Title: Bias in nutrition-health associations is not eliminated by excluding extreme reporters in empirical or simulation studies

Author Names: Nao Yamamoto[1], Keisuke Ejima[2*], Roger S. Zoh[2], Andrew W. Brown[3,4]

Author Affiliations:

[1]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA.

[2]Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA.

[3]Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR, USA

[4]Arkansas Children's Research Institute, Little Rock, AR, USA

ORCID: 0000-0002-1185-3987 (KE), 0000-0002-1758-8205 (AWB)

Corresponding Author: Keisuke Ejima, PH158A, 1025 E 7th St, Bloomington, Indiana 47405, United States. (TEL) 812-855-9172, ejimakeisuke@gmail.com

Sources of Support: NIH grants R25HL124208, R25DK099080, R25GM141507, 1R01DK132385-01, U01-CA057030-29S1. Japan Society for Promotion of Science (JSPS) KAKENHI grant 18K18146. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or any other organization.

Short running head: Utility of excluding implausible nutrition intake

Abbreviations: AARP, American Association of Retired Persons; ASA24, Automated Self-Administered 24-Hour Dietary Assessment Tool; BIO, biomarker; BW, body weight; DBP, resting diastolic blood pressure; DHQ-II, Diet History Questionnaire; DLW, doubly labeled water; EI, energy intake; EE, energy

expenditure; FFM, fat-free mass; FM, fat mass; G, Goldberg cutoffs; HO, health outcome; HR, heart rate; IDATA, Interactive Diet and Activity Tracking in American Association of Retired Persons; MSE, mean squared error; NI, nutrition intake; PoI, potassium intake; PrI, protein intake; SI, sodium intake; SBP, resting systolic blood pressure; SR, self-reported; VO2, VO_2 max; WC, waist circumference.

Data sharing: Code book and analytic code will be made publicly and freely available without restriction at http://doi.org/10.5281/zenodo.7013204. Data described in the manuscript are available through the National Cancer Institute per their terms of use.

Author contributions: Conceptualization: KE; Data curation: KE and NY; Formal Analysis: KE, NY, AWB, RZ; Funding acquisition: KE, AWB; Investigation: KE, NY, AWB, RZ; Methodology: KE, NY, AWB, RZ; Project administration: KE; Supervision: KE, AWB, RZ; Validation: KE; Visualization: NY; Writing – original draft: KE and NY; Writing – review & editing: AWB and RZ.

Competing Interests: AWB declares: in the past three years (since 2019-09-19), Dr. Brown has received travel expenses from International Food Information Council; speaking fees from Eastern North American Region of the International Biometric Society, Purchaser Business Group on Health, Purdue University, and University of Arkansas for the Medical Sciences; monetary awards from American Society for Nutrition; consulting fees from LA NORC, Pennington Biomedical Research Center, and Soy Nutrition Institute Global; and grants through his institution from Alliance for Potato Research & Education, American Egg Board, National Cattlemen's Beef Association, NIH/NHLBI, NIH/NIDDK, and NIH/NIGMS. In addition, he has been involved in research for which his institution or colleagues have received grants or contracts from Center for Open Science, Gordon and Betty Moore Foundation, Hass Avocado Board, Indiana CTSI, NIH/NCATS, NIH/NCI, NIH/NHLBI, NIH/NIA, NIH/NIGMS, NIH/NLM, and Sloan Foundation. His wife is employed by Reckitt Benckiser. NY, KE, and RZ declare that they have no competing interests.

Keywords: Nutrition Intake, Self-report, Bias, Nutrition epidemiology, Simulation study

Abstract

Self-reported nutrition intake (NI) data are prone to reporting bias that may induce bias in estimands in nutrition studies; however, they are used anyway due to high feasibility. We examined whether applying Goldberg cutoffs to remove "implausible" self-reported NI could reliably reduce bias compared to biomarkers for energy, sodium, potassium, and protein. Using IDATA data, significant bias in mean NI was removed with Goldberg cutoffs (120 among 303 participants excluded). Associations between NI and outcomes (weight, waist circumference, heart rate, systolic/diastolic blood pressure, and VO2 max) were estimated, but sample size was insufficient to evaluate bias reductions. We therefore simulated data based on IDATA. Significant bias in simulated associations using self-reported NI was reduced but not eliminated by Goldberg cutoffs in 14 of 24 nutrition-outcome pairs; bias was not reduced for remaining cases. 95% coverage probabilities were improved by applying Goldberg cutoffs in most cases but underperformed compared with biomarker data. Although Goldberg cutoffs may achieve bias elimination in estimating mean NI, bias in estimates of associations between NI and outcomes will not necessarily be reduced or eliminated after application of Goldberg cutoffs. Whether one uses Goldberg cutoffs should therefore be decided based on research purposes and not general rules.

Impact Statement: Elimination of extreme reporters using Goldberg cutoffs does not always produce unbiased estimates of associations between nutrition intakes and health outcomes.

INTRODUCTION

Measuring nutrition intake plays a pivotal role in obesity and nutrition studies. However, objectively measuring nutrition intake is not easy. Under natural settings with less restriction, two approaches are often used: self-report and biomarkers. Biomarker approaches are considered more accurate compared with self-report; however, they are typically expensive and require special equipment for each nutrient. Self-report approaches are less expensive and do not require special laboratory equipment or techniques, and can be applied to the measurement of any nutrition components from macronutrients (i.e., carbohydrates, fat, and protein) to micronutrients (i.e., minerals and vitamins); however, self-reported data are known to be fraught with bias (1, 2).

Consider the example of biomarker versus self-report energy intake. Energy intake (EI) is defined as a sum of the total metabolizable energy content of consumed foods and beverages (3). EI can be measured using biomarkers, especially by energy balance approaches in which respiratory gas analysis or doubly labeled water (DLW) are used for the calculation of energy expenditure (EE) (4). EI can be calculated using EE accounting for change in energy stores (i.e., body composition) during the period of measurement. The respiratory gas analysis typically requires the participants to wear a gas mask or stay in a metabolic chamber (5), which restricts participants' behavior; thus, the measured EI using the respiratory gas analysis may not reflect EI from habitual eating behavior. DLW requires the participants to visit the laboratory to drink labeled water enriched with stable hydrogen and oxygen isotopes and collect urine for measurement of isotope levels. It is considered to reflect habitual eating behavior and widely accepted as a standard and reliable approach in measuring EI in free-living individuals (6-8). However, given the high cost of isotopes and laboratories to analyze collected urine, DLW is not practical in large-scale epidemiological surveillance (9).

EI measured using self-report approaches (EI_{SR}) is computed by summing up the energy intake from foods and beverages reportedly consumed during the period of measurement, which can commonly range from a day to a year depending on approaches (e.g., 24-hour recall, food diary, and food frequency

questionnaire) (10-12). Given that self-report approaches are low cost and scale-up relatively easily, EI_{SR} is still widely used in epidemiological studies. However, EI_{SR} is biased compared with the EI measured by biomarker such as DLW (EI_{BIO}) (1) and is discouraged from use to estimate actual EI (2). The mechanism behind the reporting bias in EI_{SR} has been argued elsewhere. For example, EI_{SR} is computed from the information of reported foods and beverages using tables of food and beverage nutrients, composed of weights, calories, and the amount of macro- and micronutrients. Compositions of foods and beverages are different among brands, stores, and seasons (13), and thus nutrient databases cannot be exhaustive. Consequently, EI_{SR} using such databases may not be accurate (14). Further factors, including social desirability (i.e., people do not want to be seen as a "big eater" in general) (15) and weight/social status (15-19), are known to be associated with reporting bias. Thus, the difference between EI_{BIO} and EI_{SR} is due to systemic biases rather than a random error.

Despite concerns about the accuracy of self-report approaches, they have been predominantly used in large scale nutrition epidemiology due to their high feasibility. Approaches to mitigate the bias in EI_{SR} have been argued, including commonly used approaches to exclude the data of EI_{SR} that are considered unreliable. For example, one approach is to exclude those reporting EI_{SR} out of the range considered by the researchers to be plausible, such as from 500 to 3500 kcal/day (9). However, this rule does not consider individual variability and is not appropriate for those whose intake is truly above 3500 kcal/day, such as elite athletes or those whose body size is large and requires more than 3500 kcal/day to maintain their weight (20, 21).

Other approaches are to use EE prediction equations to account for individual variability in EI (22-28). If the EE predicted considering various factors (e.g., age, weight, sex) is far from EI_{SR}, those individuals are considered under/overreporting their actual EI. Indeed, such an approach has evidently reduced misclassification (29). Among proposed approaches, the Goldberg cutoffs (30) have been used in nutrition epidemiology to assess nutritional status in specific populations and investigate the association between nutritional status and health and socioeconomic outcomes (25-28, 31). Note that nutritional status in this context means dietary intakes in general; not only EI but also nutritional constituents (vitamin, fish oil, etc.) and dietary styles (Mediterranean, vegan, etc.) (32-36). Therefore, the rationale of using the Goldberg

cutoffs (and other similar approaches) to exclude data from individuals considered to have unreliable EI_{SR} is based on an assumption, "if total EI is underestimated, it is probable that the intakes of other nutrients are also underestimated" (37). The corollary for overestimation may also be true.

Bias-correcting approaches like the Goldberg cutoffs are frequently used in nutrition studies. By bias, we are referring to the difference between the estimates and their true values. Here we focus on two typical estimands in nutrition epidemiology: the mean of nutrition intake, and the associations between nutrition intake and health outcomes in a population. However, whether the Goldberg cutoffs reduce bias has not been proved theoretically or empirically. In our previous study, we demonstrated that the Goldberg cutoffs do not necessarily eliminate the bias in estimating the associations between EI and various health outcomes (38). Extending our previous study, we consider nutritional intake beyond EI to include biomarkers of sodium, potassium, and protein with the primary goal to examine if the Goldberg cutoffs eliminate or reduce the bias using a single dataset.

Evaluating the performance of the Goldberg cutoffs solely depending on a single empirical dataset is limited because the dataset is a single realization from an unobserved data-generation process. Further, if the dataset is too small, conclusions may be limited because of issues related to power to detect bias in associations or the power to detect reductions in bias. Therefore, in addition to the empirical data analyses above, we also generated data through simulation and analyzed those generated data that preserve the characteristics of the empirical data. The simulation further enables us to assess the impact of sample size on the performance of the Goldberg cutoffs.

RESULTS

1. Statistical data analysis on IDATA

The list of variables, equations, and metrics is available in **Table 1.** Abbreviations are summarized in **Supplemental Table 1**.

The impact of the Goldberg cutoffs on the reporting bias in nutrition intake

Table 2 summarizes the data included in the analyses (n=303). More than 90% of the participants are non-Hispanic whites, and their mean age was 63. The mean BMI was 27.7 kg/m^2, and 30% had obesity (BMI≥30). There was significant mean underreporting in which the biomarker nutrition intakes (NI_{BIO}) was greater than the self-reported nutrition intakes (NI_{SR}) in all four types of nutrition intake (**Table 2**).

By the Goldberg cutoffs, 120 among 303 participants were excluded (40%) (**Table 3**). There was significant underreporting in the rejected cases, whereas reporting bias (NI_{SR} - NI_{BIO}) was not significant in the accepted cases in all four NI. Further, the difference in the biases between the accepted and the rejected cases was significant. The mean NI_{SR} in the accepted cases was significantly larger than that in the rejected cases in all four NI. NI_{BIO} was significantly larger in the rejected cases in sodium intake (SI), whereas it was comparable between those groups in energy intake (EI), potassium intake (PoI), and protein intake (PrI). To further understand the dependency of the bias on NI_{BIO} and the impact of the Goldberg cutoffs on the bias, the individual data were plotted (**Figure 1**). The bias was significantly and negatively correlated with NI_{BIO} (**Figure 1 top panels**). The means of accepted cases of NI_{SR} correspond to the mean of NI_{BIO} better than the mean of NI_{SR} in the rejected cases (**Figure 1 bottom panels**).

These findings suggest that the Goldberg cutoffs reduce (and may eliminate) the bias in mean NI_{SR} by identifying predominantly underreporting individuals. The difference in self-reported nutrition intake between accepted and rejected cases is predominantly from differences in reporting and less so from differences in actual biomarker-measured intake. However, statistically significant differences in biomarker-measured intake remained between accepted and rejected cases in SI.

Associations between nutrition intake and health outcomes

The associations between EI_{BIO}, SI_{BIO}, and PrI_{BIO} and WC were statistically significant (**Supplemental Figures 1 and 2** and **Supplemental Table 2**). PoI_{BIO} was significantly associated with VO2. Assuming that the biomarker-based health outcome associations ($\beta_{NI_{BIO},HO}$) and self-report-based health outcome associations ($\beta_{NI_{SR},HO}$) are different, the associations of health outcomes with self-report after applying the Goldberg rule ($\beta_{NI_G,HO}$) are expected to be between raw self-report and biomarker-based exposures. In other words, the Goldberg cutoffs are supposed to reduce the bias: $\beta_{NI_{BIO},HO} < \beta_{NI_G,HO} < \beta_{NI_{SR},HO}$ or $\beta_{NI_{BIO},HO} > \beta_{NI_G,HO} > \beta_{NI_{SR},HO}$). We first tested whether bias exists for each nutrition-health outcome pair, and, if so, whether the bias is overestimation or underestimation, using percent bias (b). The percent bias (b) was significantly below 0% (underestimation) in only limited cases (EI and BW, and EI and WC) and we could not conclude whether the bias exists or not for the rest of the cases. Significant overestimation ($b > 0$) was not observed in any cases (the left panel in **Figure 2** and **Supplemental Table 2**). Note that significant underestimation from self-report was observed in the cases with significant associations in nutrition intake estimated from biomarkers and health outcomes. Second, we tested whether the bias remains after applying the Goldberg cutoffs, and, if so, whether it is overestimation or underestimation, using the percent remaining bias (r). The bias was reduced ($|r| < 100$) but remained ($r \neq 0$) for the association between EI and BW, and between EI and WC. We could not again conclude whether the bias exists or not after applying the Goldberg cutoffs for the rest of the cases.

From the above data analyses, we could not conclude that the Goldberg cutoffs do or do not remove the bias in most cases, because of the wide confidence intervals, which could be partially due to the sample size of the data and weak associations between some combinations of NI and HO.

2. Simulation

To strengthen the arguments on the performance of the Goldberg cutoffs, we also generated and analyzed data which preserve the characteristics of the IDATA. The simulation was designed to overcome the sample size limitations of IDATA, but keep the structure and nature of the IDATA.

Performance metrics (bias, MSE, and coverage probability) for all the combinations between NI and HO are shown in **Figure 3**. For each performance metric, the mean across the 1,000 simulation replicates is shown. We confirmed the generated biomarker data were consistent with the IDATA (**Supplemental Figure 3**). The bias for all biomarker simulations was close to 0 (red squares in **Figure 3**). For all four NI, the associations with HR were negative and those with the other outcomes were positive (**Supplemental Figures 1 and 2**). For the cases with positive correlation, the positive bias in $\beta_{NI_{SR},HO}$ was observed, and for the cases with negative correlation, the negative bias in $\beta_{NI_{SR},HO}$ was observed, suggesting that the self-reported data cause attenuation bias (i.e., $|\beta_{NI_{SR},HO}| < |\beta_{NI_{BIO},HO}|$) as observed in many studies (51). After application of the Goldberg cutoffs, the bias was reduced in 14 among 24 combinations of NI and HO. Notably, the bias was reduced in all six outcomes and reduction was large when EI is used as a predictor. When SI, PoI, and PrI are used as predictors, the bias was not much changed by applying the Goldberg cutoffs. In most cases, the MSE was also smaller for NI_{BIO} compared with NI_{SR}, and the MSE for NI_{SR} became smaller after applying the Goldberg cutoffs (**Figure 3**). The coverage probabilities (cp) were close to 95% when NI_{BIO} was used as a predictor, as was expected, whereas it was mostly lower than 95% when NI_{SR} was used. The coverage probability increased by applying Goldberg cutoffs; however, it is still lower than 95% in most cases. We repeated the simulation for other values ($n = 50, 100, 200,$ and 300) of sample size and all cases showed similar results (**Supplemental Figure 4**). We ran sensitivity analyses varying η, an error term for the regression between NI and HO, and σ, the standard deviation of the reporting bias (**Supplemental Figure 5**). Biases were not affected by varying η for the range of values we considered, suggesting that error in health outcomes (or variance in health outcomes which cannot be explained by variance in biomarker-based nutrition intake) does not influence NI-HO association estimations where NI is used as an independent variable. Bias was negative when self-reported data (NI_{SR} or NI_G) were used under the default value of σ, and the magnitude of bias increased as σ increased (i.e., the association was

attenuated). However, when the error is extremely small (e.g., σ=0.25 in **Supplemental Figure 5**), the bias becomes positive, which is because the mean error function (μ) was negative at low NI and positive at high NI, thus the association was intensified. Whether or not the bias becomes positive or negative is determined by the balance between the mean error function and random error.

DISCUSSION

Dietary self-reporting is a basic component of nutrition epidemiology, particularly for studies of nutrition-health relationships. However, reporting bias in nutrition intake is perhaps one of the greatest impediments to understanding the true effect of nutrition on disease. Failure to account for this bias in self-reports can affect the analysis and interpretation of studies designed to assess the influences of nutrition on health. Correlation or regression coefficients estimating associations between self-reported nutrition intakes and health outcomes can be subject to substantial error (52, 53). To address such reporting bias issues, there has been increased interest in the use of statistical models in conjunction with biomarker data (54, 55).

In this study, reporting bias in self-reported nutrition intake was demonstrated using both empirical data analysis and a simulation study. We first examined whether the bias in mean nutrition intake and the bias in the associations between nutrition intake and health outcomes are reduced by applying the Goldberg cutoffs to an empirical dataset, IDATA. We confirmed the bias in mean self-reported nutrition intake was eliminated by applying Goldberg cutoffs to exclude extreme reporters (mostly under-reporters). This bias can be mostly explained by differences in reporting between the accepted and rejected cases rather than a difference in actual nutrition intake between the accepted and rejected cases. We further tested whether the associations between health outcomes and nutrition intake are biased when self-reported data are used, and whether the bias is reduced by applying the Goldberg cutoffs. The significant bias for the associations between EI and BW and the association between EI and WC were reduced but remained after applying Goldberg cutoffs. We did not observe significant bias in associations for other nutrition-outcome pairs, and

therefore cannot make conclusions regarding the utility of applying the Goldberg cutoffs to reduce bias for these associations in the empirical data.

To overcome the limitations of the data (i.e., sample size and weak associations among many nutrition-outcome pairs), we generated and analyzed data preserving the characteristics of the IDATA for the assessment of the Goldberg cutoffs. We confirmed the existence of bias in estimated associations when the self-reported data were used, and these biases were reduced, but not eliminated, by applying the Goldberg cutoffs in general. We also confirmed large MSE and low coverage probability when self-reported data were used; MSE and low coverage probability were improved by using Goldberg cutoffs, but were not recovered to the levels of biomarker data. Overall, the empirical data analyses and the simulation study suggest that the Goldberg cutoffs improve the estimates in nutrition-health outcome associations; however, the biases were not completely removed.

This study has several strengths. First, although energy intake has traditionally been studied more with respect to Goldberg cutoffs, herein we also investigated three other nutrition intakes: sodium, potassium, and protein. In our previous study, we used only energy intake as a predictor (38). However, given that the Goldberg cutoffs are believed to reduce the bias in nutrition intake broadly, we expanded the approach to these other nutrition intakes, which can be measured by biomarkers (37). Second, we conducted the simulation in addition to empirical data analyses. There are two advantages of the simulation design. 1) The data generation process was based on the IDATA dataset. Thus, we were able to simulate datasets reflective of those that would be seen in the real world. Both the sample size of the generated dataset and the variables that we considered are reflective of what is often observed in nutrition studies. 2) Simulation can consider different models for the data-generating process. Although our objective was not to disentangle the underlying data generation mechanisms, we were able to generate data that preserve the characteristics of the IDATA.

The limitations of this study are worth noting. In the analysis of IDATA, we used only bivariate associations for nutrition-health outcomes. It is possible that including various covariates or different models

may influence the existence or lack of bias in nutrition-health associations. In the data-generation process for the simulations, assumptions were made on parameter settings. The reporting bias was set to be normally distributed. Modeling the bias using other distributions (such as a heavier tailed distribution that may be more consistent with a zero-bounded intake) may improve model fit, and thus generated more realistic data. We simulated the bias with constant variance, consistent with the apparent stability of variance across the intake quantiles in **Supplemental Figure 6** and **Supplemental Table 4**; however, modeling the variance as a function of intake quantile may alter the data-generating process. We used the self-reported nutrition intake reported by ASA24, as it reflects recent dietary intake. The bias we observed in this study may be different for other self-report methods. Finally, the simulation approach, although based on the IDATA, was nonetheless generated from a limited sample size. We are not aware of thresholds for sufficient sample size for empirical-data-based simulation, and we have had success with smaller plasmode based sampling in the past(56, 57); however, it is possible a larger, more representative, or more robust empirical data set may influence both the empirical and simulated results. Further as the data was from a single population in which white and relatively older people are predominant, generalizability of the study finding should be carefully considered.

In conclusion, we found that the Goldberg cutoffs reduce and eliminate the bias in mean nutrition intake in the IDATA dataset; however, the bias in the associations between nutrition intake and health outcomes, when they exist, can sometimes be reduced but not eliminated by the Goldberg cutoffs in both empirical and simulated results. Investigators considering excluding extreme reporters in epidemiological studies should consider whether such approaches are likely to be helpful in answering their specific research questions, as our study has shown that such cutoffs are more useful in estimating mean nutrition intakes than in studies of associations between nutrition intake and health outcomes.

MATERIALS AND METHODS

We conducted a set of empirical analyses to assess the reporting bias due to self-report and examine whether the Goldberg cutoffs reduce reporting bias using the data from The Interactive Diet and Activity Tracking in the American Association of Retired Persons (IDATA) study (39). The reporting bias in the mean of nutrition intakes and the associations (regression coefficients) between nutrition intakes and health outcomes were assessed. We also conducted a simulation study to investigate the performance of Goldberg cutoffs using the results from the empirical analyses as ground truth. Note that in data analyses, we assumed the estimates from biomarkers are the true values, since the true values (i.e., population parameters) are unobservable.

Among various nutrition intakes (NI), we considered sodium intake (SI), potassium intake (PoI), and protein intake (PrI) in addition to energy intake (EI). These NI were selected because they can be measured by urine samples as well as self-reports. We denote self-reported NI by NI_{SR}: EI_{SR}, SI_{SR}, PoI_{SR}, and PrI_{SR}. We denote biomarker-based NI by NI_{BIO}: EI_{BIO}, SI_{BIO}, PoI_{BIO}, and PrI_{BIO}. The NI_{SR} data remaining after applying the Goldberg cutoffs (hereafter, we call those 'accepted cases') are denoted by NI_G: EI_G, SI_G, PoI_G, and PrI_G. Health outcomes (HO) that may be associated with NI were selected from the data dictionaries; body weight (BW), waist circumference (WC), heart rate after fitness test (HR), resting systolic blood pressure (SBP), resting diastolic blood pressure (DBP), and VO_2 max (VO2). The estimated regression coefficients are denoted by the subscript of the explanatory variables and the health outcomes (e.g., $\beta_{EI_{SR},BW}$ represents the association between self-reported EI and body weight).

Goldberg cutoffs

The Goldberg cutoffs are an approach to identify those who report low or high energy intake that, if sustained across long periods of time, would be implausible to sustain life (for underreporting) or present weight (for overreporting), by comparing the reported EI and the energy expenditure (EE) predicted from body composition and physical activity level (30). First, EE is predicted as a product of physical activity

level (PAL) and basal metabolic rate (BMR): EE=PAL*BMR. PAL is assumed to be 1.75 (40). BMR is predicted using fat-free mass (41): BMR = 370 + 21.6 * FFM. Second, the ratio of EI_{SR} and predicted EE is computed and examined whether it is in a reasonable range: $e^{-2*\frac{S}{100}}<EI_{SR}: EE < e^{2*\frac{S}{100}}$, where S is defined as: $\sqrt{\frac{CV_{wEI}^2}{d} + CV_{wB}^2 + CV_{tP}^2}$. CV_{wEI}, CV_{wB}, and CV_{tP} are the within-subject variation in EI_{SR}, the within-subject variation in BMR, and the total variation in PAL, and d is the number of days of diet assessment. We set $CV_{wEI} = 23\ (\%), CV_{wB} = 8.5\ (\%), CV_{tP} = 15\ (\%)$, and $d = 7$ following the previous study (42). Finally, if individuals are not in the range, they were considered under- or overreporters and were excluded from further analyses.

Data

Overview of IDATA

The IDATA study was designed "to evaluate and compare the measurement error structure of diet and physical activity assessment tools against reference biomarkers." (39) The study participants "were recruited from a list of AARP [American Association of Retired Persons] members aged 50-74 years residing in and around Pittsburgh, Pennsylvania" (39) and were screened for eligibility by phone interview or clinic visit. The eligibility criteria were speaking English, having internet access, having no major medical issues (such as diabetes, renal/heart failure, and any conditions affecting fluid balance), having no mobility issues, and having BMI between 18.5 and 40.0 (kg/m^2). The participants "visited the study center three times and also completed assigned activities at home over a 12-month period" (39) from early 2012 through late 2013 (39, 43). Self-reported dietary/physical activity information, biomarker data to estimate dietary intake, and the data from physical objective monitors to estimate physical activity were collected. In addition, demographic information (age, sex, race/ethnicity) and anthropometric measurements (weight, height, and waist circumference) were obtained at a clinical visit or phone screening.

Self-reported nutrition intake: NI_{SR}

For self-report dietary assessment (EI_{SR}, SI_{SR}, PoI_{SR}, and PrI_{SR}), four different approaches were used in IDATA: Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24) (44), Diet History Questionnaire (DHQ-II), 7-day food checklist, and 4-day food record. These SR approaches were conducted at different times during the IDATA study. We used the data of ASA24 in our analyses because they were collected closest in time to the DLW measurements. ASA24 is an online self-administered recall system asking about a 24-hour dietary recall for the previous day, from midnight to midnight, using a dynamic user interface.

Biomarker-based nutrition intake: NI_{BIO}

Three approaches were used for biomarker assessment of NI: DLW, urine, and saliva (not used herein). EI_{BIO} was computed from DLW data, and SI_{BIO}, PoI_{BIO}, and PrI_{BIO} were calculated from 24-hour urine samples. EI_{BIO} was estimated as a sum of daily total energy expenditure estimated from DLW and daily change in energy stores during the DLW period. Total energy expenditure was calculated using the approach proposed by Schoeller et al. (6), where the respiratory quotient was assumed as 0.86. Daily change in energy stores during the DLW period (about 2 weeks) was computed from the average daily weight change during the period assuming the energy density of body weight was 2380 kcal/kg (45). The urinary values obtained from 24-hour urine samples were converted into nutrition intakes assuming that 81%, 80%, and 86% of consumed nitrogen, potassium, and sodium are excreted in the urine, respectively. Dietary protein was calculated assuming that 16% of protein is nitrogen. A more comprehensive explanation of measuring EI_{BIO}, SI_{BIO}, PoI_{BIO}, and PrI_{BIO} is reported elsewhere (46).

Health outcomes

From health outcomes (HO) available in IDATA, we selected objective variables that were potentially associated with nutrition intake. BW (kg) and WC (cm) were obtained at clinical visits; HR (beat/min after fitness test), SBP (mmHg), DBP (mmHg), and VO_2 max (L/min) were obtained by the modified Canadian Aerobic Fitness Test (mCAFT) (47).

Summary of the data

In IDATA, NI_{SR}, NI_{BIO}, and HO were repeatedly measured at different time points. For fair comparison among them, we selected the data measured in the same month; thus, the data of Month 0 and Month 11 were used from Groups 1&3 and Groups 2&4, respectively (39). From 1,082 participants, we excluded 779 participants because we did not have information necessary to apply the Goldberg cutoffs (no fat-free mass information [n=725] or NI_{SR} [n=240]) or there was no urine-based nutrition intake data [n=231]. The remaining 303 participants' data were analyzed.

IDATA data were accessible through the Cancer Data Access System (https://biometry.nci.nih.gov/cdas/idata/; downloaded 11/22/2017) after our project proposal was reviewed and approved by the National Cancer Institute (https://biometry.nci.nih.gov/cdas/approved-projects/1702/).

Statistical analysis of IDATA

Whether the bias in the mean NI_{SR} compared to NI_{BIO} (e.g., calculated as EI_{SR} - EI_{BIO}) would be reduced by the Goldberg cutoffs was tested. We first tested whether the mean bias in each NI_{SR} is significantly different from zero. Then, we repeated the same test for the data of the accepted cases for each NI_G (e.g., EI_G - EI_{BIO}). We further tested whether the mean NI_{SR} are different between those who were removed (rejected cases; NI_{SR} excluding NI_G) and the accepted cases (NI_G).

Further, we tested whether the bias in the estimates of associations between NI and HO would be reduced by the Goldberg cutoffs. We computed regression coefficients ($\beta_{NI,HO}$) between HO and each of the three different NI measuring approaches: NI_{BIO}, NI_{SR}, and NI_G. Note that we did not adjust the analyses for any other covariates, because the purpose of this study is to understand if the Goldberg cutoffs could reliably reduce bias rather than to refine associations accounting for covariates. To comprehensively assess the reporting bias and bias reduction in different outcomes with different units, we used standardized metrics for reporting bias and bias reduction, which was originally proposed in our previous study (38). In brief, we used the percent bias (b) for the assessment of the magnitude of the bias and the percent remaining bias (r) for the assessment of the magnitude of the bias remaining after applying the Goldberg cutoffs, which are defined as follows:

$$b = \frac{\beta_{NI_{SR},HO} - \beta_{NI_{BIO},HO}}{\beta_{NI_{BIO},HO}} * 100\ (\%)$$

$$r = \frac{\beta_{NI_{G},HO} - \beta_{NI_{BIO},HO}}{\beta_{NI_{BIO},HO}} * 100\ (\%)$$

A more in-depth interpretation of the metric is available in our previous study (38). Jackknife estimation (leave-one-out) (48) was used to compute 95% CIs of those metrics. All analyses were performed separately for each combination of nutrition intakes and outcomes. Two-tailed Student's *t* test was used to test whether mean of a single variable is different from zero. Two-tailed independent Welch's *t* test was used to test the mean difference of a single variable from two independent groups. The jackknife method was used to test the mean difference of a single variable from two non-independent groups (i.e., whole data vs the data of accepted cases). The type I error rate was fixed at 0.05 (2-tailed).

Simulation

To strengthen the arguments on the performance of the Goldberg cutoffs, we also generated and analyzed data that preserve the characteristics of the IDATA.

Data generation

We generated data composed of four variables: NI_{BIO} (biomarker), NI_{SR} (self-reported), HO (health outcomes: BW, WC, HR, SBP, DBP, and VO2), and FFM (fat-free mass). NI_{BIO} and FFM were resampled from the IDATA. HO is computed using a linear model using NI_{BIO} as a predictive variable:

$$HO = a_0 + a_1 NI_{BIO} + \epsilon, \epsilon \sim N(0, \eta^2),$$

where a_0, a_1, and η^2 are the model parameters estimated from the IDATA. NI_{SR} is generated by adding an error term (i.e., reporting error), e, to NI_{BIO}, adapting the approach proposed by Ward et al. (49):

$$NI_{SR} = NI_{BIO} + e,$$

where e is determined by the percentile of NI_{BIO} in the empirical distribution, $p(NI_{BIO})$. The reporting error e is assumed to follow a normal distribution parametrized by percentile-specific mean, $\mu(p(NI_{BIO}))$, and a

394 constant standard deviation, σ: $e \sim N(\mu, \sigma^2)$. μ and σ were estimated by fitting polynomial models to the
395 IDATA: $\mu\left(p(NI_{BIO})\right) = \sum_{k=0}^{K} m_k p(NI_{BIO})^k$, where K is the order of the polynomial. The order of the
396 polynomial function was varied from 1 to 5 and we selected the value of K that yields the smallest MSE .
397 We used k-fold cross-validation with 10 folds to calculate the test MSE for each model. This resulted in
398 $K = 1, 3, 3, 5$ for EI, SI, PoI, and PrI, respectively (**Supplemental Table 3** and **Supplemental Figure 6**).
399 For example, assuming EI_{BIO} is 3000 kcal, which corresponds to 90.4 percentile, the mean of the reporting
400 error, μ, was -749.6 and the standard deviation of the reporting error, σ, was 759.1. Therefore his/her e is
401 randomly sampled from the normal distribution: $N(-749.6, 759.1)$. We further confirmed our assumption of
402 homoskedastic σ is reasonable by the Goldfeld Quandt test(50) (**Supplemental Table 4**). The Goldfeld
403 Quandt test compares variance in high and low values of a variable; we discarded the middle 20% of the
404 total observations (eliminated the center 61 observations among 303 in total) to define the high and low
405 groups. The simulated data were truncated so that $NI_{SR}>0$. The variable NI_G was generated by applying the
406 Goldberg cutoffs to EI_{SR} using FFM (thus, NI_G is a subset of NI_{SR}). The above process was repeated for each
407 NI. We generated $n = 100$ individuals, and the data generation was repeated 1000 times, leading to 1000
408 replicates. The data generating process is summarized in **Figure 4**.

409 **Measurement of the performance of the Goldberg cutoffs**

410 The point estimates of the regression coefficients for an NI and an HO from the ith simulation are
411 denoted as $\hat{\beta}_{NI,HO,i}$ (e.g., $\hat{\beta}_{EI_{SR},BW,i}$ is the ith regression coefficient for self-reported energy intake and body
412 weight). The performance of the Goldberg cutoffs was assessed by three metrics: bias, mean squared error
413 (MSE), and coverage probability:

$$Bias_{NI,HO} = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{NI,HO,i}\right) - \beta_1,$$

$$MSE_{NI,HO} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_{NI,HO,i} - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{NI,HO,i}\right)^2 + \left(\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{NI,HO,i}\right) - \beta_1\right)^2,$$

$$P_{NI,HO} = \frac{\sum_{i=1}^{n} I\left[\left(\hat{\beta}_{NI,HO,i,low} < \beta_1\right) \cap \left(\beta_1 < \hat{\beta}_{NI,HO,i,high}\right)\right]}{n},$$

where $NI = \{EI_{SR}, EI_{BIO}, EI_G, SI_{SR}, SI_{BIO}, SI_G, PoI_{SR}, PoI_{BIO}, PoI_G, PrI_{SR}, PrI_{BIO}, PrI_G\}$, $HO = \{BW, WC, HR, SBP, DBP, VO2\}$, $I(\cdot)$ is an indicator function, low and $high$ represent the lower and upper 95% confidence intervals, and β_1 is the true parameter value, which we define as the point estimate from the IDATA. Note that bias is used differently here than in the empirical analysis because the parameters remain unknown in the empirical analysis whereas we define the parameters in the simulation.

Sensitivity analyses were performed varying the sample size and two other parameters: error term for the regression between health outcome and biomarker-based nutrition intakes ($\boldsymbol{\eta}$), and the standard deviation of the reporting error ($\boldsymbol{\sigma}$). All simulations and analyses were performed using the statistical computing software R (version 4.0.1); code book and analytic code are publicly and freely available without restriction at http://doi.org/10.5281/zenodo.7013204. The variables, data generation models, and metrics of performance of the Goldberg cutoffs are summarized in **Table 1**. STROBE guideline was followed, and the statement is available as supplementary material.

References

1. Schoeller DA, Bandini LG, Dietz WH. Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method. Can J Physiol Pharmacol 1990;68(7):941-9. doi: 10.1139/y90-143.
2. Dhurandhar NV, Schoeller D, Brown AW, Heymsfield SB, Thomas D, Sorensen TI, Speakman JR, Jeansonne M, Allison DB. Energy balance measurement: when something is not better than nothing. Int J Obes 2015;39(7):1109-13. doi: 10.1038/ijo.2014.199.
3. Astrup A. 3 - Energy metabolism and obesity. Edtion ed. In: Mela DJ, ed. Food, Diet and Obesity: Woodhead Publishing, 2005:58-75.
4. Ferrannini E. The theoretical bases of indirect calorimetry: a review. Metabolism 1988;37(3):287-301.
5. Leonard WR. Laboratory and field methods for measuring human energy expenditure. Am J Hum Biol 2012;24(3):372-84. doi: https://doi.org/10.1002/ajhb.22260.

6. Schoeller DA, Ravussin E, Schutz Y, Acheson KJ, Baertschi P, Jequier E. Energy expenditure by doubly labeled water: validation in humans and proposed calculation. Am J Physiol 1986;250(5 Pt 2):R823-30. doi: 10.1152/ajpregu.1986.250.5.R823.
7. Speakman J. Doubly labelled water: theory and practice: Springer Science & Business Media, 1997.
8. Schoeller DA, van Santen E. Measurement of energy expenditure in humans by doubly labeled water method. J Appl Physiol Respir Environ Exerc Physiol 1982;53(4):955-9. doi: 10.1152/jappl.1982.53.4.955.
9. Banna JC, McCrory MA, Fialkowski MK, Boushey C. Examining Plausibility of Self-Reported Energy Intake Data: Considerations for Method Selection. Front Nutr 2017;4(45):45. doi: 10.3389/fnut.2017.00045.
10. Leung DH, Heltshe SL, Borowitz D, Gelfond D, Kloster M, Heubi JE, Stalvey M, Ramsey BW, Baby O, Nutrition Study Investigators of the Cystic Fibrosis Foundation Therapeutics Development N. Effects of Diagnosis by Newborn Screening for Cystic Fibrosis on Weight and Length in the First Year of Life. JAMA Pediatr 2017;171(6):546-54. doi: 10.1001/jamapediatrics.2017.0206.
11. Sugiyama T, Tsugawa Y, Tseng C, Kobayashi Y, Shapiro MF. Different time trends of caloric and fat intake between statin users and nonusers among us adults: Gluttony in the time of statins? JAMA Internal Medicine 2014;174(7):1038-45. doi: 10.1001/jamainternmed.2014.1927.
12. Oftedal S, Davies PS, Boyd RN, Stevenson RD, Ware RS, Keawutan P, Benfer KA, Bell KL. Body composition, diet, and physical activity: a longitudinal cohort study in preschoolers with cerebral palsy. Am J Clin Nutr 2017;105(2):369-78. doi: 10.3945/ajcn.116.137810.
13. Martelli F, Giacomozzi C, Fadda A, Frazzoli C. Understanding Seasonal Changes to Improve Good Practices in Livestock Management. Frontiers in Public Health 2018;6. doi: 10.3389/fpubh.2018.00175.
14. McHenry EW, Ferguson HP, Gurland J. Sources of Error in Dietary Surveys. Canadian Journal of Public Health / Revue Canadienne de Sante'e Publique 1945;36(9):355-61.
15. Heymsfield SB, Darby PC, Muhlheim LS, Gallagher D, Wolper C, Allison DB. The calorie: myth, measurement, and reality. Am J Clin Nutr 1995;62(5 Suppl):1034S-41S. doi: 10.1093/ajcn/62.5.1034S.
16. Johansson L, Solvoll K, Bjorneboe GE, Drevon CA. Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. Am J Clin Nutr 1998;68(2):266-74. doi: 10.1093/ajcn/68.2.266.
17. Macdiarmid J, Blundell J. Assessing dietary intake: Who, what and why of under-reporting. Nutr Res Rev 1998;11(2):231-53. doi: 10.1079/NRR19980017.
18. Johansson G, Wikman A, Ahren AM, Hallmans G, Johansson I. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. Public Health Nutr 2001;4(4):919-27. doi: https://doi.org/10.1079/PHN2001124.
19. Klesges RC, Eck LH, Ray JW. Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition Examination Survey. J Consult Clin Psychol 1995;63(3):438-44. doi: http://dx.doi.org/10.1037/0022-006X.63.3.438.
20. Santos DA, Silva AM, Matias CN, Magalhaes JP, Fields DA, Minderico CS, Ekelund U, Sardinha LB. Validity of a combined heart rate and motion sensor for the measurement of free-living energy expenditure in very active individuals. J Sci Med Sport 2014;17(4):387-93. doi: 10.1016/j.jsams.2013.09.006.
21. DeLany JP, Kelley DE, Hames KC, Jakicic JM, Goodpaster BH. High energy expenditure masks low physical activity in obesity. Int J Obes 2013;37(7):1006-11. doi: 10.1038/ijo.2012.172.
22. McCrory MA, McCrory MA, Hajduk CL, Roberts SB. Procedures for screening out inaccurate reports of dietary energy intake. Public Health Nutr 2002;5(6A):873-82. doi: 10.1079/PHN2002387.
23. Huang TTK, Roberts SB, Howarth NC, McCrory MA. Effect of screening out implausible energy intake reports on relationships between diet and BMI. Obes Res 2005;13(7):1205-17. doi: DOI 10.1038/oby.2005.143.

24. Vinken AG, Bathalon GP, Sawaya AL, Dallal GE, Tucker KL, Roberts SB. Equations for predicting the energy requirements of healthy adults aged 18–81 y. Am J Clin Nutr 1999;69(5):920-6.
25. Livingstone MB, Robson PJ. Measurement of dietary intake in children. Proc Nutr Soc 2000;59(2):279-93.
26. Xie B, Gilliland FD, Li YF, Rockett HR. Effects of ethnicity, family income, and education on dietary intake among adolescents. Prev Med 2003;36(1):30-40.
27. Briefel RR, McDowell MA, Alaimo K, Caughman CR, Bischof AL, Carroll MD, Johnson CL. Total energy intake of the US population: the third National Health and Nutrition Examination Survey, 1988-1991. Am J Clin Nutr 1995;62(5 Suppl):1072S-80S. doi: 10.1093/ajcn/62.5.1072S.
28. Huang TT, Howarth NC, Lin BH, Roberts SB, McCrory MA. Energy intake and meal portions: associations with BMI percentile in U.S. children. Obes Res 2004;12(11):1875-85. doi: 10.1038/oby.2004.233.
29. Tooze JA, Krebs-Smith SM, Troiano RP, Subar AF. The accuracy of the Goldberg method for classifying misreporters of energy intake on a food frequency questionnaire and 24-h recalls: comparison with doubly labeled water. Eur J Clin Nutr 2012;66(5):569-76. doi: 10.1038/ejcn.2011.198.
30. Goldberg GR, Black AE, Jebb SA, Cole TJ, Murgatroyd PR, Coward WA, Prentice AM. Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. Eur J Clin Nutr 1991;45(12):569-81.
31. Martin JC, Moran LJ, Teede HJ, Ranasinha S, Lombard CB, Harrison CL. Diet Quality in a Weight Gain Prevention Trial of Reproductive Aged Women: A Secondary Analysis of a Cluster Randomized Controlled Trial. Nutrients 2018;11(1). doi: 10.3390/nu11010049.
32. Grimes CA, Riddell LJ, Campbell KJ, Nowson CA. Dietary salt intake, sugar-sweetened beverage consumption, and obesity risk. Pediatrics 2013;131(1):14-21. doi: 10.1542/peds.2012-1628.
33. Molina-Montes E, Sanchez MJ, Buckland G, Bueno-de-Mesquita HB, Weiderpass E, Amiano P, Wark PA, Kuhn T, Katzke V, Huerta JM, et al. Mediterranean diet and risk of pancreatic cancer in the European Prospective Investigation into Cancer and Nutrition cohort. Br J Cancer 2017;116(6):811-20. doi: 10.1038/bjc.2017.14.
34. New SA, Robins SP, Campbell MK, Martin JC, Garton MJ, Bolton-Smith C, Grubb DA, Lee SJ, Reid DM. Dietary influences on bone mass and bone metabolism: further evidence of a positive link between fruit and vegetable consumption and bone health? Am J Clin Nutr 2000;71(1):142-51. doi: 10.1093/ajcn/71.1.142.
35. Shaheen SO, Sterne JA, Thompson RL, Songhurst CE, Margetts BM, Burney PG. Dietary antioxidants and asthma in adults: population-based case-control study. Am J Respir Crit Care Med 2001;164(10 Pt 1):1823-8. doi: 10.1164/ajrccm.164.10.2104061.
36. O'Sullivan A, Gibney MJ, Brennan L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. Am J Clin Nutr 2011;93(2):314-21. doi: 10.3945/ajcn.110.000950.
37. Livingstone MB, Black AE. Markers of the validity of reported energy intake. J Nutr 2003;133 Suppl 3(3):895S-920S. doi: 10.1093/jn/133.3.895S.
38. Ejima K, Brown AW, Schoeller DA, Heymsfield SB, Nelson EJ, Allison DB. Does exclusion of extreme reporters of energy intake (the "Goldberg cutoffs") reliably reduce or eliminate bias in nutrition studies? Analysis with illustrative associations of energy intake with health outcomes. Am J Clin Nutr 2019;110(5):1231-9. doi: 10.1093/ajcn/nqz198.
39. National Cancer Institute. Cancer Data Access System. Internet: https://biometry.nci.nih.gov/cdas/idata/ (accessed 11 December 2018).
40. Dugas LR, Harders R, Merrill S, Ebersole K, Shoham DA, Rush EC, Assah FK, Forrester T, Durazo-Arvizu RA, Luke A. Energy expenditure in adults living in developing compared with industrialized countries: a meta-analysis of doubly labeled water studies. Am J Clin Nutr 2011;93(2):427-41. doi: 10.3945/ajcn.110.007278.
41. Cunningham JJ. Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation. Am J Clin Nutr 1991;54(6):963-9. doi: 10.1093/ajcn/54.6.963.

42. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. Int J Obes 2000;24:1119. doi: 10.1038/sj.ijo.0801376.
43. Matthews CE, Kozey Keadle S, Moore SC, Schoeller DS, Carroll RJ, Troiano RP, Sampson JN. Measurement of Active and Sedentary Behavior in Context of Large Epidemiologic Studies. Med Sci Sports Exerc 2018;50(2):266-76. doi: 10.1249/MSS.0000000000001428.
44. Nordestgaard LT, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Body Mass Index and Risk of Alzheimer's Disease: A Mendelian Randomization Study of 399,536 Individuals. The Journal of Clinical Endocrinology and Metabolism 2017;102(7):2310-20. doi: 10.1210/jc.2017-00195.
45. Bhutani S, Kahn E, Tasali E, Schoeller DA. Composition of two-week change in body weight under unrestricted free-living conditions. Physiol Rep 2017;5(13). doi: 10.14814/phy2.13336.
46. Park Y, Dodd KW, Kipnis V, Thompson FE, Potischman N, Schoeller DA, Baer DJ, Midthune D, Troiano RP, Bowles H, et al. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018;107(1):80-93. doi: 10.1093/ajcn/nqx002.
47. Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. Proc Natl Acad Sci U S A 2018;115(11):2571-7. doi: 10.1073/pnas.1708282114.
48. Efron B, Stein C. The Jackknife Estimate of Variance. Ann Stat 1981;9(3):586-96. doi: DOI 10.1214/aos/1176345462.
49. Ward ZJ, Bleich SN, Cradock AL, Barrett JL, Giles CM, Flax C, Long MW, Gortmaker SL. Projected U.S. State-Level Prevalence of Adult Obesity and Severe Obesity. N Engl J Med 2019;381(25):2440-50. doi: 10.1056/NEJMsa1909301.
50. Goldfeld SM, Quandt RE. Some Tests for Homoscedasticity. Journal of the American Statistical Association 1965;60(310):539-47. doi: 10.1080/01621459.1965.10480811.
51. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. Structure of Dietary Measurement Error: Results of the OPEN Biomarker Study. Am J Epidemiol 2003;158(1):14-21. doi: 10.1093/aje/kwg091 %J American Journal of Epidemiology.
52. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. J Natl Cancer Inst 2011;103(14):1086-92. doi: 10.1093/jnci/djr189.
53. Gibson RS, Charrondiere UR, Bell W. Measurement Errors in Dietary Assessment Using Self-Reported 24-Hour Recalls in Low-Income Countries and Strategies for Their Prevention. Adv Nutr 2017;8(6):980-91. doi: 10.3945/an.117.016980.
54. Freedman LS, Kipnis V, Schatzkin A, Tasevska N, Potischman N. Can we use biomarkers in combination with self-reports to strengthen the analysis of nutritional epidemiologic studies? Epidemiol Perspect Innov 2010;7(1):2. doi: 10.1186/1742-5573-7-2.
55. Brown AW, Aslibekyan S, Bier D, Ferreira da Silva R, Hoover A, Klurfeld DM, Loken E, Mayo-Wilson E, Menachemi N, Pavela G, et al. Toward more rigorous and informative nutritional epidemiology: The rational space between dismissal and defense of the status quo. Crit Rev Food Sci Nutr 2021:1-18. doi: 10.1080/10408398.2021.1985427.
56. Ejima K, Brown AW, Smith DL, Jr., Beyaztas U, Allison DB. Murine genetic models of obesity: type I error rates and the power of commonly used analyses as assessed by plasmode-based simulation. Int J Obes (Lond) 2020;44(6):1440-9. doi: 10.1038/s41366-020-0554-2.
57. Alfaras I, Ejima K, Vieira Ligo Teixeira C, Di Germanio C, Mitchell SJ, Hamilton S, Ferrucci L, Price NL, Allison DB, Bernier M, et al. Empirical versus theoretical power and type I error (false-positive) rates estimated from real murine aging research data. Cell Rep 2021;36(7):109560. doi: https://doi.org/10.1016/j.celrep.2021.109560.

Tables

Table 1. Variables, data generation models, and metrics of performance of the Goldberg cutoffs

Variables and parameters	Description
NI_{SR}	Self-reported nutrition intake, where NI can be energy intake (EI), sodium intake (SI), potassium intake (PoI), or protein intake (PrI)
NI_{BIO}	Biomarker nutrition intake
NI_G	Self-reported nutrition intake accepted by the Goldberg cutoffs
HO	Health outcomes, where HO can be body weight (BW), waist circumference (WC), heart rate after fitness test (HR), resting systolic blood pressure (SBP), resting diastolic blood pressure (DBP), and VO2 max (VO2)
$\beta_{NI,HO}$	Estimated coefficients in the regression model using NI and HO as a dependent and an independent variables.
β_1	True coefficients, the point estimate from the IDATA (used in simulation)
Data generation models	**Description**
$HO = a_0 + a_1 NI_{BIO} + \epsilon$	ϵ is the error of health outcomes $\sim N(0, \eta^2)$
$NI_{SR} = NI_{BIO} + e$	e is the reporting error $\sim N(\mu, \sigma^2)$
$\mu\big(p(NI_{BIO})\big) = \sum_{k=0}^{K} m_k p(NI_{BIO})^k$	μ is the mean error function, determined by a polynomial function of percentiles of NI_{BIO}
Metrics of performance of the Goldberg cutoffs (data analysis)	**Description**
$b = \frac{\beta_{NI_{SR},HO} - \beta_{NI_{BIO},HO}}{\beta_{NI_{BIO},HO}} * 100$	Percent bias (%) in the estimation of NI-HO association when using $\beta_{NI_{SR},HO}$.
$r = \frac{\beta_{NI_G,HO} - \beta_{NI_{BIO},HO}}{\beta_{NI_{BIO},HO}}$	Percent remaining bias (%) in the estimation of NI-HO association when using $\beta_{NI_G,HO}$ instead of $\beta_{NI_{SR},HO}$.
Metrics of preference of the Goldberg cutoffs (simulation)	**Description**
$Bias_{NI,HO} = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{NI,HO,i}\right) - \beta_1$	Bias in the estimation of NI-HO association. $\hat{\beta}_{NI,HO,i}$ is the point estimate from the ith simulation
$MSE_{NI,HO} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_{NI,HO,i} - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{NI,HO,i}\right)^2 + \left(\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{NI,HO,i}\right) - \beta_1\right)^2$	Mean Squared Error in the estimation of NI-HO association.
$P_{NI,HO} = \frac{\sum_{i=1}^{n} I\left[\left(\hat{\beta}_{NI,HO,i,low} < \beta_1\right) \cap \left(\beta_1 < \hat{\beta}_{NI,HO,i,high}\right)\right]}{n}$	Coverage probability in the estimation of NI-HO association. $\hat{\beta}_{NI,HO,i,low}$ and $\hat{\beta}_{NI,HO,i,high}$ are lower and upper 95% confidence intervals.

Table 2. Baseline characteristics of the analyzed data[#]

Variable	IDATA	P-value
Total Number	303	
Age (years)	63.0 ± 5.9	
Male	124 (40.9%)	
Race		
Non-Hispanic White	283 (93.4%)	
African American	19 (6.3%)	
Asian	1 (0.3%)	
Weight (kg)	79.4 ± 17.1	
Height (cm)	168.8± 9.0	
Waist circumference (cm)	92.0± 14.2	
BMI (kg/m2)	27.7 ± 4.7	
Fat-free mass (kg)	48.9 ± 10.9	
Daily EI estimated from ASA24 (EI_{SR}; kcal/day)	2048.0± 783.4	
Daily EI estimated from DLW (EI_{BIO}; kcal/day)	2400.3 ± 492.8	
Reporting bias in EI (EI_{SR} - EI_{BIO})	-352.3 ± 811.0	<0.001
Daily SI estimated from ASA24 (SI_{SR}; mg/day)	3457.9 ± 1440.8	
Daily SI estimated from urine (SI_{BIO}; mg/day)	4015.0 ± 1995.7	
Reporting bias in SI (SI_{SR} - SI_{BIO})	-557.0 ± 2075.2	<0.001
Daily PoI estimated from ASA24 (PoI_{SR}; mg/day)	2931.4 ± 1136.9	
Daily PoI estimated from urine (PoI_{BIO}; mg/day)	3210.9 ± 1253.9	
Reporting bias in PoI (PoI_{SR} - PoI_{BIO})	-279.5 ± 1404.8	<0.001
Daily PrI estimated from ASA24 (PrI_{SR}; mg/day)	83.0 ± 38.8	
Daily PrI estimated from urine (PrI_{BIO}; mg/day)	94.5 ± 38.3	
Reporting bias in PrI (PrI_{SR} - PrI_{BIO})	-11.5 ± 41.6	<0.001

Note: #: Values are mean ± SD or n (%).

Table 3. Summary of nutrition intakes of the accepted and the rejected cases by the Goldberg cutoffs[#]

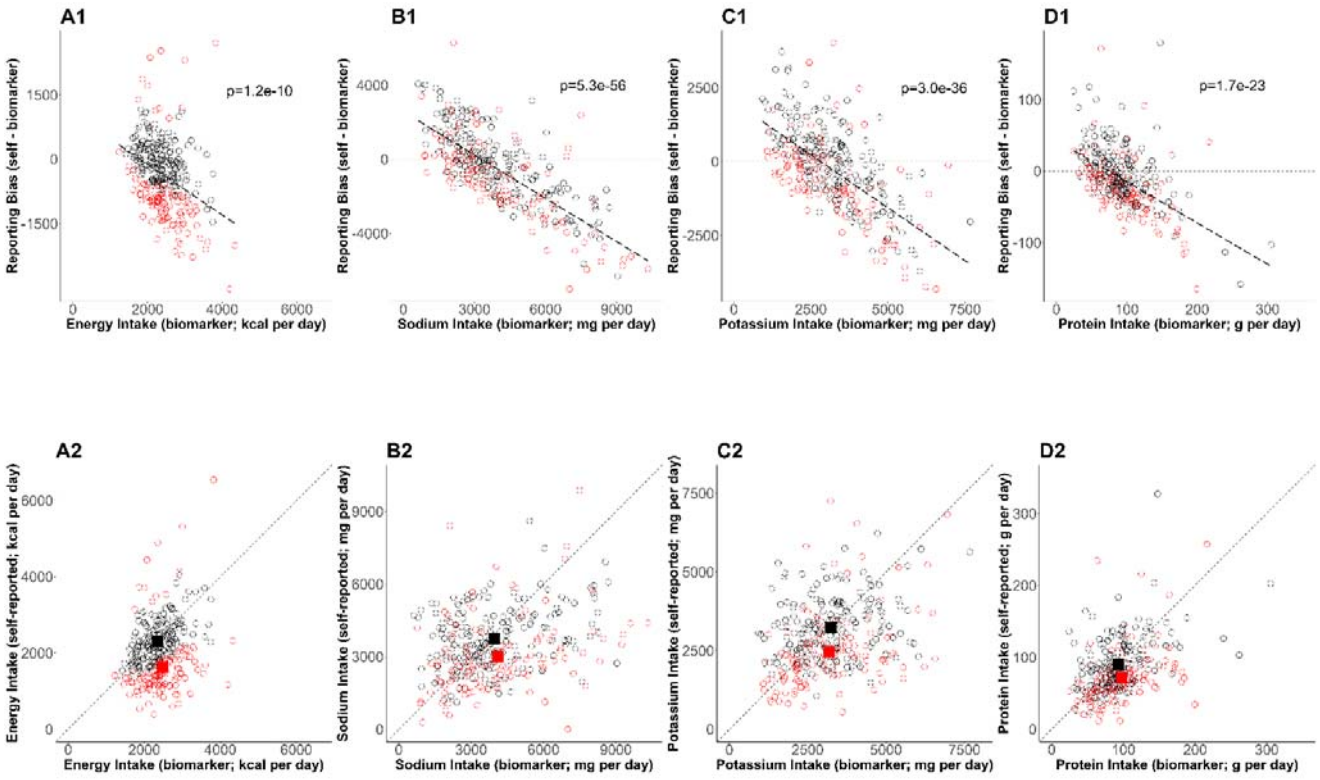
		Accepted	Rejected	P-value[$]
Number		183	120	
Bias	**Energy intake (kcal/d)**	-26 ± 33	-850 ± 89	<0.001
	Sodium intake (mg/d)	-201 ± 142	-1100 ± 200	<0.001
	Potassium intake (mg/d)	3 ± 98	-710 ± 129	<0.001
	Protein intake (g/d)	-2 ± 3	-26 ± 4	<0.001
Self-reported	**Energy intake (kcal/d)**	2320 ± 37	1633 ± 86	<0.001
	Sodium intake (mg/d)	3757 ± 93	3002 ± 145	<0.001
	Potassium intake (mg/d)	3240 ± 71	2461 ± 112	<0.001
	Protein intake (g/d)	91 ± 3	71 ± 4	<0.001
Biomarker	**Energy intake (kcal/d)**	2346 ± 34	2482 ± 49	0.023
	Sodium intake (mg/d)	3958 ± 141	4102 ± 195	<0.001
	Potassium intake (mg/d)	3237 ± 88	3171 ± 124	0.666
	Protein intake (g/d)	93 ± 3	100 ± 4	0.367

Note: #: Values are mean ± SD or n. $ Mean difference between the accepted cases and the rejected cases was tested.

595

Figures

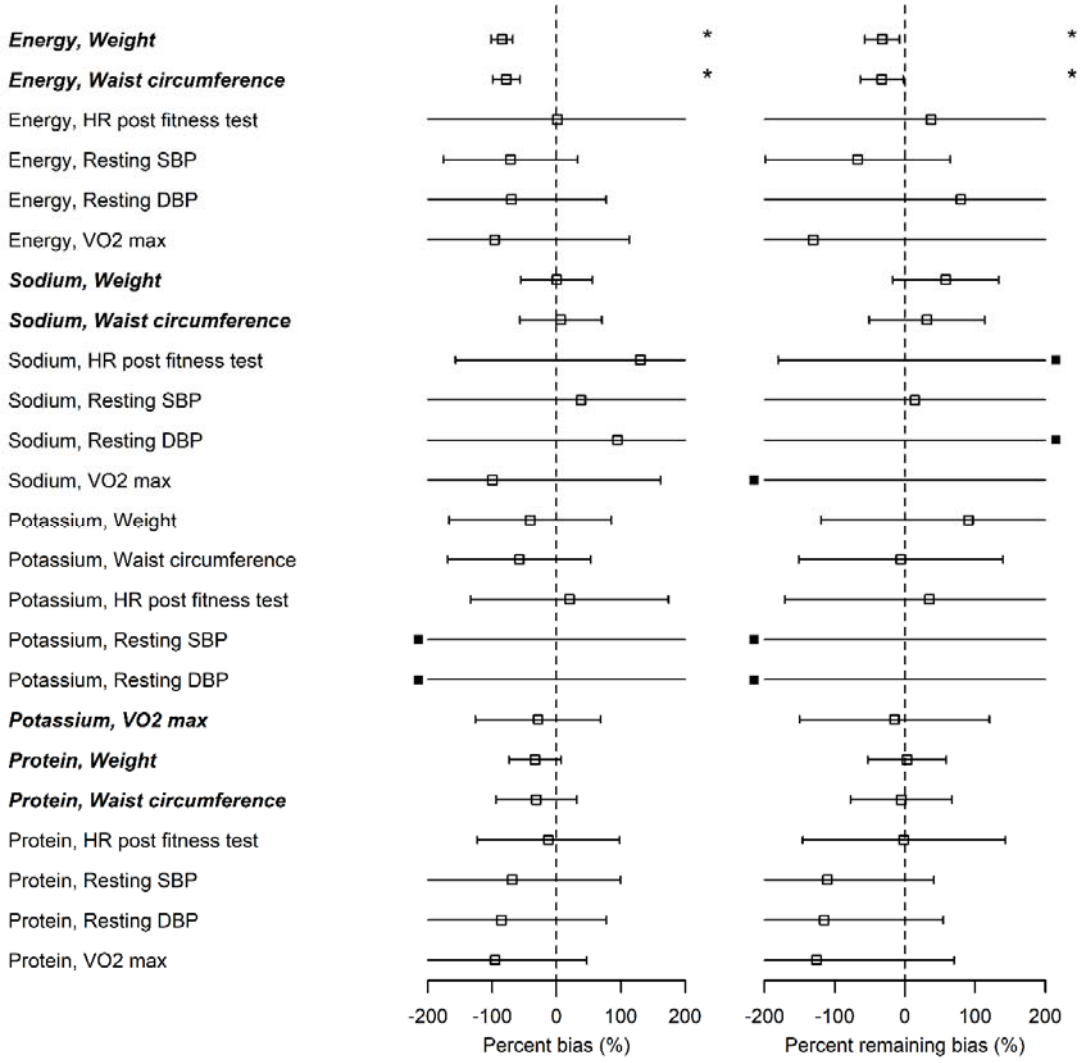
Figure 1: Bias in self-reported nutrition intakes



(Upper panels) The error in mean of self-reported nutrition intake (A: energy intake, B: sodium intake, C: potassium intake, D: protein intake) and relevant nutrition intake measured by biomarkers are plotted. The regression lines are plotted with dashed lines and the corresponding p-values are stated. The dotted horizontal lines at zero indicate there is no error in nutrition intake. The red circles and black circles are the rejected cases and the accepted cases, respectively. **(Bottom panels)** Self-reported nutrition intake and biomarker-based nutrition intake measured are plotted. Closed red and black squares are the mean of nutrition intake in the rejected cases and the accepted cases (NI_G), respectively.

596

Figure 2: Bias in self-reported nutrition intakes

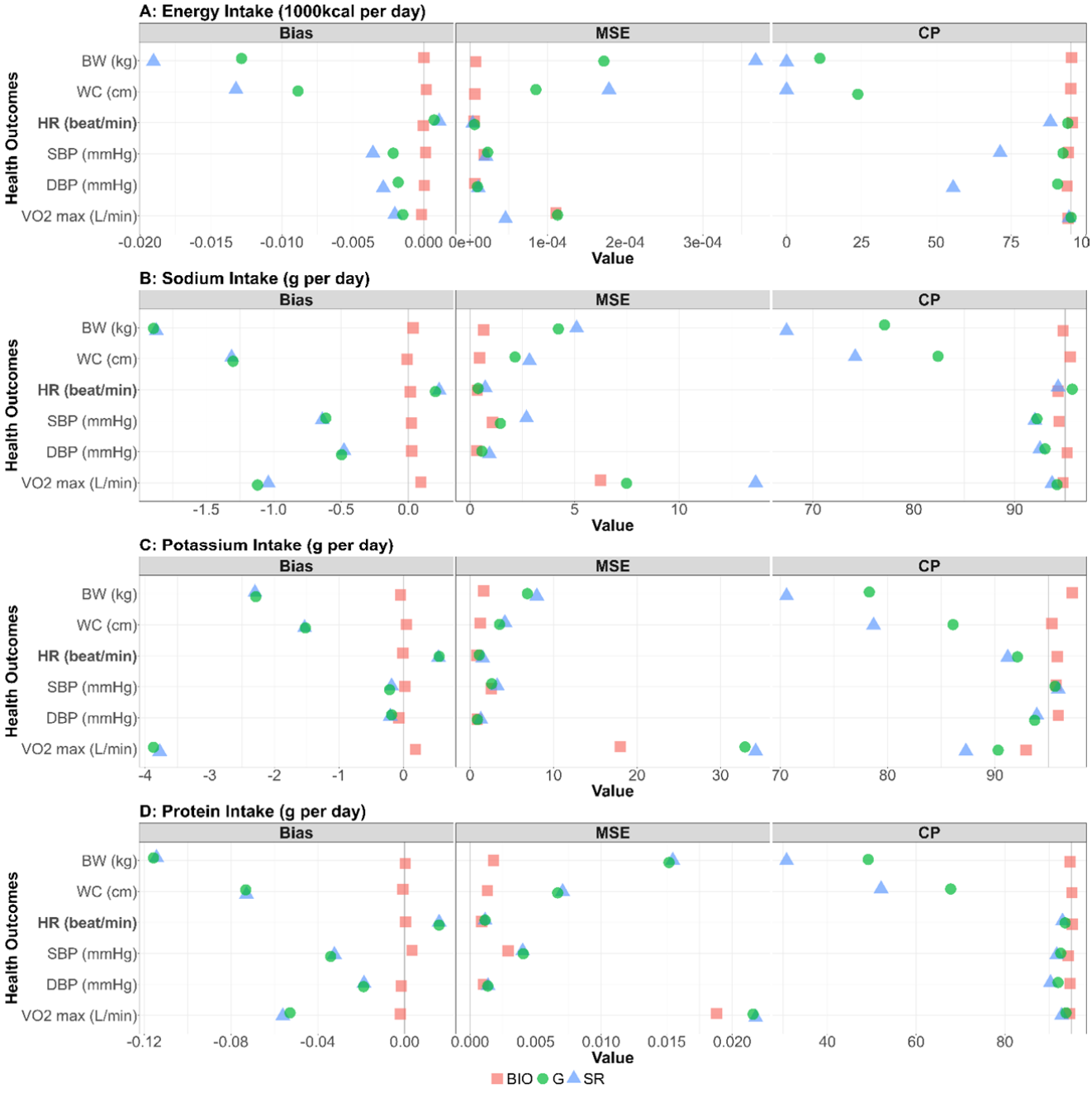


Italic bold font denotes significant associations between nutrition intake measured by biomarkers and the outcome. Open squares correspond to the maximum likelihood estimators and the bars are 95% CIs. Closed squares are plotted at the left end or right end of the panel when the point estimate is beyond the x-axis limits. Using three types of regression coefficients (β_{SR}: self-reported data, β_{BIO}: biomarker data, β_G: Goldberg accepted data), three metrics were defined. **(Left panel)** Percent bias of the linear regression coefficient, $b_\beta = (\beta_{SR} - \beta_{BIO})/\beta_{BIO} * 100(\%)$, was computed. *: Significant bias was observed. **(Right panel)** Percent remaining bias of the linear regression coefficient, $d_\beta = (\beta_G - \beta_{BIO})/\beta_{BIO} * 100\,(\%)$, was

computed. #: Significant bias reduction was observed (i.e., bias reduction 95%CI is within -100 to 100). *: Significant remaining bias was observed.

597

Figure 3: Bias, MSE, and coverage probability in the simulation study

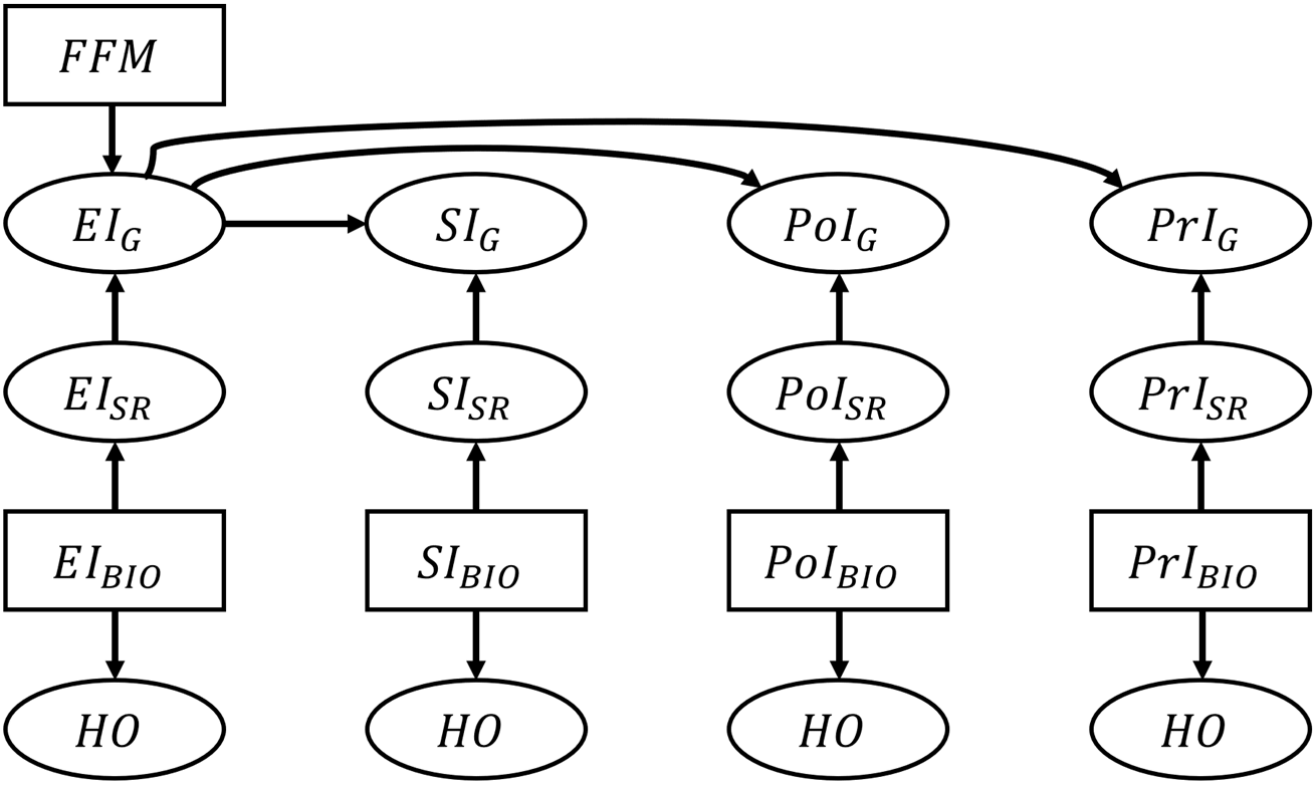


The bias, mean squared error (MSE), and the coverage probability in regression coefficients between four nutrition intakes (A: energy intake, B: sodium intake, C: potassium intake, D: protein intake) and six health outcomes (body weight, waist circumference, HR [heart rate] post fitness test, resting SBP [systolic blood pressure], resting DBP [diastolic blood pressure], and VO_2 max) for 1000 replicates. Bold font denotes negative associations between nutrition intake measured by biomarkers and the outcome. Red square, blue

triangle, and green circle represent biomarker-based nutrition intake, self-reported nutrition intake, and Goldberg accepted nutrition intake, respectively. **(Left panel)** The bias between estimated and true regression coefficients are plotted. The grey vertical line at zero indicates there is no bias between the true regression coefficient and the mean of the 1000 replicates. For each combination, if the green dot is closer to 0 than the blue dot, then that indicates Goldberg cutoff rule reduced the bias. **(Middle panel)** The MSE between estimated and true regression coefficients are plotted. The grey vertical line at zero indicates there is no bias between the true regression coefficient and the mean of the 1000 replicates. **(Right panel)** The coverage probability for simulation studies. The grey vertical line indicates coverage probability consistent with 95% confidence intervals.

598

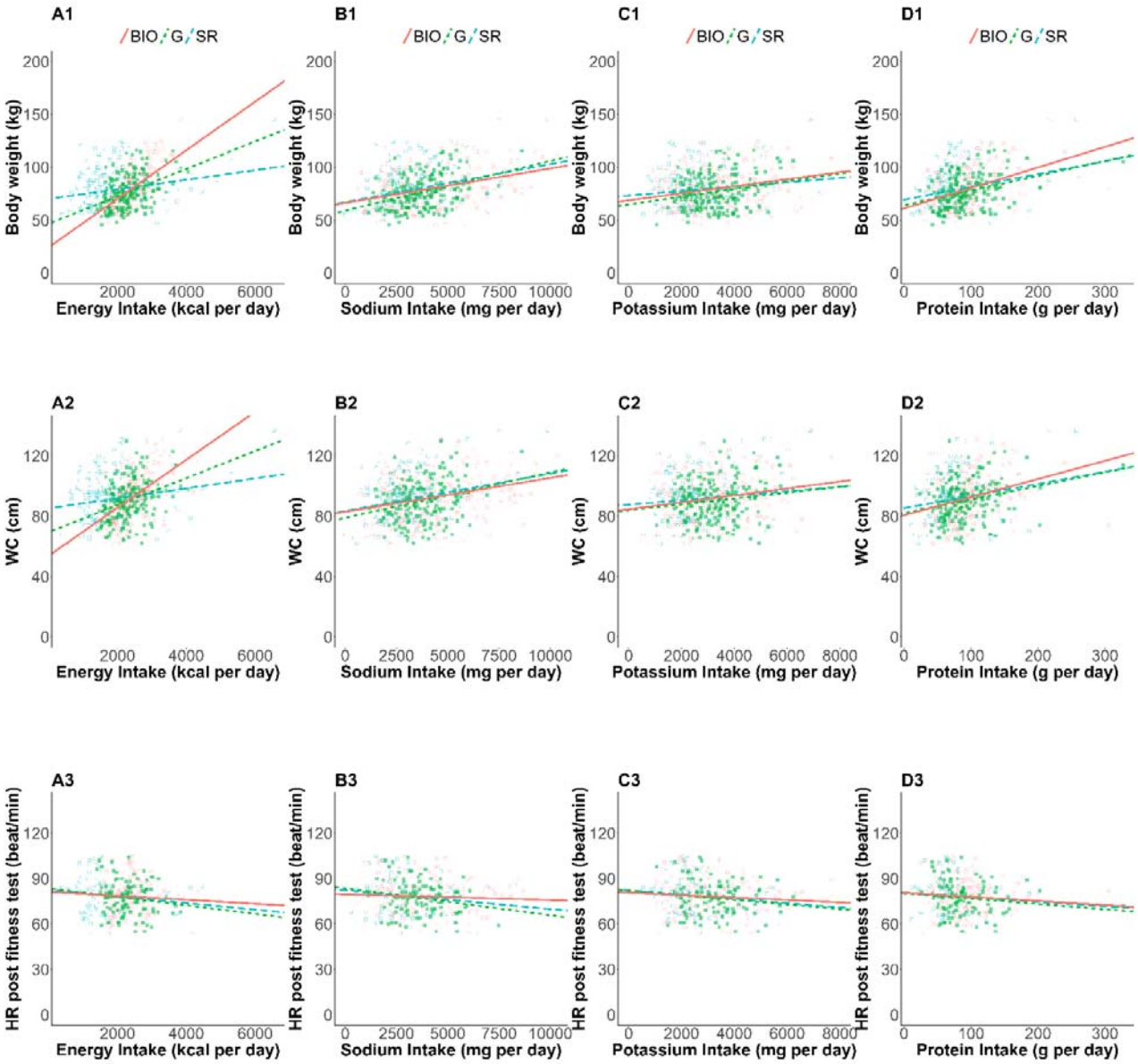
Figure 4: Schematic illustration of the data generation process for the nutrition intake and health outcomes



599 The associations between nutrition intake (EI: Energy Intake, SI: Sodium Intake, PtI: Potassium Intake, PrI:
600 Protein Intake) and health outcomes (HO: body weight, waist circumference, heart rate after fitness test,
601 resting systolic blood pressure, resting diastolic blood pressure, and VO_2 max). The subscripts "SR", "BIO",
602 and "G" denote self-reported NI, biomarker-based NI, and self-reported NI after applying the Goldberg
603 cutoffs, respectively. Fat free mass is denoted by FFM and used to calculate the Goldberg cutoff threshold.
604 An arrow from one node, A, to another, B, means "B is generated by A". Rectangles represent the variables
605 that are resampled from the empirical distribution, and ellipses are for the variables generated from the
606 models.

Supplementary Online Content

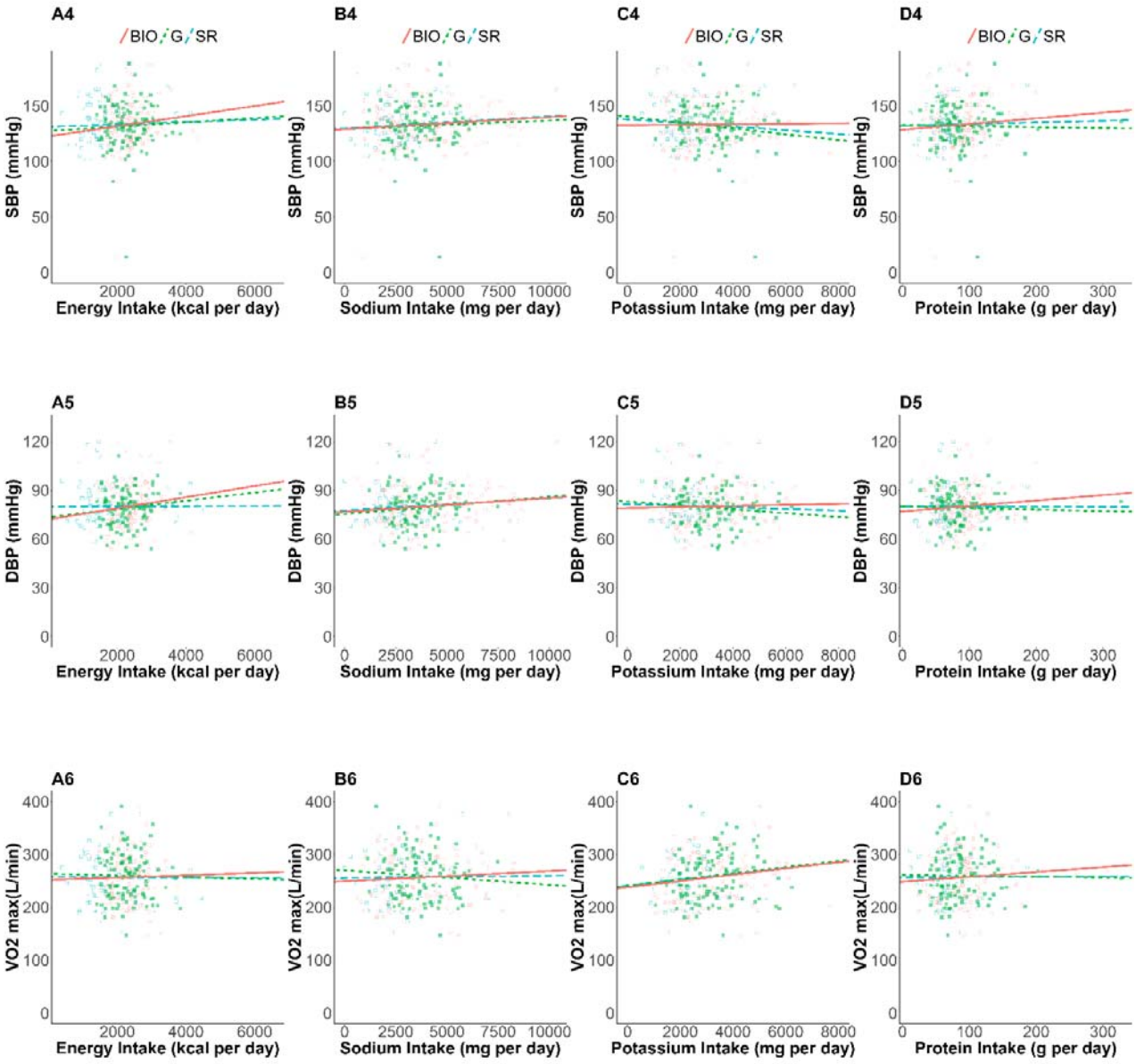
Supplemental Figures



Supplemental Figure 1. Linear regressions of health outcomes on nutrition intakes (1)

The estimated linear regression: associations between (A) EI, (B) SI, (C) PoI, (D) PrI and (1) body weight (kg), (2) waist circumference (cm), and (3) heart rate after the fitness test (beat/min). Open blue squares, filled green squares, and open red circles correspond to self-reported nutrition intake of the whole cases, self-reported nutrition intake of the accepted cases, and nutrition intake measured by biomarkers,

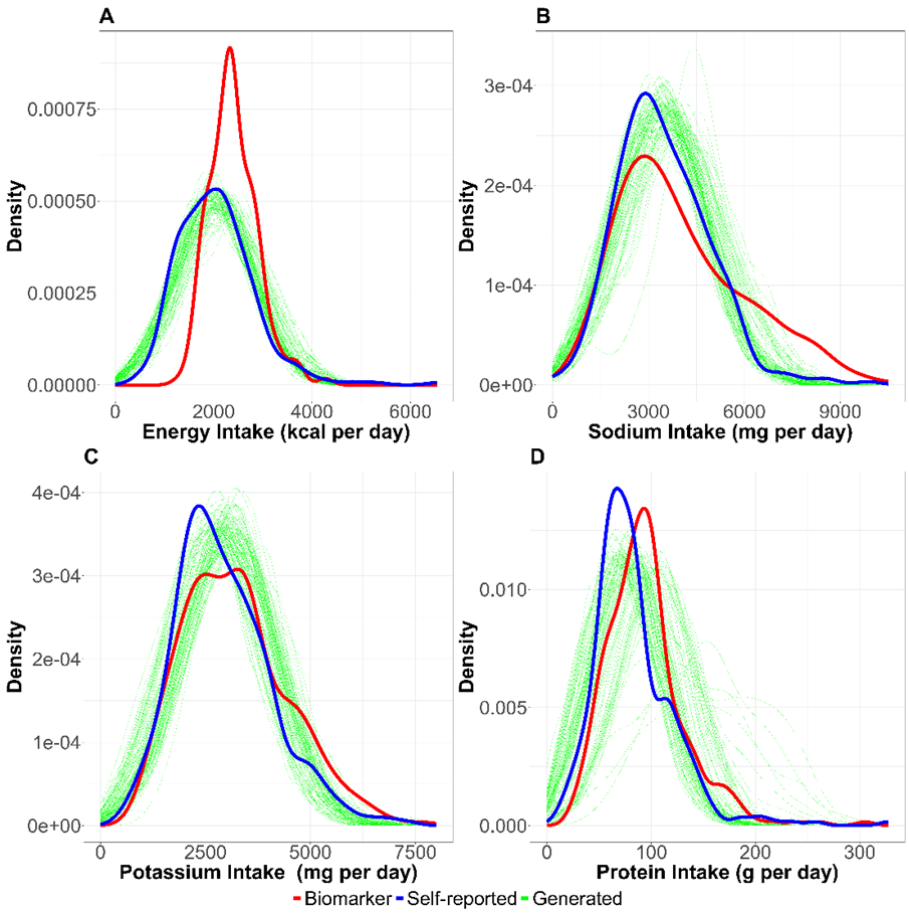
615 respectively. Dashed blue, dotted green, and solid red lines are estimated regression lines using the data with
616 the corresponding color, respectively



617
618 **Supplemental Figure 2. Linear regressions of health outcomes on nutrition intakes (2)**

619 The estimated linear regression: associations between (A) EI, (B) SI, (C) PoI, (D) PrI and (4) resting systolic
620 blood pressure (mmHg), (5) resting diastolic blood pressure (mmHg), and (6) maximal oxygen uptake
621 (L/min). Open blue squares, filled green squares, and open red circles correspond to self-reported nutrition
622 intake of the whole cases, self-reported nutrition intake of the accepted cases, and nutrition intake measured

623 by biomarkers, respectively. Dashed blue, dotted green, and solid red lines are estimated regression lines
624 using the data with the corresponding color, respectively.

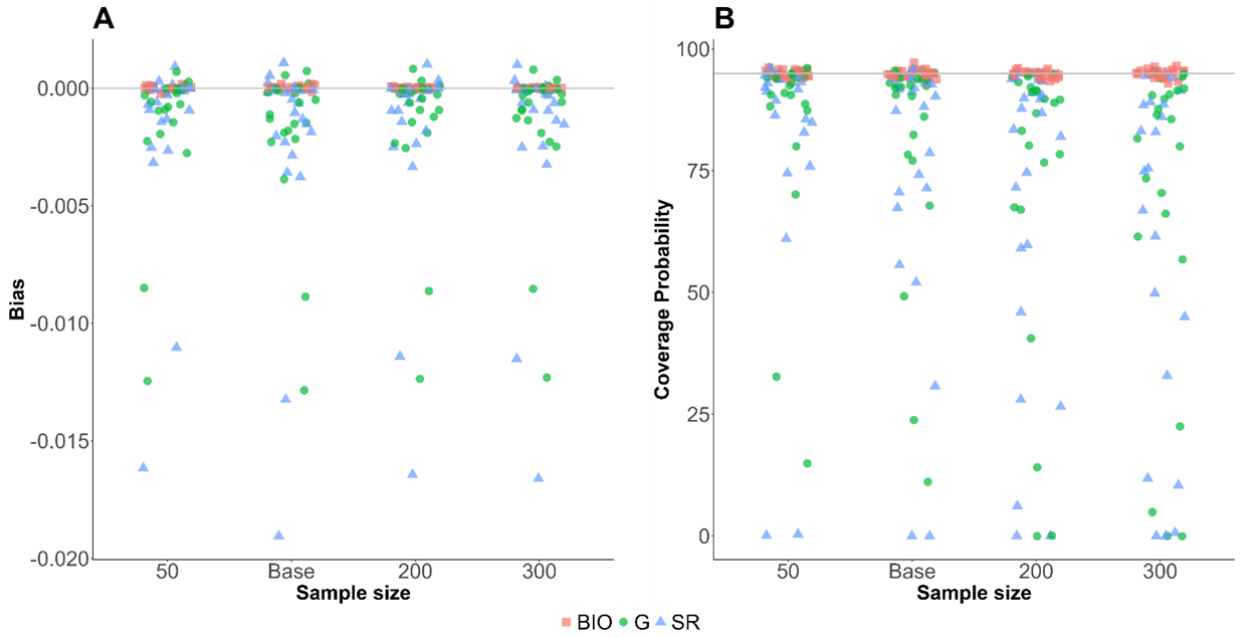


625

626 **Supplemental Figure 3. Distribution of nutrition intake**

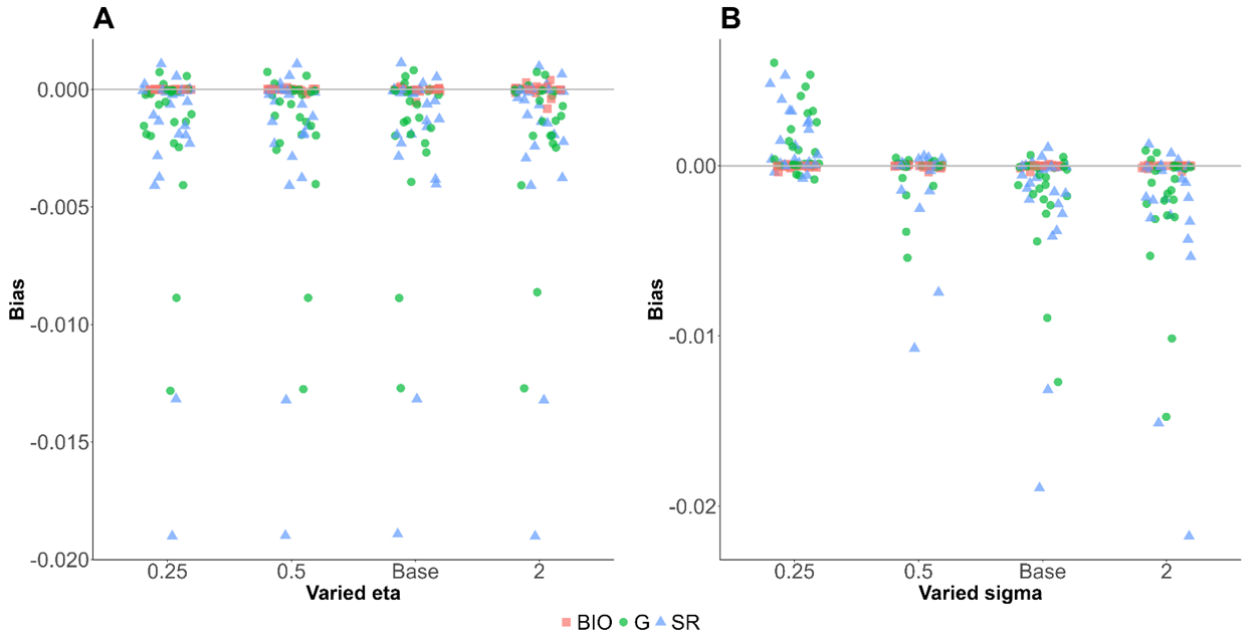
627 The distributions of nutrition intake (A: energy intake, B: sodium intake, C: potassium intake, and D: protein
628 intake) are shown. Red, blue, and green lines correspond to biomarker-based nutrition intake, self-reported
629 nutrition intake, and generated (self-reported) nutrition intake, respectively. Each simulation generating
630 nutrition intake was with n=100 and the simulation was repeated 1000 times in this figure.

631



Supplemental Figure 4. Sensitivities of the sample size relative to the base value

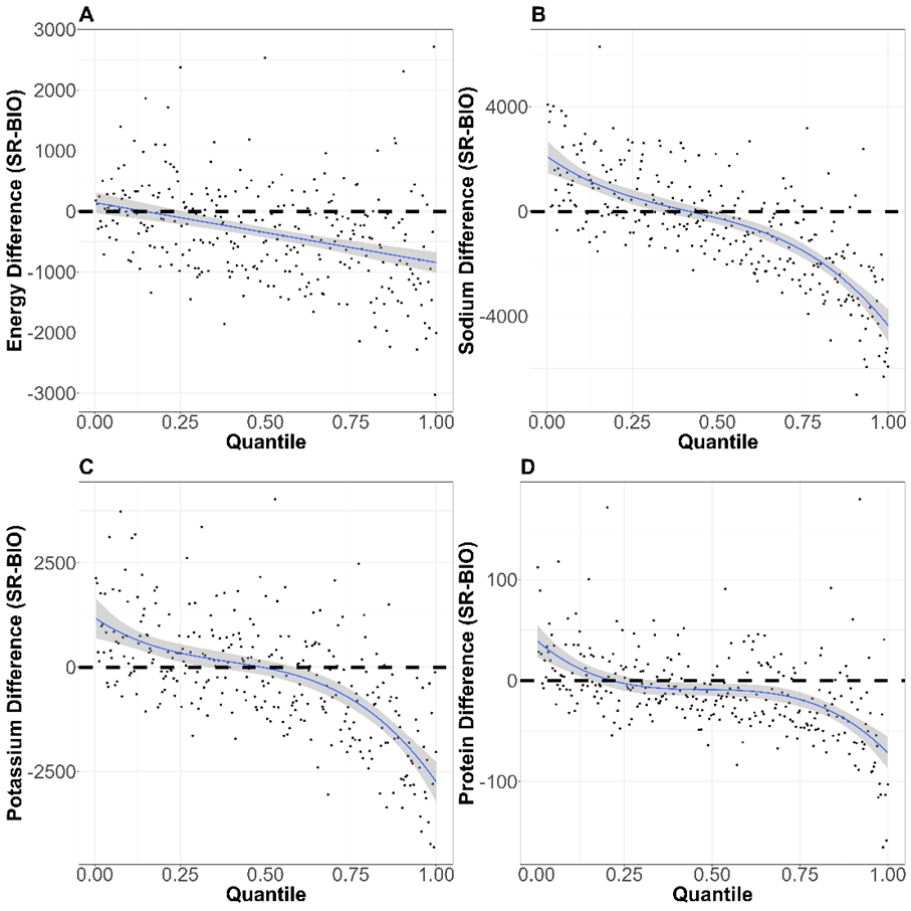
Sensitivity analysis of the sample size. The sample size was varied between 0.5 and 3 times of the baseline values ($n = 50,\ 100,\ 200,$ and 300). Red square, blue triangle, and green circle represent bias (**A**) and coverage probability (**B**) for biomarker-based nutrition intake, self-reported nutrition intake, and Goldberg accepted nutrition intake, respectively.



Supplemental Figure 5. Sensitivities of parameters relative to the base values

Sensitivity analysis of the selected parameters. The parameters were varied between 0.25 and 2 times of the baseline values. Red square, blue triangle, and green circle represent bias for biomarker-based nutrition intake, self-reported nutrition intake, and Goldberg accepted nutrition intake, respectively.

(A) Error term for the regression between health outcome and biomarker-based nutrition intake (η) was varied. (B) Variance of the bias between self-reported nutrition intake and biomarker-based nutrition intake (σ) was varied.



Supplemental Figure 6. Distribution of the reporting error

The distributions of reporting error on quantiles of different nutrition intakes (A: energy intake, B: sodium intake, C: potassium intake, and D: protein intake) are shown. The horizontal dotted lines correspond with no reporting error. The blue lines and the shaded areas are the fitted polynomials (degrees of polynomials are 1, 3, 3, and 5 for energy intake, sodium intake, potassium intake, and protein intake, respectively) and the 95% CI of the polynomials.

Supplemental Tables

Supplemental Table 1. List of abbreviation

Abbreviation	Full form
AARP	American Association of Retired Persons
ASA24	Automated Self-Administered 24-Hour Dietary Assessment Tool
BIO	Biomarker
BW	Body weight
DBP	resting diastolic blood pressure
DHQ-II	Diet History Questionnaire
DLW	Doubly labelled water
EI	Energy intake
EE	Energy expenditure
FFM	Fat-free mass
FM	Fat mass
G	Goldberg cutoffs
HO	Health outcomes
HR	Heart rate
IDATA	Interactive Diet and Activity Tracking in American Association of Retired Persons
MSE	Mean squared error
NI	Nutrition intake
PoI	Potassium intake
PrI	Protein intake
SI	Sodium intake
SBP	Resting systolic blood pressure
SR	Self-reported
VO2	VO_2 max
WC	Waist circumference

Supplemental Table 2. Estimated regression coefficients of the 3 analyses (per Mcal/day for EE and g/day for the others) with 95% CI computed by the jackknife method

Predictors	Outcomes	β_{SR}	β_{G}	β_{BIO}	Percent bias (%)[1,2]	Percent remaining bias (%)[1,3]
Energy intake (Mcal/d)	Weight (kg)	3.09 (-0.38 to 6.56)	13.72 (8.84 to 18.6)*	20.31 (16.36 to 24.27)*	-84.80 (-101.70 to -67.90)*	-32.40 (-57.20 to -7.70)*
	Waist circumference (cm)	2.98 (0 to 5.96)	9.12 (5.05 to 13.18)*	13.63 (9.84 to 17.43)*	-78.10 (-99.20 to -57.10)*	-33.10 (-63.70 to -2.50)*
	HR post fitness test (beat/min)	-2.64 (-4.98 to -0.31)*	-3.59 (-7.62 to 0.44)	-2.62 (-6.21 to 0.98)	1.30 (-204.10 to 206.60)	37.40 (-204.40 to 279.10)
	SBP (mmHg)	1.12 (-2.31 to 4.55)	1.29 (-4.14 to 6.73)	3.94 (-1.50 to 9.38)	-71.60 (-175.90 to 32.70)	-67.20 (-198.70 to 64.30)
	DBP (mmHg)	0.58 (-1.79 to 2.94)	3.45 (-0.15 to 7.05)	1.92 (-1.87 to 5.71)	-70.00 (-216.90 to 76.90)	79.80 (-260.20 to 419.80)
	VO2 max (L/min)	0.18 (-9.10 to 9.45)	-1.44 (-20.01 to 17.12)	4.76 (-7.70 to 17.21)	-96.30 (-305.70 to 113.00)	-131.00 (-583.80 to 321.90)
Sodium intake (g/d)	Weight (kg)	3.32 (1.67 to 4.97)*	5.23 (3.16 to 7.29)*	3.31 (2.14 to 4.47)*	0.50 (-55.30 to 56.30)	58.10 (-17.40 to 133.50)
	Waist circumference (cm)	2.50 (1.14 to 3.86)*	3.08 (1.30 to 4.86)*	2.34 (1.24 to 3.44)*	6.60 (-57.30 to 70.40)	31.40 (-50.90 to 113.80)
	HR post fitness test (beat/min)	-1.53 (-2.84 to -0.22)*	-2.19 (-3.89 to -0.49)*	-0.66 (-1.48 to 0.15)	130.20 (-157.10 to 417.50)	230.20 (-180.70 to 641.20)
	SBP (mmHg)	1.17 (-0.80 to 3.15)	0.97 (-1.81 to 3.74)	0.86 (-0.94 to 2.66)	38.00 (-1230.80 to 1306.80)	14.20 (-1337.00 to 1365.50)
	DBP (mmHg)	0.56 (-0.61 to 1.73)	0.93 (-0.62 to 2.48)	0.29 (-0.54 to 1.13)	94.80 (-546.50 to 736.00)	220.70 (-658.20 to 1099.50)
	VO2 max (L/min)	0.01 (-4.98 to 5.01)	-2.99 (-10.40 to 4.42)	2.12 (-2.00 to 6.24)	-99.50 (-360.90 to 161.80)	-241.80 (-808.90 to 325.20)
Potassium intake (g/d)	Weight (kg)	1.16 (-1.42 to 3.74)	3.75 (0.64 to 6.86)*	1.97 (-0.22 to 4.16)	-41.00 (-167.20 to 85.30)	90.50 (-119.60 to 300.60)
	Waist circumference (cm)	0.78 (-1.41 to 2.97)	1.74 (-0.97 to 4.44)	1.85 (-0.08 to 3.78)	-57.90 (-169.00 to 53.10)	-5.80 (-151.30 to 139.60)
	HR post fitness test (beat/min)	-1.36 (-2.92 to 0.20)	-1.53 (-3.75 to 0.69)	-1.13 (-2.56 to 0.30)	20.30 (-133.30 to 174.00)	34.80 (-170.90 to 240.40)
	SBP (mmHg)	-1.57 (-5.06 to 1.93)	-2.34 (-6.86 to 2.19)	0.46 (-2.34 to 3.27)	-510.50 (-41767.00 to 40746.10)	-686.40 (-47014.60 to 45641.90)
	DBP (mmHg)	-0.58 (-2.11 to 0.94)	-1.08 (-2.77 to 0.61)	0.04 (-1.25 to 1.32)	-1951.10 (-253484.00 to 249581.80)	-3527.90 (-457168.90 to 450113.20)
	VO2 max (L/min)	4.75 (-2.03 to 11.53)	5.67 (-3.51 to 14.85)	6.67 (0.82 to 12.52)*	-28.70 (-126.30 to 68.90)	-14.90 (-150.20 to 120.30)
Protein intake	Weight (kg)	0.10 (0.04 to 0.16)*	0.15 (0.09 to 0.21)*	0.15 (0.07 to 0.22)*	-33.60 (-74.00 to 6.90)	2.80 (-53.10 to 58.60)

(g/d)	**Waist circumference (cm)**	0.06 (0 to 0.13)	0.09 (0.02 to 0.15)*	0.09 (0.01 to 0.17)*	-31.40 (-94.10 to 31.30)	-5.30 (-77.50 to 66.90)
	HR post fitness test (beat/min)	-0.03 (-0.07 to 0.01)	-0.03 (-0.08 to 0.01)	-0.03 (-0.07 to 0.01)	-12.40 (-123.10 to 98.30)	-1.50 (-145.90 to 142.90)
	SBP (mmHg)	0.02 (-0.06 to 0.09)	-0.01 (-0.09 to 0.07)	0.06 (-0.02 to 0.14)	-68.90 (-237.50 to 99.80)	-110.40 (-261.80 to 41.10)
	DBP (mmHg)	0 (-0.03 to 0.04)	0 (-0.04 to 0.040)	0.03 (-0.02 to 0.07)	-85.90 (-249.00 to 77.20)	-115.00 (-284.50 to 54.50)
	VO2 max (L/min)	0 (-0.13 to 0.14)	-0.02 (-0.20 to 0.15)	0.10 (-0.04 to 0.23)	-95.40 (-237.80 to 46.90)	-125.90 (-321.80 to 70.00)

Note: The numbers in parentheses are 95% CIs. The 95% CIs were computed using the jackknife method. *: Significant association or bias. [1]: Computed by jackknife estimation. [2]: $\frac{\beta_{NI_{SR},HO}-\beta_{NI_{BIO},HO}}{\beta_{NI_{BIO},HO}} * 100$ [3]: $\frac{\beta_{NI_{G},HO}-\beta_{NI_{BIO},HO}}{\beta_{NI_{BIO},HO}} * 100$.

Supplemental Table 3. MSE of fitted polynomial functions for the reporting error

	Order of the polynomial functions				
	1	**2**	**3**	**4**	**5**
Energy	582729*	590639	598972	606127	614911
Sodium	2056408	1996615	1953706*	1960605	1974721
Potassium	1229786	1181882	1164808*	1170122	1184684
Protein	1371	1385	1336	1337	1322*

* The order with the smallest MSE.

669 **Supplemental Table 4. Statistical tests on the heteroskedasticity of the reporting error**

Nutrition Intake	P-value of the Goldfeld Quandt test
Energy	0.8666
Sodium	0.7374
Potassium	0.9686
Protein	0.8900

670