The **CO**coa **S**upplement and **M**ultivitamin **O**utcomes **S**tudy (**COSMOS**) Statistical Analysis Plan

Investigators:

JoAnn E. Manson, MD, DrPH and Howard D. Sesso, ScD, MPH, Brigham and Women's Hospital
Garnet L. Anderson, PhD, Fred Hutchinson Cancer Research Center

February 4, 2022

ClinicalTrials.gov Identifier: NCT02422745

Infrastructure support provided by an investigator-initiated grant from Mars Symbioscience.

Cocoa Extract supplements and matching placebo provided by Mars Symbioscience.

Multivitamins and matching placebo provided by Pfizer Inc.

Partial infrastructure support provided through the National Heart Lung and Blood Institute's funding for the Women's Health Initiative Clinical Coordinating Center and Regional Center contracts.

COSMOS Statistical Analysis Plan

In this 2x2 factorial trial, the primary aim is to compare the main effects of intention-to-treat with cocoa extract on incident cardiovascular disease (CVD) events, and multivitamins on rates of diagnoses of new cancers. The primary analysis be based on the intention-to-treat principle and use the Cox proportional hazards model for time-to event data(1) and estimate the hazard ratio (HR) for each intervention using indicators for the specific treatment, controlling for the second intervention, and stratifying the baseline hazard by (at minimum) sex, age group, and recruitment center (Women's Health Initiative (WHI) or Brigham and Women's Hospital (BWH)). P-values will be based on a score tests for each intervention HR. Because the cohort consists of older women and men, competing risks due to deaths from other causes will also be incorporated. Specifically, the primary analyses will estimate the cause-specific hazard and HR comparing intervention groups for each outcome by censoring individuals with deaths due to competing causes. We will estimate the cumulative incidence function for each comparison (multivitamin versus placebo, cocoa extract versus placebo) and plot the subdistribution of each endpoint over time.(2, 3)

Cocoa extract-specific analyses

Cox proportional hazards models estimated HRs using an indicator variable for cocoa extract treatment and stratifying the baseline hazard functions by sex, age group, recruitment source, and multivitamin intervention arm. COSMOS was designed for 22,000 participants with ≥80% power to detect a 11% relative hazard reduction in total CVD, and >95% power to detect the same reduction for the secondary outcome of CVD plus all-cause mortality. We also had ≥90% power to detect a 14% reduction in total cancer. Models were constructed for each clinical outcome, where person-time for each outcome was counted as days from randomization to the first post-randomization diagnosis of the designated outcome. Follow-up was censored at date of last contact, death, or end of the trial on December 31, 2020, whichever came first.

Kaplan-Meier cumulative incidence curves, cumulative HRs, interactions between randomization groups with trial-time, and analyses that excluded the first 1 and 2 years of follow-up assessed whether treatment effects varied over time (4). Nine subgroup analyses examined effect modification by *a priori* (concurrent multivitamin randomization assignment, sex, age, and use of statins or aspirin) and *post hoc* (history of CVD, smoking status, number of CVD risk factors, and chocolate consumption) factors. Secondary outcomes (n=14) mostly constituted subtypes of CVD or cancer. Statistical significance ($P \leq 0.05$) was assessed with two-sided p-values. We did not adjust P-values or confidence intervals for multiple testing. Consequently, results for secondary outcomes, other outcomes, subgroup analyses, and other analyses should be interpreted cautiously and considered hypothesis generating. At the nominal 0.05 level, we would expect less than one interaction and one secondary outcome to be significant by chance alone.

Per-protocol analyses censored follow-up when the participant discontinued trial pills, began outside non-study use of a cocoa supplement, and/or took <75% of study pills. HRs and 95% CIs were estimated using Cox regression models, weighted by the inverse probability of dependent-censoring for non-compliance (5). Additional analyses compared self-reports of non-monitored outcomes or potential side effects by intervention group. Analyses were performed using SAS 9.4.

Multivitamin-specific data analyses

Baseline characteristics by treatment groups were compared using chi-squared tests of association, a t-test contrasted age at randomization, and Wilcoxon rank-sum tests compared non-normally distributed variables such as body mass index and exercise. Statistical significance ($P \leq 0.05$) was assessed with two-sided p-values. Our primary analyses were based on the intention-to-treat principle for time-to first event data.(1) Cox proportional hazards models estimated HRs using an indicator variable for MVM assignment and stratifying the

baseline hazard functions by sex, age group, recruitment source, and cocoa extract intervention arm. The trial was designed for 22,000 participants with ≥90% power to detect a 14% relative hazard reduction in total cancer. Models were constructed for each clinical outcome, where person-time for each outcome was counted as days from randomization to the first post-randomization diagnosis of the designated outcome. Follow-up was censored at date of last contact, death, or end of the trial on December 31, 2020, whichever came first.

Kaplan-Meier cumulative incidence curves, cumulative HRs, interactions between randomization groups with trial-time, and analyses that excluded the first 2 years (for cancer endpoints) or 1 year (for CVD endpoints) of follow-up assessed whether treatment effects varied over time.(4) Ten subgroup analyses examined effect modification by *a priori* (concurrent cocoa extract randomization assignment, sex, age, history of cancer, use of statins or aspirin) and *post hoc* (smoking status, fruit and vegetable intake at baseline, use of dietary supplements or multivitamins prior to randomization) factors. We did not adjust P-values or confidence intervals for multiple testing. Secondary outcomes mostly constituted subtypes of cancer or CVD, and numbered fewer than twenty. Results for secondary and exploratory outcomes and those for subgroup analyses should therefore be interpreted with caution and considered hypothesis generating. At the nominal 0.05 level, we would expect less than one interaction and one secondary outcome to be significant by chance alone.

Per-protocol analyses censored follow-up when the participant discontinued trial pills, began outside (non-study) use of a MVM, took outside (non-study) vitamin D > 1000 IU/day, and/or took approximately <75% of study pills (missed >8 days of study pills per month). Hazard ratios (HRs) and 95% confidence intervals (CIs) were then estimated using Cox regression models, weighted by the inverse probability of dependent-censoring for non-compliance and developed post hoc.(5)

For the biomarker analysis, first we explored the distribution of the biomarkers. The distribution of baseline 25-hydroxy-vitamin D (25(OH)D) (n = 399) was mostly symmetric and so

data was analyzed on the untransformed scale. The distribution of baseline vitamin B_{12} was skewed so data was analyzed on the log-scale. There were 13 participants had vitamin B_{12} assay values of ">1500" assumed to be above 1500 pg/mL; these values were assigned a value of 1501 pg/mL. For serum folate, 48.9% of the cohort had values explicitly coded as ">22.3" ng/mL at baseline. Therefore, analyses examined the influence of MVM on serum folate coded folate as a binary variable (≤22.3 ng/mL vs. >22.3 ng/mL).

Finally, we compared self-reports of non-monitored outcomes and potential side effects by intervention group. Analyses were performed using SAS 9.4.

References

1. Cox DR. Regression models and life-tables. J Royal Statist Soc B 1972;34:187-220.
2. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 1999;18(6):695-706.
3. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med 2007;26(11):2389-430.
4. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer, 2000.
5. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000;56(3):779-88. doi: 10.1111/j.0006-341x.2000.00779.x.