The overlooked impact of background diet and adherence in nutrition trials

Javier I. Ottaviani[1], Hagen Schroeter[1], Dennis M. Bier[2], John W. Erdman, Jr[3], Howard D. Sesso[4], JoAnn E. Manson[4], Gunter G. C. Kuhnle*[5]

[1]) Mars, Inc., McLean, VA; [2]) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA [3]) Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA.[4]) Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School. [5]) Department of Food and Nutritional Sciences, University of Reading, Reading, UK;

Corresponding author: Gunter G. C. Kuhnle

Email: g.g.kuhnle@reading.ac.uk

Keywords: Randomized Clinical Trials, Nutrition, Biomarkers, Flavanols

Abstract

Randomized controlled trials in nutrition (RCTN) face unique challenges, including the considerable influence of the background diet and the challenge of assuring intervention adherence by participants. The impact of these factors on the outcome of RCTNs has been difficult to quantify, but nutritional biomarkers represent a valuable tool to address these challenges. Using flavanols as a model dietary intervention and a set of recently validated flavanol biomarkers, we here investigated the impact of background diet and adherence on the outcomes of the COcoa Supplement and Multivitamin Outcomes Study (COSMOS, NCT 02422745). We found that 20% of participants in the placebo and cocoa-extract intervention arms had a flavanol background intake as high as the intervention, and only 5% did not consume any flavanols. Approximately 33% of participants in the intervention group did not achieve expected biomarker levels from the assigned intervention – more than the 15% estimated with pill-taking questionnaires usually implemented in RCTN. Taking these factors into account resulted in a larger effect size for all observed endpoints (HR (95% CI)) estimated using intention-to-treat vs. per-protocol vs. biomarker-based analyses: total cardiovascular disease (CVD) events 0.83 (0.65; 1.07); 0.79 (0.59; 1.05); 0.65 (0.47; 0.89) – CVD mortality 0.53 (0.29; 0.96); 0.51 (0.23; 1.14); 0.44 (0.20; 0.97) – all-cause mortality 0.81 (0.61; 1.08); 0.69 (0.45; 1.05); 0.54 (0.37; 0.80) — major CVD events 0.75 (0.55; 1.02); 0.62 (0.43; 0.91); 0.48 (0.31; 0.74). These results highlight the importance of taking background diet and adherence into consideration in RCTN to obtain more reliable estimates of outcomes through nutritional biomarker-based analyses.

Introduction

Randomized controlled trials (RCTs) are generally considered the gold standard for the scientific study of drugs, treatments and dietary interventions in medical [1-4] and nutrition research [5-7], respectively**.** However, nutrition trials (RCTN) face unique challenges [8-10] that are often overlooked**.** In particular, the impact of background diet of participants on trial endpoints and the objective assessment of adherence are key to the interpretation of trial outcomes. Unlike in medical RCTs where the uncontrolled and unknown exposure to a test drug or treatment can almost never occur, participants in RCTN will almost always be exposed to foods, nutrients, or dietary constituents that are similar to or indistinguishable from the study intervention [8]. Very often this exposure from background diet cannot be quantified by investigators and it is therefore not included in outcomes analyses [11]. The same applies to adherence, which can be more easily monitored with pharmacological interventions where specific and unambiguous biomarkers exist [12-14], whereas most RCTN have to rely exclusively on self-reported information [15-20] – approaches that carry a higher risk of misclassification of adherence. Taken together, these two challenges can significantly affect the outcomes and mask differences between intervention and control groups, leading to incorrect interpretations. Considering the cost, efforts, and high impact of, especially large-scale RCTN, it is important to quantify the potential error introduced by these two often overlooked factors, and to identify and use better methods to address these limitations. This will allow to generate RCTN outcomes of higher scientific rigor and evidence quality, more comparable to RCT in medical research, and contribute to increase reliability and confidence in nutrition research by providing clearer messages to practitioners and the general public

Validated nutritional biomarkers provide the opportunity to address both challenges in RCTN [21, 22]. Nutritional biomarkers measure the systemic presence of the dietary compound under investigation and thereby can provide objective information on both, background diet and adherence to the intervention. However, because there are so few nutritional biomarkers that have been validated [22], integration in large-scale RCTN has been rare and the actual impact is currently unknown.

Flavanols are a group of food bioactives for which a beneficial effect on health has been established[23, 24]. The particular group of flavanols present in cocoa, namely cocoa flavanols (CF), have been investigated recently in a large RCTN, COSMOS (COcoa Supplement and Multivitamin Outcomes Study), which showed that CF intake reduces cardiovascular disease (CVD) risk [25]. There exist two validated biomarkers for the intake of flavanols, namely urinary 5-(3□,4□-dihydroxyphenyl)-γ-valerolactone metabolites ($gVLM_B$) and structurally related (–)-epicatechin metabolites ($SREM_B$) [26, 27]. These biomarkers have been successfully applied in observational studies to assess flavanol intake [28], including the intake of those specific flavanols that are part of the CF evaluated as part of COSMOS intervention [26, 27].The COSMOS trial included the collection of spot urine samples at baseline and during follow-up allowing objective, biomarker-based assessments of both participants' background diets and adherence to the intervention using $gVLM_B$ and $SREM_B$ as validated nutritional biomarkers [26, 27]. Hence, CF and COSMOS represent an excellent model to assess the impact that dietary background and adherence could have on the outcomes of a large RCTN.

Materials and Methods

Study design

This study consisted of a post-hoc, secondary analysis of a subcohort (n=6,532) of COSMOS (NCT02422745; Figure S1 in Supplementary Information), a recently completed RCTN in 21,442 participants (including 8776 males ≥60 y and 12,666 females ≥65 y) in the US [25] . The interventions included capsules containing cocoa extract that provided 500 mg/d cocoa flavanol (CF) or placebo, and a Centrum Silver daily multivitamin or placebo (provided by Pfizer Consumer Healthcare, now Haleon) in a 2x2 factorial design (see Supplementary Materials for details). Participants of the COSMOS Biomarker Cohort provided spot urine samples at baseline during the run-in phase of the study prior to randomization and again at 1, 2 and/or 3-year follow-up. Participants completed a self-reported adherence assessment every 6 months by answering a series of questions related to the number of days taking study pills [29]. Enrolment period for the entire group of participants in COSMOS extended from April 2016 to March 2018, and the intervention was completed on December 31, 2020, with a median treatment period of 3.6 years. All participants provided written informed consent, and study approvals were obtained by the Institutional Review Board (IRB) at Mass General Brigham.

Flavanol biomarker quantification

Urinary levels of 5-(3□,4□-dihydroxyphenyl)-γ-valerolactone metabolites ($gVLM_B$) and structurally related (–)-epicatechin metabolites ($SREM_B$) were used as biomarkers of flavanol intake. $gVLM_B$ represented the sum of the urinary concentrations of 5-(4□-hydroxyphenyl)-γ-valerolactone-3□-sulfate and 5-(4□-hydroxyphenyl)-γ-valerolactone-3□-glucuronide (5-(3□-hydroxyphenyl)-γ-valerolactone-4□-sulfate and 4□-glucuronide were not included as these represented minor metabolites [30] and did not affect the performance of $gVLM_B$ as a biomarker [28]). $SREM_B$ represented the sum of the urinary concentrations of (–)-epicatechin-3□-glucuronide, (–)-epicatechin-3□-sulfate and 3□-O-methyl(–)-epicatechin-5-sulfate. $gVLM_B$ and $SREM_B$ were validated as nutritional biomarkers of flavanol intake [26, 27], and were successfully implemented in observational studies [28]. While $gVLM_B$ informs on the intake of flavanols in general (particularly flavanols including a catechin or epicatechin moiety in their structure), $SREM_B$ is a specific biomarker of the intake of (–)-epicatechin, one of the main bioactive flavanol compounds in the 500 mg/d CF tested in COSMOS. $SREM_B$ and $gVLM_B$ have different systemic half-lives [30], thus a combination of both biomarkers allows capturing different periods after flavanol intake. $gVLM_B$ and $SREM_B$ were quantified using validated LC-MS methods [28] and performance of the quantification are provided in Supplementary Information. Unadjusted biomarker concentrations were used as urinary creatinine is associated with CVD risk [31] and specific gravity did not materially change the outcome (see Supplementary Data for further details).

Flavanol intake assessment using flavanol biomarkers

The quantification of $gVLM_B$ and $SREM_B$ in spot urine serve as concentration biomarkers of flavanol intake and thus, these biomarkers could not be used to determine absolute flavanol intake [21]. Instead, we have used $gVLM_B$ and $SREM_B$ concentrations to classify participants into having flavanol intakes either below, or equal and above the 500 mg/d tested in COSMOS. Thresholds for $gVLM_B$ and $SREM_B$ concentrations were derived from a dose-escalation study conducted as part of the validation of these flavanol biomarkers [26, 27], in which $gVLM_B$ and $SREM_B$ were expressed in µM concentrations in

urine. Considering the inter-individual variability in biomarker levels and to reduce the likelihood of false negatives, respective thresholds were conservatively defined as the bottom 95% CI limit of the expected concentration of $gVLM_B$ and $SREM_B$ after the intake of 500 mg of flavanols (**Figure 1**), using a linear regression model with log_2-transformed concentration data. These thresholds were thus determined as 18.2 µM for $gVLM_B$ and 7.8 µM for $SREM_B$. Given the importance of thresholds selected on the outcomes of this study, we conducted sensitivity analyses using different thresholds for $gVLM_B$ and $SREM_B$ and investigate their impact on estimated outcomes.

Endpoints

Endpoints of this study were i) total cardiovascular disease (CVD) events, the primary outcome in the COSMOS main trial that included myocardial infarction (MI), stroke, coronary revascularization, cardiovascular mortality, carotid artery surgery, peripheral artery surgery, and unstable angina requiring hospitalization; ii) CVD mortality, a secondary outcome in the COSMOS main trial; ii) all-cause mortality, a secondary outcome in the COSMOS main trial; and iv) major CVD events (MI, stroke, and CVD mortality), which are a recognized CVD outcomes, but was not included among pre-registered COSMOS outcomes. Participants reporting an outcome signed a release form to request related medical records for evaluation and processing according to standardized WHI and BWH study procedures. Self-reported primary and secondary outcomes were confirmed by medical record review by a committee of physicians and investigators blinded to treatment assignment. Further details on endpoint adjudication were published previously [25]. Median follow-up time was 3.6 years.

Statistical analysis

All analyses were conducted with R version 4.3.2 [32] in RStudio 2023.12.1 using the packages *rms* [33] for regression analyses and *ggplot2* [34] for the generation of graphics. Missing values were assumed to be missing at random and imputed using multiple imputation (including Nelson-Aalen estimator [35]). Unless indicated otherwise, results are shown with 95% confidence intervals (CI). Associations between endpoints and study groups were investigated using Cox proportional hazards models. The proportional hazard assumption was tested and confirmed with the *cox.zph* function. Statistical models were selected *a priori* based on analyses conducted previously [25]. Continuous variables were transformed using restricted cubic splines with 3 knots (outer quantiles 0.1 and 0.9). Model 1 was adjusted by sex, age at randomization and recruitment cohort (Women's Health Initiative or other). Model 2 was further adjusted by BMI (kg/m^2), smoking status (never, ever, current), and aspirin use (yes/no). Model 3 was additionally adjusted by alternative Healthy Eating Index (aHEI, continuous), family history of hypertension, high cholesterol and stroke (categorical) and history of hypertension (categorical). Model 4 was like model 2, but additionally adjusted by randomization (Placebo or CF intervention), model 5 was like model 2, but additionally adjusted by aHEI. These different models were used to conduct sensitivity analysis of the outcomes obtained. All results were shown as estimated hazard ratio (HR) (95% CI) between control and intervention. For per protocol analyses, participants were censored on the date when participants self-reported missing either more than 8 pills per month or that they were unsure how many pills they have missed, or began outside non-study use of cocoa supplement. Cumulative incidence plots were created using the *survplot* function and data from model 2 (including age at randomization, sex, BMI, smoking status, aspirin use and recruitment cohort).

Data and code availability

Code is available from https://gitlab.act.reading.ac.uk/xb901875/biomarker-based-intervention. Data and associated documentation will be available to users only under a data-sharing agreement. Details on the availability of the study data to other investigators will be on our study website at https://cosmostrial.org/.

Results

COSMOS Biomarker cohort

A total of 6,509 out of the 21,442 participants randomized in COSMOS provided spot urine samples for the quantification of flavanol biomarkers ($gVLM_B$ and $SREM_B$; see Figure S1 and Table S1 for more details).

Background diet and adherence assessment using flavanol biomarkers

Biomarker concentrations at baseline and follow up in the COSMOS Biomarker cohort are shown in Figure 2. At baseline, biomarker concentration in approximately 20% of participants were consistent with flavanol intake of at least 500 mg/d as part of their background diet, with little differences between participants on the intervention and control group (Table 1). Only 5% of participants had biomarker concentrations below the limit of quantification of the analytical method, and thus indicate a negligible or very low exposure to flavanols from the background diet.

Among 2,051 participants with follow-up spot urine samples, 67% of those assigned to the intervention group had a flavanol consumption consistent with the intake of at least 500 mg/d of flavanols (Table 1) and were thus considered adherent to study pill intake. This figure was lower than the adherence determined using self-reported methods, which was 85% [25].

Biomarker-based intervention groups

We used flavanol biomarker data to identify participants as either part of the *biomarker active* or *biomarker control* groups, analogous to the intervention-based *intervention* and *control* group. *Biomarker active* participants had a biomarker estimated flavanol intake of at least 500 mg/d either at baseline or follow-up. Those in the *biomarker control* group did not achieve these biomarker concentrations at baseline and at follow up. Participants without a follow up sample in the CF intervention group who had a background diet providing less than 500 mg/d of flavanols were excluded from the analysis (n=1,774; 27% of Biomarker Cohort) as adherence could not be determined in this group.

Only 62% of participants in the *biomarker active* group had been randomized into the intervention group, whereas 10% of participants in the *biomarker control* group had been randomized into the intervention. In addition to differences in the number of participants receiving placebo and intervention in the biomarker active and biomarker control groups, additional differences in the baseline characteristics of these participants are shown in Table S2.

CVD events in biomarker-based intervention groups

Table 2 shows the associations between intervention groups and disease risk among those in the COSMOS Biomarker cohort. There were substantial differences between HRs when using biomarker-based groups and the randomized groups. The *biomarker active* group showed statistically significant reduced risks for all endpoints, while the associations were weaker when using groups based on randomization assignment. In the intention-to-treat (ITT) analysis, which compares intervention and placebo and does not take adherence and background diet into consideration, a significant reduction in risk was found only for CVD mortality, consistent with the main COSMOS trial findings [25]. Taking self-reported adherence into consideration (per protocol analysis, PP), we also found significant risk reductions for major CVD events and all-cause mortality, but not total CVD events.

The cumulative incidence curves for total CVD events in the COSMOS trial and in the COSMOS biomarker cohort when using groups based on randomization showed a divergence in hazard rates from approximately one year of follow-up (Figure 3B). In contrast, when using *biomarker-based* intervention groups, divergence starts earlier (Figure 3A). This difference could be the consequence of baseline dietary flavanols mediating an effect in those participants that already had biomarker levels consistent with an intake of flavanol equal or above 500 mg/d before commencement of COSMOS trial.

Sensitivity analyses

A series of sensitivity analyses were conducted to determine robustness of the findings for biomarker-based intervention groups (Tables 3). These analyses showed that adjusting for diet quality did not alter the results obtained, suggesting that flavanol intake, rather than a broader beneficial dietary pattern, may be responsible for the differences in CVD events between biomarker-based groups. Additional adjustment by randomization group attenuated the estimated association for major cardiovascular events), which is not surprising as most participants in the *biomarker active* group are also in the CF intervention group.

Given the importance of thresholds selected to define the biomarker-based CF active and control groups, we also investigated the impact of using different thresholds for $gVLM_B$ and $SREM_B$ to identify participant with flavanol intake of at least 500 mg/d. We have estimated associations between biomarker status (above threshold for $gVLM_B$ or $SREM_B$ [biomarker active] vs. below threshold for $gVLM_B$ and $SREM_B$ [biomarker control]) for thresholds varying between 1 µM and 50 µM for $gVLM_B$ and 1 µM and 30 µM for $SREM_B$. The results show that thresholds for $gVLM_B$ varying between 3 to 20 µM and thresholds for $SREM_B$ varying between 5 to 30 µM would not have resulted in a material change of observed estimates (Figure 4). Thus, the thresholds of 18.2 µM for $gVLM_B$ and 7.8 µM for SREM determined for this study fall well within these ranges, and even considering potential inaccuracies in the approach chosen to determine the thresholds used in this study would have not changed the outcomes of this analysis.

Discussion

In this study, we investigated two main limitations of RCTs in nutrition: background diet and adherence, and how these limitations could affect the outcomes interpretation of such studies. For this, we used nutritional biomarkers to objectively assess background diet and adherence, and using flavanols, flavanols biomarkers and the COSMOS trial [25] as a model system [24], [31]. Our results show that both background diet and adherence have a substantial impact on the results of RCTNs:

1. One in five participants in both the placebo and intervention arms were already consuming flavanols as part of their background diet in amounts that were similar or higher to those tested in the intervention arm, and only one in twenty participants were not consuming any flavanols at baseline.
2. Approximately one third of participants assigned to the intervention did not attain the flavanol levels as set forth in this study for this intervention.
3. Considering both biomarker-estimated dietary background and adherence to study intervention resulted in insights on the effect size of the intervention that are more objective, reliable, and less impacted by methodological limitations compared to the general approach in RCTNs, which neither consideres background intake nor biomarker-estimated adherence.

The use of biomarkers allows the objective assessment of intake as part of the background diet and adherence to the intervention. This approach offers significant benefits compared to methods based on dietary and adherence questionnaires and food content databases, which are not only subjective in nature but known to incur significant levels of uncertainty [36, 37]. In contrast to biomarkers that reflect the systemic presence, self-reported data does not capture food and nutrient interactions [38, 39] and the effects of food processing [40] that can affect systemic exposure.

When using nutritional biomarkers, it is important to consider interindividual differences in absorption, distribution, metabolism, and excretion (ADME) as potential limitations. In the case of the flavanol biomarkers used here, the impact of ADME has been previously addressed and discussed during the validation of $gVLM_B$ and $SREM_B$ as biomarkers [26, 27, 31]. In this study, the impact of interindividual differences in ADME was also considered at the time of defining the thresholds levels of $gVLM_B$ and $SREM_B$, which was the basis for the classification of participants into the biomarker-based active and biomarker-based control groups. In this context, thresholds were defined as the lower 95% confidence level of $gVLM_B$ and $SREM_B$ level expected after the intake of 500 mg of CF (Figure 1), which aimed at reducing the risk of false negatives in the classification of participants into the biomarker-based active group. Sensitivity analyses confirmed the robustness of this approach as varying thresholds levels did not alter outcomes (Figure 4). Flavanol biomarker levels were assessed in spot urines and not 24 h urines, as spot urines are more commonly collected in large scale RCTs such as COSMOS. While the use of spot urines introduces uncertainty in flavanol biomarker level related to the time of sample collection, the combination of $gVLM_B$ and $SREM_B$ as complementary biomarkers helped addressing this limitation. In this context, $gVLM_B$ and $SREM_B$ have different systemic half-lives, for which the combination of these biomarkers can inform on both past and recent intakes. Nevertheless, $gVLM_B$ and $SREM_B$ better estimate intake around the day and time in which samples were collected. In this context, for a subset of participants we combined biomarkers levels estimated between 1- and 3-year follow-up, thus providing a wider time range for the biomarker-estimated flavanol intake assessment. Overall,

nutritional biomarkers represent an objective and valid approach to assess flavanol intake, especially compared to self-report dietary and-pill intake questionnaires [36, 37]. Furthermore, biomarkers allow a better estimation of systemic exposure, which is key for the investigation of the health effect of dietary components. Similarly, the use of flavanol biomarkers facilitated the extrapolations of findings beyond the supplements tested in COSMOS. As shown in the case of biomarker-estimated assessment of the background diet, $gVLM_B$ and $SREM_B$ levels are also expected to increase after the consumption of flavanol-containing foods normally found in the diet, and thus, provide benchmarks for biomarkers levels that would be expected to mediate an effect.

We segmented participants into biomarker-based active and control groups, thus not relying on the original randomization of this study to create these two groups. This approach resulted in differences in baseline characteristics of participants in biomarker active and biomarker control groups at baseline (Table S2). However, adjusting for these baseline differences (model 3, Table S3) showed that this had little impact on estimated outcomes, and that biomarker-based segmentation still showed significant changes in HRs for all endpoints. Moreover, all were still larger than those determined via intervention-based ITT and PP analyses (Table 2).

COSMOS was used as a case study to explore the impact of background diet and adherence on RCTN outcomes. The findings from this work further confirm and strengthen the results reported in COSMOS main trial. Indeed, they suggest that there could be even stronger effects linked with flavanol intake than in the ITT and PP analysis in the COSMOS main trial [25]. In addition, biomarker-estimated flavanol intake as part of background diet demonstrated that the amount of flavanols tested in COSMOS and similar to current recommendations [24] is achievable via a regular diet. In this context, further work will be needed to understand current flavanol intake levels in a US representative population. Such work will help assess the impact that increasing flavanol intake may have in the overall population.

Conclusions

RCTN have an important role in the development of public health guidance, as they are considered the *gold standard* of evidence. In this study we have shown that the use of validated nutritional biomarkers is crucial to further strengthen the impact of RCTNs and increase scientific rigor and evidence quality. Including biomarker-based assessments in RCTNs requires the need of rethinking current study designs commonly implemented in these studies, including i) the trade-off between number of volunteers recruited and possibility of collecting samples in the totality of the cohort for biomarker analysis, ii) the possibility of extending ITT and PP analyses to include biomarker-based endpoints as pre-registered outcomes in RCTNs, and iii) consider biomarker levels at baseline as part of inclusion/exclusion criteria [41]. At the same time, more validated nutritional biomarkers are critically needed both for individual dietary compounds and dietary patterns for integration into RCTN designs. This calls for further work aiming at the development and validation of nutritional biomarkers, which also includes investigating the ADME of dietary compounds under investigation and the analytical tools necessary to carry forward this research. There is a high cost in terms of funding, time and personnel needed to carry out RCTNs at scale. Their success in testing hypotheses relies strongly on knowledge of background diet and accuracy in assessing adherence to the intervention. More funding to develop and validate nutritional biomarkers would enhance

the scientific significance and outcomes validity of RCTNs and subsequently their influence on diet and health recommendations and policy.

Acknowledgments

We are deeply indebted to the 21,442 COSMOS participants for their steadfast and conscientious collaboration, and to the COSMOS Research Group for their scientific (Brigham and Women's Hospital (BWH), Fred Hutchinson Cancer Research Center (FHCRC), Women's Health Initiative (WHI), Data Safety and Monitoring Board (DSMB), Mars Edge) and logistical (BWH, FHCRC, DSMB, Mars Edge, Contract Pharmacal Corp, Pfizer Consumer Healthcare (now Haleon)) contributions.

Statements and declarations

Funding

COSMOS is supported by an investigator-initiated research grant from Mars Edge (JEM, HDS), a segment of Mars dedicated to nutrition research and products, which included infrastructure support and the donation of study pills and packaging. Pfizer Consumer Healthcare (now Haleon) provided support through the partial provision of study pills and packaging (JEM, HDS). COSMOS is also supported in part by grants AG050657, AG071611, EY025623, and HL157665 from the US National Institutes of Health (NIH), Bethesda, MD. The Women's Health Initiative (WHI) program is funded by the National Heart, Lung, and Blood Institute, NIH, US Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, and 75N92021D00005.

Competing Interest Statement

J.I.O. and H.S. are employed by Mars, Incorporated, a company engaged in flavanol research and flavanol-related commercial activities. H.D.S. has received investigator-initiated research support from Mars Edge, Pure Encapsulations, and Pfizer Inc., and honoraria and/or travel for lectures from the Council for Responsible Nutrition, BASF, NIH, and American Society of Nutrition during the conduct of the study. J.E.M. has received investigator-initiated research support from Mars Edge. G.G.C.K. has received an unrestricted grant from Mars. COSMOS is supported by an investigator-initiated grant from Mars Edge, a segment of Mars dedicated to nutrition research and products, which included infrastructure support and the donation of study pills and packaging.

Author Contributions

Designed research: JIO, GGCK, HS; performed research: JIO, GGCK; analyzed data: JIO, GGCK; wrote the paper: JIO, GGCK, HS, JWE, DMB, HDS, JEM

Ethics approval

The COSMOS trial and ancillary studies were reviewed and approved by Institutional Review Boards at Brigham and Women's Hospital/Mass General Brigham. The COSMOS website is www.cosmostrial.org. All participants provided informed consent. No identifiable information were used in this study.

References

1. U.S. Food and Drug Administration, Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance for Industry.*Journal*, 2019.
2. AMA Manual of Style Committee, *AMA Manual of Style: A Guide for Authors and Editors*, Oxford University Press, New York, 11th ed edn., 2020.
3. Guyatt G, Rennie D, Meade MO and C. DJ, *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*, McGraw-Hill Education, 3rd edn., 2015.
4. P. B. Burns, R. J. Rohrich and K. C. Chung, The levels of evidence and their role in evidence-based medicine, *Plast Reconstr Surg*, 2011, **128**, 305-310.
5. C. Achterberg, A. Astrup, D. M. Bier, J. C. King, R. M. Krauss, N. Teicholz and J. S. Volek, An analysis of the recent US dietary guidelines process in light of its federal mandate and a National Academies report, *PNAS Nexus*, 2022, **1**, pgac107.
6. J. F. Trepanowski and J. P. A. Ioannidis, Perspective: Limiting Dependence on Nonrandomized Studies and Improving Randomized Trials in Human Nutrition Research: Why and How, *Adv Nutr*, 2018, **9**, 367-377.
7. Dietary Guidelines Advisory Committee, *Scientific Report of the 2020 Dietary Guidelines Advisory Committee: Advisory Report to the Secretary of Agriculture and the Secretary of Health and Human Services.* , U.S. Department of Agriculture, Agricultural Research Service, Washington, DC., 2020.
8. C. M. Weaver and J. W. Miller, Challenges in conducting clinical nutrition research, *Nutr Rev*, 2017, **75**, 491-499.
9. A. H. Lichtenstein, K. Petersen, K. Barger, K. E. Hansen, C. A. M. Anderson, D. J. Baer, J. W. Lampe, H. Rasmussen and N. R. Matthan, Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials, *Adv Nutr*, 2021, **12**, 4-20.
10. C. M. Weaver, A. H. Lichtenstein and P. M. Kris-Etherton, Perspective: Guidelines Needed for the Conduct of Human Nutrition Randomized Controlled Trials, *Adv Nutr*, 2021, **12**, 1-3.
11. D. Stern, D. B. Ibsen, C. J. MacDonald, Y. H. Chiu, M. Lajous and D. K. Tobias, Improving nutrition science begins with asking better questions, *Am J Epidemiol*, 2024, DOI: 10.1093/aje/kwae110.
12. L. Eliasson, S. Clifford, A. Mulick, C. Jackson and B. Vrijens, How the EMERGE guideline on medication adherence can improve the quality of clinical trials, *Br J Clin Pharmacol*, 2020, **86**, 687-697.
13. S. De Geest, L. L. Zullig, J. Dunbar-Jacob, R. Helmy, D. A. Hughes, I. B. Wilson and B. Vrijens, ESPACOMP Medication Adherence Reporting Guideline (EMERGE), *Ann Intern Med*, 2018, **169**, 30-35.
14. W. Y. Lam and P. Fresco, Medication Adherence Measures: An Overview, *Biomed Res Int*, 2015, **2015**, 217047.
15. L. J. Appel, T. J. Moore, E. Obarzanek, W. M. Vollmer, L. P. Svetkey, F. M. Sacks, G. A. Bray, T. M. Vogt, J. A. Cutler, M. M. Windhauser, P. H. Lin and N. Karanja, A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group, *N Engl J Med*, 1997, **336**, 1117-1124.
16. B. V. Howard, L. Van Horn, J. Hsia, J. E. Manson, M. L. Stefanick, S. Wassertheil-Smoller, L. H. Kuller, A. Z. LaCroix, R. D. Langer, N. L. Lasser, C. E. Lewis, M. C. Limacher, K. L. Margolis, W. J. Mysiw, J. K. Ockene, L. M. Parker,

M. G. Perri, L. Phillips, R. L. Prentice, J. Robbins, J. E. Rossouw, G. E. Sarto, I. J. Schatz, L. G. Snetselaar, V. J. Stevens, L. F. Tinker, M. Trevisan, M. Z. Vitolins, G. L. Anderson, A. R. Assaf, T. Bassford, S. A. Beresford, H. R. Black, R. L. Brunner, R. G. Brzyski, B. Caan, R. T. Chlebowski, M. Gass, I. Granek, P. Greenland, J. Hays, D. Heber, G. Heiss, S. L. Hendrix, F. A. Hubbell, K. C. Johnson and J. M. Kotchen, Low-fat dietary pattern and risk of cardiovascular disease: the Women's Health Initiative Randomized Controlled Dietary Modification Trial, *JAMA*, 2006, **295**, 655-666.

17. R. Estruch, E. Ros, J. Salas-Salvado, M. I. Covas, D. Corella, F. Aros, E. Gomez-Gracia, V. Ruiz-Gutierrez, M. Fiol, J. Lapetra, R. M. Lamuela-Raventos, L. Serra-Majem, X. Pinto, J. Basora, M. A. Munoz, J. V. Sorli, J. A. Martinez, M. Fito, A. Gea, M. A. Hernan, M. A. Martinez-Gonzalez and P. S. Investigators, Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts, *N Engl J Med*, 2018, **378**, e34.
18. J. E. Manson, N. R. Cook, I. M. Lee, W. Christen, S. S. Bassuk, S. Mora, H. Gibson, D. Gordon, T. Copeland, D. D'Agostino, G. Friedenberg, C. Ridge, V. Bubes, E. L. Giovannucci, W. C. Willett, J. E. Buring and V. R. Group, Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease, *N Engl J Med*, 2019, **380**, 33-44.
19. , Dietary supplementation with n-3 polyunsaturated fatty acids and vitamin E after myocardial infarction: results of the GISSI-Prevenzione trial. Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico, *Lancet*, 1999, **354**, 447-455.
20. Age-Related Eye Disease Study 2 Research Group, Lutein + zeaxanthin and omega-3 fatty acids for age-related macular degeneration: the Age-Related Eye Disease Study 2 (AREDS2) randomized clinical trial, *JAMA*, 2013, **309**, 2005-2015.
21. R. Kaaks, E. Riboli and R. Sinha, Biochemical markers of dietary intake, *IARC Scientific Publication No. 84*, 1997, 103-126.
22. C. Cuparencu, T. Bulmus-Tuccar, J. Stanstrup, G. La Barbera, H. M. Roager and L. O. Dragsted, Towards nutrition with precision: unlocking biomarkers as dietary assessment tools, *Nat Metab*, 2024, DOI: 10.1038/s42255-024-01067-y.
23. G. Raman, E. E. Avendano, S. Chen, J. Wang, J. Matson, B. Gayer, J. A. Novotny and A. Cassidy, Dietary intakes of flavan-3-ols and cardiometabolic health: systematic review and meta-analysis of randomized trials and prospective cohort studies, *Am J Clin Nutr*, 2019, **110**, 1067-1078.
24. K. M. Crowe-White, L. W. Evans, G. G. C. Kuhnle, D. Milenkovic, K. Stote, T. Wallace, D. Handu and K. E. Senkus, Flavan-3-ols and Cardiometabolic Health: a Guideline Recommendation by the Academy of Nutrition and Dietetics, *Adv Nutr*, 2022, DOI: 10.1093/advances/nmac105.
25. H. D. Sesso, J. E. Manson, A. K. Aragaki, P. M. Rist, L. G. Johnson, G. Friedenberg, T. Copeland, A. Clar, S. Mora, M. V. Moorthy, A. Sarkissian, W. R. Carrick, G. L. Anderson and C. R. Group, Effect of cocoa flavanol supplementation for the prevention of cardiovascular disease events: the COcoa Supplement and Multivitamin Outcomes Study (COSMOS) randomized clinical trial, *Am J Clin Nutr*, 2022, **115**, 1490-1500.
26. J. I. Ottaviani, R. Fong, J. Kimball, J. L. Ensunsa, N. Gray, A. Vogiatzoglou, A. Britten, D. Lucarelli, R. Luben and P. B. Grace, Evaluation of (−)-epicatechin

metabolites as recovery biomarker of dietary flavan-3-ol intake, *Scientific reports*, 2019, **9**, 1-10.
27. J. I. Ottaviani, R. Fong, J. Kimball, J. L. Ensunsa, A. Britten, D. Lucarelli, R. Luben, P. B. Grace, D. H. Mawson and A. Tym, Evaluation at scale of microbiome-derived metabolites as biomarker of flavan-3-ol intake in epidemiological studies, *Scientific Reports*, 2018, **8**, 1-11.
28. J. I. Ottaviani, A. Britten, D. Lucarelli, R. Luben, A. A. Mulligan, M. A. Lentjes, R. Fong, N. Gray, P. B. Grace, D. H. Mawson, A. Tym, A. Wierzbicki, N. G. Forouhi, K.-T. Khaw, H. Schroeter and G. G. C. Kuhnle, Biomarker-estimated flavan-3-ol intake is associated with lower blood pressure in cross-sectional analysis in EPIC Norfolk, *Scientific reports*, 2020, **10**, 17964.
29. P. M. Rist, H. D. Sesso, L. G. Johnson, A. K. Aragaki, L. Wang, S. Rautiainen, A. Hazra, D. K. Tobias, M. S. LeBoff, H. Schroeter, G. Friedenberg, T. Copeland, A. Clar, L. F. Tinker, R. P. Hunt, S. S. Bassuk, A. Sarkissian, D. C. Smith, E. Pereira, W. R. Carrick, E. S. Wion, J. Schoenberg, G. L. Anderson, J. E. Manson and C. R. Group, Design and baseline characteristics of participants in the COcoa Supplement and Multivitamin Outcomes Study (COSMOS), *Contemp Clin Trials*, 2022, **116**, 106728.
30. J. I. Ottaviani, G. Borges, T. Y. Momma, J. P. E. Spencer, C. L. Keen, A. Crozier and H. Schroeter, The metabolome of [2-14C](−)-epicatechin in humans: implications for the assessment of efficacy, safety, and mechanisms of action of polyphenolic bioactives, *Scientific reports*, 2016, **6**, 1-10.
31. J. I. Ottaviani, H. Schroeter and G. G. C. Kuhnle, Measuring the intake of dietary bioactives: Pitfalls and how to avoid them, *Mol Aspects Med*, 2023, **89**, 101139.
32. R. C. Team, R: A Language and Environment for Statistical Computing.*Journal*, 2024.
33. F. E. Harrell Jr, rms: Regression Modeling Strategies, 2017.
34. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York, 2009.
35. I. R. White and P. Royston, Imputing missing covariate values for the Cox model., *Statistics in medicine*, 2009, **28**, 1982-1998.
36. J. I. Ottaviani, V. Sagi-Kiss, H. Schroeter and G. G. C. Kuhnle, Reliance on self-reports and estimated food composition data in nutrition research introduces signi fi cant bias that can only be addressed with biomarkers, *Elife*, 2024, **13**.
37. A. F. Subar, L. S. Freedman, J. A. Tooze, S. I. Kirkpatrick, C. Boushey, M. L. Neuhouser, F. E. Thompson, N. Potischman, P. M. Guenther, V. Tarasuk, J. Reedy and S. M. Krebs-Smith, Addressing Current Criticism Regarding the Value of Self-Report Dietary Data, *J Nutr*, 2015, **145**, 2639-2645.
38. J. I. Ottaviani, J. L. Ensunsa, R. Y. Fong, J. Kimball, V. Medici, G. G. C. Kuhnle, A. Crozier, H. Schroeter and C. Kwik-Uribe, Impact of polyphenol oxidase on the bioavailability of flavan-3-ols in fruit smoothies: a controlled, single blinded, cross-over study, *Food Funct*, 2023, **14**, 8217-8228.
39. J. I. Ottaviani, R. Y. Fong, G. Borges, J. Kimball, J. L. Ensunsa, V. Medici, L. K. Pourshahidi, E. Kane, K. Ward, R. Durkan, S. Dobani, R. Lawther, G. O'Connor, C. I. R. Gill, H. Schroeter and A. Crozier, Flavan-3-ol-methylxanthine interactions: Modulation of flavan-3-ol bioavailability in volunteers with a functional colon and an ileostomy, *Free Radic Biol Med*, 2023, **196**, 1-8.
40. J. I. Ottaviani, T. Y. Momma, C. Heiss, C. Kwik-Uribe, H. Schroeter and C. L. Keen, The stereochemical configuration of flavanols influences the level and

metabolism of flavanols in humans and their biological activity in vivo, *Free Radic Biol Med*, 2011, **50**, 237-244.

41. A. M. Brickman, L. K. Yeung, D. M. Alschuler, J. I. Ottaviani, G. G. C. Kuhnle, R. P. Sloan, H. Luttmann-Gibson, T. Copeland, H. Schroeter, H. D. Sesso, J. E. Manson, M. Wall and S. A. Small, Dietary flavanols restore hippocampal-dependent memory in older adults with lower diet quality and lower habitual flavanol consumption, *Proc Natl Acad Sci U S A*, 2023, **120**, e2216932120.

Figures and Tables

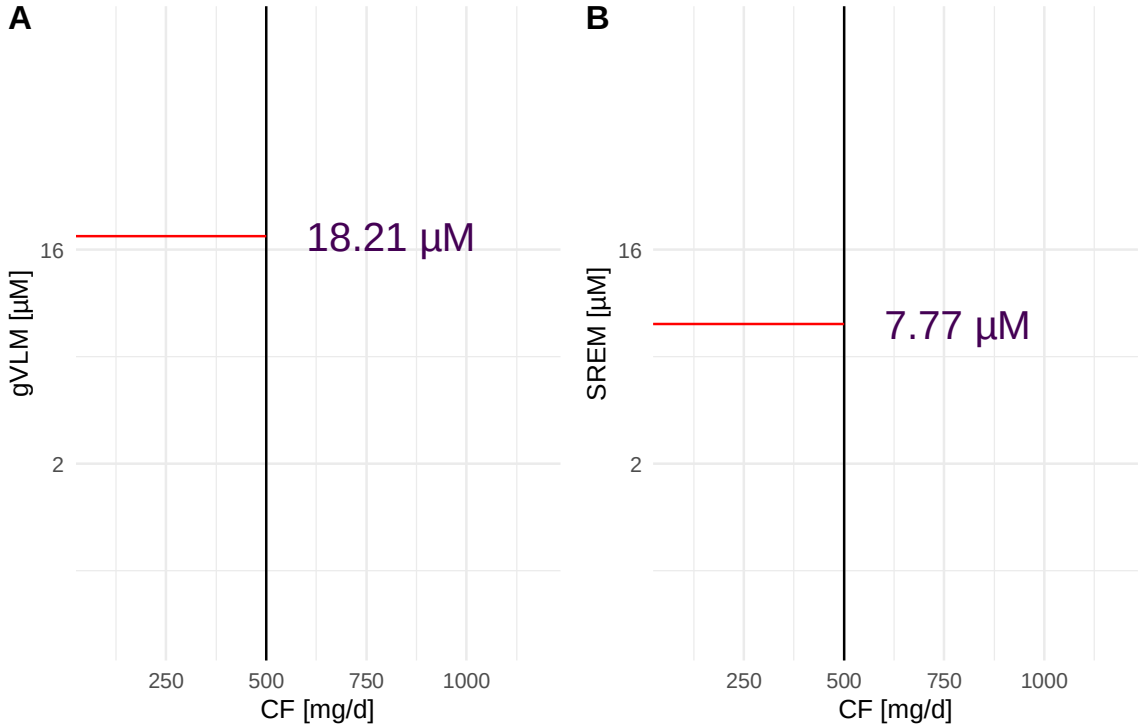

Figure 1: Concentration of $gVLM_B$ (A) and $SREM_B$ (B) in urine as a function of the amount of cocoa flavanols (CF) consumed. Flavanol biomarker data were log-transformed. $gVLM_B$ and $SREM_B$ thresholds were defined as the bottom 95%CI limit of the expected concentration of $gVLM_B$ and $SREM_B$ after the intake of 500 mg of CF. Data were obtained from data collected previously [26, 27]. $gVLM_B$: 5-(3□,4□-dihydroxyphenyl)-γ-valerolactone metabolites; $SREM_B$: structurally related (–)-epicatechin metabolites.

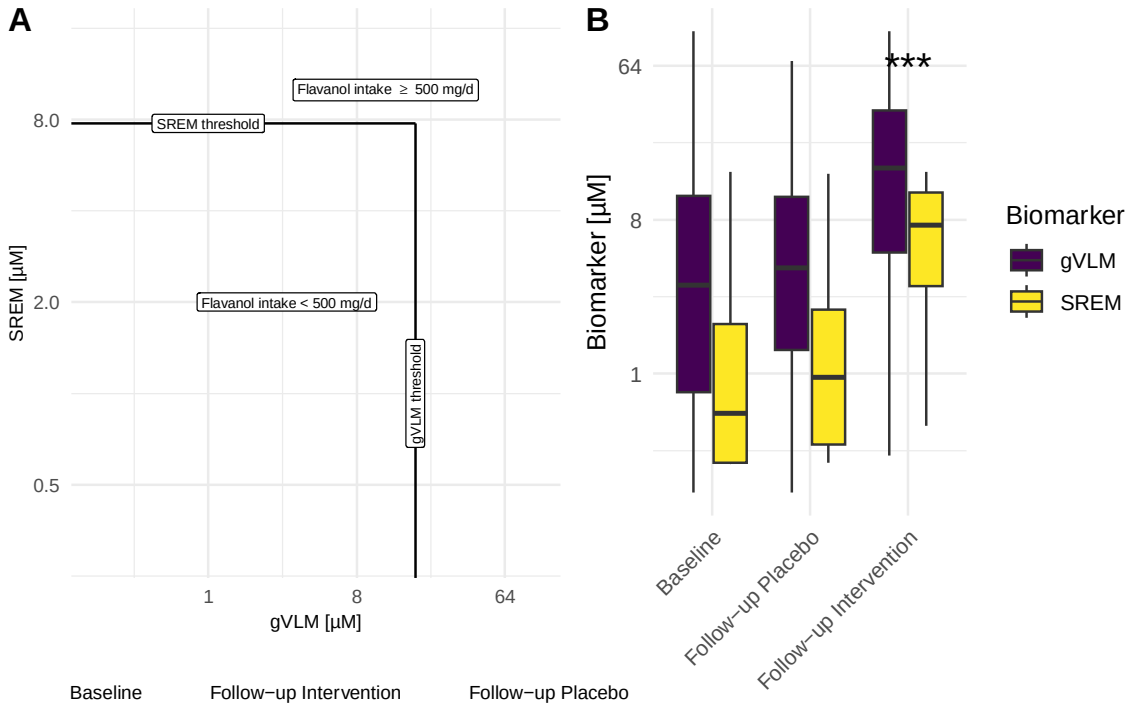

Figure 2. A) Concentration of $SREM_B$ and $gVLM_B$ in individual participants (n=6,509) at baseline and follow-up in intervention and placebo groups. Black line represents thresholds for $SREM_B$ and gVLMB. Participants having $gVLM_B$ or $SREM_B$ levels above the corresponding thresholds were considered to have a flavanol intake equal or higher than 500 mg/d. B) Concentration of $SREM_B$ and $gVLM_B$ in participants (n=6,509) at baseline and follow-up in intervention and placebo groups. Data are shown as box-plots, with the box representing the inter-quartile range. Biomarker concentration at follow-up in the intervention group were significantly higher (Tukey's Honestly Significant Difference).

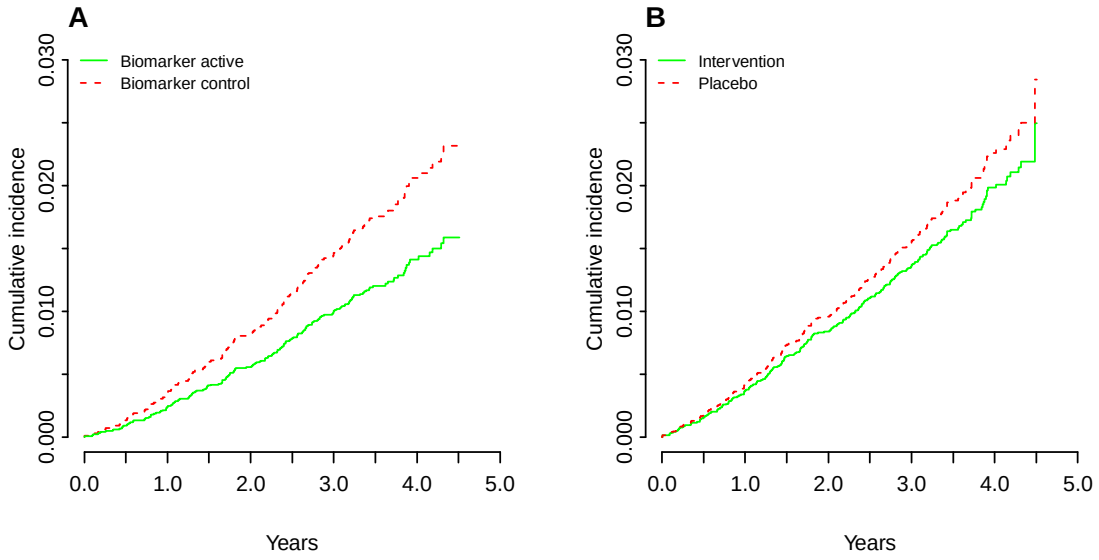

Figure 3. Cumulative incidence of total cardiovascular disease (CVD) events according to biomarker-based groups (A) and randomized assignment in per-protocol analysis (B). Results shown are adjusted for age at randomization (75), sex (male), BMI (25 kg/m^2), smoking status (never), aspirin use (no) and recruitment cohort (WHI). Total CVD events was a composite of myocardial infarction, stroke, CVD death, coronary artery bypass graft and percutaneous coronary intervention, unstable angina including hospitalization, carotid artery surgery, and peripheral artery surgery.

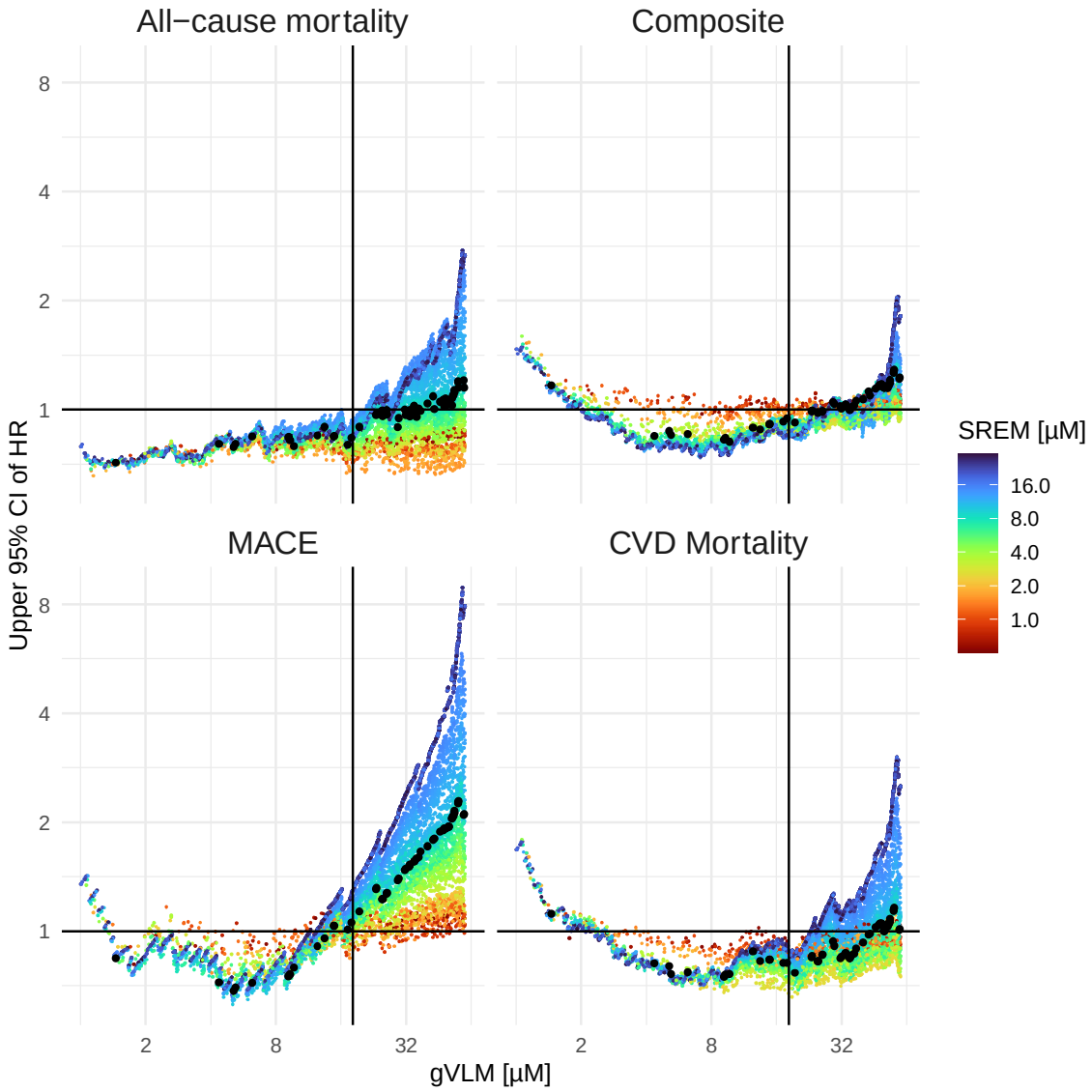

Figure 4: Comparison of 95% upper confidence limit of HR for different CVD events between biomarker-active and biomarker-control group when using different concentration thresholds. Summary statistics were from Cox regression models by age, sex, BMI, smoking status, and aspirin use. CIs were not adjusted for multiple comparisons. CVD events included total CVD events (which was a composite of myocardial infarction, stroke, CVD death, coronary artery bypass graft and percutaneous coronary intervention, unstable angina including hospitalization, carotid artery surgery, and peripheral artery surgery), CVD deaths, all-cause deaths and major adverse CVD events (which was a composite of myocardial infarction, stroke, and CVD death). Vertical lines and black dots indicate the gVLM and SREM thresholds used in this study. Horizontal line shows HR of 1.

Table 1. Number of participants with flavanol biomarker levels consistent with a flavanol intake of 500 mg/d or higher at baseline and follow-up.

	Biomarker-estimated flavanol intake ≥ 500 mg/d	
	Baseline (background diet)	Follow-up (adherence)
Participants with spot urine at baseline (n=6,509)	1,249 (19%)	—
Placebo (n=3,257)	603 (19%)	—
CF intervention (n=3,252)	646 (20%)	—
Participants with spot urine at baseline and follow-up (n=2,051)	414 (20%)	881 (43%)
Placebo (n=991)	186 (19%)	175 (18%)
CF intervention (n=1,060)	228 (22%)	706 (67%)

Table 2. HRs and 95% CIs for cardiovascular (CVD) outcomes, according to randomized assignment in intention-to-treat (ITT) and per-protocol (PP) analyses and according to biomarker-based groups. Summary statistics were from Cox regression models adjusted by age, sex, BMI, smoking status, recruitment cohort and aspirin use (model 2). CIs were not adjusted for multiple comparisons.

	Intervention-based groups		Biomarker-based groups
	ITT[†] (n=6,532)	PP[‡] (n=6,532)	Biomarker (n=4,735)
All-cause mortality	0.81 (0.61; 1.08)	0.69 (0.45; 1.05)	0.54 (0.37; 0.80)
Total CVD events	0.83 (0.65; 1.07)	0.79 (0.59; 1.05)	0.65 (0.47; 0.89)
CVD mortality	0.53 (0.29; 0.96)	0.51 (0.23; 1.14)	0.44 (0.20; 0.97)
Major CVD events	0.75 (0.55; 1.02)	0.62 (0.43; 0.91)	0.48 (0.31; 0.74)

Total CVD events is a composite of myocardial infarction, stroke, CVD death, coronary artery bypass graft and percutaneous coronary intervention, unstable angina including hospitalization, carotid artery surgery, and peripheral artery surgery. Major CVD events was a composite of myocardial infarction, stroke, and CVD death. CVD, cardiovascular disease; [†]Intention to treat; [‡]per-protocol

Table 3: Sensitivity analyses. HRs and 95% CIs for cardiovascular disease (CVD) outcomes using biomarker-based groups and different models. Summary statistics from Cox regression models show HRs comparing *biomarker control* and *biomarker active* group for each outcome. CIs were not adjusted for multiple comparisons.

	Model 1†	Model 2†	Model 3†	Model 4†	Model 5†
All-cause mortality	0.52 (0.35; 0.77)	0.54 (0.37; 0.80)	0.57 (0.39; 0.85)	0.65 (0.41; 1.02)	0.55 (0.37; 0.82)
Total CVD events	0.64 (0.47; 0.88)	0.65 (0.47; 0.89)	0.68 (0.49; 0.93)	0.70 (0.48; 1.02)	0.66 (0.48; 0.90)
CVD mortality	0.42 (0.19; 0.93)	0.44 (0.20; 0.97)	0.46 (0.21; 1.02)	0.64 (0.26; 1.54)	0.45 (0.20; 0.98)
Major CVD events	0.48 (0.31; 0.73)	0.48 (0.31; 0.74)	0.50 (0.33; 0.78)	0.55 (0.33; 0.91)	0.49 (0.32; 0.75)

Total CVD events is a composite of myocardial infarction, stroke, CVD death, coronary artery bypass graft and percutaneous coronary intervention, unstable angina including hospitalization, carotid artery surgery, and peripheral artery surgery. Major CVD events was a composite of myocardial infarction, stroke, and CVD death. CVD, cardiovascular disease; †Model 1 was adjusted by sex, age at randomization and recruitment cohort (Women's Health Initiative or other). Model 2 was further adjusted by BMI (kg/m^2), smoking status (never, ever, current), and aspirin use (yes/no). Model 3 was additionally adjusted by alternative Healthy Eating Index (aHEI, continuous), family history of hypertension, high cholesterol and stroke (categorical) and history of hypertension (categorical). Model 4 was like model 2, but additionally adjusted by randomization (Placebo or CF intervention), model 5 was like model 2, but additionally adjusted by aHEI.