials.[33] Splitting the dataset into training and

testing subsets for estimating and evaluating purposes, which is necessary, leads to a loss of large number of samples. Cross-fitting enables a use of every sample for the evaluation purpose, thereby increasing the statistical efficiency. In brief, the approach repeats the CATE estimation and prediction process for every fold so that CATE is predicted for every test data **(Figure 3.5)**.

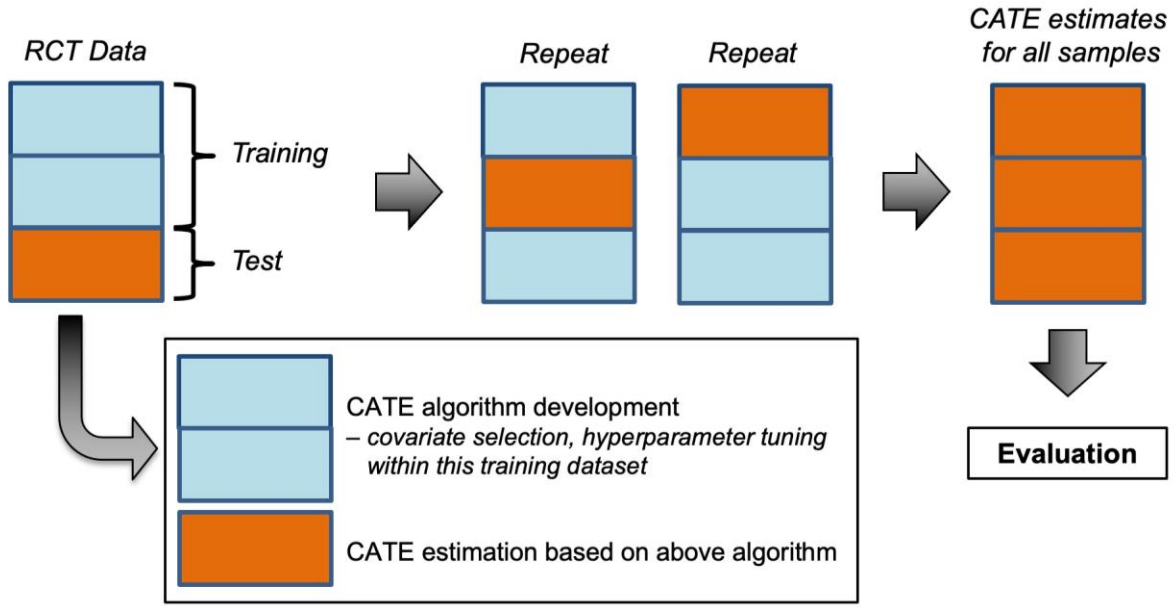


Figure 3.5: Cross-fitting in simple approaches

Figure illustrates a 3-fold cross-fitting approach to develop a model estimating CATE while estimate CATE for all samples using one RCT. First split the dataset into 3-fold, in which two is used for CATE algorithm development (training dataset) and the other one is for CATE estimation (test dataset). Repeat the process by switching the roles, and CATE can be estimated for all samples while avoiding overfitting due to using same population for model development and the estimation.

Cross-fitting in simple algorithms

First, suppose that in an application of the T-learner the trial data is first divided into K mutually exclusive and exhaustive subsets, known as "folds". K=5 is commonly used but some studies suggest larger K values (such as 200) for obtaining stable results.[10,95] This partitioning

separates the data used for model estimation (training) from the data used for prediction (test). Once K is determined, the following steps are performed:

1: Model estimation: Fit the prediction models that estimate CATE, such as a meta-learner or causal forest, setting aside one fold or subset it turn to serve as the test data (i.e. training data). Covariate selection or hyperparameter tuning, often through cross-validation, should be conducted separately each time within the training data.

2: Prediction: Using each developed model, estimate CATE for the observations in the held-out or test data. In the example, each of the five subsets, which were not used for the training, will be used in turn.

3: Cross-fitting: Repeat 1-2 for K times, obtaining CATE estimates for all samples. In this process each observation is included in one of the held-out data sets, and treated as a test observation.

Cross-fitting in DML

Cross-fitting may be applicable to DML algorithms including DR-learner and R-learner, but it gets a bit complex since these algorithms generally need additional cross-validation. **Figure 3.6** illustrates an efficient process with R-learner, which is applicable to DR-learner too. Researchers may need to set the numbers of folds for cross-fitting (to estimate CATE for all samples) and cross-validation (within the DML algorithm), to say K_1=4 and K_2=3, respectively. Then perform the following steps:

1: Nuisance model development: set aside one fold from K_1 for test purpose (i.e. estimate CATE). Using the remaining 3/4 of the total data, which we call training data, we further split it into K_2 folds. Setting aside one fold from K_2 within the training data (i.e. 2/4 of the total data or $Data_1$), we develop nuisance models, corresponding to a model predicting outcome and another model predicting intervention status.

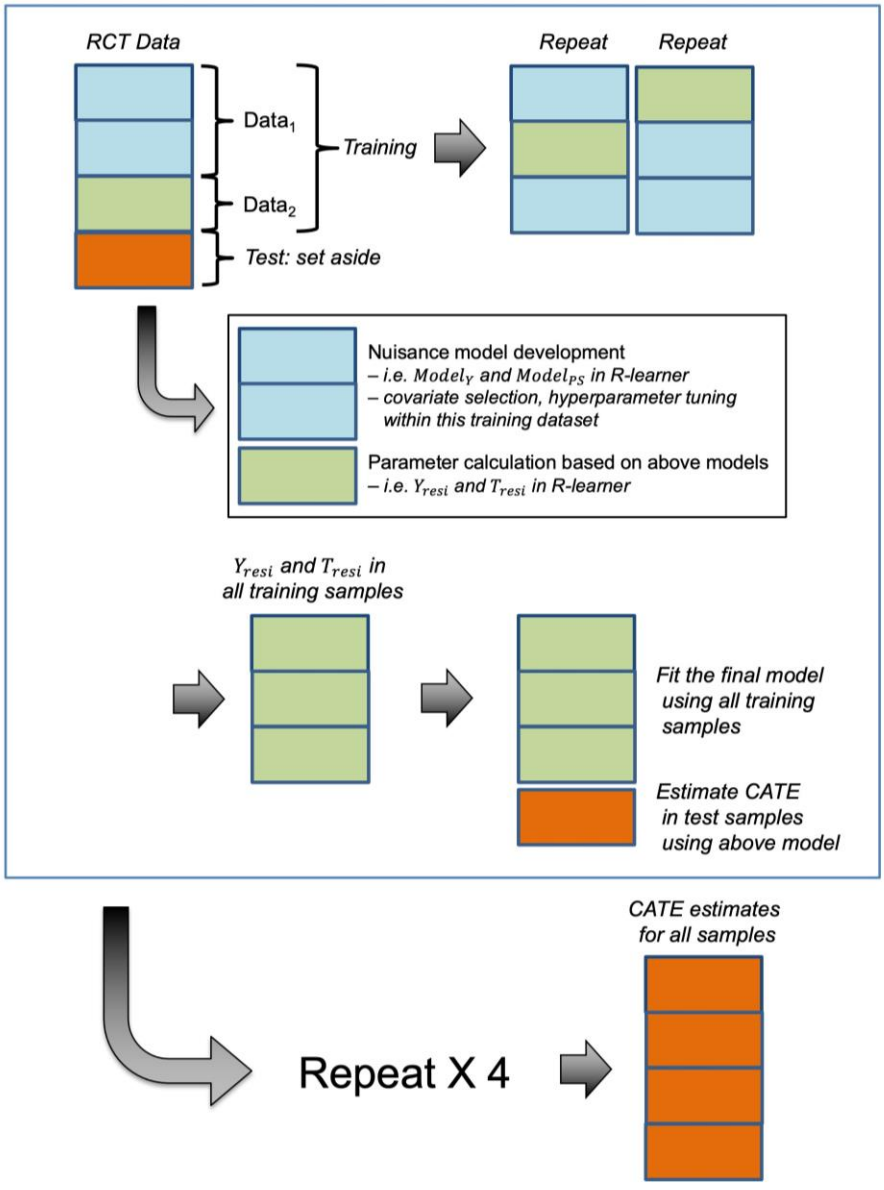


Figure 3.6: Cross-fitting in approaches needing cross-validation in their frameworks

Figure 3.6 (continued)

Figure illustrates a 4-fold cross-fitting approach to develop a model estimating CATE using DR-learner or R-learner while estimate CATE for all samples using one RCT. First split the dataset into 4-fold, in which three is used for CATE algorithm development (training dataset) and the other one is for CATE estimation (test dataset). Further split the training dataset into 3-fold for example. Two ($Data_1$) are used to develop nuisance models (e.g. $Model_Y$ and $Model_{PS}$ in R-learner), and one ($Data_2$) is used to calculate parameters necessary for CATE model development (e.g. Y_{resi} and T_{resi} in R-learner). Repeat the process by switching the roles of $Data_1$ and $Data_2$, and these parameters are estimated in all training samples. Then fit the final model using all samples in training dataset to estimate CATE, and estimate CATE in test dataset. Repeat the process by switching the roles of training and test dataset, and CATE can be estimated for all samples while avoiding overfitting due to using same population for model development and the estimation.

2: Residual calculation: In the held-out data within training data (1/4 of the total data or $Data_2$), the residuals for outcome and intervention status are calculated using the models developed in $Data_1$.

3: CATE estimation model development: Repeat 1-2 for K_2 times, obtaining residuals for every sample in the training data. Fit the final model in the whole training data (i.e. 3/4 of the total data), which is the model to estimate CATE.

4: CATE estimation: Using the test data (1/4 of the total data or test data), estimate CATE using the model developed in step 3.

5: Cross-fitting: Repeat 1-4 for K_1 times, obtaining CATE estimates for all samples.

By employing cross-fitting, the CATE estimates are derived from models that are independent of the test dataset, thereby preventing overfitting and avoiding the complexity conditions necessary to obtain a consistent estimator. As all samples are used for evaluation, the statistical precision is improved compared with a simple training/test splitting approach.

However, in case of an evaluation of ITRs, to appropriately account for the statistical gain in precision necessitates a complex computation of the variance estimator of the ITR evaluation indices, which is described in a later section of this paper.[10,85]

3.5 Part C: Evaluation approaches to an ITR

An ITR refers to a data-driven rule indicating who to treat or not such that the rule maximizes the treatment effect (i.e. benefit) in the target population. As detailed in **Section 1**, developing an effective ITR can be a major aim of HTE-related studies. Once CATE is estimated, a simple ITR can be easily developed – if CATE is preferable (such as CATE < 0 for incident CVD outcome), then the subject should be treated, and otherwise not. An ITR may be developed based on a variable or a scoring rule other than CATE. In addition, more complex treatment rules can be made, for example, considering HTE for ≥2 distinct outcomes or considering the limit of offering treatment (called budget constraints)[106], but here we focus on the evaluation of a simple ITR. For this purpose, we would want to observe two potential outcomes[86,94] per patient, either if treated or if not treated, in order to directly evaluate the ITR performance. However, this is impossible, and thus we need evaluation metrics specifically designed for ITR evaluation. Such metrics should reflect the performance of an ITR with respect to the overall treatment effect in the target population. In addition, the evaluation methods should be generalizable to any ITR development approaches in practice. We introduce the population average value (PAV)[10,11], the population average prescription effect (PAPE)[10], the area under the prescriptive effect curve (AUPEC), the sorted group average treatment effect (GATE), and the rank-weighted average treatment effect (RATE) metrics.[85,95,107] Visual explanation on PAV,

PAPE, and AUPEC with reference to ATE is provided in **Figure 3.7**. The implementation of some these evaluation methods can be easily done with an R package 'evalITR' or 'grf'.[85,107] Note, the valid evaluation metrics for an ITR are not limited to these measuress[95] and are under rigorous investigation.

Population average value (PAV)

A standard metric to evaluate an ITR is the PAV, which describes the average outcome (of the super-population) if the population is treated according to an ITR.[10,11,108] If a smaller value is better for the outcome (such as incident CVD), an ITR that produces the smallest PAV (biggest reduction) is considered as the most effective treatment strategy. Of note, PAV indicates an average outcome if everyone were treated according to an ITR, not a treatment effect; therefore, it is not reasonable to compare PAV (average outcome) and ATE (treatment effect), for example. In a case of cross-fitting, the PAV estimand and the variance estimator become a bit complex, accounting for the estimation uncertainty of the ITR by averaging over the random sampling of training sets.[10]

Population average prescription effect (PAPE)

A good ITR is always expected to outperform a non-individualized treatment rule that does not use individuals' information, and a budget constraint should be considered since intervention is often costly. To reflect these concepts, PAPE compares the overall average outcome in the population under an ITR versus under random assignment of the intervention with a fixed proportion of the intervention assignment as in the ITR.[10] In this metric, if an ITR recommends 30% of the target population to be treated, then the comparator is a non-

individualized treatment rule in which 30% of the population is randomly assigned to the intervention arm. Thus, PAPE explains how much average outcome could be improved by just changing the assignment of intervention from random to an organized way (i.e. ITR). Similar to PAV, the best ITR should confer the most beneficial PAPE, when the budget constraints are accounted for.

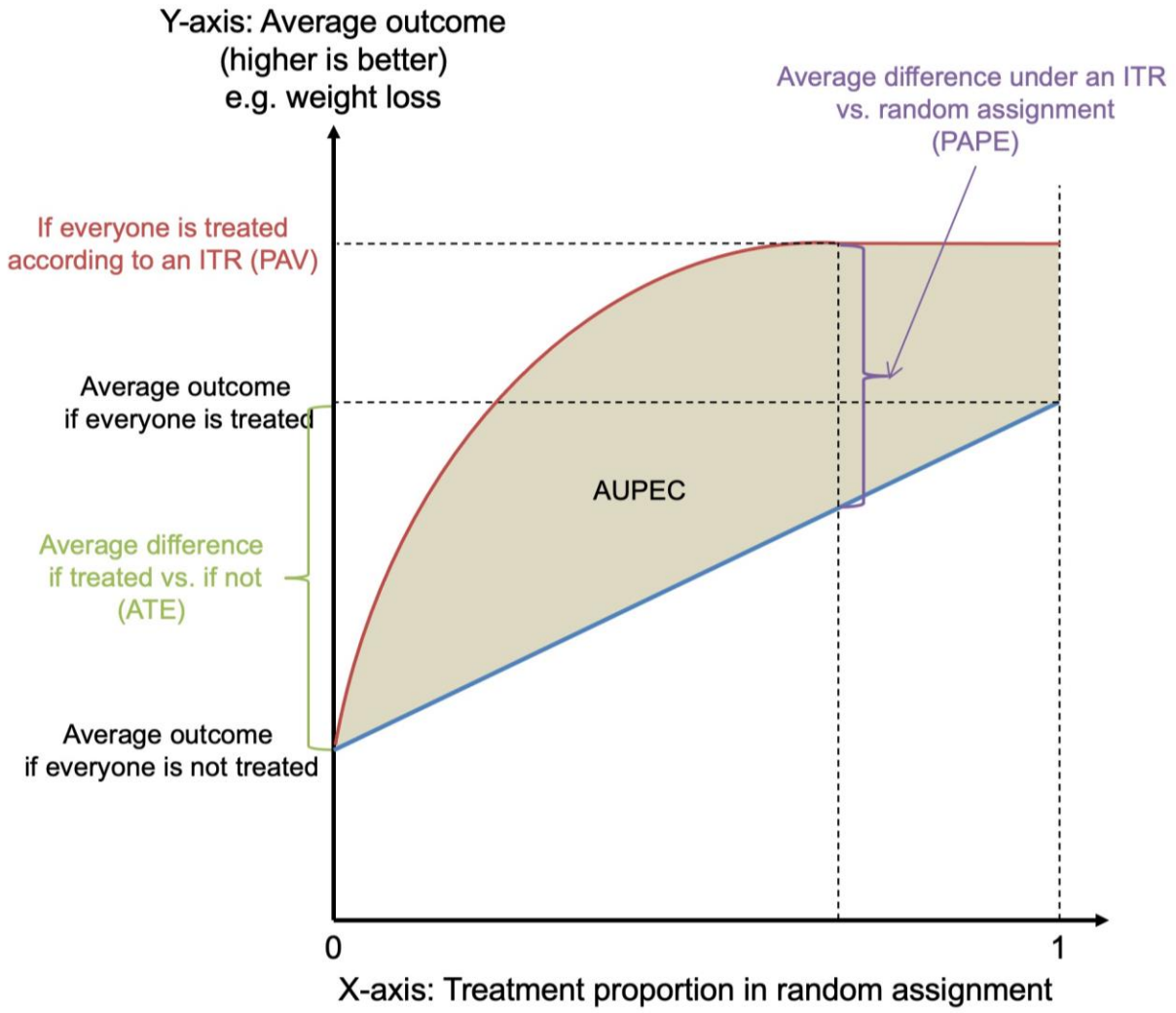


Figure 3.7: Graphical illustration of PAV, PAPE, and AUPEC
Individualized treatment rule (ITR) may be evaluated in unique evaluation metrics when the ITR is developed and evaluated in the same trial using cross-fitting. X-axis indicates treatment proportion in random assignment and Y-axis is average outcome (e.g. weight loss). Blue line indicates average outcome with the treatment proportion (assuming random assignment) equal to the value in X-axis.

Figure 3.7 (continued)
Red line indicates population average value (PAV), an average outcome if everyone is treated according to the ITR, which should be larger than random assignment. Average difference under an ITR and random assignment at the treatment proportion according to the ITR is called population average prescription effect (PAPE). Finally, the area under the prescription effect curve (AUPEC) is the shaded area between red and blue line.

Area under the prescriptive effect curve (AUPEC)

Figure 3.7 effectively shows the relationships between PAV, PAPE, ATE, and average potential outcome if treated and if not treated. In this example, average potential outcome (e.g. weight loss) if treated is better than that if not treated, and thus the difference between if randomly treated (i.e. treatment proportion > 0%) and if not treated (i.e. treatment proportion = 0%) is higher if the proportion of intervention assignment (x-axis) is higher. The difference between the treatment proportion = 100% and 0% indicates ATE. The ITR in this figure necessitates that 70% of the population should be treated, and so PAV is plateaued if x-axis ≥70%. PAPE is the difference between PAV and average outcome if 70% of the population is randomly treated. AUPEC is the shaded area between the curve indicating PAV (red curve) and the straight line indicating the average outcome under random assignment of the intervention (blue line). Thus, the AUPEC represents the average performance of an ITR compared with the random allocation of intervention over the range of the intervention proportion.

The sorted group average treatment effect (GATE)

With good estimates of CATE, we should be able to correctly stratify the individuals according to the estimated outcome value for the actual observed outcome; i.e. individuals with lower CATE should be associated with lower observed outcome value, on average, compared with those with higher CATE. This represents a concept of GATE, which describes the groups

according to sorted estimated CATE values.[85,95] Tertile, quartile, or quintile categorization approaches may be used, depending on the size of the dataset. Estimating GATEs is fairly straightforward; with estimated CATEs in the test dataset, categorize individuals into, say, quartiles according to the CATE values, such that the least quartile group has the smallest CATE values and vice versa. Then, for each category separately, estimate the treatment effect as the average of the observed outcomes in the intervention arm minus those in the control arm. There are a few distinct ways to calculate the variance estimator and to generalize the method to cross-fitting.[85,95] In addition, a valid statistical test for effect heterogeneity and monotonicity across GATEs have been proposed.[85]

The rank-weighted average treatment effect (RATE) metrics

Given very limited sample sizes for exploring HTE in medical/epidemiological RCTs, a high power test not requiring division of data into subgroups (as in GATE) is warranted. RATE metrics, including the targeting operator characteristic (TOC) and area under the targeting operator characteristic (AUTOC), are such evaluation approaches (**Figure 3.8**).[107] For the calculation, the rank according to estimated CATEs (or another continuous metric used to tailor the intervention) is used instead of estimated CATEs themselves. Then the TOC is defined as average effect among those with top X quantile (i.e. ranks) of CATE minus ATE, with X determined arbitrarily. We expect that if X is smaller (i.e. the cutoff rank of CATE is higher), the TOC becomes higher, since we are then targeting a very selective population with expected large positive treatment effect. TOC becomes 0 when X is 1. TOC curve is defined by TOC values when X varies between 0 to 1 (i.e. 100 to 0 percentile), and AUTOC is the area under the TOC curve. AUTOC thus evaluates the performance of CATE-based ranks with respect to ATE, not

necessitating the arbitrary grouping. Of note, AUTOC differs from AUPEC in two aspects; the x-axis is based on ranks based on CATE versus CATE estimates themselves, and the y-axis represents average effects versus average outcome in AUTOC and AUPEC, respectively. The RATE metrics can be powerful since it ignores the CATE estimates themselves. However, the interpretation of AUTOC may be challenging – which represents the average performance of the ranking approaches (often based on CATE) compared with ATE over the range of the cutoffs for the ranking as to who to treat.

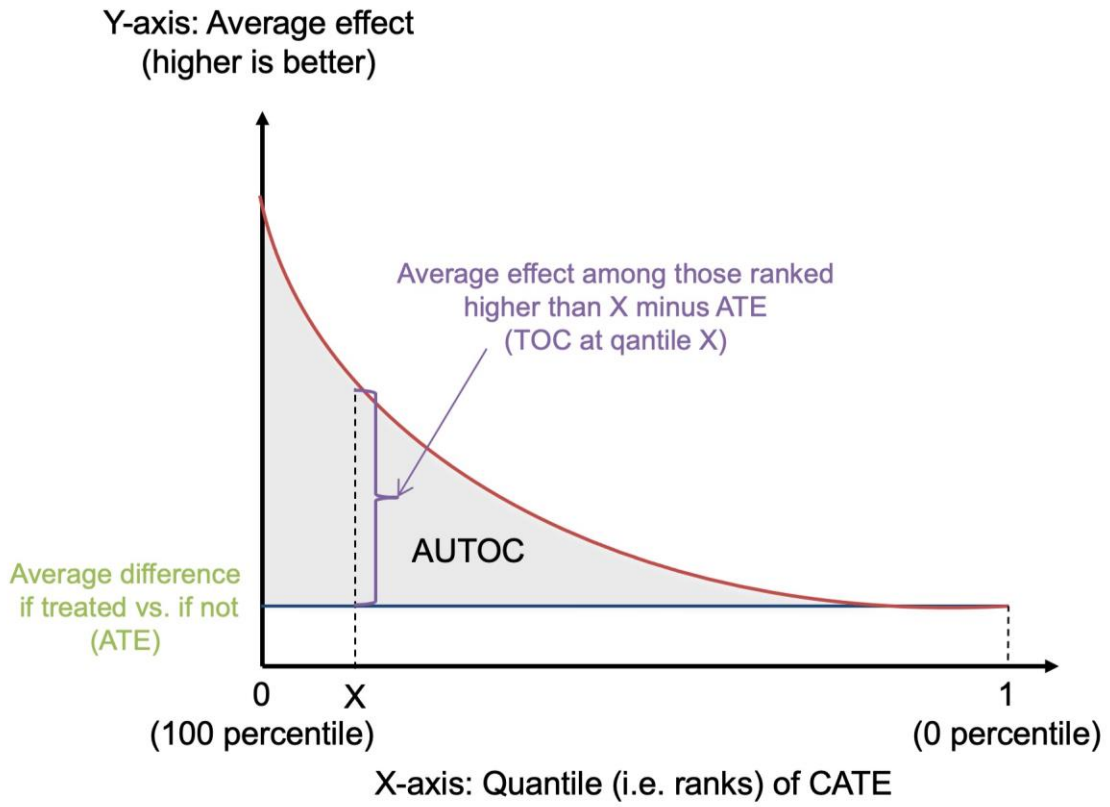


Figure 3.8: Graphical illustration of TOC and AUTOC
Targeting operator characteristic (TOC) and area under the targeting operator characteristic (AUTOC) are ITE evaluation approaches based on CATE-based ranks. X-axis indicates quantiles of CATE (or other metrics indicative for heterogeneous treatment effect), or percentiles. Y-axis is the average treatment effect among the target population. Red curve indicates the average effect among those ranked higher

Figure 3.8 (continued)
than the quantile at the x-axis that varies from 0 to 1. When X-axis becomes 1, the red curve is equal to average treatment effect (ATE) because then the target population is everyone. TOC is the difference between a point on the red curve at arbitral x-axis level (X) and ATE. Blue line denotes ATE at any X. Finally, AUTOC is the shaded area between red and blue line.

3.6 Part D: How to avoid common pitfalls

There are several pitfalls that researchers can make in studies focusing on exploring HTE and development of an ITR in medical and epidemiological research. In this section, we explain why these are problematic and how to avoid them.

Split the dataset into training and test

Some studies use a whole dataset to train HTE algorithms, estimate CATE in the same dataset using the algorithms, and make inferences. This practice is problematic partly because the algorithm is very likely to be overfitted to the dataset. Complex ML algorithms can always be overfitted to the data used for the model training, and that is why practicing training/test data splitting is always necessary to evaluate valid prediction accuracy. An inference that is made using an apparently overfitted algorithm to the target dataset does not provide any useful information. Splitting training/test is an easy way to avoid complex assumptions needed to obtain valid estimates of CATE, and that is why DML or DR-learner is built upon cross-validation.[100,101]

A practice of training/test splitting is a standard in the computer science community since developing prediction algorithms aims for generalizable usage, for which accounting for

overfitting is essential. However, in medical and epidemiological communities, a practice of data splitting is less common, in particular when estimating subgroup effects. This potentially reflects a widespread usage of simple linear or logistic regressions for the inference, which are less sensitive to overfitting issues. Such practice also reflects that generally high dimensionality is an inevitable issue in medical and epidemiological research. In HTE-related research, it is generally recommended not only to split the data but also to conduct cross-fitting process for statistical efficiency. Applied researchers should be aware of the necessity of such methods and how to conduct them. Of note, data splitting is essential for internal validity, and external validity for the generalizability of an ITR should be tested using another independent dataset, as in popular simple prediction practices.

In addition, any ML training procedures, including hyper-parameter tuning, should be done strictly in the training dataset in each cross-fitting process. Use of the information on a test dataset in the training phase will lead to "data leakage", a major source overfitting.[109] Therefore, in each fold, the same CATE estimation algorithms, including the pre-specified approaches of covariate selection and hyperparameter tuning, should be trained respectively.

Do not consider variables with high variable importance in causal forest algorithm as causally important for estimating HTE

Inference on causality of an effect modification is difficult. To be an effect modifier, the covariate may just need to be associated with outcome.[110] Suppose there is one true causal effect modifier; then a highly correlated covariate could show similar effect modification, which is called a surrogate effect modifier. In general, it is not possible to estimate the causal contribution

of one covariate on an effect modification using a data-driven approach. VI measures in causal forest algorithm provide useful information on the extent to which variables contribute to the prediction algorithm, but nothing more. The measures may be used to partly interpret the "black-box" algorithm but may not provide causal interpretation on the effect modification. Furthermore, the important variables may be likely different in other HTE algorithms such as meta-learners, and thus generalizable information may not be obtained from exploring variable importance in a causal forest algorithm alone. In addition, simple interaction terms between intervention status and a single covariate would offer more interpretable information on the importance of the effect modification.

Do not evaluate the ITR performance with an interaction in simple outcome regression

In medical and epidemiology literatures, it is very common to evaluate effect modification based on the coefficient of the interaction term between intervention status and a covariate in simple outcome regressions.[81] This is a good practice for a pre-specified list of potential effect modifiers in an RCT, but not for a data-driven approach such as an ITR. Fundamentally, ITR development relies on the covariate distribution of the derivation data (in this case an RCT), and the evaluation metric needs to account for it. Interaction analysis between intervention status and CATE is incorrect, since the estimated CATE is not a linear predictor of the true CATE.[95] For this purpose, investigators should rely on a method to estimate the best linear predictor of CATE, for example.[95] Interaction analysis between intervention status and a categorical variable based on CATE (GATE for example) is also incorrect, since the standard variance estimator does not account for the uncertainty stemming from estimating the cutoff points.[85] Finally, in the cross-fitting settings, the variation across training datasets and efficiency

gain from the cross-validation procedure should be considered when estimating the variance of GATE, which may not be achievable by simple outcome regression. For these reasons, the evaluation approaches for an ITR should be different from conventional statistical methods in medical and epidemiological literatures.

3.7 Part E: An applied example

The Preventing Overweight Using Novel Strategies (POUNDS Lost) trial was conducted to fill the knowledge gap of whether overweight people have a better response in the long term to reduced-calorie diets that emphasized a specific macronutrient composition[12]. The 2x2 factorial RCT of high vs. low fat and high vs. low protein, which consequently also compared high vs. low carbohydrates, resulted in similar weight loss, blood pressure, and glucose metabolism over 2 years. However, the treatment effects were very heterogeneous and many effect measure modifications have been identified primarily by genotypes.[111–114] Given the variety of potential effect modifiers, it is of particular interest to determine which individuals are most likely to benefit from which reduced-calorie diets, and to develop an ITR such that the rule maximizes the overall effectiveness of the dietary interventions – corresponding to the study type 2 in the Section 3. In this example, we sought to estimate CATE for the dietary interventions in the POUNDS Lost trial and develop a clinically meaningful CATE-based ITR.

Methods

Population

The POUNDS Lost Trial is a randomized intervention trial in which 811 individuals with a body mass index (BMI) between 25 to 40 kg/m^2 were assigned to one of 4 energy-reduced diets varying in macronutrient compositions of fat, protein, and carbohydrate to compare their effects on body weight and composition over 2 years.[12] The interventions included (1) low-fat, average-protein diet (% of fat/protein/carbohydrate: 20/15/65); (2) low-fat, high-protein diet (20/25/55); (3) high-fat, average protein diet (40/15/45); (4) high-fat, high-protein diet (40/25/35). Hence, the trial constituted a 2-by-2 factorial design contrasting low (20%) versus high (40%) fat, and average (15%) versus high (25%) protein diet. More details of this trial were described elsewhere.[12] The study was conducted from October 2004 through December 2007 at 2 sites: Harvard T.H. Chan School of Public Health and Brigham and Women's Hospital in Boston, MA, and the Pennington Biomedical Research Center of Louisiana State University System, in Baton Rouge, LA. The study was approved by the human subjects committee at both institutions and by a data safety monitoring board appointed by the National Heart, Lung, and Blood Institute. All participants gave written informed consent.

Intervention arm, outcomes, and covariates

Intervention/control arms were defined for high-fat/low-fat diet and high-protein/average-protein diet, respectively. These two interventions were evaluated separately. The primary outcome of the study was defined as, in accordance with the primary report[12], the change in body weight over 2 years. N=645 out of total 811 participants completed the weight assessment at 2-year, and for persons who withdrew from the study early (after at least 6 months of participation), their weight at the end of the trial was imputed on the basis of a rate of 0.3 kg per month of regained weight after withdrawal.[12] ITRs were developed to maximize the effects of

each intervention on the primary outcome. Covariates included age, gender, height, weight, waist, and income.

Statistical analysis

For CATE estimation, we applied the following approaches; Causal forest, T-learner with linear regression as the base-learner, S-learner with random forest as the base-learner, R-learner with random forest as the base-learner and second-stage model, and DR-learner with random forest as the base-learner and second-stage model. An ITR was defined as a treatment rule in which participants with estimated CATE<0 are offered interventions and others are assigned to controls. PAV, PAPE, and GATE were used for the ITR evaluation. In the present context, the estimate of PAV is the average 2-year weight change if participants are treated under an ITR. The estimate of PAPE is the difference in average 2-year weight change under the ITR compared with the random assignment of the intervention with the same proportion of intervention arm in the ITR. The estimates of GATEs represent a difference in 2-year weight change between intervention arm and control arm in groups according to the tertile of the estimated CATE. A 5-fold cross-fitting process was applied and the valid variance for the estimates of the evaluation metrics was calculated. The statistical estimates were calculated based on the R package 'evalITR'. We also explored the CATE algorithm using SMD across GATE-based groups and a regression tree employed for the estimated CATE.[91] All the statistical analysis was conducted using R 4.2.1.

Results

The average 2-year weight reduction (95% CI) was 3.3 kg (2.8, 3.8) in the total population, 3.3 kg (2.6, 4.0) in high-fat group, 3.3 kg (2.6, 4.0) in low-fat group, 3.6 kg (2.9, 4.3) in high-protein group, and 3.0 kg (2.3, 3.7) in average-protein group. The ATEs (as weight reduction) were -0.034 kg [95% CI: -0.99, 0.92] in the fat intervention (low fat preferable) and 0.60 kg [95% CI: -0.35, 1.6] in the protein intervention (high protein preferable), neither leading to a significant difference on average.

Table 3.2 summarizes the GATEs, PAV, and PAPE from the five approaches of CATE estimation. Note, the values indicate weight loss. In fat intervention analyses, GATE1 is a subgroup with expected smaller weight change by high-fat intervention (i.e. expected greater weight loss by high-fat intervention); in contrast, GATE3 is a subgroup with expected greater weight loss by low-fat intervention. GATE1 in the protein intervention is a subgroup with expected greater weight loss by high-protein intervention, and GATE3 is the opposite. Overall, effect heterogeneity was well detected for the fat intervention but not for the protein intervention. In the fat intervention, the R-learner algorithm conferred the biggest weight loss by the intervention shown as beneficial PAV and PAPE. In the R-learner, those with the highest estimated benefit from the low-fat (i.e. GATE3) had an actual 2.68 kg (SD: 0.97) net weight gain by high-fat compared with low-fat intervention, or 2.68 kg net weight loss by low-fat intervention. In contrast, those in other GATE groups had 1.28 kg and 1.45 kg net weight reduction by reversing intervention (by offering high-fat compared with low-fat intervention). The resulting ITR algorithm indicates 448 participants (55.2%) should be treated with high-fat intervention. A fat intervention following the ITR would lead to a 4.13 kg (SD: 0.43) weight

reduction in the population on average (interpretation of PAV), and changing the assignment according to the ITR from random resulted in 0.84 kg (SD: 0.32) weight reduction (interpretation of PAPE). The best model was also R-learner for the protein intervention, and the resulting ITR showed an overall average weight reduction of 3.8 kg (SD: 0.06), which was not very different from the high-protein group. While the groups were ordered in the development data, the GATEs did not show a rank-consistent pattern (smaller treatment effect for GATE1 and higher for GATE3) in the testing data.

A. Fat intervention

	GATEs			***ITR evaluation***	
CATE estimation approach	GATE1	GATE2	GATE3	PAV	PAPE
Causal Forest	1.32 (0.77)	0.02 (0.86)	-1.28 (2.51)	3.89 (0.39)	0.65 (0.47)
T-learner/ Linear Regression	1.04 (1.35)	-0.39 (0.82)	-0.60 (2.02)	3.45 (0.37)	0.20 (0.53)
S-learner/ Random Forest	0.69 (1.66)	0.53 (1.81)	-1.17 (1.06)	3.76 (0.37)	0.52 (0.36)
R-learner/ Random Forest	1.28 (0.85)	1.45 (1.83)	-2.68 (0.97)	4.13 (0.43)	0.84 (0.32)
DR-learner/ Random Forest	2.36 (1.85)	-0.03 (0.93)	-2.28 (1.52)	4.08 (0.67)	0.78 (0.49)

B. Protein intervention

	GATEs			***ITR evaluation***	
CATE estimation approach	GATE1	GATE2	GATE3	PAV	PAPE
Causal Forest	-0.01 (2.35)	1.62 (0.84)	0.35 (2.60)	3.67 (0.42)	0.04 (0.16)
T-learner/ Linear Regression	-0.79 (2.01)	2.59 (0.95)	0.16 (1.27)	3.37 (0.36)	-0.11 (0.30)
S-learner/ Random Forest	0.37 (1.88)	1.51 (1.65)	0.07 (1.48)	3.61 (0.63)	0.18 (0.51)
R-learner / Random Forest	-0.03 (1.54)	1.93 (0.99)	0.06 (1.82)	3.75 (0.60)	0.34 (0.32)
DR-learner/ Random Forest	0.60 (1.71)	1.33 (1.30)	0.03 (1.65)	3.40 (0.36)	0.00 (0.27)

Table 3.2. Evaluation of ITRs in POUNDS Lost trial

Numbers are estimated weight loss (SD).
For GATEs, the number indicates the effect of high-fat diet on the 2-year weight reduction (kg) minus that of low-fat diet (A), and the effect of high-protein diet on the 2-year weight reduction minus that of average-protein diet (B). Positive values favor either high-fat or high-protein diet intervention, and negative values favor low-fat or average-protein diet intervention.
GATE1 is a subgroup with expected smaller weight change by high-fat or high-protein intervention (i.e. expected greater weight loss by high-fat or high-protein intervention); in contrast, GATE3 is a subgroup with expected greater weight loss by low-fat or average-protein intervention.

PAV stands for the average 2-year weight reduction under the ITR. PAPE indicates the additional weight reduction by changing the allocation of the intervention according to the ITR compared with random assignment. Higher values indicate greater effects on weight reduction of the ITR.

Finally, we aimed to interpret the algorithm of the best ITR for the fat intervention. **Table 3.3** describes the patient characteristics according to the GATEs. In the R-learner model, those with estimated higher treatment effect by low fat (GATE 3) were characterized by high proportion of male and larger body size,. **Figure 3.9** illustrates a treatment algorithm developed by applying the CART algorithm to the estimated CATE.

	CATE based on R-learner			
	Tertile 1 (GATE 1) N = 271	Tertile 2 (GATE 2) N = 270	Tertile 3 (GATE 3) N = 270	SMD
Age, years	51.5 (8.1)	51.7 (8.6)	49.4 (10.6)	0.16
Male, %	56 (20.7)	56 (20.7)	184 (68.1)	0.73
Height, cm	165 (8)	166 (7)	174 (8)	0.80
Weight, kg	88 (17)	89 (13)	102 (12)	0.75
Waist, cm	99 (14)	100 (11)	111 (10)	0.70
Income, %				0.66
>150K	25 (9.2)	41 (15.2)	54 (20.0)	
100-150K	72 (26.6)	45 (16.7)	47 (17.4)	
50-100K	66 (24.4)	110 (40.7)	147 (54.4)	
<50K	108 (39.9)	74 (27.4)	22 (8.1)	

Table 3.3. Interpretation of ITR for fat intervention in POUNDS Lost trial

CATE is estimated effect of high-fat vs. low-fat diet on 2-year weight loss. GATE1 is a subgroup with expected smaller weight change by high-fat or high-protein intervention (i.e. expected greater weight loss by high-fat or high-protein intervention); in contrast, GATE3 is a subgroup with expected greater weight loss by low-fat or average-protein intervention.
Numbers are mean (SD) for continuous variables and frequency (%) for categorical variables.

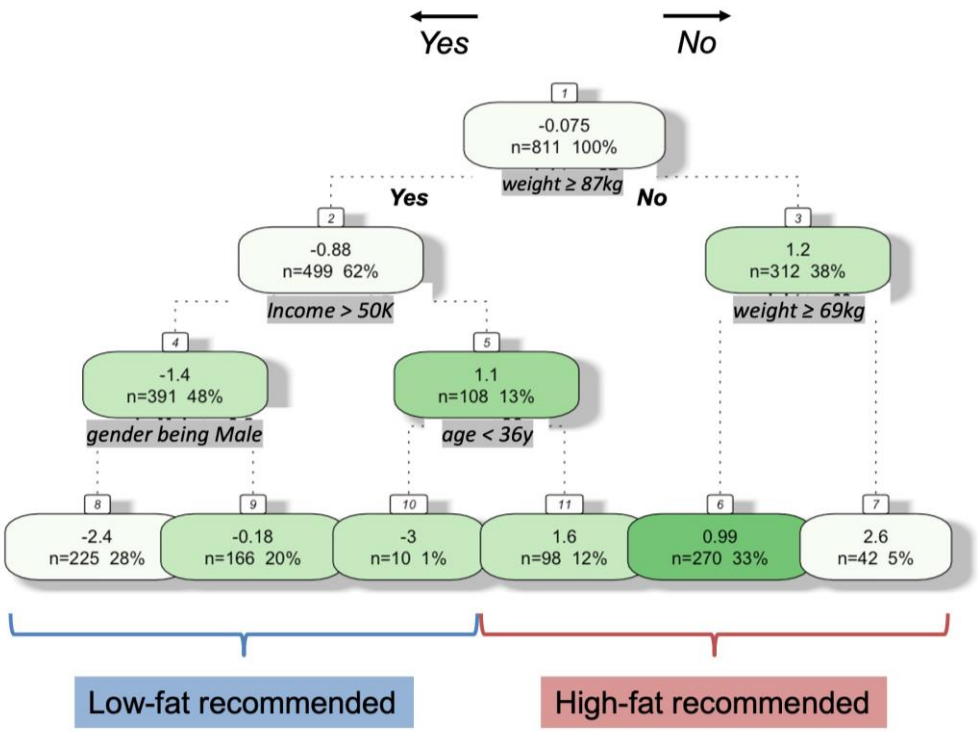


Figure 3.9: Decision tree explaining the ITR of individualized fat intervention in the POUNDS Lost trial

An ITR of high fat/ low fat intervention is developed in POUNDS LOST using R-learner, and a regression tree is fit to predict estimated CATE. Values indicate estimated CATE, average weight loss by high-fat intervention compared with low-fat intervention (i.e. positive values mean greater weight loss by high-fat intervention compared with low-fat).

Discussion

In the present post-hoc analysis of the POUNDS Lost trial, we applied machine-learning algorithms on baseline covariates to identify the subpopulations in which the effects of distinct macronutrient-based dietary interventions on 2-year weight loss were heterogeneous. We identified effect heterogeneity of the fat intervention but not the protein intervention. The most striking algorithm was the R-learner, in which those with the highest estimated treatment effect had 2.7 kg net weight reduction with low-fat compared with high-fat intervention, although the

ATE was almost null. The ITR based on the causal forest algorithm showed a 4.1 kg weight reduction in the total population, and a 0.84 kg weight reduction compared with random assignment of the fat intervention. Those with estimated greater response to low-fat diet was characterized by a higher proportion male, of tall height, and high weight and waist. Although the overall effect size was modest, our machine-learning approach successfully developed an effective ITR based on the baseline covariates.

Precision nutrition has been a major focus in nutritional epidemiology in which dietary interventions may be best tailored to individuals according to their characteristics.[115] Although there are many approaches to advance precision nutrition, the present approaches based on CATE and ITR using an RCT could directly provide evidence for this purpose. The CATE-based ITR can be quite flexible and may better capture the complex effect modification across multiple variables. The drawbacks of this approach include that an exploration of CATE generally always be under-powered given that an RCT is only powered to detect ATE. However, our application in the POUNDS Lost successfully resulted in a discovery of a novel treatment algorithm for a low-fat intervention based on baseline characteristics, informing the precision nutrition approach. Of note, the observed sex difference is in line with previous studies showing female mice more resistant to high-fat diet-induced obesity onset than males[116,117], as well as with human trials.[118,119]

There are several limitations of our study. First, we imputed the outcome values with a stable rate for weight regain in accordance with the primary study. This imposed an assumption of such natural changes of participant weight, which was untestable. Second, we applied

intention-to-treat analysis, not accounting for the adherence to the assigned interventions. However, the present CATE estimation and ITR development implicitly took account for the adherence by the correlation with covariates, since individuals with low adherence would lead to little changes in weight. Third, the study consists of 80% White participants and thus the generalizability is limited. Fourth, the lack of observed heterogeneity for the protein intervention may be due to lack of HTE or the type 2 error. Combining RCTs with the same or similar treatment arms would increase the chance of detecting meaningful signals. Finally, our treatment algorithm should be tested in another RCT to assess generalizability.

3.8 Concluding Remark

In this paper, general approaches to HTE-related research in medical and epidemiological RCT settings. The methodological approaches of HTE-related study depend on the aims of the study. Meta-learners and the causal forest are the popular approaches to estimate CATEs that can be easily implemented, while DR-learner and R-learner are powerful in a setting of high-dimensionality. Cross-fitting is recommended to gain statistical efficiency. Statistically valid evaluation metrics for an ITR include PAV, PAPE, AUPEC, GATE, and RATE metrics. In the POUNDS Lost application, the R-learner algorithm detected significant HTE by high vs. low-fat diet according to baseline characteristics. The present approach and observations may thus contribute to the growing evidence for precision nutrition by directly suggesting practical treatment strategies.

Chapter 4

Association of plasma branched chain amino acid with biomarkers of inflammation and lipid metabolism in women

Author: Rikuta Hamaya, MD, MSc[1,2], Samia Mora, MD, MHS[1,3], Patrick R Lawler, MD, MPH[4], Nancy R Cook, ScD[1,5], Paul M Ridker, MD, MPH[1], Julie E Buring, ScD[1,5], I-Min Lee, MBBS, ScD[1,5], JoAnn E Manson, MD, DrPH[1,5,6], Deirdre K Tobias, ScD[1,2]

[1]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[2]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA

[3]Center for Lipid Metabolomics and Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[4]Peter Munk Cardiac Centre, University Health Network, and Heart and Stroke Richard Lewar Centre of Excellence in Cardiovascular Research, University of Toronto, Toronto, ON, Canada

[5]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[6]Mary Horrigan Connors Center for Women's Health and Gender Biology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

4.1 Abstract

Backgrounds: Branched-chain amino acids (BCAAs; isoleucine, leucine and valine) correlate with insulin resistance and poor glucose control, which may in part explain associations between type 2 diabetes (T2D) and cardiovascular disease (CVD). However, the relationships of BCAAs with other cardiometabolic pathways, including inflammation and dyslipidemia, are unclear. We hypothesized that plasma BCAAs would correlate with multiple pathways of cardiometabolic dysfunction.

Methods: We conducted a cross-sectional analysis among 19,472 participants (mean age=54.9 years, SD=7.2 years) in the Women's Health Study without a history of T2D, CVD, or cancer. We quantified the concentrations of individual biomarkers of inflammation and lipids, across quartiles of BCAAs, adjusting for age, smoking, BMI, physical activity, and other established CVD risk factors at blood draw.

Results: Women in the highest vs. lowest quartiles of plasma BCAAs had higher inflammatory markers including high-sensitivity C-reactive protein (multivariable-adjusted means: 1.96 vs. 1.43 mg/L), fibrinogen (367 vs. 362 mg/dL), soluble intercellular cell adhesion molecule-1 (361 vs. 353 ng/mL), and glycoprotein acetylation (407 vs. 371 μmol/L) (p-trend=0.0002 for fibrinogen; $p<0.0001$ for others). Similarly for lipids, women with higher BCAAs had lower HDL-c (49.0 vs. 55.0 mg/dL), and higher triglycerides (143 vs. 114 mg/dL), LDL-c (133 vs. 124 mg/dL), and lipoprotein insulin resistance score (52.6 vs. 37.3) (all: $p<0.0001$). Similar associations with these biomarkers were observed in isoleucine, leucine and valine, respectively.

Conclusions: Higher circulating BCAA concentrations are associated with adverse profiles of biomarkers of inflammation and dyslipidemia independent of established CVD risk factors, and thus may reflect poorer cardiometabolic health through multiple pathways.

4.2 Introduction

Type 2 diabetes (T2D) is one of the most prevalent chronic diseases, which is strongly linked to the development of cardiovascular disease (CVD). However, mechanisms underlying these interrelated diseases are poorly understood. Characterizing the metabolite traits shared by T2D and CVD years prior to their diagnosis may allow the identification of high-risk individuals, increase opportunities for early intervention and prevention, and uncover shared pathways for potential novel therapeutic targets.

Branched-chain amino acids (BCAAs; isoleucine, leucine, and valine) are essential amino acids that are preserved in muscle and utilized to synthesize proteins and perform various metabolic/physiological functions[120]. The degradation of BCAAs occurs mainly in mitochondria, eventually producing acetyl-CoA or succinyl-CoA, which enters Krebs cycle. This process occurs outside of liver, which lacks the expression of mitochondrial branched-chain aminotransferase. Dysfunction at each step in this process can lead to an accumulation of plasma BCAAs, such as in maple syrup urine disease, a deficiency of branched-chain α-ketoacid dehydrogenase complex. Recent evidence shows lowered branched-chain keto acid dehydrogenase activity[121], muscle breakdown[122], as well as excess adiposity[15,123], contribute to higher circulating BCAAs. Circulating BCAAs are highly predictive of incident T2D[124,125], and we have previously demonstrated the positive association between BCAAs with incident CVD risk.[126] Mendelian randomization studies suggest a causal role of impaired BCAA metabolism in the disease process of T2D[127,128] although evidence is inconclusive. However, the relationships

of BCAAs with other cardiometabolic traits predictive of CVD incidence have not been explored, which may contribute to our understanding of the T2D/CVD relationship.

Evidence supports T2D as an inflammatory disease in terms of hypoxia, cell death, or various inflammatory cytokines/chemokines, which may partly explain the consequent development of CVD.[129] A variety of molecules are involved in systemic inflammation, some represented as biomarkers including C-reactive protein (CRP), fibrinogen, soluble intercellular adhesion molecule-1 (sICAM-1), or glycoprotein acetylation (GlycA), which may be related to impaired glucose metabolism and incident T2D.[130–133] Dyslipidemia may also contribute to the relationship between T2D and CVD risk. Various lipid and lipoprotein abnormalities are associated with impaired glucose metabolism, including higher triglycerides and lower HDL cholesterol, primarily triggered by the overproduction of triglyceride-rich VLDL particle (VLDL-p) mediated by insulin.[134,135] A recently derived lipoprotein insulin resistance score (LPIR) is a composite biomarker based on six lipid metabolite features[136] reflecting risk for T2D.[136–138] However, the relationships between circulating BCAA levels with each of these cardiometabolic pathways represented by these individual biomarkers are largely unknown.

We therefore aimed to evaluate the interrelationship of BCAAs with established cardiometabolic traits to further characterize BCAAs as metabolites of T2D and CVD risks. We conducted cross-sectional analyses for the associations of plasma BCAAs with inflammatory and lipid biomarkers in the Women's Health Study (WHS), which recruited a large cohort of US women. We hypothesized that higher plasma total or individual BCAAs would be correlated with cardiometabolic biomarkers representing lipid or inflammatory profiles, possibly

independent of concurrent traits of glycemic control as measured by hemoglobin A1C (HbA1c). We also examined effect modification by BMI, a consistent predictor of higher total plasma BCAAs.[139–141]

4.3 Methods

Data described in the manuscript, code book, and analytic code will be made available upon request pending application and approval. Written informed consent was obtained from all participants and the study protocol was approved by the Institutional Review Board of the Brigham and Women's Hospital (Boston, Massachusetts). The full methods are now available as supplemental data.

Study design and participants

Our analysis included participants in the WHS, a completed, randomized, placebo-controlled, factorial trial of low-dose aspirin and vitamin E for the primary prevention of cardiovascular disease and cancer (ClinicalTrials.gov identifier: NCT00000479), whose participants are currently being followed observationally. The trial randomized 39,876 female US health professionals who were aged 45 years or older without a history of cancer (except non-melanoma skin cancer) or cardiovascular disease to low dose aspirin (100 mg every other day) and vitamin E (600 IU every other day).[142] Our analyses included participants who were fasting (at least 8 hours since last eating or drinking) at baseline blood draw, with assay data available for isoleucine, leucine and valine concentrations. We excluded samples from participants reporting non-fasting status at blood draw to minimize extraneous variability from

recent food intake.[143] We excluded those reporting a history of diabetes at baseline, which left 19,472 women eligible for the analysis.

Plasma BCAA measurements

Blood samples were collected voluntarily prior to randomization, shipped to the laboratory on ice via overnight courier, processed, and stored at −170°C in vapor liquid nitrogen until biomarker measurements were performed. Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy (LipoScience, now LabCorp; Raleigh, North Carolina) quantified isoleucine, leucine, and valine.[144] The amino acid signal amplitudes deconvoluted from the NMR spectra were converted into μmol/L as previously described.[144] The inter-assay reproducibility among stored Women's Health Study (WHS) plasma samples was previously estimated with good intra- and inter-assay coefficient of variations (CVs) for isoleucine (5.9-6.1%), leucine (4.5-4.9%), and valine (1.5-2.1%).

Inflammatory biomarkers

Assay laboratory methods for the inflammatory biomarkers have been previously described in detail[145–149]. Briefly, high-sensitivity CRP (hsCRP) was assayed on a Hitachi 917 auto- analyzer with a high-sensitivity immunoturbidimetric assay (Denka Seiken, Tokyo, Japan).[146] Fibrinogen was measured using an immunoturbidimetric assay with internal standards (Kamiya Biomedical, Seattle, Washington).[147] LipoScience (now LabCorp; Raleigh, North Carolina) measured GlycA from 400 MHz plasma proton (1H) NMR spectra. Signals were quantified through deconvolution analysis from signal amplitudes that originated from N-acetyl methyl group protons of the N-acetylglucosamine moieties of specific serum proteins and

reflects serum concentration and glycosylation state of main acute-phase reactants such as α1-acid glycoprotein, haptoglobin, α1-antitrypsin, α1-antichymotrypsin and transferrin[145,148]. sICAM-1 was assayed with the R&D assay via a standard quantitative sandwich enzyme immunoassay technique by enzyme-linked immunosorbent assay that does not detect the modified form of sICAM-1 in African Americans (R&D Systems; Minneapolis, Minnesota).[149] Laboratory inter-assay coefficients of variation for hsCRP, fibrinogen, GlycA, and sICAM-1were 3.0%, 1.17%, 1.9%, and 7.4%, respectively.

Lipid biomarkers

LDL cholesterol (LDL-c), HDL cholesterol (HDL-c), and triglycerides were measured by direct assays (Roche Diagnostics, Indianapolis, IN, USA). Lipoprotein subclasses, including large, medium and small VLDL; large, medium and small HDL; and large and small LDL concentrations, were measured using targeted metabolomics approach (Liposcience, Inc. now LabCorp, Raleigh, NC, USA) by detecting the proton NMR spectroscopy methyl group signal[136,150]. Mean VLDL, LDL, and HDL particle sizes derive from weighted averages of each subclass diameter relative to its mass percentage. LPIR, a composite weighted score of six lipoprotein parameters with homeostasis model assessment of insulin resistance (HOMA-IR), was calculated based on the profiles of lipoprotein subclasses as previously described.[136] LPIR interassay repeatability from 80 replicate analyses of 8 plasma pools over 20 days had a CV of 6% within-run and 9% between-run.[136]

Covariates

At baseline, questionnaires captured information on demographics, health status, reproductive history, and lifestyle characteristics, as previously described.[142] A semi-quantitative food frequency questionnaire (FFQ) was used to evaluate usual dietary intake, and the Alternative Health Eating Index 2010 (aHEI-2010) dietary pattern score was derived as an estimate of diet quality of habitual diet.[151]

Statistical analysis

Baseline characteristics were compared across the quartiles of summed concentrations of plasma BCAAs (isoleucine [μmol/L] + leucine [μmol/L] + valine [μmol/L]). Spearman correlation coefficients were derived to compare the associations between total plasma BCAAs, inflammatory/lipid biomarkers, and HbA1c. A correlation network was leveraged for this visualization using qgraph R package.

We used linear regression models to evaluate the associations between total BCAAs as the independent variable and individual cardiometabolic biomarkers as the dependent variables. Total plasma BCAAs were analyzed categorically using quartiles. We analyzed the biomarkers of inflammation and lipid continuously after excluding the top and bottom 0.5% of the distributions to minimize the influence of outliers, and high-sensitivity CRP (hsCRP) values >20 mg/L to exclude probable acute infections. Biomarkers other than LDL cholesterol (LDL-c) and LPIR were logarithmically transformed to improve approximation of a normal distribution.

Models were adjusted for the assignment to aspirin and to vitamin E group, and age at blood draw (continuous) (age-adjusted model), and additionally for body mass index (BMI)

(kg/m^2, continuous), white race, family history of T2D, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural or non-natural menopause], unsure), use of postmenopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (0, 1, 2, ≥3), leisure-time physical activity as total MET-hour/week (quintiles), the Alternative Health Eating Index 2010 (aHEI-2010) (quintiles), alcohol consumption (none, <10g/day, 10 to <20g/day, ≥20g/day), and the use of cholesterol-lowering drugs. The missing values in covariates were all <1%, and we imputed the median values for continuous characteristics and coded as 'never' or 'unsure' categories for categorical factors. We calculated the adjusted least-square means and 95% confidence intervals (CIs) of each biomarker across total BCAA quartiles, in separate models. Tests for linear trends of inflammatory/lipid biomarkers across increasing plasma BCAA levels were conducted by assigning participants the median BCAA level in each BCAA quartile and modeling as a continuous variable. In addition, we explored whether any associations of BCAAs with inflammatory/lipid biomarkers were independent of impaired glucose metabolism by further adjusting models for HbA1c.

We also evaluated the relationship between continuous total BCAAs with each biomarker standardized to a common scale per 1 standard deviation of its distribution in the WHS to facilitate visual comparison and contrast their magnitudes of associations. We performed analyses stratified by BMI at blood draw (<25.0, 25.0 to 29.9, ≥30.0 kg/m^2) to examine potential effect modification by body weight. The statistical interaction between BCAA and BMI was assessed for each biomarker using a 2 degrees of freedom log-likelihood ratio test comparing models with and without the categorical interaction terms between total continuous BCAAs and categorical BMI categories. Secondary analyses were conducted repeating the above models for

quartiles of individual plasma BCAAs, isoleucine, leucine, and valine. In sensitivity analyses we analyzed women age<60 years and ≥60 years separately.

All analyses were conducted using R (The R Foundation). Bonferroni correction was used to adjust the threshold of two-sided p-value for the multiple comparisons, which was 0.006 given the eight biomarkers of interest.

4.4 Results

Baseline characteristics

In 19,472 women in the WHS included in this analysis, the mean (SD) age was 54.9 (7.2) years with median [IQR] BMI of 24.8 [22.4, 28.3] kg/m^2. Median [IQR] concentration of summed BCAAs was 396 [349, 450] μmol/L (the distribution is shown in **Supplemental Figure 4.1**). The baseline characteristics of the participants according to the quartiles of plasma BCAA levels are summarized in **Table 4.1**. Higher BCAA was associated with older age, higher BMI, lower physical activity level, lower diet quality represented by AHEI-2010, lower alcohol intake, lower prevalence of smoking, use of cholesterol-lowering drugs, and post-menopausal status.

	BCAA quartile			
	Quartile 1 n=4868	Quartile 2 n=4868	Quartile 3 n=4868	Quartile 4 n=4868
Age, years	54.7 (7.3)	55.0 (7.4)	55.2 (7.1)	54.9 (7.0)
Race, %				
White, Non-Hispanic	4640 (96.0)	4643 (96.1)	4622 (95.9)	4569 (94.7)
Hispanic	41 (0.8)	43 (0.9)	48 (1.0)	58 (1.2)
African American	81 (1.7)	67 (1.4)	86 (1.8)	89 (1.8)
Asian/Pacific Islander	51 (1.1)	61 (1.3)	52 (1.1)	83 (1.7)
Other/unknown	19 (0.4)	19 (0.4)	13 (0.2)	25 (0.5)
Smoking, %				

Never	2380 (48.9)	2521 (51.8)	2612 (53.7)	2613 (53.7)
Past use	1844 (37.9)	1800 (37.0)	1729 (35.5)	1698 (34.9)
Current use	644 (13.2)	547 (11.2)	527 (10.8)	557 (11.4)
Family history of T2D, %	1065 (21.9)	1148 (23.6)	1223 (25.1)	1462 (30.0)
BMI, %				
<25	3418 (70.2)	2921 (60.0)	2356 (48.4)	1535 (31.5)
≥25, <30	1100 (22.6)	1422 (29.2)	1621 (33.3)	1845 (37.9)
≥30	350 (7.2)	525 (10.8)	891 (18.3)	1488 (30.6)
Menopausal status, %				
Premenopausal	1469 (30.2)	1306 (26.8)	1230 (25.3)	1208 (24.8)
Postmenopausal, natural	1898 (39.0)	1929 (39.6)	1926 (39.6)	1812 (37.2)
Postmenopausal, non-natural	679 (13.9)	778 (16.0)	845 (17.4)	876 (18.0)
Uncertain	822 (16.9)	855 (17.6)	867 (17.8)	972 (20.0)
PA, MET-h/w	10.5 [3.4, 23.0]	10.0 [3.4, 21.2]	9.1 [2.9, 20.5]	6.9 [2.2, 17.5]
aHEI-2010	49 [42, 56]	48 [42, 55]	48 [42, 55]	48 [41, 54]
Alcohol intake, g/day	1.2 [0.0, 6.5]	1.1 [0.0, 5.7]	0.86 [0.0, 4.6]	0.86 [0.0, 2.9]
Cholesterol drugs	120 (2.5)	146 (3.0)	171 (3.5)	201 (4.1)
Total BCAAs	317 [292, 334]	373 [361, 385]	421 [408, 434]	493 [468, 533]
hsCRP, mg/L	1.3 [0.54, 2.8]	1.7 [0.69, 3.6]	2.1 [0.93, 4.3]	2.8 [1.4, 5.2]
Fibrinogen, mg/dL	337 [298, 387]	345 [306, 394]	356 [313, 407]	365 [319, 418]
GlycA,µmol/L	353 [316, 395]	373 [332, 416]	389 [349, 434]	410 [366, 455]
sICAM-1, ng/mL	333 [294, 381]	337 [297, 386]	344 [303, 394]	356 [311, 408]
Triglyceride,mg/dL	94 [69, 134]	106 [76, 151]	118 [83, 168]	140 [99, 193]
LDL cholesterol, mg/dL	116 [97, 138]	122 [102, 144]	125 [104, 148]	128 [108, 152]
HDL cholesterol, mg/dL	57 [48, 68]	55 [46, 65]	52 [43, 61]	47 [40, 56]
LPIR score	25 [15, 44]	33 [18, 52]	41 [22, 61]	56 [35, 71]
HbA1c, %	5.0 [4.8, 5.1]	5.0 [4.8, 5.1]	5.0 [4.9, 5.2]	5.1 [4.9, 5.3]

Table 4.1: Characteristics of 19,472 Women's Health Study participants at baseline blood draw according to quartiles of plasma BCAA level

Correlation between BCAA and inflammation, lipid, and HbA1c

Figure 4.1 illustrates the spearman correlation network between total BCAAs, inflammatory biomarkers, and lipid biomarkers. Total plasma BCAAs were significantly correlated with all biomarkers ($p<0.0001$), with highest correlations observed with LPIR ($\rho=0.35$) and GlycA ($\rho=0.30$) (hsCRP: $\rho=0.24$, fibrinogen: $\rho=0.13$, sICAM-1: $\rho=0.11$, triglyceride: $\rho=0.26$, HDL: $\rho=-0.27$, LDL: $\rho=0.14$, HbA1c: $\rho=0.17$). hsCRP with GlycA had the highest correlations between inflammatory biomarkers ($\rho=0.58$, $p<0.0001$). The correlation matrix is shown in **Supplemental Table 4.1**.

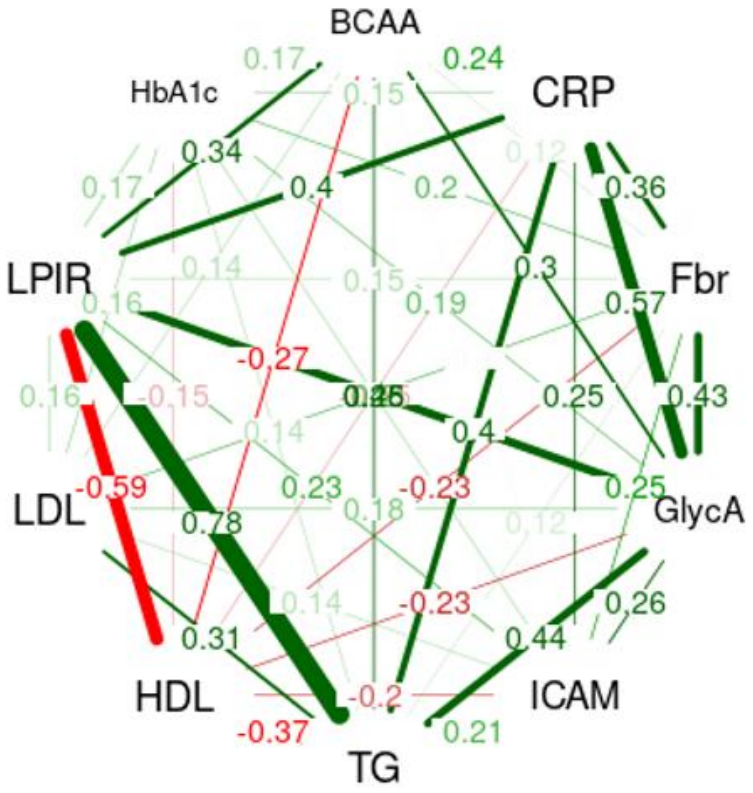


Figure 4.1: Correlation network between BCAA, inflammatory biomarkers, lipid biomarkers, and HbA1c

Correlation network between BCAA, hsCRP, fibrinogen, GlycA, sICAM-1, triglyceride, HDL-c, LDL-c, LPIR, and HbA1c is shown. Spearman correlation coefficients (ρ) were calculated in each pair of biomarkers and only associations with ρ>0.10 are visualized as lines. Thicker lines indicate stronger correlations and green and red lines represent positive and negative correlations, respectively.

Associations of BCAA and inflammation/lipid

Table 4.2 summarizes the associations between BCAA and cardiometabolic biomarkers. Women in the highest vs. lowest quartiles of plasma BCAAs had higher hsCRP (adjusted mean [95% CI]: 1.96 [1.85, 2.07] vs. 1.43 [1.35, 1.51] mg/L), fibrinogen (367 [363, 371] vs. 362 [358, 366] mg/dL), sICAM-1 (361 [357, 365] vs. 353 [349, 357] ng/mL), and GlycA (407 [403, 410] vs. 371 [368, 375] μmol/L) (p-trend=0.0002 for fibrinogen; p<0.0001 for others).

	BCAA Q1	BCAA Q2	BCAA Q3	BCAA Q4	†p for linear trend
Inflammation	Geometric mean [95% confidence interval]				
hsCRP, mg/L					
Age-adjusted	1.22 [1.19, 1.26]	1.54 [1.49, 1.58]	1.92 [1.86, 1.98]	2.51 [2.44, 2.59]	<0.0001
*Multivariable-adjusted	1.43 [1.35, 1.51]	1.62 [1.53, 1.71]	1.8 [1.7, 1.91]	1.96 [1.85, 2.07]	<0.0001
Fibrinogen, mg/dL					
Age-adjusted	340 [338, 341]	347 [345, 349]	356 [354, 358]	365 [363, 367]	<0.0001
*Multivariable-adjusted	362 [358, 366]	365 [361, 369]	368 [364, 372]	367 [363, 371]	0.0002
sICAM-1, ng/mL					
Age-adjusted	338 [336, 340]	340 [338, 342]	346 [344, 348]	358 [356, 360]	<0.0001
*Multivariable-adjusted	353 [349, 357]	353 [349, 358]	356 [352, 360]	361 [357, 365]	<0.0001
GlycA, µmol/L					
Age-adjusted	354 [352, 355]	371 [369, 373]	387 [386, 389]	406 [404, 408]	<0.0001
*Multivariable-adjusted	371 [368, 375]	385 [382, 389]	397 [393, 400]	407 [403, 410]	<0.0001
Lipid					
Triglyceride, mg/dL					
Age-adjusted	99 [98, 100]	108 [107, 110]	119 [117, 120]	139 [137, 141]	<0.0001
*Multivariable-adjusted	114 [111, 117]	121 [118, 124]	128 [125, 132]	143 [140, 147]	<0.0001
HDL-c, mg/dL					
Age-adjusted	56.9 [56.5, 57.3]	54.3 [53.9, 54.7]	51.6 [51.2, 51.9]	47.3 [47.0, 47.6]	<0.0001
*Multivariable-adjusted	55.0 [54.2, 55.7]	53.1 [52.4, 53.9]	51.5 [50.8, 52.3]	49.0 [48.3, 49.6]	<0.0001
LDL-c, mg/dL					
Age-adjusted	119 [118, 120]	124 [124, 125]	127 [126, 128]	131 [130, 132]	<0.0001
*Multivariable-adjusted	124 [122, 126]	129 [127, 131]	131 [129, 133]	133 [131, 135]	<0.0001
LPIR score					
Age-adjusted	30.5 [29.9, 31.2]	35.9 [35.3, 36.5]	42 [41.4, 42.6]	52.7 [52.1, 53.4]	<0.0001
*Multivariable-adjusted	37.3 [36.1, 38.6]	41 [39.8, 42.3]	45.1 [43.9, 46.3]	52.6 [51.4, 53.9]	<0.0001

Table 4.2: Adjusted means of inflammatory and lipid biomarkers by the quartiles of BCAA level

Numbers are adjusted geometric means [95% confidence intervals] based on linear regression.
*Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), cholesterol lowering drugs, and BMI (continuous).
Test for trend was based on a variable containing the median value for each quartile.
†P-trend threshold was 0.006 after Bonferroni correction.

Further adjustment for HbA1c attenuated these associations, and the association with fibrinogen became not statistically significant after consideration of multiple comparisons (**Table 4.3**).

	BCAA Q1	BCAA Q2	BCAA Q3	BCAA Q4	*p-trend
	Geometric mean [95% confidence interval]				
Inflammation					
hsCRP, mg/dL	1.41 [1.33, 1.5]	1.6 [1.51, 1.7]	1.78 [1.68, 1.89]	1.9 [1.8, 2.02]	<0.0001
Fibrinogen, mg/dL	361 [357, 365]	364 [360, 368]	367 [363, 371]	364 [360, 368]	0.0090
sICAM-1, ng/mL	352 [348, 356]	352 [348, 357]	355 [351, 359]	358 [354, 362]	<0.0001
GlycA, µmol/L	371 [367, 374]	384 [381, 388]	395 [392, 399]	404 [400, 408]	<0.0001
Lipid					
Triglyceride, mg/dL	113 [110, 116]	121 [117, 124]	128 [124, 131]	142 [138, 146]	<0.0001
HDL-c, mg/dL	55.1 [54.3, 55.9]	53.3 [52.6, 54.0]	51.7 [51.0, 52.4]	49.3 [48.6, 50.0]	<0.0001
LDL-c, mg/dL	124 [122, 126]	129 [127, 131]	131 [129, 133]	132 [130, 134]	<0.0001
LPIR score	37.2 [35.9, 38.4]	40.8 [39.6, 42]	44.9 [43.6, 46.1]	52 [50.8, 53.3]	<0.0001

Table 4.3: Adjusted means of inflammation/lipid biomarkers by the quartiles of BCAA level after adjustment for HbA1c

Numbers are adjusted geometric means [95% confidence intervals] based on linear regression. Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, BMI (continuous), and HbA1c (continuous).
*Test for trend was based on a variable containing the median value for each quartile. P-trend threshold was 0.006 after Bonferroni correction.

Higher BCAA was associated with elevated triglycerides (143 [140, 147] vs. 114 [111, 117] mg/dL), LDL-c (133 [131, 135] vs. 124 [122, 126] mg/dL) and LPIR (52.6 [51.4, 53.9] vs. 37.3 [36.1, 38.6] unit), and lower HDL cholesterol (HDL-c) (49.0 [48.3, 49.6] vs. 55.0 [54.2,

55.7] mg/dL) in the multivariable-adjusted models (all: p-trend<0.0001) (**Table 4.2**). Further adjustment for HbA1c did not substantially attenuate these associations (**Table 4.3**). Standardized differences of each cardiometabolic biomarker per SD difference of BCAA levels after adjustment of confounders were illustrated in **Figure 4.2**. In inflammatory biomarkers, the strongest association with BCAA was seen for GlycA (0.203 [95%CI: 0.19. 0.217] per SD difference of BCAA). Among lipid biomarkers, the association with BCAA was strongest in LPIR (0.254 [95%CI: 0.241. 0.267] per SD difference of BCAA).

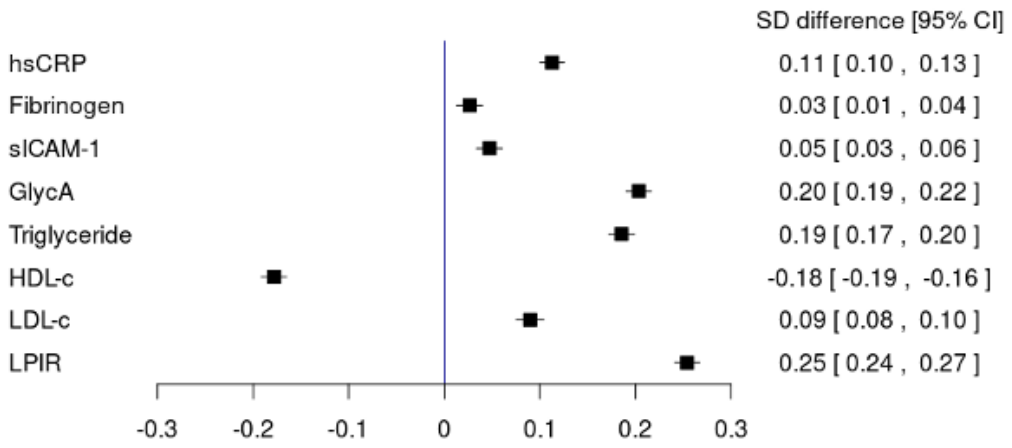


Figure 4.2: Standardized differences of cardiometabolic biomarkers per SD changes of BCAA levels

Linear regressions of standardized biomarkers constructed by standardized continuous total BCAA levels and covariates [age at randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or non-white), family history of diabetes, smoking history (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous)]. Standardized differences [95% confidence interval] per SD of BCAAs are shown.

Stratified analysis by BMI

Supplemental Table 4.2 (least square means) and Supplemental Figure 4.2 (standardized differences) illustrates the associations between BCAA concentrations and inflammation/lipid biomarkers stratified by BMI. Overall, these associations differed across BMI categories, but the patterns were not consistent. In inflammatory biomarkers, after consideration of multiple comparisons, the associations of BCAAs and sICAM-1 were significantly different according to BMI (p-interaction<0.0001). In particular, sICAM-1 was robustly associated with BCAA concentration in women with BMI≥25 kg/m^2 (p-trend<0.0001) but not in those with BMI<25 kg/m^2 (p-trend=0.19). Among lipid biomarkers, LDL-c and LPIR score were differentially related to BCAA levels according to BMI categories (p-interaction=0.0004 and 0.0002, respectively). There were no significant interactions between BMI and hsCRP or HDL-c with BCAAs.

Associations between individual BCAAs and biomarkers

We also assessed the relationships between inflammatory/lipid biomarkers and the individual BCAAs, isoleucine (**Supplemental Table 4.3 and Supplemental Figure 4.3**), leucine (**Supplemental Table 4.4 and Supplemental Figure 4.4**), and valine (**Supplemental Table 4.5 and Supplemental Figure 4.5**). The associations between fibrinogen and isoleucine and valine were not significant in the multivariable-adjusted models. The other associations were significant and the directions were the same as the relationships with summed BCAA levels.

Sensitivity analysis

The results were similar when we stratified by age <60 or ≥60 years (**Supplemental Tables 4.6 and Supplemental Figure 4.6)**, with all p-values for interaction non-significant. In

the adjusted models, circulating BCAA concentration was significantly associated with all of the inflammatory and lipid biomarkers except for fibrinogen. Additional sensitivity analyses by fasting status showed similar results for nonfasting BCAA measurements (not shown).

4.5 Discussion

In this large cross-sectional study of US women, plasma BCAA concentrations were associated with biomarkers of inflammation and dyslipidemia, indicative of their correlation with an overall poorer cardiometabolic health profile. Among inflammatory biomarkers, plasma BCAAs were moderately associated with hsCRP and GlycA. Higher BCAA levels were also moderately associated with the LPIR score. The interactions of BCAAs and BMI varied according to cardiometabolic biomarkers. The findings for the individual BCAAs, isoleucine, leucine and valine and inflammatory/lipid biomarkers were similar to total BCAAs.

Few studies have investigated the relationships between circulating BCAA metabolites and inflammation in humans. In a study of 286 Finnish twins, hsCRP levels were modestly correlated with isoleucine and leucine, but not with valine.[152] Among 611 Chinese adults, hsCRP was not significantly associated with serum BCAAs after adjustment for age, sex, smoking and alcohol consumption ($p=0.064$).[153] However, this was a small and diverse population that included participants with ages ranging from 21 to 110 years, and without exclusion for prevalent T2D at blood draw, which may introduce variability. We investigated circulating BCAAs in relation to biomarkers representing various inflammatory pathways. Associations of BCAAs with hsCRP and GlycA persisted even after adjusting for HbA1c, a marker of glycemic

control, suggesting BCAAs may be related to cardiometabolic risk independent of this T2D-related glycemic trait. GlycA is a nuclear magnetic resonance (NMR) signal basically reflecting the glycosylation and abundance of α1-acid glycoprotein, haptoglobin, α1-antitrypsin, α1-antichymotrypsin and transferrin.[145] Evidence indicates that GlycA itself[132], as well as the major contributors including α1-acid glycoprotein[154], α1-antitrypsin[155] and transferrin[156], are associated with incident T2D. The association of BCAAs with fibrinogen was attenuated after adjusting for HbA1c, suggesting that BCAAs are not likely to be independently related to this inflammatory marker.

Higher circulating BCAAs were associated with biomarkers of dyslipidemia. The relationship between BCAAs and the LPIR score persisted with adjustment for HbA1c and was similar across BMI strata. These findings are consistent with known relationships between BCAAs and insulin resistance, which correlates highly with LPIR[124,136,138]. Of note, dyslipidemia, as well as inflammation, may potentially precede the development of insulin resistance, as is supported by prior evidence that elevated LPIR[137] and hsCRP[132,157] are involved upstream of T2D progression; in addition, both biomarkers were strongly associated with incident CHD[158,159]; therefore, impaired BCAA metabolism, capturing multiple aspects of inflammation and dyslipidemia, may represent a shared pathology predisposing to T2D and CVD. However, the temporal relationship of these correlations cannot be established given the cross-sectional nature of this analysis.

Strengths and limitations

Strengths of this study include the large sample size, measured inflammatory/lipid biomarkers, and detailed demographic, lifestyle, and health information to carefully control for potential confounders. Our study has limitations, however, including the cross-sectional design, which precludes the ability to establish the temporality between biomarkers. The female participants were predominantly white with higher socioeconomic status[160], thus limiting generalizability of the study findings. Also, we cannot rule out residual confounding by unmeasured confounding by other determinants of BCAAs and cardiometabolic risk, including other biomarkers that were not included in our investigation.

Conclusion

In a large cohort of US women without T2D or CVD, plasma BCAAs were associated with biomarkers of inflammation (hsCRP, sICAM-1 and GlycA) and dyslipidemia (triglyceride, LDL-c, HDL-c and LPIR), indicative of an overall poorer cardiometabolic health profile. BCAAs remained positively associated with some of these pathways independent of impaired glucose metabolism, supporting elevated BCAAs may be an independent component of cardiometabolic risk.

REFERENCES

1. Mensah GA, Fuster V, Murray CJL, Roth GA, Mensah GA, Abate YH, Abbasian M, Abd -Allah Foad, Abdollahi A, Abdollahi M, Abdulah DM, Abdullahi A, Abebe AM, Abedi A, Abedi A, Abiodun OO, Ali HA, Abu -Gharbieh Eman, Abu -Rmeileh Niveen M. E., Aburuz S, Abushouk AI, Abu -Zaid Ahmed, Adane TD, Adderley NJ, Adebayo OM, Aden B, Adeyeoluwa TE, Adeyomoye OI, Sakilah AQE, Afrashteh F, Afyouni S, Afzal S, Agasthi P, Agodi A, Aguilera ACE, Agyemang -Duah Williams, Ahinkorah BO, Ahmad A, Ahmad D, Ahmad F, Ahmad MM, Ahmed A, Ahmed H, Ahmed MB, Ahmed SA, Ajami M, Akinosoglou K, Ala M, Ali A-ATM, Alalalmeh SO, Al -Aly Ziyad, Alam N, Al -amer Rasmieh Mustafa, Alashi A, Albashtawy M, AlBataineh MT, Alema HB, Alemi S, Alemu YM, Saeed A-GAA, Alhabib KF, Naji AFA, Ali MU, Ali R, pursuing P, Shujait ASS, Alicandro G, Alikhani R, Aljunid SM, Alla F, Almahmeed W, Al -Marwani Sabah, Alonso J, Al -Raddadi Rajaa M., Alvi FJ, Alvis -Guzman Nelson, Alvis -Zakzuk Nelson J., Alwafi H, Aly H, Amegbor PM, Amin TT, Amindarolzarbi A, Amini -Rarani Mostafa, Amiri S, Ammirati E, Anand T, Ancuceanu R, Anderlini D, Anil A, Ansari G, Anyanwu PE, Anyasodor AE, Carace AGL, Arabloo J, Arafat M, Aravkin AY, Aremu O, Armocida B, et al. Global Burden of Cardiovascular Diseases and Risks, 1990-2022. *Journal of the American College of Cardiology*. 2023;82:2350–2473.

2. Huang Y, Van Horn L, Tinker LF, Neuhouser ML, Carbone L, Mossavar-Rahmani Y, Thomas F, Prentice RL. Measurement error corrected sodium and potassium intake estimation using 24-hour urinary excretion. *Hypertension*. 2014;63:238–244.

3. Siddique J, Daniels MJ, Carroll RJ, Raghunathan TE, Stuart EA, Freedman LS. Measurement error correction and sensitivity analysis in longitudinal dietary intervention studies using an external validation study. *Biometrics*. 2019;75:927–937.

4. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, Ramirez A, Schlaich M, Stergiou GS, Tomaszewski M, Wainford RD, Williams B, Schutte AE. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. *Hypertension*. 2020;75:1334–1357.

5. Niiranen TJ, Suvila K, Cheng S. Early-Onset Hypertension: Under-Recognized, Under-Treated, and Under-Estimated in Risk. *J Am Coll Cardiol*. 2020;75:2931–2933.

6. Niiranen TJ, McCabe EL, Larson MG, Henglin M, Lakdawala NK, Vasan RS, Cheng S. Heritability and risks associated with early onset hypertension: multigenerational, prospective analysis in the Framingham Heart Study. *BMJ*. 2017;357:j1949.

7. Wager S, Athey S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*. 2018;113:1228–1242.

8. Künzel SR, Sekhon JS, Bickel PJ, Yu B. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*. 2019;116:4156–4165.

9. Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centred evidence. *Int J Epidemiol*. 2016;45:2184–2193.

10. Imai K, Li ML. Experimental Evaluation of Individualized Treatment Rules. *Journal of the American Statistical Association*. 2021;1–15.

11. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. *The Annals of Statistics*. 2011;39:1180–1210.

12. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, McManus K, Champagne CM, Bishop LM, Laranjo N, Leboff MS, Rood JC, de Jonge L, Greenway FL, Loria CM, Obarzanek E, Williamson DA. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med*. 2009;360:859–873.

13. Guertin KA, Moore SC, Sampson JN, Huang W-Y, Xiao Q, Stolzenberg-Solomon RZ, Sinha R, Cross AJ. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr*. 2014;100:208–217.

14. Ussher JR, Elmariah S, Gerszten RE, Dyck JRB. The Emerging Role of Metabolomics in the Diagnosis and Prognosis of Cardiovascular Disease. *J Am Coll Cardiol*. 2016;68:2850–2870.

15. Tobias DK, Mora S, Verma S, Lawler PR. Altered branched chain amino acid metabolism: toward a unifying cardiometabolic hypothesis. *Current opinion in cardiology*. 2018;33:558–564.

16. Tobias DK, Lawler PR, Harada PH, Demler OV, Ridker PM, Manson JE, Cheng S, Mora S. Circulating Branched-Chain Amino Acids and Incident Cardiovascular Disease in a Prospective Cohort of US Women. *Circ Genom Precis Med*. 2018;11:e002157.

17. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition. 2nd ed. New York: Chapman and Hall/CRC; 2006.

18. Al-Shaar L, Yuan C, Rosner B, Dean SB, Ivey KL, Clowry CM, Sampson LA, Barnett JB, Rood J, Harnack LJ, Block J, Manson JE, Stampfer MJ, Willett WC, Rimm EB. Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire in Men Assessed by Multiple Methods. *Am J Epidemiol*. 2021;190:1122–1132.

19. Yue Y, Yuan C, Wang DD, Wang M, Song M, Shan Z, Hu F, Rosner B, Smith-Warner SA, Willett WC. Reproducibility and validity of diet quality scores derived from food-frequency questionnaires. *Am J Clin Nutr*. 2022;115:843–853.

20. Cook NR, Kumanyika SK, Cutler JA. Effect of change in sodium excretion on change in blood pressure corrected for measurement error. The Trials of Hypertension Prevention, Phase I. *Am J Epidemiol*. 1998;148:431–444.

21. Pittman A, Stuart EA, Siddique J. Characterizing Measurement Error in Dietary Sodium in Longitudinal Intervention Studies. *Front Nutr*. 2020;7:581439.

22. Neal B, Wu Y, Feng X, Zhang R, Zhang Y, Shi J, Zhang J, Tian M, Huang L, Li Z, Yu Y, Zhao Y, Zhou B, Sun J, Liu Y, Yin X, Hao Z, Yu J, Li K-C, Zhang X, Duan P, Wang F, Ma B, Shi W, Di Tanna GL, Stepien S, Shan S, Pearson S-A, Li N, Yan LL, Labarthe D, Elliott P. Effect of Salt Substitution on Cardiovascular Events and Death. *N Engl J Med*. 2021;385:1067–1077.

23. He FJ, MacGregor GA. Role of salt intake in prevention of cardiovascular disease: controversies and challenges. *Nat Rev Cardiol*. 2018;15:371–377.

24. Muntner P, Hardy ST, Fine LJ, Jaeger BC, Wozniak G, Levitan EB, Colantonio LD. Trends in Blood Pressure Control Among US Adults With Hypertension, 1999-2000 to 2017-2018. *JAMA*. 2020;324:1190–1200.

25. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, Chen J, He J. Global Disparities of Hypertension Prevalence and Control: A Systematic Analysis of Population-Based Studies From 90 Countries. *Circulation*. 2016;134:441–450.

26. Ma Y, He FJ, Sun Q, Yuan C, Kieneker LM, Curhan GC, MacGregor GA, Bakker SJL, Campbell NRC, Wang M, Rimm EB, Manson JE, Willett WC, Hofman A, Gansevoort RT, Cook NR, Hu FB. 24-Hour Urinary Sodium and Potassium Excretion and Cardiovascular Risk. *N Engl J Med*. 2022;386:252–263.

27. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*. 1989;8:1051–1069; discussion 1071-1073.

28. Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, Chavarro JE. Origin, Methods, and Evolution of the Three Nurses' Health Studies. *Am J Public Health*. 2016;106:1573–1581.

29. Cook NR, Cutler JA, Obarzanek E, Buring JE, Rexrode KM, Kumanyika SK, Appel LJ, Whelton PK. Long term effects of dietary sodium reduction on cardiovascular disease outcomes: observational follow-up of the trials of hypertension prevention (TOHP). *BMJ*. 2007;334:885–888.

30. Sun Q, Bertrand KA, Franke AA, Rosner B, Curhan GC, Willett WC. Reproducibility of urinary biomarkers in multiple 24-h urine samples. *Am J Clin Nutr*. 2017;105:159–168.

31. Yuan C, Spiegelman D, Rimm EB, Rosner BA, Stampfer MJ, Barnett JB, Chavarro JE, Rood JC, Harnack LJ, Sampson LK, Willett WC. Relative Validity of Nutrient Intakes Assessed by Questionnaire, 24-Hour Recalls, and Diet Records as Compared With Urinary Recovery and Plasma Concentration Biomarkers: Findings for Women. *Am J Epidemiol*. 2018;187:1051–1063.

32. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45:1–67.

33. Jacob D. Cross-Fitting and Averaging for Machine Learning Estimation of Heterogeneous Treatment Effects [Internet]. 2020 [cited 2022 Sep 29];Available from: http://arxiv.org/abs/2007.02852

34. Willett W. Implications of Total Energy Intake for Epidemiologic Analyses [Internet]. In: Willett W, editor. Nutritional Epidemiology. Oxford University Press; 2012 [cited 2023 Feb 6]. p. 0.Available from: https://doi.org/10.1093/acprof:oso/9780199754038.003.0011

35. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework [Internet]. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: Association for Computing Machinery; 2019 [cited 2023 Feb 6]. p. 2623–2631.Available from: https://doi.org/10.1145/3292500.3330701

36. Forman JP, Curhan GC, Taylor EN. Plasma 25-hydroxyvitamin D levels and risk of incident hypertension among young women. *Hypertension*. 2008;52:828–832.

37. Ascherio A, Rimm EB, Giovannucci EL, Colditz GA, Rosner B, Willett WC, Sacks F, Stampfer MJ. A prospective study of nutritional factors and hypertension among US men. *Circulation*. 1992;86:1475–1484.

38. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, Hennekens CH, Speizer FE. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. *Am J Epidemiol*. 1986;123:894–900.

39. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? *Journal of Clinical Epidemiology*. 2020;118:93–99.

40. Li F, Chen L, Liu B, Zhong VW, Deng Y, Luo D, Gao C, Bao W, Rong S. Frequency of adding salt at the table and risk of incident cardiovascular disease and all-cause mortality: a prospective cohort study. *BMC Medicine*. 2022;20:486.

41. Ma H, Wang X, Li X, Heianza Y, Qi L. Adding Salt to Foods and Risk of Cardiovascular Disease. *J Am Coll Cardiol*. 2022;80:2157–2167.

42. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Food and Nutrition Board; Committee to Review the Dietary Reference Intakes for Sodium and Potassium. Dietary Reference Intakes for Sodium and Potassium [Internet]. Washington (DC): National Academies Press (US); 2019 [cited 2023 Jun 6]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK538102/

43. Mill JG, Baldo MP, Molina M del CB, Schmidt MI, Barreto SM, Chor D, Griep RH, Matos SM, Ribeiro ALP, Duncan BB, Aquino EM, Lotufo PA, Bensenor I. Sex-specific patterns in the association between salt intake and blood pressure: The ELSA-Brasil study. *The Journal of Clinical Hypertension*. 2019;21:502–509.

44. Danon-Hersch N, Chiolero A, Shamlaye C, Paccaud F, Bovet P. Decreasing Association Between Body Mass Index and Blood Pressure Over Time. *Epidemiology*. 2007;18:493.

45. Chiriacò M, Tricò D, Leonetti S, Petrie JR, Balkau B, Højlund K, Pataky Z, Nilsson PM, Natali A, Heine R j., Dekker J, de Rooij S, Nijpels G, Boorsma W, Mitrakou A, Tournis S, Kyriakopoulou K, Thomakos P, Lalic N, Lalic K, Jotic A, Lukic L, Civcic M, Nolan J, Yeow T p., Murphy M, DeLong C, Neary G, Colgan M p., Hatunic M, Konrad T, Böhles H, Fuellert S, Baer F, Zuchhold H, Golay A, Harsch Bobbioni E, Barthassat V, Makoundou V, Lehmann T n. o., Merminod T, Perry C, Neary F, MacDougall C, Shields K, Malcolm L, Laakso M, Salmenniemi U, Aura A, Raisanen R, Ruotsalainen U, Sistonen T, Laitinen M, Saloranta H, Coppack S w., McIntosh N, Ross J, Pettersson L, Khadobaksh P, Laville M, Bonnet F, Brac de la Perriere A, Louche-Pelissier C, Maitrepierre C, Peyrat J, Beltran S, Serusclat A, Gabriel R, Sánchez E m., Carraro R, Friera A, Novella B, Nilsson P, Persson M, Östling G, Melander O, Burri P, Piatti P m., Monti L d., Setola E, Galluccio E, Minicucci F, Colleluori A, Walker M, Ibrahim I m., Jayapaul M, Carman D, Ryan C, Short K, McGrady Y, Richardson D, Beck-Nielsen H, Staehr P, Vestergaard V, Olsen C, Hansen L, Bolli G b., Porcellati F, et al. Female Sex and Angiotensin-Converting Enzyme (ACE) Insertion/Deletion Polymorphism Amplify the Effects of Adiposity on Blood Pressure. *Hypertension*. 2022;79:36–46.

46. Gail MH, Wu J, Wang M, Yaun S-S, Cook NR, Eliassen AH, McCullough ML, Yu K, Zeleniuch-Jacquotte A, Smith-Warner SA, Ziegler RG, Carroll RJ. Calibration and seasonal adjustment for matched case–control studies of vitamin D and cancer. *Statistics in Medicine*. 2016;35:2133–2148.

47. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *Lancet*. 2021;398:957–980.

48. Muntner P, Carey RM, Gidding S, Jones DW, Taler SJ, Wright JT, Whelton PK. Potential US Population Impact of the 2017 ACC/AHA High Blood Pressure Guideline. *Circulation*. 2018;137:109–118.

49. Clark D, Colantonio LD, Min Y-I, Hall ME, Zhao H, Mentz RJ, Shimbo D, Ogedegbe G, Howard G, Levitan EB, Jones DW, Correa A, Muntner P. Population-Attributable Risk for Cardiovascular Disease Associated With Hypertension in Black Adults. *JAMA Cardiol*. 2019;4:1194–1202.

50. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, Alexander L, Estep K, Hassen Abate K, Akinyemiju TF, Ali R, Alvis-Guzman N, Azzopardi P, Banerjee A, Bärnighausen T, Basu A, Bekele T, Bennett DA, Biadgilign S, Catalá-López F, Feigin VL, Fernandes JC, Fischer F, Gebru AA, Gona P, Gupta R, Hankey GJ, Jonas JB, Judd SE, Khang Y-H, Khosravi A, Kim YJ, Kimokoti RW, Kokubo Y, Kolte D, Lopez A, Lotufo PA, Malekzadeh R, Melaku YA, Mensah GA, Misganaw A, Mokdad AH, Moran AE, Nawaz H, Neal B, Ngalesoni FN, Ohkubo T, Pourmalek F, Rafay A, Rai RK, Rojas-Rueda D, Sampson UK, Santos IS, Sawhney M, Schutte AE, Sepanlou SG, Shifa GT, Shiue I, Tedla BA, Thrift AG, Tonelli M, Truelsen T, Tsilimparis N, Ukwaja KN, Uthman OA,

Vasankari T, Venketasubramanian N, Vlassov VV, Vos T, Westerman R, Yan LL, Yano Y, Yonemoto N, Zaki MES, Murray CJL. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *JAMA*. 2017;317:165–182.

51. Valenzuela PL, Carrera-Bastos P, Gálvez BG, Ruiz-Hurtado G, Ordovas JM, Ruilope LM, Lucia A. Lifestyle interventions for the prevention and treatment of hypertension. *Nat Rev Cardiol*. 2021;18:251–275.

52. Neter JE, Stam BE, Kok FJ, Grobbee DE, Geleijnse JM. Influence of weight reduction on blood pressure: a meta-analysis of randomized controlled trials. *Hypertension*. 2003;42:878–84.

53. Graudal NA, Hubeck-Graudal T, Jurgens G. Effects of low sodium diet versus high sodium diet on blood pressure, renin, aldosterone, catecholamines, cholesterol, and triglyceride. *Cochrane Database Syst Rev*. 2020;12:CD004022.

54. Filippou CD, Tsioufis CP, Thomopoulos CG, Mihas CC, Dimitriadis KS, Sotiropoulou LI, Chrysochoou CA, Nihoyannopoulos PI, Tousoulis DM. Dietary Approaches to Stop Hypertension (DASH) Diet and Blood Pressure Reduction in Adults with and without Hypertension: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Adv Nutr*. 2020;11:1150–1160.

55. Roerecke M, Kaczorowski J, Tobe SW, Gmel G, Hasan OSM, Rehm J. The effect of a reduction in alcohol consumption on blood pressure: a systematic review and meta-analysis. *Lancet Public Health*. 2017;2:e108–e120.

56. Whelton SP, Chin A, Xin X, He J. Effect of aerobic exercise on blood pressure: a meta-analysis of randomized, controlled trials. *Ann Intern Med*. 2002;136:493–503.

57. Forman JP, Stampfer MJ, Curhan GC. Diet and lifestyle risk factors associated with incident hypertension in women. *JAMA*. 2009;302:401–411.

58. Wang C, Yuan Y, Zheng M, Pan A, Wang M, Zhao M, Li Y, Yao S, Chen S, Wu S, Xue H. Association of Age of Onset of Hypertension With Cardiovascular Diseases and Mortality. *J Am Coll Cardiol*. 2020;75:2921–2930.

59. Colafella KMM, Denton KM. Sex-specific differences in hypertension and associated cardiovascular disease. *Nat Rev Nephrol*. 2018;14:185–201.

60. Suvila K, Langén V, Cheng S, Niiranen TJ. Age of Hypertension Onset: Overview of Research and How to Apply in Practice. *Curr Hypertens Rep*. 2020;22:68.

61. Zucker DM. Restricted Mean Life with Covariates: Modification and Extension of a Useful Survival Analysis Method. *Journal of the American Statistical Association*. 1998;93:702–709.

62. Al-Shaar L, Pernar CH, Chomistek AK, Rimm EB, Rood J, Stampfer MJ, Eliassen AH, Barnett JB, Willett WC. Reproducibility, Validity, and Relative Validity of Self-Report

Methods for Assessing Physical Activity in Epidemiologic Studies: Findings From the Women's Lifestyle Validation Study. *Am J Epidemiol*. 2022;191:696–710.

63. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch Intern Med*. 2008;168:713–20.

64. Tobias DK, Zhang C, Chavarro J, Bowers K, Rich-Edwards J, Rosner B, Mozaffarian D, Hu FB. Prepregnancy adherence to dietary patterns and lower risk of gestational diabetes mellitus. *Am J Clin Nutr*. 2012;96:289–95.

65. Niiranen TJ, Larson MG, McCabe EL, Xanthakis V, Vasan RS, Cheng S. Prognosis of Prehypertension Without Progression to Hypertension. *Circulation*. 2017;136:1262–1264.

66. Niiranen TJ, McCabe EL, Larson MG, Henglin M, Lakdawala NK, Vasan RS, Cheng S. Risk for hypertension crosses generations in the community: a multi-generational cohort study. *Eur Heart J*. 2017;38:2300–2308.

67. Wacholder S, Benichou J, Heineman EF, Hartge P, Hoover RN. Attributable risk: advantages of a broad definition of exposure. *Am J Epidemiol*. 1994;140:303–9.

68. McCaw ZR, Yin G, Wei L-J. Using the Restricted Mean Survival Time Difference as an Alternative to the Hazard Ratio for Analyzing Clinical Cardiovascular Studies. *Circulation*. 2019;140:1366–1368.

69. Forman JP, Stampfer MJ, Curhan GC. Non-narcotic analgesic dose and risk of incident hypertension in US women. *Hypertension*. 2005;46:500–7.

70. Yano Y, Reis JP, Colangelo LA, Shimbo D, Viera AJ, Allen NB, Gidding SS, Bress AP, Greenland P, Muntner P, Lloyd-Jones DM. Association of Blood Pressure Classification in Young Adults Using the 2017 American College of Cardiology/American Heart Association Blood Pressure Guideline With Cardiovascular Events Later in Life. *Jama*. 2018;320:1774–1782.

71. Shah AM, Claggett B, Folsom AR, Lutsey PL, Ballantyne CM, Heiss G, Solomon SD. Ideal Cardiovascular Health During Adult Life and Cardiovascular Structure and Function Among the Elderly. *Circulation*. 2015;132:1979–89.

72. Wilding JPH, Batterham RL, Calanna S, Davies M, Van Gaal LF, Lingvay I, McGowan BM, Rosenstock J, Tran MTD, Wadden TA, Wharton S, Yokote K, Zeuthen N, Kushner RF, STEP 1 Study Group. Once-Weekly Semaglutide in Adults with Overweight or Obesity. *N Engl J Med*. 2021;384:989–1002.

73. Mukamal KJ, Conigrave KM, Mittleman MA, Camargo CA, Stampfer MJ, Willett WC, Rimm EB. Roles of drinking pattern and type of alcohol consumed in coronary heart disease in men. *N Engl J Med*. 2003;348:109–118.

74. He FJ, Li J, MacGregor GA. Effect of longer term modest salt reduction on blood pressure: Cochrane systematic review and meta-analysis of randomised trials. BMJ : British Medical Journal. 2013;346:f1325.

75. He J, Huang JF, Li C, Chen J, Lu X, Chen JC, He H, Li JX, Cao J, Chen CS, Bazzano LA, Hu D, Kelly TN, Gu DF. Sodium Sensitivity, Sodium Resistance, and Incidence of Hypertension: A Longitudinal Follow-Up Study of Dietary Sodium Intervention. *Hypertension*. 2021;78:155–164.

76. Razavi MA, Bazzano LA, Nierenberg J, Huang Z, Fernandez C, Razavi AC, Whelton SP, He J, Kelly TN. Advances in Genomics Research of Blood Pressure Responses to Dietary Sodium and Potassium Intakes. *Hypertension*. 2021;78:4–15.

77. Al-Shaar L, Pernar CH, Chomistek AK, Rimm EB, Rood J, Stampfer MJ, Eliassen AH, Barnett JB, Willett WC. Reproducibility, Validity, and Relative Validity of Self-Report Methods for Assessing Physical Activity in Epidemiologic Studies: Findings from the Women's Lifestyle Validation Study. *Am J Epidemiol*. 2022;

78. Grobbee DE, Rimm EB, Giovannucci E, Colditz G, Stampfer M, Willett W. Coffee, caffeine, and cardiovascular disease in men. *N Engl J Med*. 1990;323:1026–1032.

79. Giovannucci E, Colditz G, Stampfer MJ, Rimm EB, Litin L, Sampson L, Willett WC. The assessment of alcohol consumption by a simple self-administered questionnaire. *Am J Epidemiol*. 1991;133:810–817.

80. Wong BHW, Lee J, Spiegelman D, Wang M. Estimation and inference for the population attributable risk in the presence of misclassification. *Biostatistics*. 2021;22:805–818.

81. Kent DM, Paulus JK, van Klaveren D, D'Agostino R, Goodman S, Hayward R, Ioannidis JPA, Patrick-Lake B, Morton S, Pencina M, Raman G, Ross JS, Selker HP, Varadhan R, Vickers A, Wong JB, Steyerberg EW. The Predictive Approaches to Treatment effect Heterogeneity (PATH) Statement. *Ann Intern Med*. 2020;172:35–45.

82. Imai K, Ratkovic M. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*. 2013;7:443–470.

83. Abrevaya J, Hsu Y-C, Lieli RP. Estimating Conditional Average Treatment Effects. *Journal of Business & Economic Statistics*. 2015;33:485–505.

84. Hahn PR, Murray JS, Carvalho C. Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects [Internet]. 2019 [cited 2022 Nov 7];Available from: http://arxiv.org/abs/1706.09523

85. Imai K, Li ML. Statistical Inference for Heterogeneous Treatment Effects Discovered by Generic Machine Learning in Randomized Experiments [Internet]. 2022 [cited 2023 Mar 23];Available from: http://arxiv.org/abs/2203.14511

86. Hernan MA, Robins JM. Causal Inference: What If. *Boca Raton: Chapman & Hall/CRC*. 2020;

87. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology*. 2017;28:54–59.

88. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9:48–55.

89. Vegetabile BG. On the Distinction Between "Conditional Average Treatment Effects" (CATE) and "Individual Treatment Effects" (ITE) Under Ignorability Assumptions [Internet]. 2021 [cited 2023 Apr 20];Available from: http://arxiv.org/abs/2108.04939

90. Fan Q, Hsu Y-C, Lieli RP, Zhang Y. Estimation of Conditional Average Treatment Effects with High-Dimensional Data [Internet]. 2021 [cited 2023 Apr 20];Available from: http://arxiv.org/abs/1908.02399

91. Logan BR, Sparapani R, McCulloch RE, Laud PW. Decision making and uncertainty quantification for individualized treatments using Bayesian Additive Regression Trees. *Stat Methods Med Res*. 2019;28:1079–1093.

92. Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions [Internet]. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.; 2017 [cited 2023 Aug 15]. Available from: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

93. Verstraete K, Gyselinck I, Huts H, Das N, Topalovic M, Vos MD, Janssens W. Estimating individual treatment effects on COPD exacerbations by causal machine learning on randomised controlled trials. *Thorax* [Internet]. 2023 [cited 2023 Jul 11];Available from: https://thorax.bmj.com/content/early/2023/04/02/thorax-2022-219382

94. Rubin DB. Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association*. 2005;100:322–331.

95. Chernozhukov V, Demirer M, Duflo E, Fern'andez-Val I. Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments [Internet]. arXiv.org; 2020 [cited 2022 Jan 17]. Available from: https://ideas.repec.org/p/arx/papers/1712.04802.html

96. Nie X, Wager S. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*. 2021;108:299–319.

97. Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects. *Proc Natl Acad Sci U S A*. 2016;113:7353–7360.

98. Green DP, Kern HL. Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees. *Public Opinion Quarterly*. 2012;76:491–511.

99. Hill JL. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*. 2011;20:217–240.

100. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*. 2018;21:C1–C68.

101. Kennedy EH. Towards optimal doubly robust estimation of heterogeneous causal effects [Internet]. 2022 [cited 2023 May 23];Available from: http://arxiv.org/abs/2004.14497

102. Foster DJ, Syrgkanis V. Orthogonal Statistical Learning [Internet]. 2023 [cited 2023 Jul 10];Available from: http://arxiv.org/abs/1901.09036

103. Hirano K, Imbens G, Ridder G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica*. 2003;71:1161–1189.

104. Robins JM, Mark SD, Newey WK. Estimating Exposure Effects by Modelling the Expectation of Exposure Conditional on Confounders. *Biometrics*. 1992;48:479–495.

105. Athey S, Wager S. Estimating Treatment Effects with Causal Forests: An Application [Internet]. 2019 [cited 2022 Nov 7];Available from: http://arxiv.org/abs/1902.07409

106. Athey S, Wager S. Policy Learning With Observational Data. *Econometrica*. 2021;89:133–161.

107. Yadlowsky S, Fleming S, Shah N, Brunskill E, Wager S. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects [Internet]. 2023 [cited 2023 Dec 1];Available from: http://arxiv.org/abs/2111.07966

108. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *J Am Stat Assoc*. 2012;107:1106–1118.

109. Kapoor S, Narayanan A. Leakage and the Reproducibility Crisis in ML-based Science [Internet]. 2022 [cited 2023 Apr 20];Available from: http://arxiv.org/abs/2207.07048

110. VanderWeele TJ, Robins JM. Four types of effect modification: a classification based on directed acyclic graphs. *Epidemiology*. 2007;18:561–568.

111. Huang T, Huang J, Qi Q, Li Y, Bray GA, Rood J, Sacks FM, Qi L. PCSK7 genotype modifies effect of a weight-loss diet on 2-year changes of insulin resistance: the POUNDS LOST trial. *Diabetes Care*. 2015;38:439–444.

112. Qi Q, Bray GA, Smith SR, Hu FB, Sacks FM, Qi L. Insulin receptor substrate 1 gene variation modifies insulin resistance response to weight-loss diets in a 2-year randomized trial: the Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) trial. *Circulation*. 2011;124:563–571.

113. Xu M, Qi Q, Liang J, Bray GA, Hu FB, Sacks FM, Qi L. Genetic determinant for amino acid metabolites and changes in body weight and insulin resistance in response to weight-loss diets: the Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) trial. *Circulation*. 2013;127:1283–1289.

114. Chen Y, Zhou T, Sun D, Li X, Ma H, Liang Z, Heianza Y, Pei X, Bray GA, Sacks FM, Qi L. Distinct genetic subtypes of adiposity and glycemic changes in response to weight-loss diet intervention: the POUNDS Lost trial. *Eur J Nutr*. 2021;60:249–258.

115. Rodgers GP, Collins FS. Precision Nutrition—the Answer to "What to Eat to Stay Healthy." *JAMA*. 2020;324:735–736.

116. Arcones AC, Cruces-Sande M, Ramos P, Mayor F, Murga C. Sex Differences in High Fat Diet-Induced Metabolic Alterations Correlate with Changes in the Modulation of GRK2 Levels. *Cells*. 2019;8:1464.

117. Guerra-Cantera S, Frago LM, Collado-Pérez R, Canelles S, Ros P, Freire-Regatillo A, Jiménez-Hernaiz M, Barrios V, Argente J, Chowen JA. Sex Differences in Metabolic Recuperation After Weight Loss in High Fat Diet-Induced Obese Mice. *Front Endocrinol (Lausanne)*. 2021;12:796661.

118. Zhu R, Craciun I, Bernhards-Werge J, Jalo E, Poppitt SD, Silvestre MP, Huttunen-Lenz M, McNarry MA, Stratton G, Handjiev S, Handjieva-Darlenska T, Navas-Carretero S, Sundvall J, Adam TC, Drummen M, Simpson EJ, Macdonald IA, Brand-Miller J, Muirhead R, Lam T, Vestentoft PS, Færch K, Martinez JA, Fogelholm M, Raben A. Age- and sex-specific effects of a long-term lifestyle intervention on body weight and cardiometabolic health markers in adults with prediabetes: results from the diabetes prevention study PREVIEW. *Diabetologia*. 2022;65:1262–1277.

119. Beavers KM, Neiberg RH, Kritchevsky SB, Nicklas BJ, Kitzman DW, Messier SP, Rejeski WJ, Ard JD, Beavers DP. Association of Sex or Race With the Effect of Weight Loss on Physical Function: A Secondary Analysis of 8 Randomized Clinical Trials. *JAMA Netw Open*. 2020;3:e2014631.

120. Brosnan JT, Brosnan ME. Branched-chain amino acids: enzyme and substrate regulation. *The Journal of nutrition*. 2006;136:207s–11s.

121. Burrage LC, Nagamani SC, Campeau PM, Lee BH. Branched-chain amino acid metabolism: from rare Mendelian diseases to more common disorders. *Hum Mol Genet*. 2014;23:R1-8.

122. Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, Yuan C, Bao Y, Townsend MK, Tworoger SS, Davidson SM, Papagiannakopoulos T, Yang A, Dayton TL, Ogino S, Stampfer MJ, Giovannucci EL, Qian ZR, Rubinson DA, Ma J, Sesso HD, Gaziano JM, Cochrane BB, Liu S, Wactawski-Wende J, Manson JE, Pollak MN, Kimmelman AC, Souza A, Pierce K, Wang TJ, Gerszten RE, Fuchs CS, Vander Heiden MG, Wolpin BM. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med*. 2014;20:1193–1198.

123. Liu R, Hong J, Xu X, Feng Q, Zhang D, Gu Y, Shi J, Zhao S, Liu W, Wang X, Xia H, Liu Z, Cui B, Liang P, Xi L, Jin J, Ying X, Wang X, Zhao X, Li W, Jia H, Lan Z, Li F, Wang R, Sun Y, Yang M, Shen Y, Jie Z, Li J, Chen X, Zhong H, Xie H, Zhang Y, Gu W, Deng X, Shen B, Xu X, Yang H, Xu G, Bi Y, Lai S, Wang J, Qi L, Madsen L, Wang J, Ning G, Kristiansen K, Wang W. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat Med*. 2017;23:859–868.

124. Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J, Hu FB. Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-analysis. *Diabetes Care*. 2016;39:833–846.

125. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS, Jacques PF, Fernandez C, O'Donnell CJ, Carr SA, Mootha VK, Florez JC, Souza A, Melander O, Clish CB, Gerszten RE. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17:448–453.

126. Tobias DK, Clish C, Mora S, Li J, Liang L, Hu FB, Manson JE, Zhang C. Dietary Intakes and Circulating Concentrations of Branched-Chain Amino Acids in Relation to Incident Type 2 Diabetes Risk Among High-Risk Women with a History of Gestational Diabetes Mellitus. *Clin Chem*. 2018;64:1203–1210.

127. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, Schmidt AF, Imamura F, Stewart ID, Perry JRB, Marney L, Koulman A, Karoly ED, Forouhi NG, Sjögren RJO, Näslund E, Zierath JR, Krook A, Savage DB, Griffin JL, Chaturvedi N, Hingorani AD, Khaw K-T, Barroso I, McCarthy MI, O'Rahilly S, Wareham NJ, Langenberg C. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS Med*. 2016;13:e1002179.

128. Xuan L, Hou Y, Wang T, Li M, Zhao Z, Lu J, Xu Y, Chen Y, Qi L, Wang W, Bi Y, Xu M. Association of branched chain amino acids related variant rs1440581 with risk of incident diabetes and longitudinal changes in insulin resistance in Chinese. *Acta Diabetol*. 2018;55:901–908.

129. Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol*. 2011;11:98–107.

130. Jain A, Gupta HL, Narayan S. Hyperfibrinogenemia in patients of diabetes mellitus in relation to glycemic control and urinary albumin excretion rate. *The Journal of the Association of Physicians of India*. 2001;49:227–30.

131. Liu C, Feng X, Li Q, Wang Y, Li Q, Hua M. Adiponectin, TNF-α and inflammatory cytokines and risk of type 2 diabetes: A systematic review and meta-analysis. *Cytokine*. 2016;86:100–109.

132. Akinkuolie AO, Pradhan AD, Buring JE, Ridker PM, Mora S. Novel protein glycan side-chain biomarker and risk of incident type 2 diabetes mellitus. *Arteriosclerosis, thrombosis, and vascular biology*. 2015;35:1544–50.

133. Blüher M, Unger R, Rassoul F, Richter V, Paschke R. Relation between glycaemic control, hyperinsulinaemia and plasma concentrations of soluble adhesion molecules in patients with impaired glucose tolerance or Type II diabetes. *Diabetologia*. 2002;45:210–6.

134. McLaughlin T, Abbasi F, Cheal K, Chu J, Lamendola C, Reaven G. Use of metabolic markers to identify overweight individuals who are insulin resistant. *Annals of internal medicine*. 2003;139:802–9.

135. Sparks JD, Sparks CE, Adeli K. Selective hepatic insulin resistance, VLDL overproduction, and hypertriglyceridemia. *Arteriosclerosis, thrombosis, and vascular biology*. 2012;32:2104–12.

136. Shalaurova I, Connelly MA, Garvey WT, Otvos JD. Lipoprotein insulin resistance index: a lipoprotein particle-derived measure of insulin resistance. *Metab Syndr Relat Disord*. 2014;12:422–429.

137. Mackey RH, Mora S, Bertoni AG, Wassel CL, Carnethon MR, Sibley CT, Goff DC. Lipoprotein particles and incident type 2 diabetes in the multi-ethnic study of atherosclerosis. *Diabetes Care*. 2015;38:628–36.

138. Harada PHN, Demler OV, Dugani SB, Akinkuolie AO, Moorthy MV, Ridker PM, Cook NR, Pradhan AD, Mora S. Lipoprotein insulin resistance score and risk of incident diabetes during extended follow-up of 20 years: The Women's Health Study. *Journal of clinical lipidology*. 2017;11:1257-1267.e2.

139. Zheng Y, Ceglarek U, Huang T, Li L, Rood J, Ryan DH, Bray GA, Sacks FM, Schwarzfuchs D, Thiery J, Shai I, Qi L. Weight-loss diets and 2-y changes in circulating amino acids in 2 randomized intervention trials. *The American journal of clinical nutrition*. 2016;103:505–11.

140. Lips MA, Van Klinken JB, van Harmelen V, Dharuri HK, t Hoen PA, Laros JF, van Ommen GJ, Janssen IM, Van Ramshorst B, Van Wagensveld BA, Swank DJ, Van Dielen F, Dane A, Harms A, Vreeken R, Hankemeier T, Smit JW, Pijl H, Willems van Dijk K. Roux-en-Y gastric bypass surgery, but not calorie restriction, reduces plasma branched-chain amino acids in obese women independent of weight loss or the presence of type 2 diabetes. *Diabetes Care*. 2014;37:3150–6.

141. Hanvold SE, Vinknes KJ, Bastani NE, Turner C, Løken EB, Mala T, Refsum H, Aas AM. Plasma amino acids, adiposity, and weight change after gastric bypass surgery: are amino acids associated with weight regain? *European journal of nutrition*. 2018;57:2629–2637.

142. Ridker PM, Cook NR, Lee I-M, Gordon D, Gaziano JM, Manson JE, Hennekens CH, Buring JE. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *N Engl J Med*. 2005;352:1293–1304.

143. Tobias DK, Mora S, Verma S, Billia F, Buring JE, Lawler PR. Fasting status and metabolic health in relation to plasma branched chain amino acid concentrations in women. *Metabolism: clinical and experimental*. 2020;117:154391.

144. Wolak-Dinsmore J, Gruppen EG, Shalaurova I, Matyus SP, Grant RP, Gegen R, Bakker SJL, Otvos JD, Connelly MA, Dullaart RPF. A novel NMR-based assay to measure circulating concentrations of branched-chain amino acids: Elevation in subjects with type 2 diabetes mellitus and association with carotid intima media thickness. *Clinical biochemistry*. 2018;54:92–99.

145. Otvos JD, Shalaurova I, Wolak-Dinsmore J, Connelly MA, Mackey RH, Stein JH, Tracy RP. GlycA: A Composite Nuclear Magnetic Resonance Biomarker of Systemic Inflammation. *Clin Chem*. 2015;61:714–723.

146. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. *N Engl J Med*. 2002;347:1557–1565.

147. Whitton CM, Sands D, Hubbard AR, Gaffney PJ. A collaborative study to establish the 2nd International Standard for Fibrinogen, Plasma. *Thromb Haemost*. 2000;84:258–262.

148. Akinkuolie AO, Buring JE, Ridker PM, Mora S. A novel protein glycan biomarker and future cardiovascular disease events. *J Am Heart Assoc*. 2014;3:e001221–e001221.

149. Albert MA, Glynn RJ, Buring JE, Ridker PM. Differential effect of soluble intercellular adhesion molecule-1 on the progression of atherosclerosis as compared to arterial thrombosis: a prospective analysis of the Women's Health Study. *Atherosclerosis*. 2008;197:297–302.

150. Mora S, Otvos JD, Rifai N, Rosenson RS, Buring JE, Ridker PM. Lipoprotein particle profiles by nuclear magnetic resonance compared with standard lipids and apolipoproteins in predicting incident cardiovascular disease in women. *Circulation*. 2009;119:931–9.

151. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr*. 2012;142:1009–1018.

152. Bogl LH, Kaye SM, Rämö JT, Kangas AJ, Soininen P, Hakkarainen A, Lundbom J, Lundbom N, Ortega-Alonso A, Rissanen A, Ala-Korpela M, Kaprio J, Pietiläinen KH. Abdominal obesity and circulating metabolites: A twin study approach. *Metabolism: clinical and experimental*. 2016;65:111–21.

153. Sun L, Hu C, Yang R, Lv Y, Yuan H, Liang Q, He B, Pang G, Jiang M, Dong J, Yang Z. Association of circulating branched-chain amino acids with cardiometabolic traits differs between adults and the oldest-old. *Oncotarget*. 2017;8:88882–88893.

154. Schmidt MI, Duncan BB, Sharrett AR, Lindberg G, Savage PJ, Offenbacher S, Azambuja MI, Tracy RP, Heiss G. Markers of inflammation and prediction of diabetes mellitus in adults (Atherosclerosis Risk in Communities study): a cohort study. *Lancet*. 1999;353:1649–52.

155. Sandström CS, Ohlsson B, Melander O, Westin U, Mahadeva R, Janciauskiene S. An association between Type 2 diabetes and alpha-antitrypsin deficiency. *Diabetic medicine : a journal of the British Diabetic Association*. 2008;25:1370–3.

156. Fumeron F, Péan F, Driss F, Balkau B, Tichet J, Marre M, Grandchamp B. Ferritin and transferrin are both predictive of the onset of hyperglycemia in men and women over 3 years: the data from an epidemiological study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes Care*. 2006;29:2090–4.

157. Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *Jama*. 2001;286:327–34.

158. Ridker PM. From C-Reactive Protein to Interleukin-6 to Interleukin-1: Moving Upstream To Identify Novel Targets for Atheroprotection. *Circ Res*. 2016;118:145–156.

159. Dugani SB, Moorthy MV, Li C, Demler OV, Alsheikh-Ali AA, Ridker PM, Glynn RJ, Mora S. Association of Lipid, Inflammatory, and Metabolic Biomarkers With Age at Onset for Incident Coronary Heart Disease in Women. *JAMA cardiology*. 2021;

160. Fryar CD, Kruszon-Moran D, Gu Q, Ogden CL. Mean Body Weight, Height, Waist Circumference, and Body Mass Index Among Adults: United States, 1999-2000 Through 2015-2016. *National health statistics reports*. 2018;1–16.

Appendix A. Supplemental data to Chapter 1

Prediction of 24-hour urinary sodium excretion using machine-learning algorithms

Appendix 1: Description of machine-learning (ML) algorithms used in the study

Supplemental Figure 1.1: Comparison of each ML-predicted and FFQ-based sodium intake with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS based on full list of predictors

Supplemental Figure 1.2: Bland-Altman plots of ML-predicted and FFQ-based sodium intake compared with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS

Supplemental Figure 1.3: Comparison of each ML-predicted and FFQ-based sodium intake with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS based on selected list of predictors

Supplemental Figure 1.4: Bland-Altman plots of ML-predicted and FFQ-based sodium intake compared with averaged 24-hour urinary sodium excretion in N=1,423 participants in TOHP-I

Supplemental Figure 1.5: Comparison of each ML-predicted energy and FFQ-based total calorie intake with DLW-based energy expenditure in N=1,085 participants from NHS, NHS-II, and HPFS

Supplemental Figure 1.6: Bland-Altman plots of each ML-predicted energy and FFQ-based total calorie intake with DLW-based energy expenditure in N=1,085 participants from NHS, NHS-II, and HPFS

Supplemental Figure 1.7: Bland-Altman plots of ML-predicted and FFQ-based energy-adjusted sodium measures compared with DLW-based energy expenditure-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants from NHS, NHS-II, and HPFS

Supplemental Table 1.1. Baseline characteristics according to the quartile of the DLW-based energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants whose doubly-labelled water (DLW) information is available

Supplemental Table 1.2. Interpretation of ML models based on selected covariate list predicting averaged 24-hour urinary sodium excretion in N=3,454 participants

Supplemental Table 1.3. Interpretation of ML models predicting DLW-based energy expenditure in N=1,085 participants

Supplemental Table 1.4. Interpretation of ML models predicting DLW-based energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants

Appendix 1. Description of machine-learning (ML) algorithms used in the study

1. Elastic net

Elastic Net is a regularization method employed in linear regression models to avoid overfitting and enhance model generalization. This technique blends two well-known regularization approaches: (L1 regularization (used in Lasso Regression) and L2 regularization (used in Ridge Regression).

L1 Regularization (Lasso):

Lasso (Least Absolute Shrinkage and Selection Operator) is a regularization method that employs L1 norm as the penalty term. It incorporates the absolute values of the coefficients into the cost function (i.e. measure of the performance of the model), making some coefficients to become precisely zero. Consequently, this leads to feature selection, as less significant features are effectively eliminated from the model.

L2 Regularization (Ridge):

Ridge regression is another regularization method that uses L2 norm as the penalty term. It incorporates the squared coefficients into the cost function, shrinking the coefficients but not setting them to exactly zero, while minimizing multicollinearity issues and prevents overfitting.

Elastic Net:

Elastic Net is a hybrid of L1 and L2 regularization methods, merging both Lasso and Ridge penalties. A mixing parameter, α (alpha), is used to balance the contributions of L1 and L2 penalties. The Elastic Net cost function can be expressed as:

$$\text{Cost} = \text{MSE} + \lambda[(1-\alpha)/2 * \Sigma(\beta i^2) + \alpha * \Sigma|\beta i|]$$

Here, λ (lambda) represents the regularization strength, and α is the mixing parameter ranging between 0 and 1. When $\alpha = 0$, Elastic Net is equivalent to Ridge regression, and when $\alpha = 1$, it becomes Lasso regression.

Elastic Net addresses some limitations of Lasso and Ridge regression. By merging both penalties, it can carry out feature selection like Lasso while managing multicollinearity like Ridge. In the hyper-parameter tuning, both λ and α need to be optimized.

2. eXtreme Gradient Boosting (XGBoost)

eXtreme Gradient Boosting (XGBoost) is a tree-based machine learning algorithm based on the gradient boosting framework. XGBoost is very popular in machine learning competitions since its performance is better than numerous other algorithms.

Gradient Boosting Framework is an ensemble technique that combines weak learners (usually shallow decision trees) to form a strong learner. This is achieved by iteratively adding weak learners to the ensemble, with each new learner aiming to correct the errors of the previous ones. The new learners are fitted to the residual errors (gradients) of the earlier learners.

The objective function in XGBoost consists of two components: a loss function and a regularization term. The loss function quantifies the discrepancy between the true labels and the predicted values, while the regularization term, including both both L1 (Lasso) and L2 (Ridge) regularization, regulates the complexity of the model. The objective function can be expressed as:

$$Obj = Loss(y, y\_pred) + \Omega(f)$$

where y and y_pred denote the true and predicted labels, respectively, and $\Omega(f)$ represents the regularization term for the base learners (decision trees).

XGBoost trains the model additively, sequentially incorporating decision trees into the ensemble. During each iteration, the algorithm calculates the gradients and the Hessian (second-order derivatives) of the loss function concerning the predictions. These values are utilized to identify the best split points and leaf values for the new tree. The process is repeated for a specified number of rounds or until a stopping criterion, such as no further improvement in the loss function, is met.

The capabilities to handle large datasets, missing values, and its support for early stopping and cross-validation make XGBoost a versatile and powerful tool for a wide range of machine learning problems.

In the hyper-parameter tuning, the following parameters were optimized:
nrounds: the total number of iterations.
max_depth: the maximum depth of a tree and is used to control over-fitting.
eta: the weighting of new trees added to the model, with a range of 0 (no contribution) to 1 (full contribution).
gamma: a regularization parameter that controls the minimum loss reduction required to make a split.
colsample_bytree: the fraction of the features (columns) to be used in each decision tree or iteration, chosen by random sampling.
min_child_weight: the minimum sum of weights of all observations required in a child.
subsample: the fraction of the total data set that's used for each iteration or tree.

3. Bayesian regularization for feed-forward neural net (brnn)

BRNN is a technique that applies Bayesian learning principles to regularize and improve the generalization performance of neural networks. Feed-Forward Neural Networks are a class of artificial neural networks where the connections between nodes do not form a cycle. These networks consist of an input layer, one or more hidden layers, and an output layer. The information flows from the input layer through the hidden layers to the output layer without any feedback loops.

In the context of neural networks, Bayesian regularization aims to find an optimal set of weights by treating them as random variables and estimating their posterior probability distribution given the training data. The regularization term is derived from the prior distribution of the weights, and it encodes our beliefs about the possible values of the weights before observing the data.

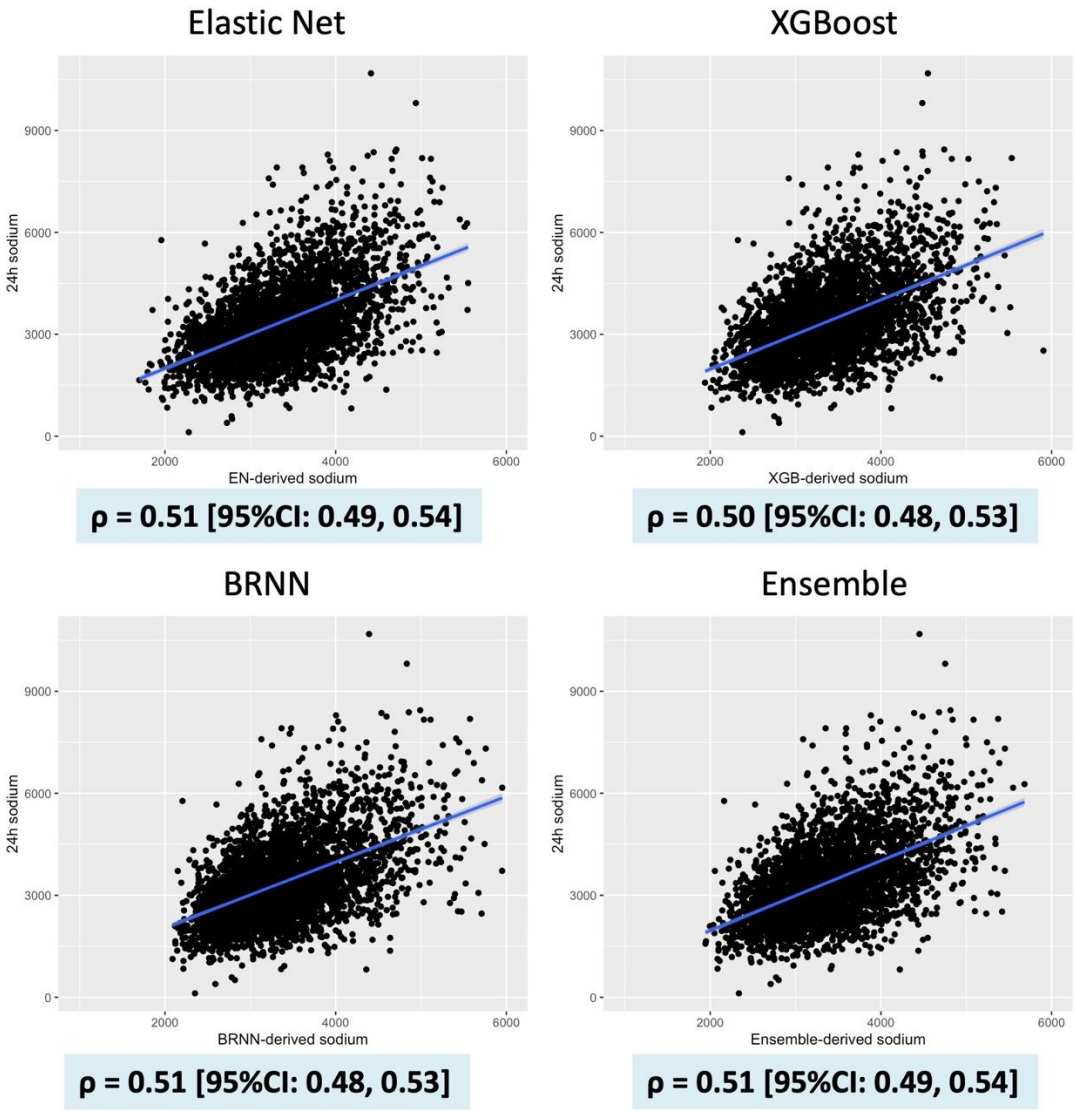
The objective function for Bayesian regularization in neural networks can be written as:

$$Obj = Loss(y, y\_pred) + \lambda * \Omega(w)$$

Here, y and y_pred represent the true and predicted labels, respectively, w is the vector of weights, $\Omega(w)$ is the regularization term, and λ is the regularization strength.

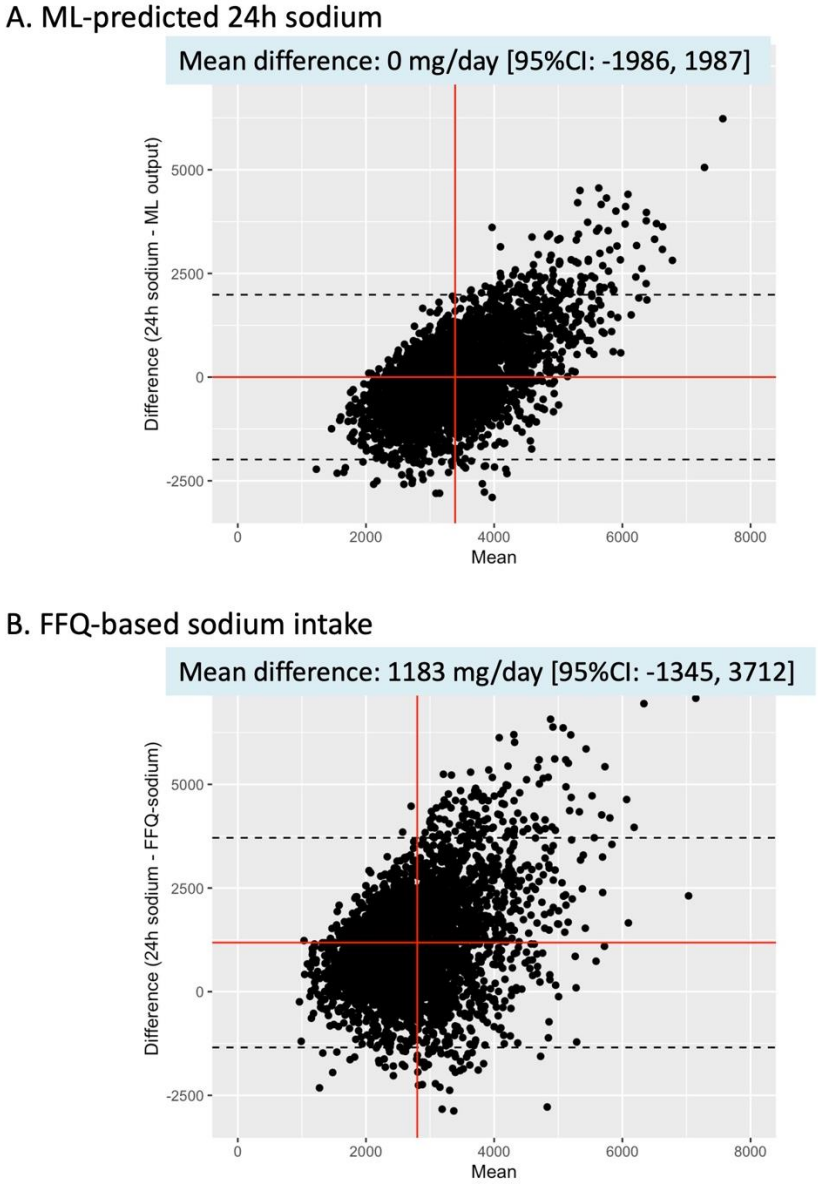
In brnn function in R caret package, only one hyperparameter, neuron, has to be tuned. This design is for usability. The neurons hyperparameter defines the complexity of the neural network. The more neurons there are in the hidden layers, the more complex the patterns the network can learn. Bayesian regularization inherently handles some of the issues that other hyperparameters usually manage in other types of neural networks. For example, the regularization part of Bayesian regularization acts as a form of weight decay, reducing the necessity for an explicit learning rate hyperparameter. Other parameters are just set to default or empirically determined values within the function's implementation.

Supplemental Figure 1.1: Comparison of each ML-predicted and FFQ-based sodium intake with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS based on full list of predictors



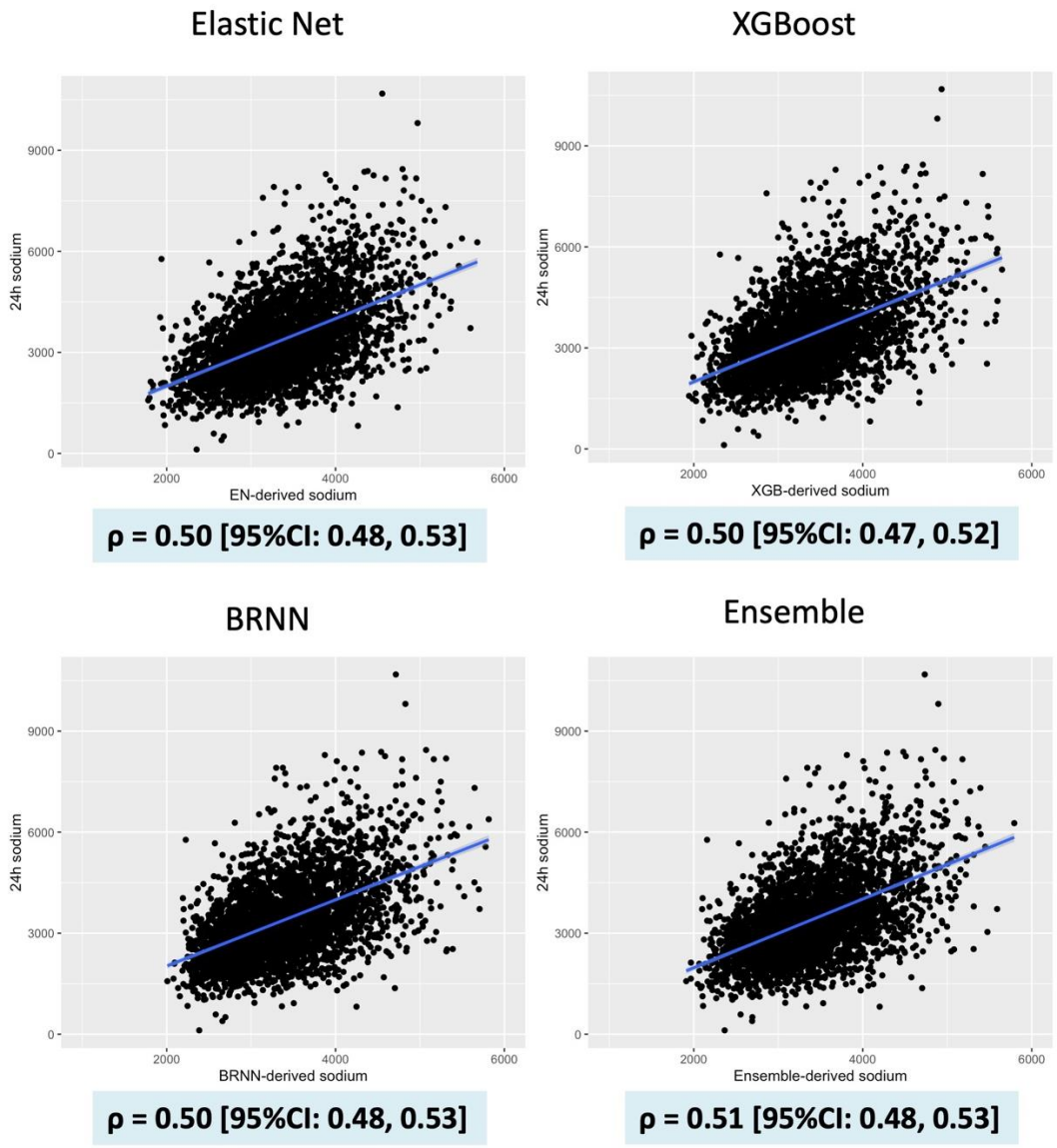
Scatter plots describing the association between each ML-predicted sodium intake with respect to averaged 24-hour urinary sodium excretion in N=3,454 participants. ML-predicted sodium intakes were based on out-of-fold prediction based on 36 predictors using Elastic net, XGBoost, BRNN, and ensemble over these ML algorithms. Each dot represents each participant. Blue straight lines are the calibration curves. ρ represents Spearman correlation coefficient.

Supplemental Figure 1.2: Bland-Altman plots of ML-predicted and FFQ-based sodium intake compared with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS



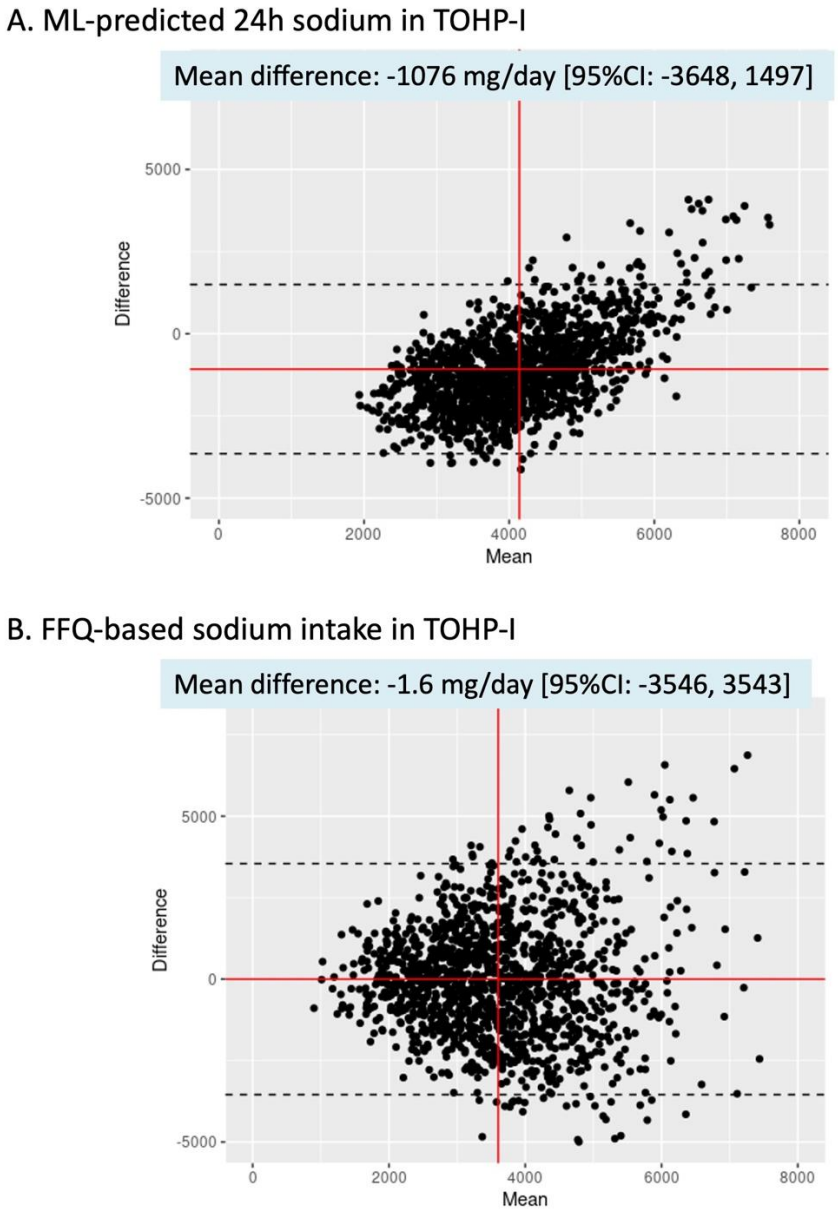
Bland-Altman plots describing the association between ML-predicted sodium intake (A) and FFQ-based sodium intake (B) with respect to averaged 24-hour urinary sodium excretion in N=3,454 participants. ML-predicted sodium intake was based on out-of-fold prediction based on 36 predictors using ensemble over three ML algorithms. Each dot represents each participant. Red lines indicate the mean values of the means and differences, and dashed black lines represent the 95% confidence intervals of the mean differences.

Supplemental Figure 1.3: Comparison of each ML-predicted and FFQ-based sodium intake with averaged 24-hour urinary sodium excretion in N=3,454 participants from NHS, NHS-II, and HPFS based on selected list of predictors



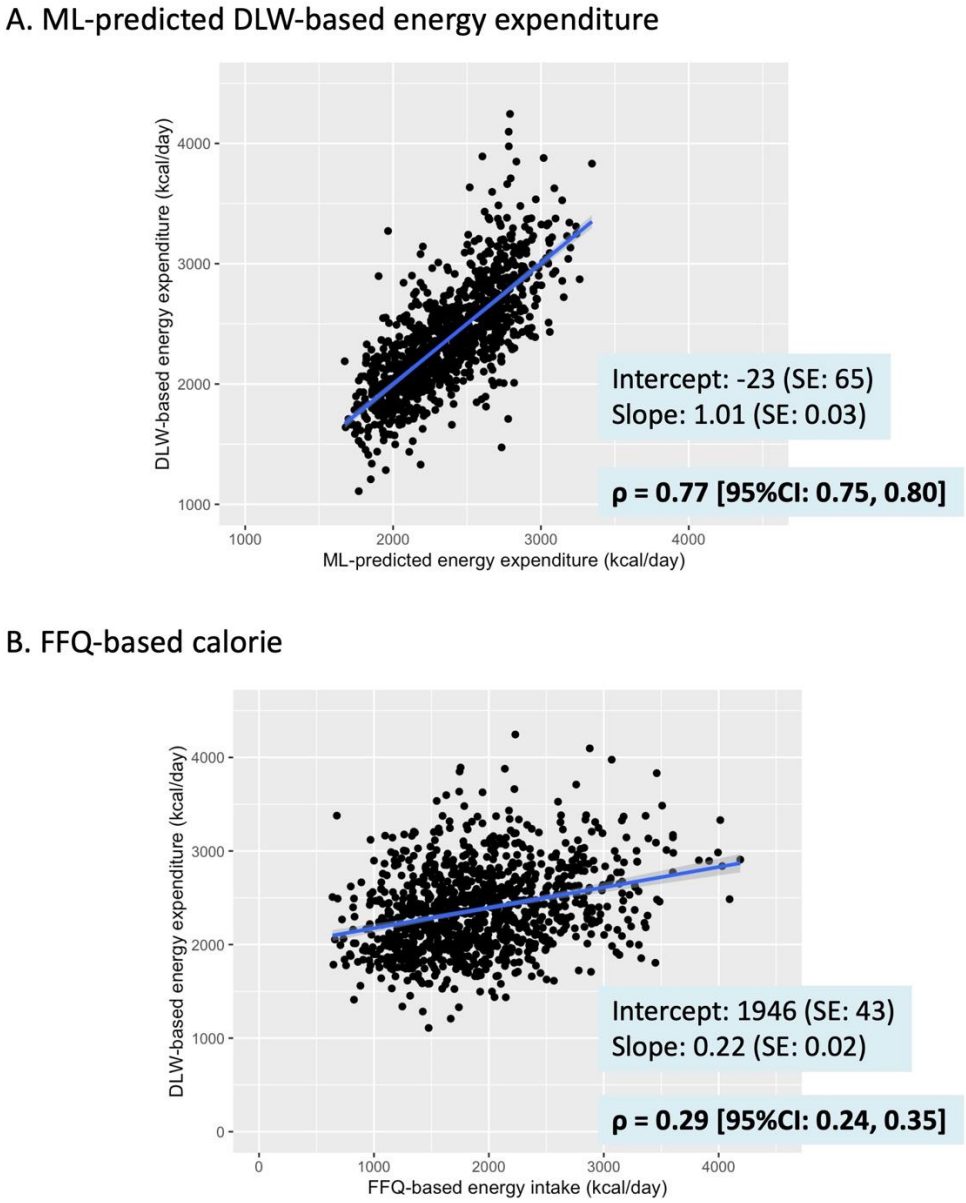
Scatter plots describing the association between each ML-predicted sodium intake with respect to averaged 24-hour urinary sodium excretion in N=3,454 participants. ML-predicted sodium intakes were based on out-of-fold prediction based on 23 predictors using Elastic net, XGBoost, BRNN, and ensemble over these ML algorithms. Each dot represents each participant. Blue straight lines are the calibration curves. ρ represents Spearman correlation coefficient.

Supplemental Figure 1.4: Bland-Altman plots of ML-predicted and FFQ-based sodium intake compared with averaged 24-hour urinary sodium excretion in N=1,423 participants in TOHP-I



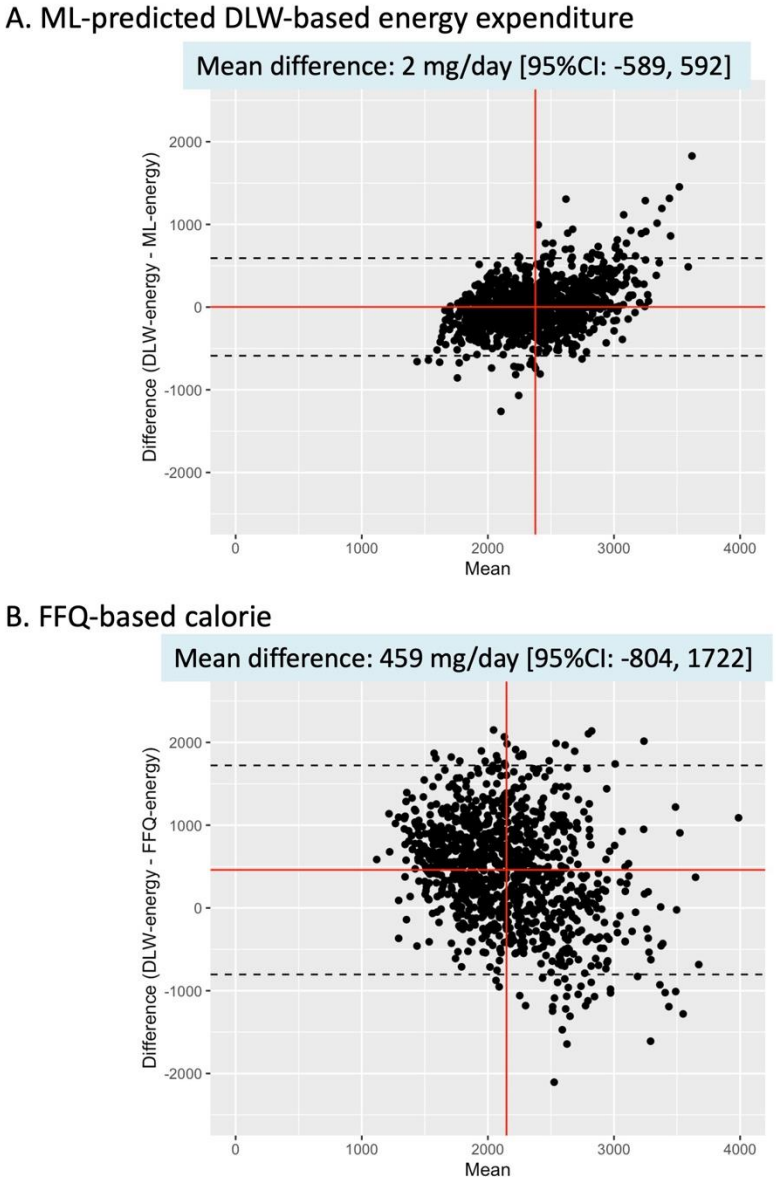
Bland-Altman plots describing the association between ML-predicted sodium intake (A) and FFQ-based sodium intake (B) with respect to averaged 24-hour urinary sodium excretion in N=1,423 participants. ML-predicted sodium intake was based using NHS, NHS-II, and HPFS based on 22 predictors using ensemble over three ML algorithms. Each dot represents each participant. Red lines indicate the mean values of the means and differences, and dashed black lines represent the 95% confidence intervals of the mean differences.

Supplemental Figure 1.5: Comparison of each ML-predicted energy and FFQ-based total calorie intake with DLW-based energy expenditure in N=1,085 participants from NHS, NHS-II, and HPFS



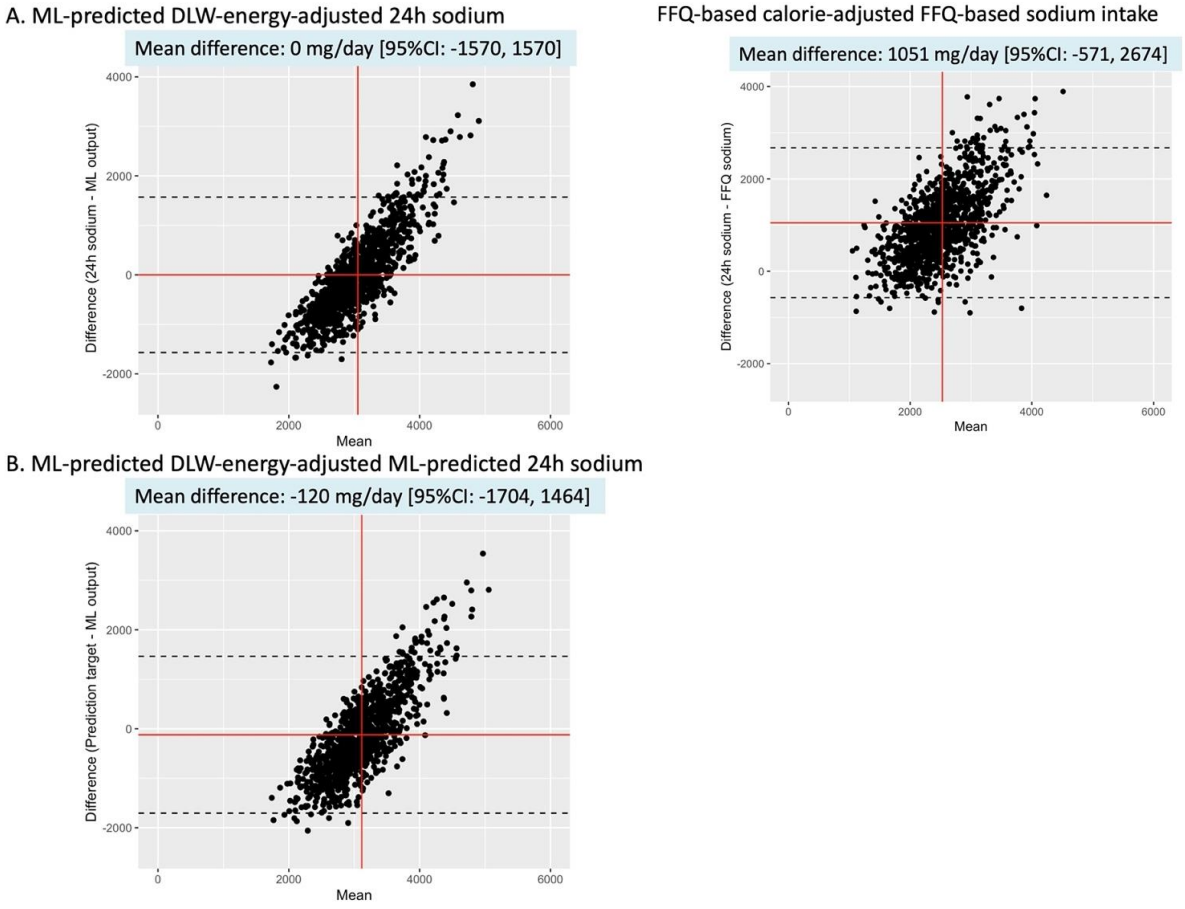
Scatter plots describing the association between ML-predicted DLW-based energy expenditure (A) and FFQ-based total calorie intake (B) with respect to actual DLW-based energy expenditure in N=1,085 participants. ML-predicted sodium intakes were based on out-of-fold prediction based on 6 predictors using Elastic net, XGBoost, BRNN, and ensemble over these ML algorithms. Each dot represents each participant. Blue straight lines are the calibration curves. ρ represents Spearman correlation coefficient.

Supplemental Figure 1.6: Bland-Altman plots of each ML-predicted energy and FFQ-based total calorie intake with DLW-based energy expenditure in N=1,085 participants from NHS, NHS-II, and HPFS



Bland-Altman plots describing the association between ML-derived predicted sodium intake (A) and FFQ-based sodium intake (B) with respect to averaged 24-hour urinary sodium excretion in N=1,085 participants. ML-derived predicted sodium intake was based on out-of-fold prediction based on 6 predictors using ensemble over three ML algorithms. Each dot represents each participant. Red lines indicate the mean values of the means and differences, and dashed black lines represent the 95% confidence intervals of the mean differences.

Supplemental Figure 1.7: Bland-Altman plots of ML-predicted and FFQ-based energy-adjusted sodium measures compared with DLW-based energy expenditure-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants from NHS, NHS-II, and HPFS



Bland-Altman plots describing the association between ML-predicted DLW-energy-adjusted 24-hour urinary sodium excretion (ML-energy-adjusted sodium; A), ML-predicted DLW-energy-adjusted ML-predicted 24-hour urinary sodium excretion (ML-energy-adjusted ML-sodium; B), and FFQ-based calorie-adjusted FFQ-based sodium intake (C) with respect to DLW-energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants. ML-sodium was based on out-of-fold prediction based on N=3,454 participants using ensemble over three ML algorithms, while ML-energy and ML-energy-adjusted sodium was derived from N=1,085 participants. Each dot represents each participant. Red lines indicate the mean values of the means and differences, and dashed black lines represent the 95% confidence intervals of the mean differences.

Supplemental Table 1.1. Baseline characteristics according to the quartile of the DLW-based energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants whose doubly-labelled water (DLW) information is available

	Quartile 1 N = 272	Quartile 2 N = 271	Quartile 3 N = 271	Quartile 4 N = 271
Cohort				
1: HPFS (Men only)	121 (44.5)	102 (37.6)	88 (32.5)	83 (30.6)
2: NHS (Women only)	76 (27.9)	75 (27.7)	83 (30.6)	76 (28.0)
3: NHS-II (Women only)	75 (27.6)	94 (34.7)	100 (36.9)	112 (41.3)
Age, years	65.8 (9.1)	65.1 (9.4)	65.3 (9.1)	63.8 (9.6)
Body mass index, kg/m^2	25.1 (4.4)	25.6 (4.4)	25.8 (4.4)	26.6 (5.1)
Non-White race, %	18 (6.6)	24 (8.9)	16 (5.9)	27 (10.0)
History of hypertension, %	100 (36.8)	94 (34.7)	118 (43.5)	109 (40.2)
History of cancer, %	44 (16.2)	41 (15.1)	30 (11.1)	34 (12.5)
Use of antihypertensives, %	86 (31.6)	88 (32.5)	104 (38.4)	101 (37.3)
Family history of hypertension, %	127 (46.7)	136 (50.2)	115 (42.4)	136 (50.2)
Family history of CVD, %	48 (17.6)	48 (17.7)	61 (22.5)	52 (19.2)
Moderate to vigorous PA, hours/w	6.1 (5.7)	5.2 (5.9)	5.1 (5.4)	4.9 (5.3)
Smoking status				
1: Never	175 (64.3)	170 (62.7)	152 (56.1)	151 (55.7)
2: Past	89 (32.7)	96 (35.4)	115 (42.4)	115 (42.4)
3: Current	8 (2.9)	5 (1.8)	4 (1.5)	5 (1.8)
Alcohol, gram/day	12.4 (15.3)	12.4 (15.5)	11.7 (14.2)	9.9 (13.8)
DASH score without sodium component	22.0 (4.4)	21.7 (4.0)	21.2 (4.0)	20.8 (4.1)
FFQ-based sodium, mg/day	1925 (740)	2147 (788)	2079 (756)	2190 (812)
FFQ-based pottasium, mg/day	3480 (1082)	3546 (1126)	3330 (1087)	3331 (1150)
FFQ-based total calorie, kcal/day	1948 (613)	1994 (625)	1874 (599)	1855 (623)
Averaged 24-hour urinary sodium excretion, mg/day	0.9 (0.1)	1.2 (0.1)	1.4 (0.1)	1.8 (0.2)
DLW-based energy expenditure (kcal/day)	2512 (502)	2385 (429)	2329 (424)	2283 (424)
Only in women				
Parity, N	1.3 (0.7)	1.3 (0.8)	1.4 (0.7)	1.3 (0.7)
Post menopausal	115 (76.2)	130 (76.9)	151 (82.5)	148 (78.7)
Use of hormone replacement therapy				
1: Never	46 (30.5)	54 (32.0)	53 (29.0)	60 (31.9)
2: Past	67 (44.4)	85 (50.3)	94 (51.4)	79 (42.0)
3: Current	38 (25.2)	30 (17.8)	36 (19.7)	49 (26.1)
Ever use of oral contraceptives	72 (47.7)	89 (52.7)	97 (53.0)	111 (59.0)

Values are frequency (%) for categorical variables and mean (SD) or median [IQR] for continuous variables.

Supplemental Table 1.2. Interpretation of ML models based on selected covariate list predicting averaged 24-hour urinary sodium excretion in N=3,454 participants

A. Coefficients of elastic net (per SD for continuous variables)

Positive association		***Shrunk to null***		***Negative association***	
	Coef		*Coef*		*Coef*
Weight	503	Height	0	Age	-237
Men	281			FFQ-calorie	-132
FFQ-sodium	168			BMI	-125
FFQ-sodium/FFQ-calorie	138			DASH score	-45
Red meat	49			FFQ-potassium	-21
Use of diuretics	38			Alcohol	-21
FFQ-potassium/FFQ-calorie	35			PA hours	-13
Postmenopausal	30			FFQ-sodium/FFQ-potassium	-6
Use of antihypertensives	16				
Vegetable	9				
History of hypertension	2				
Tea or coffee	1				
Fruit	1				

B. Top 10 variable importance in XGBoost

	Variable importance
Weight	100
FFQ-sodium/FFQ-calorie	34
Age	30
BMI	28
Height	20
FFQ-sodium/ FFQ-potassium	19
FFQ-sodium	14
Fruit	12
Red meat	12
Men	11

Values are coefficients per category for categorical variables and per SD for continuous variables (A). Variable importance measures are scaled into 0 to 100 (B).

Supplemental Table 1.3. Interpretation of ML models predicting DLW-based energy expenditure in N=1,085 participants

A. Coefficients of elastic net (per SD for continuous variables)

	Coef
Weight	180
Men	153
Height	68
PA hours	56
FFQ-calorie	34
Age	-110

B. Variable importance in XGBoost

	Variable importance
Weight	100
PA hours	32
Age	32
Height	30
FFQ-calorie	29
Men	0

Values are coefficients per category for categorical variables and per SD for continuous variables (A). Variable importance measures are scaled into 0 to 100 (B).

Supplemental Table 1.4. Interpretation of ML models predicting DLW-based energy-adjusted averaged 24-hour urinary sodium excretion in N=1,085 participants

A. Coefficients of elastic net (per SD for continuous variables)

Positive association		***Shrunk to null***		***Negative association***	
	Coef		*Coef*		*Coef*
FFQ-sodium/FFQ-calorie	54	Married	0	Fruit	-22
FFQ-sodium/ FFQ-potassium	35	Family history of hypertension	0	FFQ-calorie	-17
Use of table salt	34			PA hours	-15
Hamburger	19			DASH score	-15
Red meat	16			Height	-13
FFQ-sodium	15			FFQ-potassium	-11
Use of oral contraceptives	15			Alcohol	-9
Post menopausal	14			Age	-3
Use of diuretics	14			Fried food away from home	-1
Parity	14			NHS	-1
Vegetable	13				
NHS II	13				
BMI	13				
Processed meat	13				
Current use of HRT	10				
Use of antihypertensives	9				
FFQ-potassium	9				
Tea or coffee	7				
Fried food at home	6				
History of hypertension	4				
Weight	4				
Cheese	4				
Living alone	3				
Hotdog	1				
Ever use of HRT	1				
Family history of CVD	1				

B. Top 20 variable importance in XGBoost

	Variable importance
FFQ-sodium/FFQ-calorie	100
FFQ-sodium/FFQ-potassium	57
Age	45
Vegetable	44
Fruit	42

Use of table salt	40
FFQ-calorie	37
Tea or coffee	34
DASH score	33
FFQ-potassium	32
PA hours	30
FFQ-sodium	29
BMI	26
FFQ-potassium/FFQ-calorie	26
Alcohol	24
Height	23
Weight	23
Red meat	22
Cheese	21
Hotdog	17

A: Values are coefficients per category for categorical variables and per SD for continuous variables. Coefficients were multiplied by the median of DLW-based energy expenditure (2347 kcal/day) to make the value interpretable.
B: Variable importance measures are scaled into 0 to 100.

Appendix B. Supplemental data to Chapter 2

Modifiable lifestyle factors in the primordial prevention of hypertension in three US cohorts

Supplemental Table 2.1. Sensitivity analysis of population attributable risks % for modifiable lifestyles and incident hypertension and early-onset hypertension among women and men in the Nurses' Health Studies and the Health Professionals Follow-up Study

Supplemental Table 2.1. Sensitivity analysis of population attributable risks % for modifiable lifestyles and incident hypertension and early-onset hypertension among women and men in the Nurses' Health Studies and the Health Professionals Follow-up Study

NHS

	Weight	*PA*	*Diet*	*Alcohol*
Incident hypertension at any age				
Main analyses	19.6 (18.1, 21.1)	12.3 (10.7, 14.0)	4.0 (2.5, 5.5)	1.2 (0.8, 1.6)
Outcome: self-reporting plus starting anti-hypertensives	20.2 (18.6, 21.8)	12.7 (11.0, 14.4)	3.5 (1.9, 5.1)	1.2 (0.8, 1.6)
Population: among those not using anti-hypertensives	19.5 (17.9, 21.1)	12.1 (10.4, 13.8)	4.8 (3.2, 6.4)	1.2 (0.7, 1.6)
Predicted absolute sodium intake used in DASH score	19.7 (18.1, 21.2)	12.4 (10.8, 14.1)	3.4 (2.1, 4.8)	1.2 (0.8, 1.6)
Including smoking as an exposure *PAR for smoking: 0.4 (-0.1, 0.8)	19.6 (18.1, 21.1)	12.3 (10.7, 14.0)	4.0 (2.5, 5.5)	1.3 (0.9, 1.7)
Incident early-onset hypertension				
Main analyses	30.6 (27.4, 33.7)	13.8 (9.9, 17.7)	9.6 (5.1, 14.1)	2.0 (1.2, 2.9)
Outcome: self-reporting plus starting anti-hypertensives	30.8 (27.5, 33.9)	13.3 (9.3, 17.2)	10.2 (5.6, 14.8)	1.8 (0.9, 2.7)
Population: among those not using anti-hypertensives	29.9 (26.6, 33.2)	14.0 (9.9, 18.0)	10.3 (5.6, 15.0)	1.9 (1.0, 2.8)
Predicted absolute sodium intake used in DASH score	30.7 (27.6, 33.8)	14.0 (10.1, 17.9)	7.2 (3.3, 11.2)	2.0 (1.2, 2.9)
Including smoking as an exposure *PAR for smoking: –	30.6 (27.4, 33.7)	13.7 (9.8, 17.6)	9.6 (5.1, 14.1)	2.1 (1.2, 2.9)

NHS II

	Weight	*PA*	*Diet*	*Alcohol*
Incident hypertension at any age				
Main analyses	35.5 (33.7, 37.2)	10.8 (9.5, 12.1)	14.0 (12.4, 15.6)	2.6 (2.2, 2.9)
Outcome: self-reporting plus starting anti-hypertensives	36.0 (34.2, 37.8)	10.9 (9.5, 12.3)	14.8 (13.1, 16.5)	2.4 (2.0, 2.8)
Population: among those not using anti-hypertensives	35.6 (33.9, 37.4)	10.8 (9.4, 12.2)	14.0 (12.3, 15.7)	2.5 (2.1, 2.9)
Predicted absolute sodium intake used in DASH score	35.7 (34.0, 37.4)	11.5 (10.2, 12.8)	10.1 (8.6, 11.5)	2.5 (2.2, 2.9)
Optimal BMI and waist as the weight measure	38.6 (36.1, 41.0)	10.7 (8.9, 12.5)	13.5 (11.4, 15.7)	2.9 (2.3, 3.4)
Including smoking as an exposure *PAR for smoking: –	35.5 (33.7, 37.2)	10.8 (9.4, 12.1)	14.0 (12.4, 15.6)	2.6 (2.3, 3.0)
Incident early-onset hypertension				
Main analyses	37.5 (35.6, 39.5)	11.2 (9.7, 12.8)	15.1 (13.1, 17.0)	2.1 (1.8, 2.5)
Outcome: self-reporting plus starting anti-hypertensives	38.1 (36.0, 40.1)	11.4 (9.8, 13.1)	15.8 (13.7, 17.8)	2.1 (1.7, 2.5)
Population: among those not using anti-hypertensives	37.6 (35.6, 39.5)	11.1 (9.5, 12.6)	15.1 (13.1, 17.1)	2.1 (1.7, 2.5)

Predicted absolute sodium intake used in DASH score	37.8 (35.9, 39.7)	12.0 (10.4, 13.5)	10.4 (8.6, 12.1)	2.1 (1.7, 2.5)
Optimal BMI and waist as the weight measure	40.0 (37.2, 42.6)	10.7 (8.7, 12.7)	14.6 (12.1, 17.1)	2.5 (2.0, 3.1)
Including smoking as an exposure *PAR for smoking: –	37.5 (35.6, 39.5)	11.2 (9.7, 12.8)	15.1 (13.1, 17.0)	2.2 (1.8, 2.5)

HPFS

	Weight	***PA***	***Diet***	***Alcohol***
Incident hypertension at any age				
Main analyses	21.7 (19.4, 23.9)	2.5 (1.1, 3.9)	3.4 (1.1, 5.7)	3.5 (2.6, 4.3)
Outcome: self-reporting plus starting anti-hypertensives	18.6 (15.4, 21.7)	0.8 (-1.1, 2.8)	0.1 (-3.1, 3.2)	4.4 (3.1, 5.6)
Population: among those not using anti-hypertensives	22.0 (19.7, 24.3)	2.1 (0.6, 3.6)	3.5 (1.1, 5.9)	3.6 (2.7, 4.5)
Predicted absolute sodium intake used in DASH score	21.7 (19.5, 23.9)	2.6 (1.2, 4.0)	2.5 (0.5, 4.6)	3.4 (2.6, 4.3)
Including smoking as an exposure *PAR for smoking: 0.6 (0.2, 1.0)	22.1 (19.9, 24.3)	2.7 (1.3, 4.1)	3.7 (1.5, 6.0)	4.0 (3.1, 4.8)
Incident early-onset hypertension				
Main analyses	33.6 (28.9, 38.0)	5.5 (1.9, 9.0)	7.5 (1.1, 13.9)	4.4 (2.6, 6.3)
Outcome: self-reporting plus starting anti-hypertensives	29.8 (22.3, 36.9)	8.5 (3.1, 13.9)	1.1 (-9.2, 11.4)	3.3 (0.5, 6.2)
Population: among those not using anti-hypertensives	33.9 (29.1, 38.6)	4.7 (1.1, 8.4)	5.8 (-0.9, 12.5)	4.2 (2.2, 6.1)
Predicted absolute sodium intake used in DASH score	33.6 (29.0, 38.1)	5.5 (2.0, 9.0)	7.7 (1.9, 13.4)	4.4 (2.5, 6.3)
Including smoking as an exposure *PAR for smoking: 0.5 (-0.7, 1.6)	33.6 (29.0, 38.0)	5.5 (1.9, 9.0)	7.5 (1.1, 13.9)	4.4 (2.6, 6.3)

Numbers indicate PAR % (95% CI). Early-onset hypertension was defined as incident hypertension in age < 55 years.

PAR percent were computed for each dichotomized lifestyle factor; BMI (<25 kg/m2 vs. ≥25 kg/m2); moderate-to-vigorous physical activity (≥2.5 hours/week vs. <2.5 hours/week); DASH score (40th percentile or higher vs. lower than 40th percentile); and alcohol intake (≤20 gram/day vs. >20 gram/day in male; ≤15 gram/day vs. >15 gram/day in female).

Models were mutually adjusted for four modifiable risk factors and the following covariates: age, race, family history of hypertension, smoking status, potassium intake, use of aspirin, acetaminophen, and NSAIDs, and energy expenditure in all cohorts; plus parity, menopausal status, and hormone replacement therapy in NHS and NHS II; and plus contraceptive use in NHS II.

PAR was not shown if the value was less than zero.

Appendix C. Supplemental data to Chapter 4

Association of plasma branched chain amino acid with biomarkers of inflammation and lipid metabolism in women

Supplemental Table 4.1. Spearman correlation matrix between plasma BCAA and inflammation, lipid, and HbA1c biomarkers

	BCAAs	hsCRP	Fibrinogen	GlycA	sICAM-1	TG	HDL-c	LDL-c	LPIR	HbA1c
BCAAs	1	0.24	0.12	0.30	0.10	0.26	-0.27	0.14	0.34	0.17
hsCRP		1	0.36	0.57	0.25	0.40	-0.15	0.07	0.40	0.15
Fibrinogen			1	0.43	0.25	0.12	-0.23	0.20	0.15	0.20
GlycA				1	0.26	0.44	-0.23	0.18	0.45	0.19
sICAM-1					1	0.21	-0.20	0.14	0.23	0.15
Triglyceride						1	-0.37	0.31	0.78	0.14
HDL-c							1	-0.08	-0.59	-0.15
LDL-c								1	0.16	0.16
LPIR									1	0.17
HbA1c										1

Values indicate spearman correlation coefficients.

Supplemental Table 4.2. Adjusted means of inflammation/lipid biomarkers by the quartiles of BCAA level stratified by BMI

	BCAA quartiles				
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	p-interaction
	Geometric mean [95% confidence interval]				
Inflammation					
hsCRP, mg/L					0.29
BMI: <25	1.06 [0.98, 1.16]	1.16 [1.06, 1.26]	1.26 [1.15, 1.38]	1.40 [1.27, 1.53]	
BMI: 25 to <30	1.62 [1.48, 1.79]	1.93 [1.76, 2.12]	2.18 [1.99, 2.39]	2.30 [2.11, 2.51]	
BMI: ≥30	2.90 [2.55, 3.31]	3.29 [2.9, 3.73]	3.63 [3.23, 4.09]	3.92 [3.5, 4.39]	
Fibrinogen, mg/dL					0.0083
BMI: <25	344 [339, 349]	347 [341, 352]	349 [344, 355]	349 [343, 354]	
BMI: 25 to <30	371 [364, 378]	375 [368, 382]	377 [371, 385]	378 [371, 385]	
BMI: ≥30	413 [400, 426]	409 [397, 421]	413 [401, 424]	409 [398, 420]	
sICAM-1, ng/mL					<0.0001
BMI: <25	348 [343, 354]	348 [342, 354]	350 [344, 355]	351 [345, 357]	
BMI: 25 to <30	354 [346, 361]	357 [350, 364]	361 [354, 368]	365 [358, 372]	
BMI: ≥30	360 [348, 372]	369 [358, 381]	373 [362, 384]	381 [370, 391]	
GlycA, µmol/L					0.081
BMI: <25	363 [358, 368]	374 [369, 379]	384 [379, 389]	394 [388, 399]	
BMI: 25 to <30	375 [369, 381]	393 [387, 399]	404 [398, 411]	413 [407, 420]	
BMI: ≥30	397 [387, 406]	410 [401, 420]	421 [412, 430]	431 [423, 440]	
Lipid Triglyceride, mg/dL					0.0008
BMI: <25	110 [106, 114]	115 [111, 120]	122 [117, 127]	134 [129, 140]	
BMI: 25 to <30	119 [113, 125]	129 [123, 135]	137 [131, 144]	154 [147, 161]	
BMI: ≥30	127 [118, 137]	130 [121, 140]	134 [125, 143]	149 [140, 159]	
HDL-c, mg/dL					0.12
BMI: <25	57.1 [56, 58.2]	55.5 [54.4, 56.6]	53.7 [52.6, 54.8]	51.2 [50.2, 52.3]	
BMI: 25 to <30	53.0 [51.7, 54.3]	51.3 [50.2, 52.5]	49.8 [48.7, 51]	47.2 [46.2, 48.3]	
BMI: ≥30	49.8 [48, 51.6]	48.3 [46.6, 50]	47.7 [46.2, 49.2]	45.2 [43.8, 46.6]	
LDL-c, mg/dL					0.0004
BMI: <25	124 [121, 126]	128 [125, 131]	130 [127, 133]	132 [130, 135]	
BMI: 25 to <30	127 [124, 131]	132 [129, 136]	134 [131, 137]	137 [134, 140]	

BMI: ≥30	125 [120, 131]	127 [122, 132]	129 [125, 134]	130 [126, 135]	
LPIR score					0.0002
BMI: <25	35.3 [33.7, 37]	38.1 [36.3, 39.8]	41.8 [40, 43.5]	48.3 [46.5, 50.1]	
BMI: 25 to <30	39.8 [37.5, 42.1]	44.2 [42, 46.4]	48.5 [46.3, 50.7]	56.4 [54.2, 58.5]	
BMI: ≥30	42.3 [39.1, 45.6]	47.5 [44.4, 50.6]	50.5 [47.6, 53.4]	57.8 [55, 60.6]	

Numbers are adjusted geometric means [95% confidence intervals] calculated based on multivariable linear regression.

Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous).

P-interaction was assessed using log-likelihood ratio test (two degrees of freedom test).

P-thresholds after Bonferroni correction were 0.0063.

Supplemental Table 4.3. Adjusted means of inflammatory and lipid biomarkers by the quartiles of isoleucine level

	Isoleucine quartile				
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	p-trend
Inflammation					
hsCRP, mg/L	1.52 [1.44, 1.61]	1.61 [1.52, 1.71]	1.70 [1.61, 1.8]	1.94 [1.83, 2.05]	<0.0001
Fibrinogen, mg/dL	364 [360, 368]	366 [362, 370]	365 [361, 369]	366 [362, 370]	0.26
sICAM-1, ng/mL	351 [347, 355]	352 [348, 356]	356 [352, 360]	363 [359, 367]	<0.0001
GlycA, µmol/L	380 [376, 383]	386 [383, 390]	390 [387, 394]	403 [399, 407]	<0.0001
Lipid					
TG, mg/dL	113 [110, 116]	121 [118, 124]	126 [123, 130]	147 [143, 151]	<0.0001
HDL-c, mg/dL	55.1 [54.3, 55.8]	53.4 [52.7, 54.1]	52.0 [51.3, 52.7]	48.4 [47.7, 49]	<0.0001
LDL-c, mg/dL	126 [125, 128]	129 [127, 131]	130 [128, 132]	131 [129, 133]	<0.0001
LPIR score	36.7 [35.5, 37.9]	40.7 [39.5, 41.9]	44.4 [43.2, 45.6]	54 [52.8, 55.2]	<0.0001

Numbers are adjusted geometric means [95% confidence intervals] calculated based on multivariable linear regression.

Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous).

Test for trend was based on a variable containing the median value for each quartile.

P-trend threshold was 0.0021 after Bonferroni correction.

Supplemental Table 4.4. Adjusted means of inflammatory and lipid biomarkers by the quartiles of leucine level

	Leucine quartile				
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	p-trend
Inflammation					
hsCRP, mg/L	1.49 [1.41, 1.57]	1.61 [1.53, 1.71]	1.74 [1.65, 1.84]	1.97 [1.86, 2.09]	<0.0001
Fibrinogen, mg/dL	361 [357, 364]	364 [360, 368]	368 [364, 372]	370 [366, 374]	<0.0001
sICAM-1, ng/mL	354 [350, 358]	355 [350, 359]	356 [352, 360]	360 [356, 364]	<0.0001
GlycA, μmol/L	374 [370, 377]	384 [380, 387]	394 [391, 398]	410 [406, 414]	<0.0001
Lipid					
TG, mg/dL	122 [118, 125]	122 [119, 125]	127 [123, 130]	137 [133, 141]	<0.0001
HDL-c, mg/dL	53.5 [52.7, 54.2]	52.8 [52, 53.5]	51.8 [51.1, 52.5]	50.0 [49.3, 50.7]	<0.0001
LDL-c, mg/dL	124 [122, 126]	127 [125, 129]	131 [129, 133]	135 [133, 137]	<0.0001
LPIR score	41.1 [39.8, 42.3]	42 [40.8, 43.3]	44.7 [43.4, 45.9]	49.5 [48.3, 50.8]	<0.0001

Numbers are adjusted geometric means [95% confidence intervals] calculated based on multivariable linear regression.

Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous).

Test for trend was based on a variable containing the median value for each quartile.

P-trend threshold was 0.0021 after Bonferroni correction.

Supplemental Table 4.5. Adjusted means of inflammatory and lipid biomarkers by the quartiles of valine level

	Valine quartile				
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	p-trend
Inflammation					
hsCRP, mg/L	1.43 [1.36, 1.52]	1.65 [1.56, 1.75]	1.79 [1.7, 1.9]	1.93 [1.82, 2.04]	<0.0001
Fibrinogen, mg/dL	365 [361, 368]	365 [361, 369]	367 [363, 371]	365 [361, 369]	0.34
sICAM-1, ng/mL	354 [350, 358]	355 [350, 359]	356 [352, 360]	359 [355, 363]	<0.0001
GlycA, µmol/L	375 [372, 379]	386 [383, 390]	396 [393, 400]	403 [399, 406]	<0.0001
Lipid					
TG, mg/dL	112 [109, 116]	123 [119, 126]	131 [127, 135]	142 [138, 146]	<0.0001
HDL-c, mg/dL	54.9 [54.1, 55.6]	53.1 [52.4, 53.9]	51.2 [50.5, 51.9]	49.2 [48.5, 49.9]	<0.0001
LDL-c, mg/dL	126 [124, 128]	129 [127, 131]	131 [130, 133]	131 [129, 133]	<0.0001
LPIR score	37.2 [35.9, 38.4]	41.4 [40.2, 42.7]	46.2 [45, 47.4]	51.9 [50.7, 53.1]	<0.0001

Numbers are adjusted geometric means [95% confidence intervals] calculated based on multivariable linear regression.

Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous).

Test for trend was based on a variable containing the median value for each quartile.

P-trend threshold was 0.006 after Bonferroni correction.

Supplemental Table 4.6. Adjusted means of inflammatory and lipid biomarkers by the quartiles of BCAA level stratified by age

A. 14926 women with age<60 years

	BCAA quartile				
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	p-trend
Inflammation					
hsCRP, mg/L	1.3 [1.2, 1.4]	1.5 [1.4, 1.7]	1.7 [1.6, 1.9]	1.9 [1.8, 2.0]	<0.0001
Fibrinogen, mg/dL	357 [352, 362]	360 [355, 365]	363 [358, 368]	362 [357, 367]	0.0005
sICAM-1, ng/mL	353 [348, 358]	354 [349, 359]	357 [352, 362]	361 [355, 366]	<0.0001
GlycA, μmol/L	371 [367, 375]	385 [381, 389]	396 [392, 401]	406 [402, 411]	<0.0001
Lipid					
TG, mg/dL	112 [108, 116]	120 [116, 124]	127 [122, 131]	142 [138, 147]	<0.0001
HDL-c, mg/dL	55 [54, 56]	53 [52, 54]	51 [50, 52]	49 [48, 49]	<0.0001
LDL-c, mg/dL	125 [122, 127]	130 [127, 132]	131 [129, 134]	134 [132, 136]	<0.0001
LPIR score	37 [35, 38]	41 [39, 43]	45 [43, 46]	53 [51, 54]	<0.0001

B. 4546 women with age≥60 years

	BCAA quartile				
	Quartile 1	Quartile 2	Quartile 3	Quartile 4	p-trend
Inflammation					
hsCRP, mg/L	2.4 [1.5, 3.8]	2.6 [1.6, 4.0]	2.8 [1.8, 4.4]	3.1 [2.0, 4.8]	<0.0001
Fibrinogen, mg/dL	377 [344, 413]	380 [347, 416]	383 [350, 419]	380 [347, 416]	0.16
sICAM-1, ng/mL	380 [346, 417]	377 [344, 414]	380 [346, 417]	389 [354, 427]	0.0014
GlycA, μmol/L	365 [339, 394]	378 [351, 408]	390 [362, 421]	399 [370, 430]	<0.0001
Lipid					
TG, mg/dL	125 [100, 156]	131 [105, 164]	140 [112, 174]	153 [122, 190]	<0.0001
HDL-c, mg/dL	55 [49, 62]	54 [48, 61]	52 [46, 58]	49 [44, 55]	<0.0001
LDL-c, mg/dL	128 [111, 144]	133 [117, 150]	135 [119, 151]	136 [120, 152]	<0.0001
LPIR score	48 [38, 59]	51 [40, 61]	56 [45, 66]	62 [52, 73]	<0.0001

Numbers are adjusted geometric means [95% confidence intervals] calculated based on multivariable linear regression.

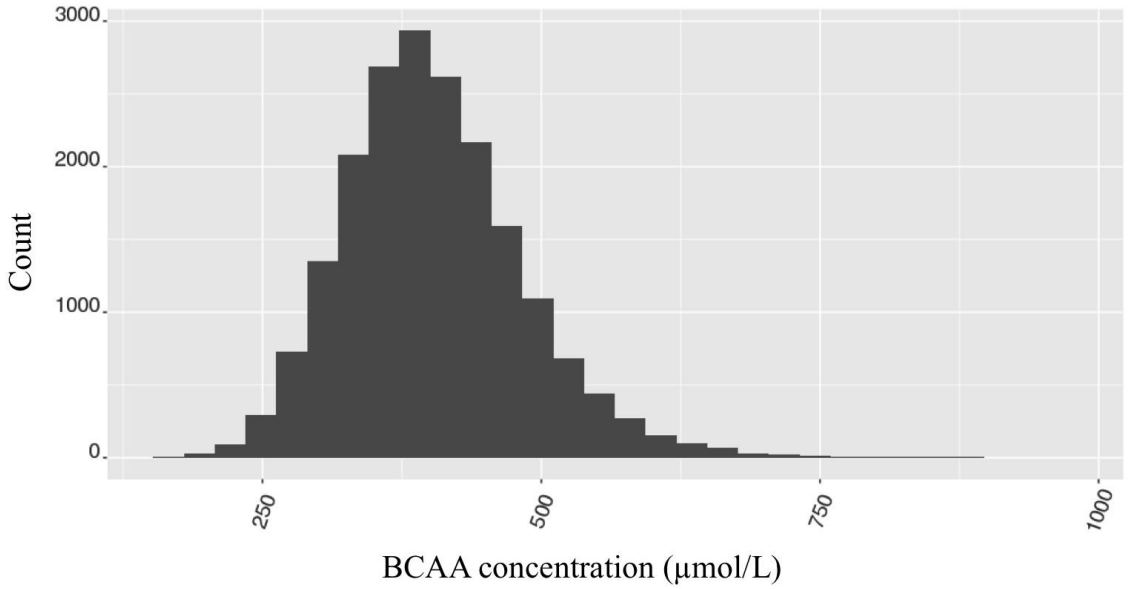
Models were adjusted for age at the randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or not), family history of diabetes, smoking (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol

consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous).

Test for trend was based on a variable containing the median value for each quartile.

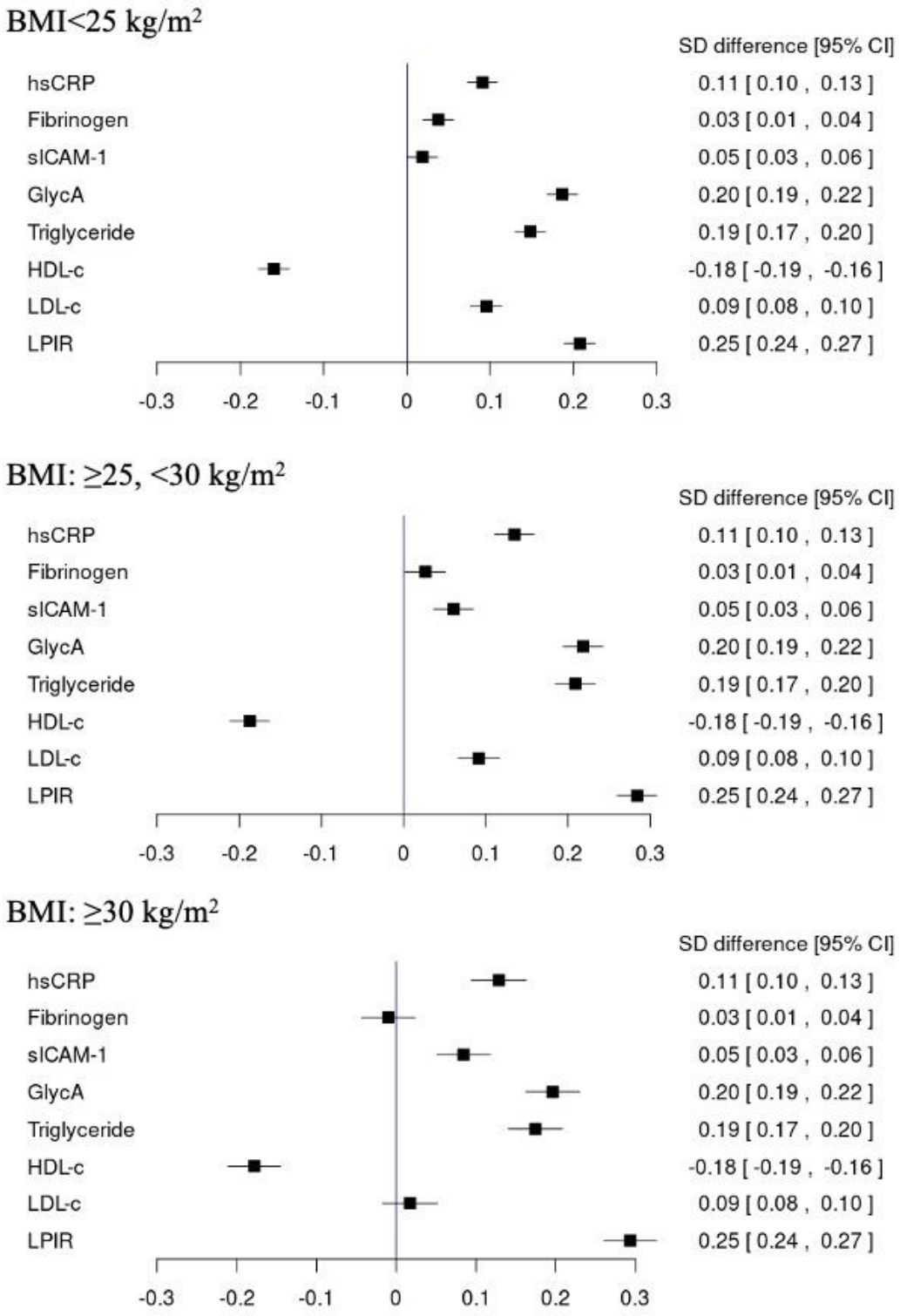
P-trend threshold was 0.006 after Bonferroni correction.

Supplemental Figure 4.1: Distribution of circulating BCAA concentrations



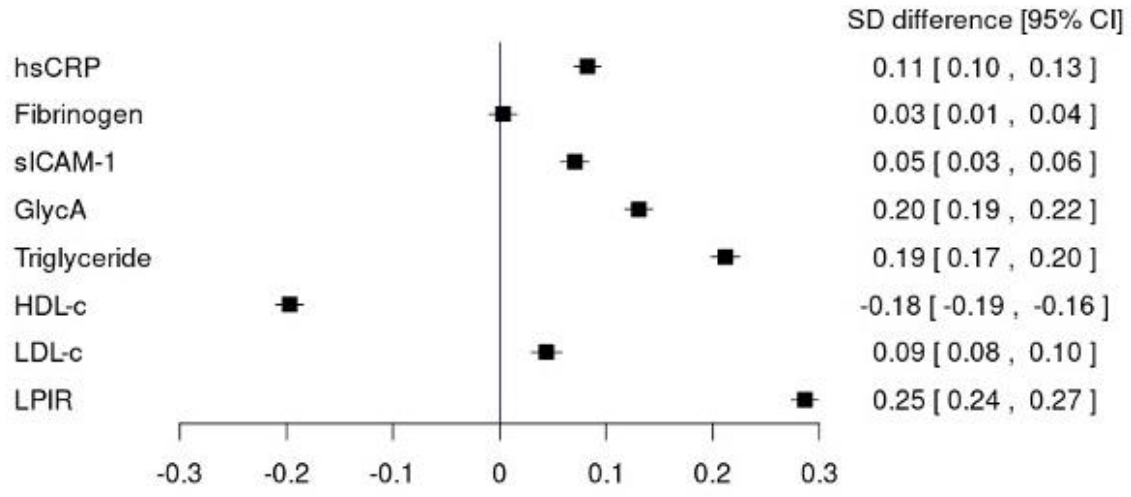
Histogram of plasma branched chain amino acid (BCAA) levels in the present cohort is shown.

Supplemental Figure 4.2: Standardized differences of cardiometabolic biomarkers per SD difference of BCAA levels, stratified by BMI



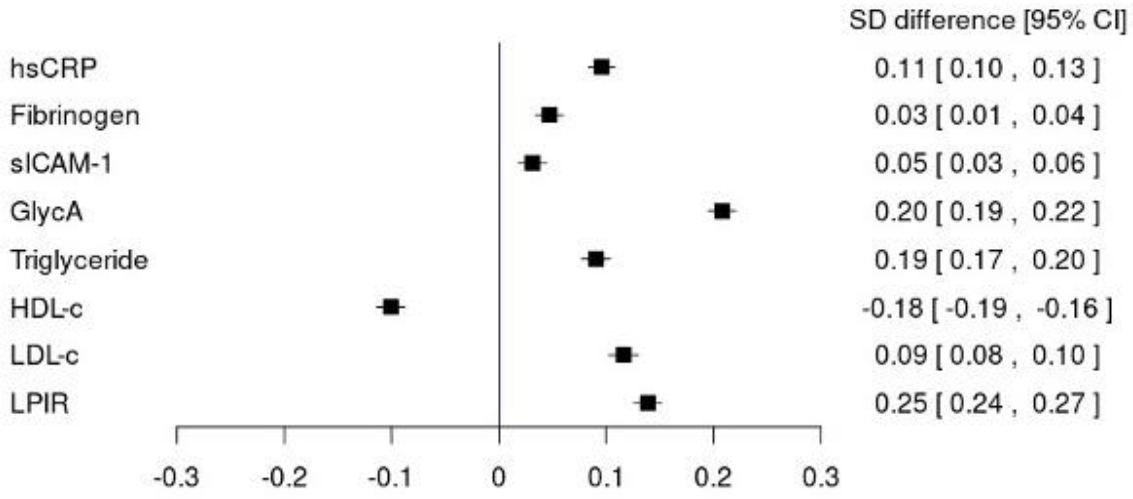
Linear regressions of standardized biomarkers constructed by standardized continuous total BCAA levels and covariates [age at randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or non-white), family history of diabetes, smoking history (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous)]. Standardized differences [95% confidence interval] per SD of BCAAs stratified by BMI (<25, 25 to <30, and ≥30 kg/m^2) are shown.

Supplemental Figure 4.3: Standardized differences of cardiometabolic biomarkers per SD difference of isoleucine levels



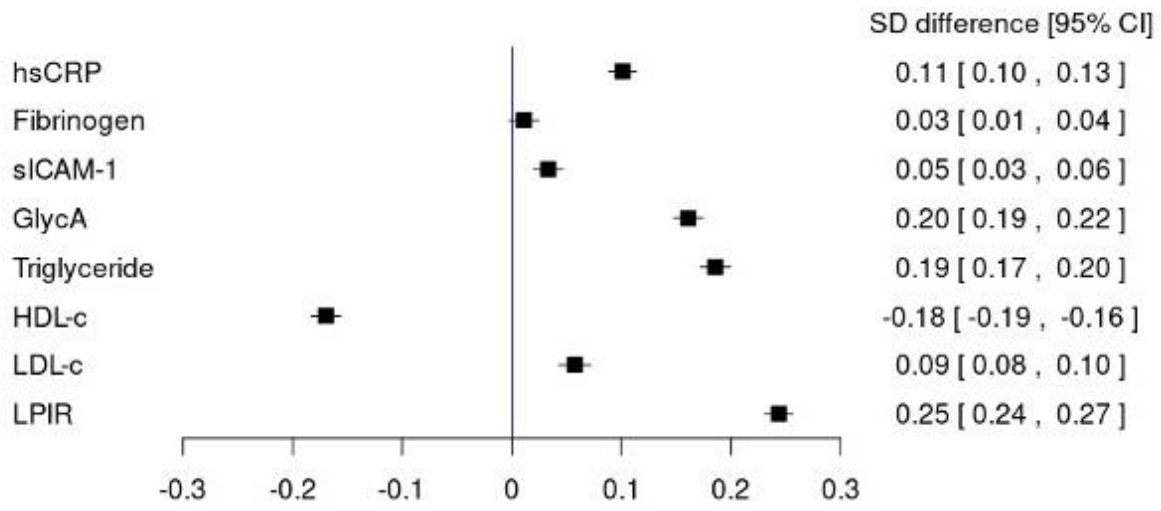
Linear regressions of standardized biomarkers constructed by standardized continuous total isoleucine levels and covariates [age at randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or non-white), family history of diabetes, smoking history (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous)]. Standardized differences [95% confidence interval] per SD of isoleucine are shown.

Supplemental Figure 4.4: Standardized differences of cardiometabolic biomarkers per SD difference of leucine levels



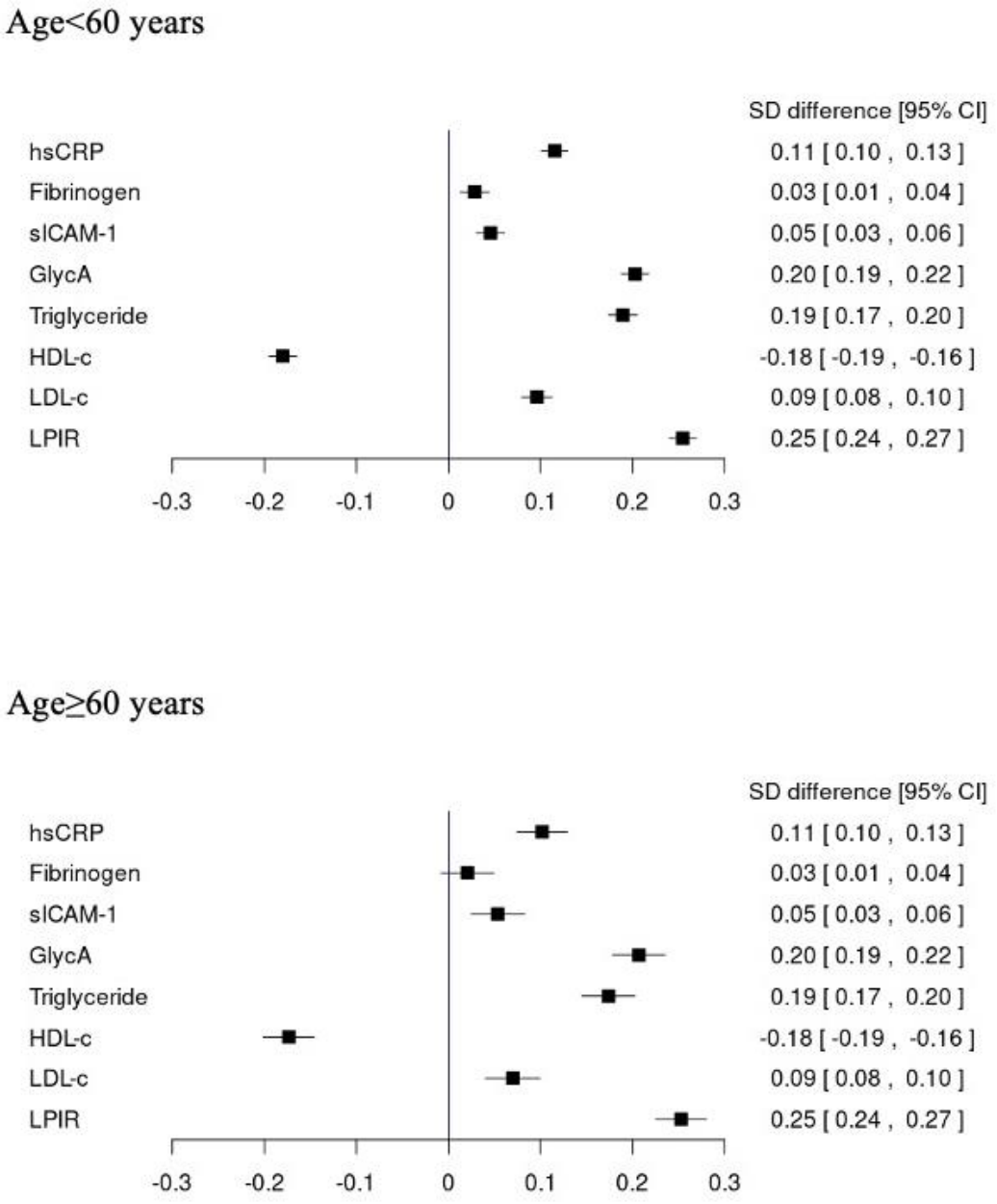
Linear regressions of standardized biomarkers constructed by standardized continuous total leucine levels and covariates [age at randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or non-white), family history of diabetes, smoking history (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous)]. Standardized differences [95% confidence interval] per SD of leucine are shown.

Supplemental Figure 4.5: Standardized differences of cardiometabolic biomarkers per SD difference of valine levels



Linear regressions of standardized biomarkers constructed by standardized continuous total valine levels and covariates [age at randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or non-white), family history of diabetes, smoking history (none, ever, current), menopausal status (premenopausal, postmenopausal [natural], postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous)]. Standardized differences [95% confidence interval] per SD of valine are shown.

Supplemental Figure 4.6: Standardized differences of cardiometabolic biomarkers per SD difference of BCAA levels stratified by age



Linear regressions of standardized biomarkers constructed by standardized continuous total BCAA levels and covariates [age at randomization (continuous), assignment to ASA group, assignment to vitamin E group, race (white or non-white), family history of diabetes, smoking history (none, ever, current), menopausal status (premenopausal, postmenopausal [natural],

postmenopausal [non-natural], unsure), use of menopausal hormone therapy (never, past, current), parity as number of pregnancies lasting ≥6 months (nulliparous, 0, 1, 2, ≥3), exercise as total MET-hour/week (quintiles), aHEI-2010 (quintiles), alcohol consumption (none, <10g/day, <20g/day, ≥20g/day), the use of cholesterol lowering drugs, and BMI (continuous)]. Standardized differences [95% confidence interval] per SD of BCAAs stratified by age (<60 and ≥60 years) are shown.