See corresponding editorial on page 416.

Estimating the effect of nutritional interventions using observational data: the American Heart Association's 2020 Dietary Goals and mortality

Yu-Han Chiu,[1,2] Jorge E Chavarro,[1,3,4] Barbra A Dickerman,[1] JoAnn E Manson,[1,4,5] Kenneth J Mukamal,[4,6] Kathryn M Rexrode,[7] Eric B Rimm,[1,3,4] and Miguel A Hernán[1,8,9]

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [2]Mongan Institute, Massachusetts General Hospital, Boston, MA, USA; [3]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA; [4]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [5]Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [6]Division of General Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA; [7]Division of Women's Health, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [8]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA; and [9]Harvard-MIT Division of Health Sciences and Technology, Boston, MA, USA

ABSTRACT

Background: Because randomized trials of sustained dietary changes are sometimes impractical for long-term outcomes, the explicit emulation of a (hypothetical) target trial using observational data may be an important tool for nutritional epidemiology.

Objectives: We describe a methodological approach that aims to emulate a target trial of dietary interventions sustained over many years using data from observational cohort studies.

Methods: We estimated the 20-y risk of all-cause mortality under the sustained implementation of the food-based goals of the American Heart Association (AHA) 2020 using data from 3 prospective observational studies of US men [Health Professionals Follow-up Study (HPFS)] and women [Nurses' Health Study (NHS) and Nurses' Health Study II (NHS II)]. We applied the parametric g-formula to estimate the 20-y mortality risk under a dietary intervention and under no dietary intervention.

Results: There were 165,411 participants who met the eligibility criteria. The mean age at baseline was 57.4 y (range, 43–82 y) in the HPFS, 52.4 y (range, 39–66 y) in the NHS, and 40.2 y (range, 30–50 y) in the NHS II. During 20 y of follow-up, 13,241 participants died. The estimated 20-y mortality risks under a dietary intervention versus no intervention were 21.9% compared with 25.8%, respectively, in the HPFS (risk difference, −3.9%; 95% CI: −4.9% to −3.2%); 10.0% compared with 12.6%, respectively, in the NHS (risk difference, −2.6%; 95% CI: −3.1% to −1.8%); and 2.1% compared with 2.5%, respectively, in the NHS II (risk difference, −0.35%; 95% CI: −0.56% to −0.09%). The corresponding risk ratios were 0.85 (95% CI: 0.81–0.88) in the HPFS, 0.79 (95% CI: 0.75–0.85) in the NHS, and 0.86 (95% CI: 0.78–0.96) in the NHS II.

Conclusions: We estimated that adherence to the food-based AHA 2020 Dietary Goals starting in midlife may reduce the 20-y risk of mortality. *Am J Clin Nutr* 2021;114:690–703.

Keywords: target trial, nutritional epidemiology, American Heart Association 2020 Dietary Goals, mortality, g-formula

Supported by National Institutes of Health (NIH) grants UM1 CA186107, UO1 CA167552, UO1 CA176726, U01 HL 145386, R01 HL034594, R01 HL35464, R01 HL088521, and P01 CA87969. Y-HC is supported by American Heart Association Grant #834106. BAD is supported by NIH grant K99 CA248335.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors assume full responsibility for analyses and interpretation of these data.

Supplemental Figure 1, Supplemental Tables 1–7, Supplemental Methods, and Supplemental Text are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

Address correspondence to Y-HC (e-mail: yuc187@mail.harvard.edu).

Abbreviation used: AHA, American Heart Association; CVD, cardiovascular disease; DASH, Dietary Approaches to Stop Hypertension; HPFS, Health Professionals Follow-up Study; ICD, International Classification of Diseases; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II; PREDIMED, Prevención con Dieta Mediterránea.

Received August 17, 2020. Accepted for publication March 9, 2021.

First published online May 27, 2021; doi: https://doi.org/10.1093/ajcn/nqab100.

Introduction

Estimating the long-term effects of dietary interventions on human health outcomes—a prominent task of nutritional epidemiology—relies, with important exceptions (1–7), on observational studies. These effect estimates may be confounded because people who eat differently tend to have different lifestyles, health histories, and health-care utilization patterns. Further, assessments of what people eat over long periods are typically self-reported, which may result in a large amount of measurement bias. As a result of the potential for confounding and measurement error, some individuals have criticized the utility of nutritional epidemiology (8–11). In a recent assessment, methodological standards designed to evaluate the quality of randomized trials were applied to observational dietary studies which , predictably, failed the test (12).

However, critics of nutritional epidemiology often do not acknowledge that lengthy, massive, randomized trials of sustained dietary changes may be impractical or face formidable logistical challenges to maintain long-term dietary adherence (13–16). Thus, these criticisms leave unanswered the question of how scientists should attempt to identify the long-term health effects of diet in human populations. A possible answer, which we do not endorse, is that nutritional epidemiology studies need to be suspended until long-term dietary assessments are improved and perhaps complemented by new data sources (e.g., grocery receipts) to help capture habitual diet and time-varying confounders. Another answer, which we offer here, is that attempting to estimate the effects of diet on human health using the currently available data remains a legitimate scientific goal.

This paper describes a 2-step approach to specifying and estimating the causal effects of dietary interventions using observational cohorts. The first step is the description of the protocol of a (hypothetical) target trial. The second step is the emulation of the target trial using the available observational data (17, 18). An explicit emulation of a target trial approach has been previously attempted for a variety of interventions (19–24), including dietary ones (25, 26).

For illustration, we estimate the 20-y mortality risk in 3 large observational cohorts under the sustained implementation of food-based goals of the American Heart Association (AHA) (27). We explicitly defined the dietary strategies, presented absolute risk estimates, and used methods that appropriately account for joint interventions and time-varying confounding.

Methods

The American Heart Association 2020 Dietary Goals

The AHA 2020 Strategic Impact Goals are a set of behavioral strategies and health factors for improving cardiovascular health (27). The goals for diet include 5 primary elements (fruits and vegetables, fish, whole grains, sodium, and sugar-sweetened beverages) and 3 secondary elements (nuts/legumes/seeds, processed meat, and saturated fat), and are consistent with the existing Dietary Approaches to Stop Hypertension (DASH) dietary patterns (6). Two of the goals are expressed in terms of nutrients (i.e., sodium and saturated fat) which, compared with food items (e.g., fish), are *1*) more prone to measurement error; and *2*) less well-defined targets for intervention (28). In an attempt to translate our estimates into more actionable recommendations, we restricted our attention to the recommendations based on food items only (**Table 1**).

A randomized trial to estimate the effect of the AHA 2020 Dietary Goals on mortality does not exist. We conceptualized our observational analyses as an attempt to explicitly emulate such a target trial (17, 18). Below, we first specify the protocol of the target trial, and then describe its emulation using observational data.

Target trial specification

Table 2 summarizes the key components of the target trial. Briefly, the trial would include individuals aged ≥25 y without a history of diabetes, cardiovascular disease (CVD), or cancer, who would be randomly assigned to 1 of the dietary strategies described below. The primary outcome of interest would be all-cause death. The secondary outcomes of interest would be death from CVD (including heart disease and stroke) and death from any type of cancer. Each eligible participant would be followed from assignment until death, incomplete follow-up, or administrative end of follow-up (20 y after assignment), whichever happens first.

Dietary strategies.

Each individual would be assigned to 1 of the strategies, which would be followed for 20 y:

- No intervention (usual diet)
- Joint intervention on all 6 food-based components of the AHA 2020 Dietary Goals: fruits and vegetables ≥4.5 servings/d; fish ≥2 servings/wk for nonvegetarians, and no intervention on fish intake for vegetarians; whole grains ≥3 servings/d; sugar-sweetened beverages ≤3 servings/wk; processed meats ≤2 servings/wk; and nuts/seeds/legumes ≥4 servings/wk.

Participants assigned to a dietary strategy would be expected to maintain their dietary intake within the range prespecified by the corresponding intervention. For example, an individual assigned to "at least 2 servings of fish per week" would have to eat 2 or more servings per week, which may be operationalized as follows (29): at the start of each week, the individual would be asked how many servings of fish she would eat if she were now reassigned to "no intervention." If the answer is 2 or more, then the individual would be instructed to make no dietary changes. If the answer is less than 2, the individual would be instructed to eat exactly 2 servings: that is, to reach the minimum threshold of servings compatible with the intervention. These so-called threshold interventions maintain diet within a prespecified range (e.g., at least 2 servings/wk of fish), while minimizing the number of individuals who require intervention (30, 31).

To determine the contribution of a specific dietary goal to the joint effects of interventions on several dietary factors (6, 32), we included 12 additional arms:

- Intervention on only 1 of the components (6 separate strategies)

TABLE 1 Food-based American Heart Association 2020 Dietary Goals

Food groups	AHA recommendations	Example foods
Whole grains	≥3 servings/d	Whole-grain breakfast cereal, other cooked breakfast cereal, cooked oatmeal, dark bread, brown rice, other grains, bran, wheat germ, popcorn
Fruits andvegetables	≥4.5 servings/d	Fruits: Raisins or grapes, prunes, bananas, cantaloupe, watermelon, fresh apples or pears, oranges, grapefruit, strawberries, blueberries, peaches or apricots or plums Vegetables: Tomatoes, tomato juice, tomato sauce, broccoli, cabbage, cauliflower, brussels sprouts, carrots, mixed vegetables, yellow or winter squash, eggplant or zucchini, yams or sweet potatoes, spinach cooked, spinach raw, kale or mustard or chard greens, iceberg or head lettuce, romaine or leaf lettuce, celery, mushrooms, beets, alfalfa sprouts, garlic, or corn
Fish	≥2 servings/wk	Canned tuna, dark-meat fish (e.g., tuna steak, mackerel, salmon, sardines, bluefish, swordfish), other fish (e.g., cod, haddock, halibut). Breaded fish is not included.
Processed meat	≤2 servings/wk	Bacon, salami, bologna, processed meat sandwiches, and other processed meats (e.g., sausage, kielbasa)
Sugar-sweetened beverages	≤3 servings/wk	Carbonated beverages with caffeine & sugar (e.g., Coke, Pepsi, Mt. Dew, Dr. Pepper), carbonated beverages with sugar (e.g., 7-Up, root beer, ginger ale, caffeine-free Coke), and other sugared beverages (e.g., punch, lemonade, sports drinks, or sugared iced tea)
Legumes, nuts, and seeds	≥4 servings/wk	Legumes: string beans, beans or lentils. Soy is not included Nuts: peanuts, peanut butter, walnuts, other nuts Seeds: flaxseed

The original AHA 2020 Dietary Goals also included sodium <1500 mg/d and saturated fat <7% of total energy. Abbreviation: AHA, American Heart Association.

- Joint intervention on 5 of the 6 components (6 separate strategies, leaving 1 component out under each strategy).

Causal contrasts and statistical analysis.

The primary (intention-to-treat) effect in the target trial would be estimated by comparing the 20-y mortality risk across groups assigned to each strategy (with adjustment for loss to follow-up, if necessary). However, the information provided by the intention-to-treat effect would be limited if, as expected, many individuals deviated from their dietary assignments during the 20-y follow-up. In this setting, a contrast of the mortality risks that would have been observed if all individuals had adhered to their assigned dietary strategy (i.e., per-protocol effect) may be more relevant for setting dietary goals (33). These risks can be estimated in the target trial using the parametric g-formula (34, 35) under the assumptions of no unmeasured confounding and selection bias (incomplete follow-up is expanded to include questionnaire nonresponses and incomplete responses to dietary questions) (33), no measurement error, and no model misspecification.

The parametric g-formula has been described elsewhere (36, 37). Briefly, the method is a generalized form of standardization in which the standardized risk of the outcome is calculated as a weighted average of the outcome risks conditional on the time-varying confounders, with the distribution of the time-varying confounders used as weights. The method starts by estimating the distribution of the time-varying confounders and of the outcome using parametric (e.g., linear, logistic) regression models with previous dietary and covariate histories as covariates. For each dietary strategy, the conditional probabilities of the outcome given past covariates are then calculated under dietary values compatible with the intervention. Under the above assumptions, the probability of the outcome that would have been observed if everyone in the population had adhered to the dietary strategy is the standardized (to confounder history) risk. The weighted average required for the standardization is approximated via a Monte Carlo simulation. These standardized probabilities can then be used to estimate the risk (cumulative incidence) of cause-specific mortality or its corresponding survival (1 minus the risk) curves. Nonparametric bootstrapping with 500 samples can be used to construct percentile-based 95% CIs of the estimated risks at the time points of interest. In the analyses of cardiovascular or cancer mortality, in which death from other causes is a competing event, we can estimate the risk (cumulative incidence) of cause-specific mortality (38).

To identify potential subgroups of patients for whom the dietary strategies may be more beneficial, analyses can be conducted separately in subsets of the study population defined at baseline according to baseline age (<50 y versus ≥50 y), baseline BMI (<25 kg/m^2 versus ≥25 kg/m^2), or baseline physical activity levels (<median versus ≥ median hours/week).

Target trial emulation

The above multi-arm trial with a 20-y follow-up is unlikely to be conducted. Therefore, we emulated it using observational data from 3 large prospective cohorts: the Health Professionals Follow-up Study (HPFS), Nurses' Health Study (NHS), and Nurses' Health Study II (NHS II).

Observational data.

The HPFS is a prospective cohort study established in 1986 with an enrolment of 51,529 male health professionals aged 40 to 75 y from 50 US states (39). The NHS is a prospective cohort study established in 1976 with an enrolment of 121,701 female registered nurses aged 30 to 55 y from 11 US states (40). The NHS II is a prospective cohort study established in 1989 with an enrollment of 116,429 female registered nurses aged 25 to 42 from 15 US states (40). In all cohorts, self-administered

TABLE 2 Emulation of a target trial of dietary interventions using observational data from the Health Professionals Follow-up Study, Nurses' Health Study, and Nurses' Health Study II

	Target trial specification	Target trial emulation
Eligibility criteria	Age ≥25 y, no history of diabetes, cardiovascular disease, and cancer.	Same. We also required complete questions on diet and covariates and report plausible energy intake (800 to 4200 kcal/d in men; 500 to 3500 kcal/d in women) at prebaseline and baseline questionnaires. • Baseline is defined as the date of return of the second dietary questionnaire (1990 for HPFS, 1986 for NHS, and 1995 for NHS II) to allow for adjustment for prebaseline diet.
Dietary strategies	Each individual would be assigned to 1 of 14 following strategies: • No intervention (usual diet) • Joint intervention on all 6 food-based components of the AHA 2020 Dietary Goals • Intervention on only 1 of the components (6 separate strategies) • Joint intervention on 5 of the 6 components (6 strategies, leaving 1 component out under each strategy) Each strategy is followed for 20 y. Fish interventions apply to nonvegetarians only. Participants assigned to a dietary strategy are expected to maintain their dietary intake within the range prespecified by the corresponding intervention.	Same. We assumed that each 4-y dietary questionnaire accurately reflects *1*) the average diet during the previous 4-y period; and *2*) the intended diet (under no intervention) that the individual would have reported at the start of the 4-y period.
Assignment	Individuals are randomly assigned to a dietary strategy.	We attempted to emulate randomized assignment by adjusting for prebaseline or baseline covariates: baseline age at enrollment; family history of myocardial infarction before 60 y; smoking index; aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); baseline diagnosis of hypertension or hypercholesterolemia; and prebaseline values of fruits and vegetables, whole grains, processed meat, fish, sugar- sweetened beverages, legumes/nuts/seeds, and alcohol; and total energy intake.
Outcome	Primary outcome: 20-y risk of all-cause mortality. Secondary outcomes: 20-y risk of death from CVD, cancer, and other causes.	Same.
Follow-up	Starts at baseline and ends at death, incomplete follow-up, or 20 y after baseline, whichever occurs first.	Same. Incomplete follow-up is defined as questionnaire nonresponse or incomplete responses to dietary questions.
Causal contrast	Intention-to-treatment effect. Per-protocol effect.	Observational analog of per-protocol effect.
Statistical analysis	Intention-to-treat analysis. Per-protocol analysis: Apply g-formula to compare 20-y risk of death between groups receiving each treatment strategy with adjustment for pre- and postbaseline prognostic factors associated with adherence to strategies and loss to follow-up.	Same as per-protocol analysis.

Abbreviations: AHA, American Heart Association; CVD, cardiovascular disease; HPFS, Health Professionals Follow-Up Study; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

questionnaires were distributed every 2 y to collect information about medical history, lifestyle, and health conditions.

Dietary information was collected using an FFQ (41–47) that was first sent to participants in the HPFS in 1986, to participants in the NHS in 1984 and 1986, to participants in the NHS-II in 1991, and to all participants every 4 y afterward. The FFQ evaluated average consumption of specified proportions of foods and beverages during the previous year (41–47). We truncated dietary intake values at their 99th percentiles to prevent extreme values from affecting the analyses. Height, race, and parental history of myocardial infarction before the age of 60 y were ascertained at the first questionnaire in each cohort.

Self-reported diagnoses at the follow-up questionnaires were verified by medical records review. Deaths were identified through a search of the vital statistics records of states and of the National Death Index, supplemented by reports from next of kin and the US postal system. A physician blinded to questionnaire data classified the cause of death by reviewing medical records and death certificates according to the International Classification of Diseases (ICD), Eighth and Ninth Revisions. CVD mortality was defined as ICD-8 codes 390–458 or ICD-9 codes 390–459,

cancer mortality was defined as ICD-8 codes 140–207 or ICD-9 codes 140–208, and mortality from external causes of injury and poisoning was defined as ICD-8 E800-E999 or ICD-9 codes E800-E999.

Modifications to the target trial protocol.

To emulate the above target trial using these observational data (see last column of Table 2), we identified individuals in the above 3 cohorts who met all eligibility criteria. We also excluded participants who reported implausible energy intake (<800 or >4200 kcal/d in men; <500 or >3500 kcal/d in women) at prebaseline or baseline. Eligible individuals were followed from the return of their baseline questionnaire until death, incomplete follow-up (nonreturn of a questionnaire or incomplete responses to dietary questions), or 20 y after baseline, whichever happened first. To allow for adjustments of prebaseline diet, we defined the baseline questionnaire as the response with the second FFQ (1990 for HPFS, 1986 for NHS, and 1995 for NHS-II). However, the available observational data impose significant assumptions for the emulation of the target trial protocol.

First, the dietary strategies described above cannot be directly emulated, because dietary data were collected every 4 y rather than weekly and because individuals were not asked about their intended diet in the absence of an intervention. Therefore, we had to assume that each dietary FFQ accurately reflects *1*) the average diet during the previous 4-y period; and *2*) the intended diet (under no intervention) that the individual would have reported at the start of the 4-y period.

Second, we attempted to emulate the randomized assignment to the dietary strategies by adjusting for potential baseline confounders: age at enrollment; parental history of myocardial infarction before the age of 60 y; smoking status; BMI; physical activity; aspirin use; diagnosis of hypertension or hypercholesterolemia; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); and prebaseline (1 questionnaire before baseline) values of alcohol, fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, and total energy intake.

We do not expect that these assumptions of accurate measurements and complete confounding adjustments will hold exactly; typically, diet will not remain constant for periods of 4 y. However, we expect that the richness and periodic updating of the available observational data will allow to approximately characterize diet over 2 decades and to adjust for much confounding. Below, we explore the sensitivity of our estimates to variations of these assumptions.

Statistical analysis.

We implemented the parametric g-formula using the above baseline covariates and the following time-varying variables at each questionnaire: BMI; physical activity; number of cigarettes smoked per day; aspirin use; menopausal status (NHS and NHS II only); menopausal hormone therapy use (NHS and NHS II only); intakes of alcohol, fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, and total energy; incidences of hypertension, hypercholesterolemia, diabetes, cancer, nonfatal coronary heart disease; and time since report of each diagnosis (see **Supplemental Table 1** for details on functional form and models used for each variable). When a time-varying covariate was not assessed in a 2-y period, we carried forward the value from the last interval in which it was measured, and accounted for this information by adding to the model a product term between the most recent measurement and the time since that measurement.

We conducted additional analyses to evaluate the sensitivity of the estimates to incomplete confounding adjustments: *1*) restricting to never-smokers at baseline; *2*) additionally adjusting for a baseline healthy behavior score (regular multivitamin use, routine physical examinations, rectal examination, mammography or sigmoidoscopy/coloscopy for screening); *3*) adjusting for covariates from previous questionnaires rather than those concurrently measured with diet in a given questionnaire; and *4*) using mortality from external causes of injury and poisoning as a negative outcome control (48).

We also conducted analyses to evaluate the sensitivity of the estimates to model misspecification: *1*) using different functional forms for covariates (replaced cubic spline with categorical variables for age); and *2*) changing the order of time-varying covariates concurrently reported in the same questionnaire. A description of the differences between our approach and conventional analyses can be found in the **Supplemental Methods**. All analyses were conducted using SAS 9.4 software (SAS Institute) and the GFORMULA macro, which is publicly available at http://www.hsph.harvard.edu/causal/software.

Consent and approval

The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and the Harvard T.H. Chan School of Public Health, and those of participating registries as required. The return of the completed self-administered questionnaires was considered to imply informed consent.

Results

Baseline characteristics of the participants

Of 165,411 eligible individuals, 32,685 were men from the HPFS, 60,635 were women from the NHS, and 72,091 were women from the NHS II (**Figure 1**). **Table 3** shows the baseline characteristics of the participants from each cohort. The mean ages were 57.4 y (range, 43–82 y) in the HPFS, 52.4 y (range, 39–66 y) in the NHS, and 40.2 y (range, 30–50 y) in the NHS II. **Supplemental Table 2** shows the proportion of individuals who met the food-based AHA recommendations at baseline.

During the 20-y follow-up, there were 6221 deaths in the HPFS (2173 from CVD, 2551 from cancer, 1497 from other causes), 5847 deaths in the NHS (1154 from CVD, 3487 from cancer, 1206 from other causes), and 1173 deaths in the NHS II (127 from CVD, 672 from cancer, 374 from other causes). The observed 20-y mortality risks were 24.9% in the HPFS, 11.8% in the NHS, and 2.3% in the NHS II. No individuals adhered to all AHA Dietary Goals throughout the entire follow-up period.

All-cause mortality

Figure 2 shows the estimated 20-y survival curves under *1*) no dietary intervention; and *2*) joint intervention on the

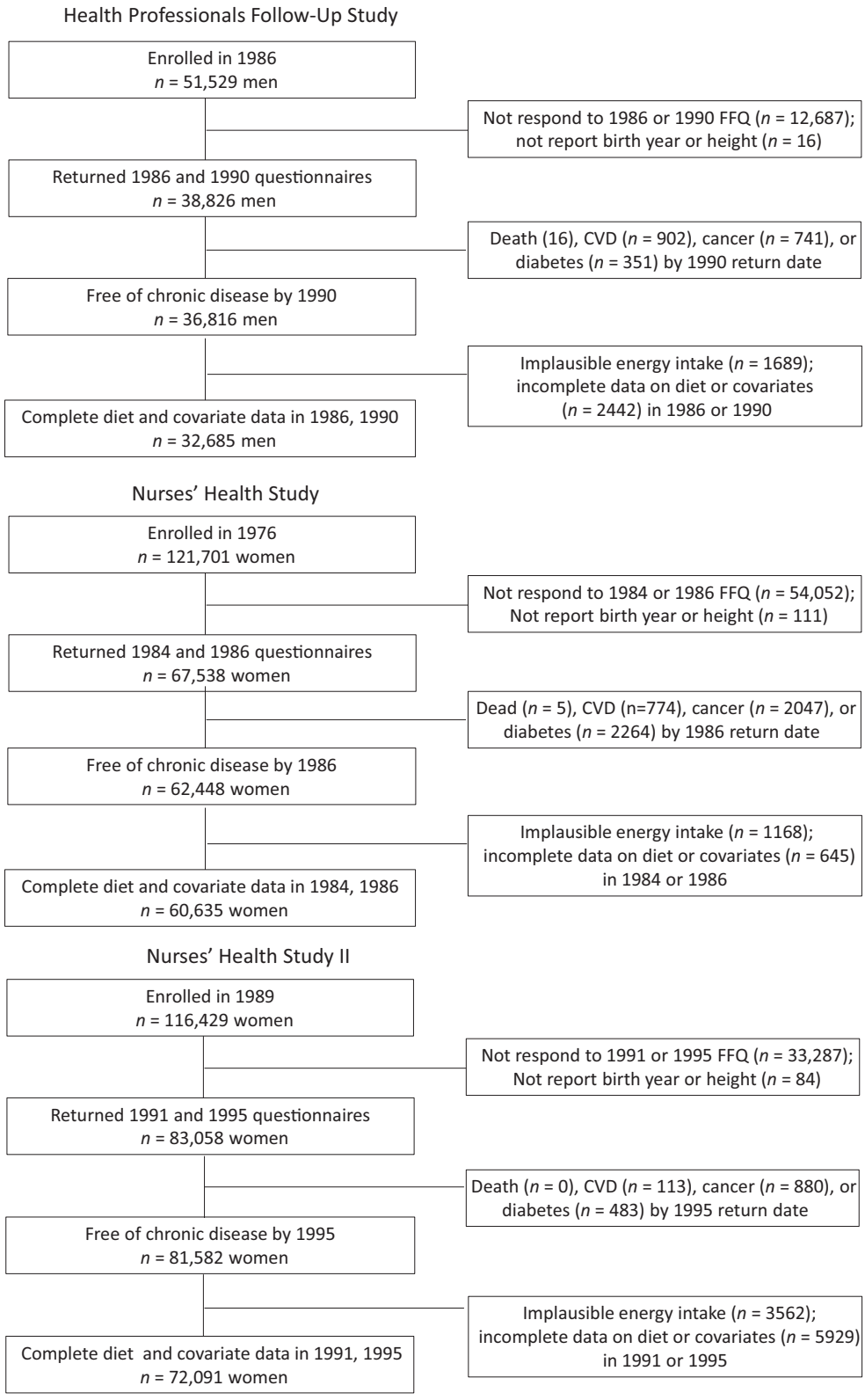


FIGURE 1 Flowchart of eligible individuals for the emulation of a target trial of dietary interventions in the Health Professionals Follow-Up study, 1990–2010; Nurses' Health Study, 1986–2006; and Nurses' Health Study II, 1995–2015. Abbreviation: CVD, cardiovascular disease.

TABLE 3 Baseline characteristics of eligible participants in the Health Professionals Follow-Up Study, Nurses' Health Study, and Nurses' Health Study II

Characteristics	HPFS (1990) $N = 32,685$	NHS (1986) $N = 60,635$	NHS II (1995) $N = 72,091$
Age, year	57.4 (9.6)	52.4 (7.2)	40.2 (4.7)
BMI, kg/m^2	25.5 (3.1)	25.1 (4.6)	25.7 (5.8)
Physical activity, hours/wk	7.2 (7.1)	2.9 (3.3)	3.2 (3.6)
Total energy intake, kcal/d	1921 (575)	1767 (517)	1806 (545)
Dietary intake, servings/d			
Fruits and vegetables	4.59 (2.38)	5.01 (2.52)	4.37 (2.41)
Fish	0.33 (0.28)	0.30 (0.24)	0.18 (0.17)
Whole grains	1.84 (1.47)	1.21 (1.05)	1.34 (1.08)
Sugar-sweetened beverages	0.31 (0.51)	0.22 (0.45)	0.43 (0.74)
Legumes/nuts/seeds	0.74 (0.55)	0.63 (0.42)	0.49 (0.36)
Processed meat	0.28 (0.32)	0.28 (0.28)	0.18 (0.20)
Caucasian/White, %	91.9	98.0	94.8
Current smoker, %	7.3	20.9	10.5
Family history of MI (at age <60 y), %	12.4	25.4	28.3
Multivitamin supplement use, %	39.0	43.0	48.0
Aspirin user, %	28.9	71.3	24.4
Premenopausal, %	—	43.0	86.6
Menopausal hormone user, %	—	17.6	7.7
Hypertension, %	25.9	23.7	9.3
Hypercholesterolemia, %	30.5	11.9	21.1

Values are mean (standard deviation) or %. Abbreviations: HPFS, Health Professionals Follow-Up Study; MI, myocardial infarction; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

6 food-based components of the AHA 2020 Dietary Goals. **Table 4** shows the estimated risk of all-cause mortality under these 2 strategies. The estimated 20-y risks of all-cause mortality under no intervention were 25.8% (95% CI: 25.5%–27.1%) in the HPFS, 12.9% (95% CI: 11.9%–12.8%) in the NHS, and 2.5% (95% CI: 2.2%–2.6%) in the NHS II. When comparing dietary intervention with no intervention, the estimated 20-y risk differences were −3.9% (95% CI: −4.9% to −3.2%) in the HPFS, −2.6% (95% CI: −3.1% to −1.8%) in the NHS, and −0.35% (95% CI: −0.56% to −0.09%) in the NHS-II. The corresponding risk ratios were 0.85 (95% CI: 0.81–0.88) in the HPFS, 0.79 (95% CI: 0.75–0.85) in the NHS, and 0.86 (95% CI: 0.78–0.96) in the NHS II. The average proportion of participants who would have been required to change their diet to adhere to the food-based AHA 2020 Dietary Goals (had they adhered through the previous period) was close to 50% at any period in each of the 3 cohorts (Table 4).

We also estimated the mortality risk under variations of the AHA 2020 Dietary Goals by omitting 1 dietary goal at a time (**Figure 3**; **Supplemental Table 3**). The estimates remained similar under most variations. The mortality risk differences for single-food interventions compared with no intervention were smaller than those for joint interventions in all 3 cohorts (**Figure 4**; **Supplemental Table 4**). Compared with no intervention, the risk difference for a single-food intervention was larger for whole grain than for other individual foods in the HPFS and NHS, and was close to null for every single-food intervention in the NHS II. The risk difference for the joint dietary intervention compared with no intervention increased with the duration of the intervention (**Supplemental Table 5**).

The estimated risks and risk differences were similar in subgroups defined by baseline BMI and physical activity (**Supplemental Table 6**). However, the estimated risk difference was larger among participants ≥50 y than those <50 y (Supplemental Table 6). For example, in the HPFS, the estimated 20-y risk difference of total mortality was −5.9% (95% CI: −7.2% to −4.7%) for participants ≥ 50 y, compared with −1.6% (95% CI: −2.4% to −0.6%) for participants <50 y.

When using a conventional analysis, the 2-y mortality HR estimates for adherence compared with no adherence to the AHA goals were 0.73 (95% CI: 0.58–0.91) in the HPFS, 0.97 (95% CI: 0.76–1.25) in the NHS, and 1.20 (95% CI: 0.67–2.14) in the NHS II (see Supplemental Methods for details).

Cause-specific mortality

Compared with no intervention, the estimated 20-y risks of CVD mortality and cancer mortality under the AHA 2020 food-based goals were lower in the HPFS and NHS participants but similar in the NHS II participants (**Table 5**). Specifically, the estimated risks of CVD mortality under the joint dietary intervention versus no intervention were 9.1% (95% CI: 8.8%–9.7%) compared with 8.0% (95% CI: 7.4%–8.9%) in the HPFS, 2.5% (95% CI: 2.3%–2.6%) compared with 1.9% (95% CI: 1.6%–2.1%) in the NHS, and 0.28% (95% CI: 0.21%–0.32%) compared with 0.24% (95% CI: 0.14%–0.32%) in the NHS II. The estimated risks of the 20-y risk of cancer mortality under a dietary intervention compared with no intervention were 8.8% (95% CI: 8.6%–9.4%) compared with 7.8% (95% CI: 7.1%–8.7%) in the HPFS, 6.4% (95% CI: 6.0%–6.6%) compared with 5.8% (95% CI: 5.3%–6.4%) in the NHS, and 1.2% (95% CI:

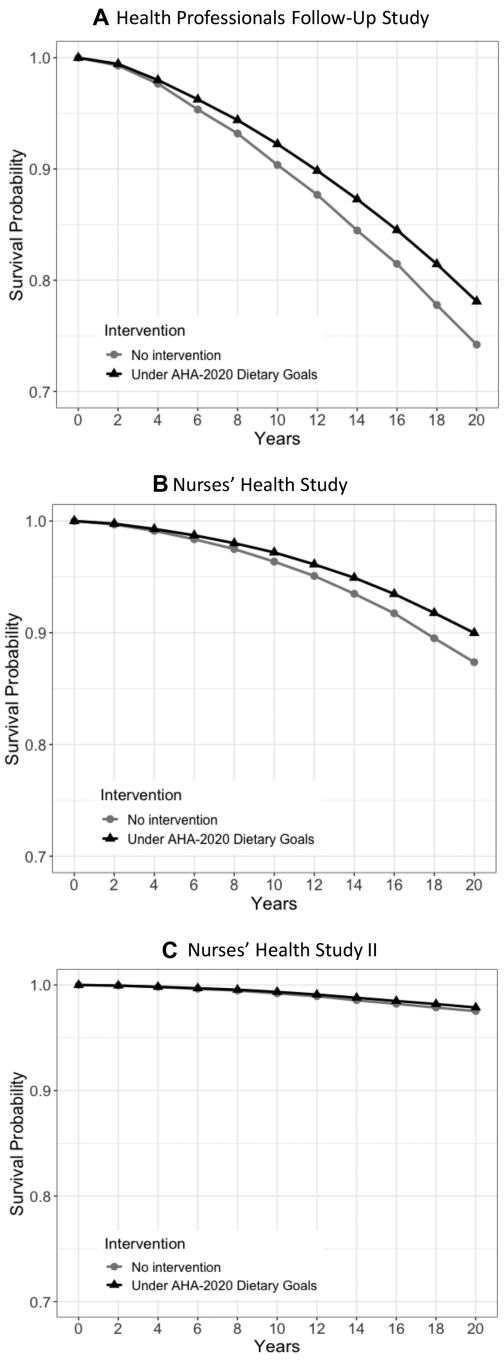


FIGURE 2 Estimated survival under hypothetical dietary interventions compared with no intervention in the Health Professionals Follow-Up study, the NHS, and the NHS II. Estimates are based on the parametric g-formula with baseline and prebaseline covariates: baseline age; BMI; smoking status; physical activities; parental history of myocardial infarction (<60 y); aspirin use; baseline diagnosis of hypertension or hypercholesterolemia; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); prebaseline intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; and time-varying covariates: BMI; cigarette smoked per day; physical activity; aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; incidences of hypertension, hypercholesterolemia, diabetes, cancer, nonfatal myocardial infarction; and time since report of each diagnosis. Abbreviations: AHA, American Heart Association; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

1.0%–1.3%) compared with 1.1% (95% CI: 0.9%–1.3%) in the NHS II. Table 5 shows the corresponding risk differences and risk ratios. When using a conventional analysis, the 2-y mortality HR estimates for adherence compared with no adherence to the AHA goals were 0.68 (95% CI: 0.46, 1.00) in the HPFS, 0.51 (95% CI: 0.23–1.15) in the NHS, and 1.03 (95% CI: 0.14–7.50) in the NHS II for CVD mortality and were 0.83 (95% CI: 0.62–1.15) in the HPFS, 1.28 (95% CI: 0.97–1.70) in the NHS, and 1.46 (95% CI: 0.71–2.98) in the NHS II for cancer mortality (see Supplemental Methods for details).

Sensitivity analyses

The estimates did not materially change under any of the sensitivity analyses (**Supplemental Table 7**). The estimated risk difference of mortality from external causes, a negative outcome control, was close to null for the 3 cohorts (Supplemental Table 7).

Discussion

We described how to specify and emulate a target trial to estimate the causal effects of dietary interventions using data from observational cohorts. For illustration, we described and attempted to emulate a target trial of the food-based AHA 2020 Dietary Goals in middle-aged healthy men and women from 3 cohorts. Our results were very compatible with decreases in 20-y mortality, ranging from between 0.09 and 0.56 percentage points (in the NHS II) to between 3.2 and 4.9 percentage points (in the HPFS) for continuous adherence to the AHA goals compared with no dietary changes. The estimated beneficial effect decreased when we considered interventions sustained over shorter periods and, for the 2 older cohorts, when we considered interventions that did not involve the intake of whole grains.

The validity of our effect estimates cannot be directly confirmed because no randomized trials have assessed the effect of implementing the AHA 2020 Dietary Goals on the risk of mortality. However, the AHA 2020 Goals are generally consistent with the Mediterranean diet: both include increased intakes of fruits and vegetables, nuts and legumes, and fish, and limited intakes of sugar-sweetened beverages and processed red meat. The PREDIMED (Prevención con Dieta Mediterránea) randomized trial compared a related intervention (Mediterranean diet supplemented with extra-virgin olive oil or supplemented with nuts) versus a control (reduced-fat) diet among participants aged 55–80 y at high cardiovascular risk (1). Over 4.8 y of follow-up, the trial found no difference in all-cause mortality in the intervention arms compared with the control arm [absolute risks for Mediterranean diet with extra-virgin oil, Mediterranean diet with nuts, and control group were 4.4% (95% CI: 3.6%–5.4%), 5.4% (95% CI: 4.4%–6.6%), and 5.4% (95% CI: 4.4%–6.7%), respectively.] If we stopped our emulated trial at 4 y, the estimated mortality risk difference would also be close to null (see Figure 2). Also note that because our study populations were younger, without chronic disease at baseline, and had a healthier baseline dietary quality, the risk reduction is expected to be smaller than that in participants from the PREDIMED.

TABLE 4 Estimated risks of all-cause mortality under dietary strategies consistent with the AHA 2020 Dietary Goals in the Health Professionals Follow-Up Study (1990–2010), Nurses' Health Study (1986–2006), and Nurses' Health Study II (1995–2015)

	20-y risk (95% CI), %	Risk difference (95% CI)	Risk ratio (95% CI), %	Difference in restricted mean survival, months	% needed to change diet[1]
Health Professionals Follow-Up Study					
No dietary intervention	25.8 (25.5–27.1)	0 (reference)	1 (reference)	0 (reference)	—
Dietary Intervention[2]	21.9 (21.1–23.3)	−3.9 (−4.9 to −3.2)	0.85 (0.81–0.88)	5.0 (4.2–6.4)	46.9%
Nurses' Health Study					
No dietary intervention	12.6 (11.9–12.8)	0 (reference)	1 (reference)	0 (reference)	—
Dietary Intervention[2]	10.0 (9.2–10.5)	−2.6 (−3.1 to−1.8)	0.79 (0.75–0.85)	2.7 (1.9–3.1)	47.7%
Nurses' Health Study-II					
No dietary intervention	2.5 (2.2–2.6)	0 (reference)	1 (reference)	0 (reference)	—
Dietary intervention[2]	2.1 (1.9–2.4)	−0.35 (−0.56 to −0.09)	0.86 (0.78–0.96)	0.43 (0.17–0.62)	49.2%

Estimates are based on the parametric g-formula with baseline and prebaseline covariates: baseline age; BMI; smoking status; physical activities; parental history of myocardial infarction (<60 y); aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); baseline diagnosis of hypertension or hypercholesterolemia; and prebaseline intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; and time-varying covariates: BMI; cigarettes smoked per day; physical activity; aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; incidences of hypertension, hypercholesterolemia, diabetes, cancer, and nonfatal myocardial infarction; and time since report of each diagnosis. Abbreviations: AHA, American Heart Association; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

[1] Average proportion of the participants who would need to change their diet in each follow-up period to keep adhering to the dietary strategy.

[2] Joint intervention on all food-based AHA 2020 Dietary Goals: fruits and vegetables ≥4.5 servings/d, whole grains ≥3 servings/d, fish ≥2 servings/wk (on nonvegetarians only), sugar-sweetened beverages ≤3 servings/wk, legumes/nuts/seeds ≥4 servings/wk, and processed meat ≤2 servings/wk.

Previous observational studies have reported a higher baseline or average dietary quality—assessed by dietary indices such as the Healthy Eating Index, Alternative Healthy Eating Index (49), Alternate Mediterranean Diet Score (50, 51), and DASH diet score (52)—was associated with lower risks of all-cause mortality, cardiovascular mortality (53, 54), and cancer mortality (55–57), albeit some reported no associations with cancer mortality (58–60). These estimates do not have a straightforward interpretation, because the dietary strategies under study and the periods over which they are implemented were not clearly defined. Therefore, these studies cannot yield estimates of absolute risk under those strategies.

The use of the target trial approach eliminates the above limitations, which in turns provides direct, actionable insight and context for discussions of confounding and measurement error (17, 18). First, specifying the target trial clarifies the question of interest as a trial would do, including minimum and maximum intakes for each food item, as well as the starting points and durations of the sustained dietary interventions and follow-up periods. Second, the target trial approach yields absolute risks under realistic dietary strategies (e.g., vegetarians are not forced to eat fish) instead of using an average time-varying HR over an unspecified period (61). Third, unlike traditional outcome regression, the g-formula appropriately adjusts for time-varying confounders affected by past exposure (34, 35). Such time-varying confounders are often present when, for example, a newly diagnosed disease (prognostic of the outcome) influences future dietary patterns, and the disease itself can also be affected by past dietary patterns. Fourth, the g-formula naturally incorporates joint estimation of several dietary components. Last, competing events are handled in an explicitly causal way (38).

But our approach does not eliminate the potential for unmeasured confounding, measurement error, and model misspecification. Like in any observational study, we cannot rule out the possibility of unmeasured confounding, despite adjustments for many potential confounders. An estimated null effect on outcomes that were not expected to be influenced by diet (e.g., accidental death) is reassuring, but not a proof of lack of confounding. Also, we cannot rule out bias due to model misspecification, though our estimates under no dietary intervention closely tracked the observed ones (**Supplemental Figure 1**), which suggested no gross model misspecification under the no intervention condition (62). In addition, we relied on FFQs to measure dietary intake, and therefore some degree of measurement error in diet was expected even though the FFQ has been validated against biomarkers and diet records (41, 42, 45–47) and can detect changes in intake (63, 64). Note that per-protocol effect estimation is also susceptible to measurement error in randomized trials, because the adherence to the assigned intervention is also measured by a questionnaire (1, 65). Finally, the results may not be applicable to populations with different dietary practices (because the effects are estimated in comparison with usual diet) or age distributions (because we found differences among cohorts that might be explained by the different age at which the hypothetical intervention started).

In summary, we estimate that adhering to the food-based AHA 2020 Impact Goals reduces the 20-y risk of all-cause mortality.

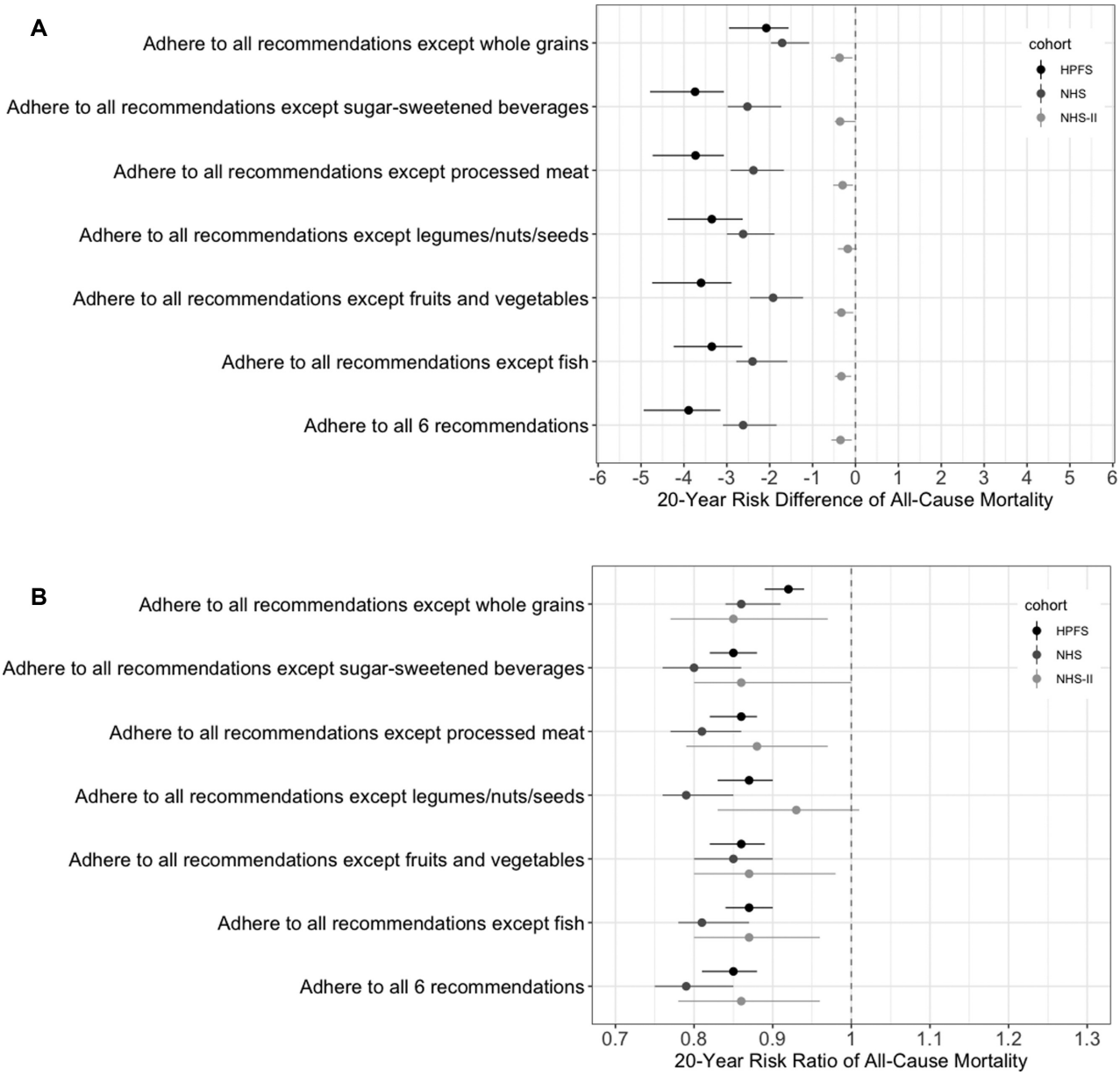
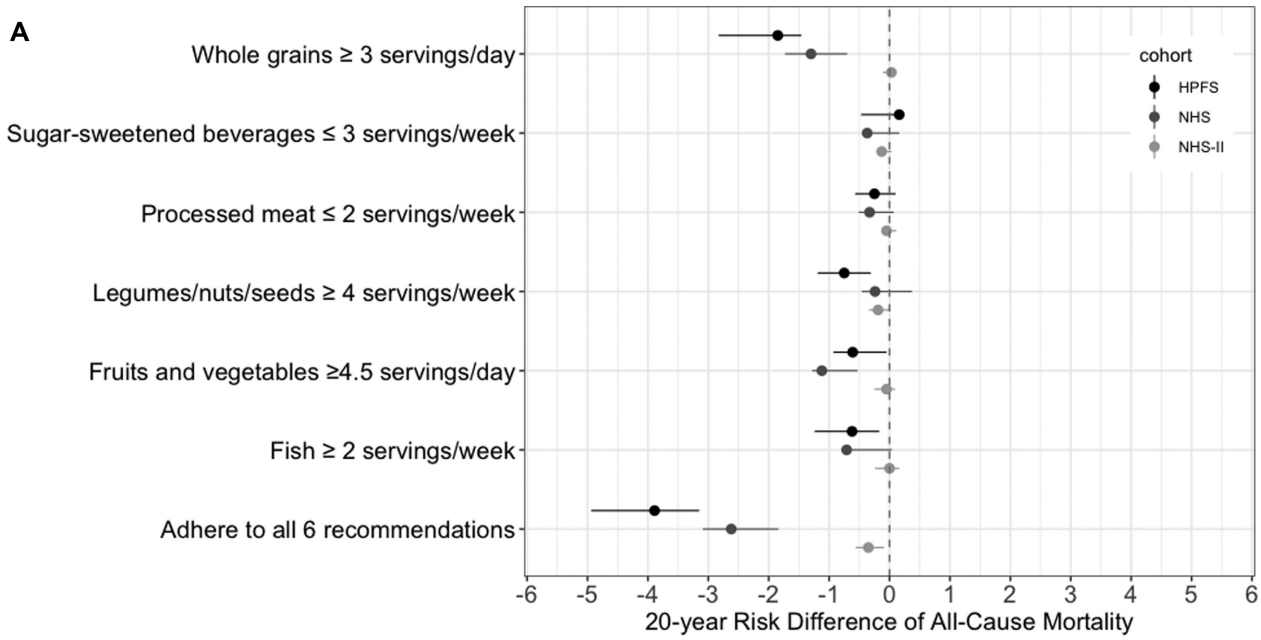


FIGURE 3 Estimated (A) risk difference and (B) risk ratio of 20-y mortality for multi-food dietary strategies derived from the AHA 2020 Dietary Goals compared with no intervention in the HPFS (1990–2010), NHS (1986–2006), and NHS II (1995–2015). Estimates are based on the parametric g-formula with baseline and prebaseline covariates: baseline age; BMI; smoking status; physical activities; parental history of myocardial infarction (<60 y); aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); baseline diagnosis of hypertension or hypercholesterolemia; and prebaseline intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; and time-varying covariates: BMI; cigarettes smoked per day; physical activity; aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; incidences of hypertension, hypercholesterolemia, diabetes, cancer, and nonfatal myocardial infarction; and time since report of each diagnosis. Abbreviations: AHA, American Heart Association; HPFS, Health Professionals Follow-Up Study; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

Our explicit emulation of a target trial helped define and compare the sort of dietary strategies that are used to inform health policy and develop guidelines. Our approach did not allow us to rule out the potential for influential unmeasured confounding and measurement error, but we could not find an alternative to rich longitudinal data from observational cohorts for estimating the effects of long-term diet.

We thank the participants and staff of the Nurses' Health Study (NHS), the Health Professionals Follow-Up Study (HPFS), and Nurses' Health Study II (NHS II) who contributed data for their valuable contributions, as well as the state cancer registries in the following states: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, and WY.



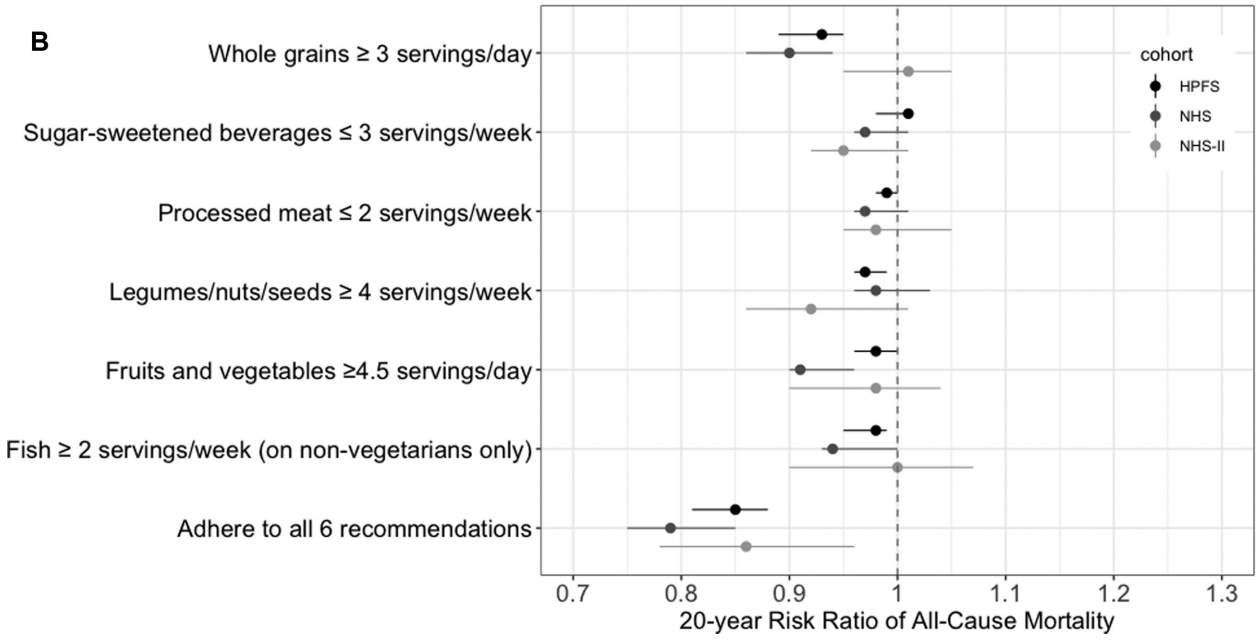


FIGURE 4 Estimated (A) risk difference and (B) risk ratio of 20-y mortality under several single-food interventions derived from the AHA 2020 Dietary Goals compared with no intervention in the HPFS (1990–2010), NHS (1986–2006), and NHS II (1995–2015). Estimates are based on the parametric g-formula with baseline and prebaseline covariates: baseline age; BMI; smoking status; physical activities; parental history of myocardial infarction (<60 y); aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); baseline diagnosis of hypertension of hypercholesterolemia; and prebaseline intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; and time-varying covariates: BMI; cigarettes smoked per day; physical activity; aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; incidences of hypertension, hypercholesterolemia, diabetes, cancer, and nonfatal myocardial infarction; and time since report of each diagnosis. Abbreviations: AHA, American Heart Association; HPFS, Health Professionals Follow-Up Study; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

The authors' responsibilities were as follows—Y-HC and MAH: designed the research, wrote the first draft of the manuscript, and had primary responsibility for the final content; JEC, JEM, KJM, KMR, and EBR: were involved in data collection; Y-HC: analyzed the data; BAD: conducted the technical review; JEC, BAD, JEM, KJM, KMR, and EBR: helped with data interpretation and provided critical revision of the manuscript for important intellectual content; and all authors: read and approved the final manuscript. The authors report no conflicts of interest.

TABLE 5 Estimated risks of cardiovascular mortality and cancer mortality under dietary strategies in the Health Professionals Follow-Up Study (1990–2010), Nurses' Health Study (1986–2006), and Nurses' Health Study II (1995–2015)

	Cardiovascular mortality			Cancer mortality		
	20-y risk (95% CI), %	Risk difference (95% CI)	Risk ratio (95% CI), %	20-y risk (95% CI), %	Risk difference (95% CI)	Risk ratio (95% CI), %
Health Professionals Follow-Up Study						
No dietary intervention	9.1 (8.8–9.7)	0 (reference)	1 (reference)	8.8 (8.6–9.4)	0 (reference)	1 (reference)
Dietary intervention[1]	8.0 (7.4–8.9)	−1.1 (−1.8 to −0.55)	0.88 (0.80–0.94)	7.8 (7.1–8.7)	−1.0 (−1.8 to −0.44)	0.89 (0.80–0.95)
Nurses' Health Study						
No dietary intervention	2.5 (2.3–2.6)	0 (reference)	1 (reference)	6.4 (6.0–6.6)	0 (reference)	1 (reference)
Dietary intervention[1]	1.9 (1.6–2.1)	−0.66 (−0.86 to−0.36)	0.74 (0.65–0.85)	5.8 (5.3–6.4)	−0.52 (−0.91 to 0.07)	0.92 (0.86–1.01)
Nurses' Health Study-II						
No dietary intervention	0.28 (0.21–0.32)	0 (reference)	1 (reference)	1.2 (1.0–1.3)	0 (reference)	1 (reference)
Dietary intervention[1]	0.24 (0.14–0.32)	−0.04 (−0.12 to 0.04)	0.84 (0.56–1.17)	1.1 (0.9–1.3)	−0.06 (−0.19 to 0.12)	0.95 (0.83–1.11)

Estimates are based on the parametric g-formula with baseline and prebaseline covariates: baseline age; BMI; smoking status; physical activities; parental history of myocardial infarction (<60 y); aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); baseline diagnosis of hypertension or hypercholesterolemia; and prebaseline intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; and time-varying covariates: BMI; cigarettes smoked per day; physical activity; aspirin use; menopausal status (NHS/NHS II); menopausal hormone therapy (NHS/NHS II); intakes of fruits and vegetables, fish, whole grains, processed meat, sugar-sweetened beverages, legumes/nuts/seeds, alcohol, and calories; incidences of hypertension, hypercholesterolemia, diabetes, cancer, and nonfatal myocardial infarction; and time since report of each diagnosis. In the NHS II, due to fewer cases, we removed diabetes and nonfatal coronary heart disease, as well as time since diabetes or nonfatal coronary heart disease diagnosis, from the models for analyses of cancer mortality; we also removed hypertension, hypercholesterolemia, diabetes, cancer, and nonfatal myocardial infarction, as well as the time since report of each diagnosis, from the analysis of cardiovascular mortality. Abbreviations: AHA, American Heart Association; NHS, Nurses' Health Study; NHS II, Nurses' Health Study II.

[1]Joint intervention on all food-based AHA 2020 Dietary Goals: fruits and vegetables ≥4.5 servings/d, whole grains ≥3 servings/d, fish ≥2 servings/wk (on nonvegetarians only), sugar-sweetened beverages ≤3 servings/wk, legumes/nuts/seeds ≥4 servings/wk, and processed meat ≤2 servings/wk.

Data Availability

An example of analytic codes is included in the Supplemental Text. The procedures to obtain and access data from the Nurses' Health Studies, Nurses' Health Studies II, and Health Professionals Follow-Up Study are described at https://www.nurseshealthstudy.org/researchers (contact e-mail: nhsaccess@channing.harvard.edu) and https://sites.sph.harvard.edu/hpfs/for-collaborators/.

References

1. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Aros F, Gomez-Gracia E, Ruiz-Gutierrez V, Fiol M, Lapetra J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. N Engl J Med 2018;378(25):e34. doi:10.1056/NEJMoa1800389.
2. de Lorgeril M, Salen P, Martin JL, Monjaud I, Delaye J, Mamelle N. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction: final report of the Lyon Diet Heart Study. Circulation 1999;99(6):779–85.
3. Esposito K, Marfella R, Ciotola M, Di Palo C, Giugliano F, Giugliano G, D'Armiento M, D'Andrea F, Giugliano D. Effect of a Mediterranean-style diet on endothelial dysfunction and markers of vascular inflammation in the metabolic syndrome: a randomized trial. JAMA 2004;292(12):1440–6.
4. Shai I, Schwarzfuchs D, Henkin Y, Shahar DR, Witkow S, Greenberg I, Golan R, Fraser D, Bolotin A, Vardi H, et al. Weight loss with a low-carbohydrate, Mediterranean, or low-fat diet. N Engl J Med 2008;359(3):229–41.
5. Appel LJ, Moore TJ, Obarzanek E, Vollmer WM, Svetkey LP, Sacks FM, Bray GA, Vogt TM, Cutler JA, Windhauser MM, et al. A clinical trial of the effects of dietary patterns on blood pressure. DASH Collaborative Research Group. N Engl J Med 1997;336(16):1117–24.
6. Sacks FM, Svetkey LP, Vollmer WM, Appel LJ, Bray GA, Harsha D, Obarzanek E, Conlin PR, Miller ER 3rd, Simons-Morton DG, et al. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. N Engl J Med 2001;344(1):3–10.
7. Vincent-Baudry S, Defoort C, Gerber M, Bernard MC, Verger P, Helal O, Portugal H, Planells R, Grolier P, Amiot-Carlin MJ, et al. The Medi-RIVAGE study: reduction of cardiovascular disease risk factors after a 3-mo intervention with a Mediterranean-type diet or a low-fat diet. Am J Clin Nutr 2005;82(5):964–71.
8. Taubes G. Epidemiology faces its limits. Science 1995;269(5221):164–9.
9. Ioannidis JPA. The challenge of reforming nutritional epidemiologic research. JAMA 2018;320(10):969–70.
10. Trepanowski JF, Ioannidis JPA. Perspective: limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: why and how. Adv Nutr 2018;9(4):367–77.
11. Ioannidis JP. Implausible results in human nutrition research. BMJ 2013;347:f6698. doi:10.1136/bmj.f6698.
12. Johnston BC, Zeraatkar D, Han MA, Vernooij RWM, Valli C, El Dib R, Marshall C, Stover PJ, Fairweather-Taitt S, Wojcik G, et al. Unprocessed red meat and processed meat consumption: dietary guideline recommendations from the Nutritional Recommendations (NutriRECS) Consortium. Ann Intern Med 2019; 171(10):756–64.
13. Mozaffarian D, Forouhi NG. Dietary guidelines and health–is nutrition science up to the task? BMJ 2018;360:k822.
14. Hu FB, Willett WC. Current and future landscape of nutritional epidemiologic research. JAMA 2018;320(20):2073–4.
15. Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. Adv Nutr 2015;6(1):5–18.
16. Satija A, Stampfer MJ, Rimm EB, Willett W, Hu FB. Perspective: are large, simple trials the solution for nutrition research? Adv Nutr 2018;9(4):378–87.
17. Hernan MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183(8):758–64.
18. Hernan MA, Sauer BC, Hernandez-Diaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol 2016;79:70–5.
19. Zhang Y, Thamer M, Kaufman J, Cotter D, Hernán MA. Comparative effectiveness of two anemia management strategies for complex elderly dialysis patients. Med Care 2014;52(Suppl 2):S132–9.
20. García-Albéniz X, Hsu J, Bretthauer M, Hernán MA. Effectiveness of screening colonoscopy to prevent colorectal cancer among Medicare beneficiaries aged 70 to 79 years: a prospective observational study. Ann Intern Med 2017;166(1):18–26.
21. Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. J Clin Epidemiol 2018;96:12–22.
22. Petito LC, García-Albéniz X, Logan RW, Howlader N, Mariotto AB, Dahabreh IJ, Hernán MA. Estimates of overall survival in patients with cancer receiving different treatment regimens: emulating hypothetical target trials in the Surveillance, Epidemiology, and End Results (SEER)-Medicare linked database. JAMA Network Open 2020;3(3):e200452.
23. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med 2019;25(10):1601–6.
24. Caniglia EC, Robins JM, Cain LE, Sabin C, Logan R, Abgrall S, Mugavero MJ, Hernández-Díaz S, Meyer L, Seng R, et al. Emulating a trial of joint dynamic strategies: an application to monitoring and treatment of HIV-positive individuals. Stat Med 2019;38(13):2428–46.
25. Lajous M, Willett WC, Robins J, Young JG, Rimm E, Mozaffarian D, Hernán MA. Changes in fish consumption in midlife and the risk of coronary heart disease in men and women. Am J Epidemiol 2013;178(3):382–91.
26. Jain P, Suemoto CK, Rexrode K, Manson JE, Robins JM, Hernán MA, Danaei G. Hypothetical lifestyle strategies in middle-aged women and the long-term risk of stroke. Stroke 2020;51(5):1381–7.
27. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, Greenlund K, Daniels S, Nichol G, Tomaselli GF, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic impact goal through 2020 and beyond. Circulation 2010;121(4):586–613.
28. Hernan MA. Does water kill? A call for less casual causal inferences. Ann Epidemiol 2016;26(10):674–80.
29. Young JG, Herńan MA, Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. Epidemiol Meth 2014;3(1):1–19.
30. Taubman SL, Robins JM, Mittleman MA, Hernán MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula, Int J Epidemiol, 2009, 38, 6, 1599–611, PMID: 19389875; PMCID: PMC2786249
31. Young JG, Hernan MA, Robins JM. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. Epidemiol Meth 2014;3(1):1–19.
32. Cespedes EM, Hu FB. Dietary patterns: from nutritional epidemiologic analysis to national guidelines. Am J Clin Nutr 2015;101(5):899–900.
33. Hernan MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017;377(14):1391–8.
34. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect. Math Model 1986;7(9–12):1393–512.
35. Hernán MA, Robins JM. Causal inference: what if. Boca Raton: Chapman & Hall/CRC Press; 2020.
36. Robins J, Hernán MA, Siebert U. In: Ezzati M, Murray CJL, Lopez AD, Rodgers A, editors. Effects of multiple interventions. In: Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Chapter 28; Vol 2. Geneva, World Health Organization, 2004, 2191–2230.
37. Taubman SL, Robins JM, Mittleman MA, Hernan MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. Int J Epidemiol 2009;38(6):1599–611.
38. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernan MA. A causal framework for classical statistical estimands in failure-time settings with competing events. Stat Med 2020;39(8):1199–236.
39. Rimm EB, Giovannucci EL, Willett WC, Colditz GA, Ascherio A, Rosner B, Stampfer MJ. Prospective study of alcohol consumption and risk of coronary disease in men. Lancet North Am Ed 1991;338(8765):464–8.

40. Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, Chavarro JE. Origin, methods, and evolution of the three Nurses' Health Studies. Am J Public Health 2016;106(9):1573–81.
41. Yuan C, Spiegelman D, Rimm EB, Rosner BA, Stampfer MJ, Barnett JB, Chavarro JE, Subar AF, Sampson LK, Willett WC. Validity of a dietary questionnaire assessed by comparison with multiple weighed dietary records or 24-hour recalls. Am J Epidemiol 2017;185(7):570–84.
42. Yuan C, Spiegelman D, Rimm EB, Rosner BA, Stampfer MJ, Barnett JB, Chavarro JE, Rood JC, Harnack LJ, Sampson LK, et al. Relative validity of nutrient intakes assessed by questionnaire, 24-hour recalls, and diet records as compared with urinary recovery and plasma concentration biomarkers: findings for women. Am J Epidemiol 2018;187(5):1051–63.
43. Salvini S, Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Food-based validation of a dietary questionnaire: the effects of week-to-week variation in food consumption. Int J Epidemiol 1989;18(4):858–67.
44. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of a expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135(10):1114–26.
45. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135(10):1114–26; discussion 27–36.
46. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122(1):51–65.
47. Feskanich D, Rimm EB, Giovannucci EL, Colditz GA, Stampfer MJ, Litin LB, Willett WC. Reproducibility and validity of food intake measurements from a semiquantitative food frequency questionnaire. J Am Diet Assoc 1993;93(7):790–6.
48. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 2010;21(3):383–8.
49. Chiuve SE, Fung TT, Rimm EB, Hu FB, McCullough ML, Wang M, Stampfer MJ, Willett WC. Alternative dietary indices both strongly predict risk of chronic disease. J Nutr 2012;142(6):1009–18.
50. Fung TT, Rexrode KM, Mantzoros CS, Manson JE, Willett WC, Hu FB. Mediterranean diet and incidence of and mortality from coronary heart disease and stroke in women. Circulation 2009;119(8):1093–100.
51. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med 2003;348(26):2599–608. doi:10.1056/NEJMoa025039.
52. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. Arch Intern Med 2008;168(7):713–20.
53. Sofi F, Macchi C, Abbate R, Gensini GF, Casini A. Mediterranean diet and health status: an updated meta-analysis and a proposal for a literature-based adherence score. Public Health Nutr 2014;17(12):2769–82.
54. Schwingshackl L, Bogensberger B, Hoffmann G. Diet quality as assessed by the healthy eating index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension score, and health outcomes: an updated systematic review and meta-analysis of cohort studies. J Acad Nutr Diet 2018;118(1):74–100.e11.
55. George SM, Ballard-Barbash R, Manson JE, Reedy J, Shikany JM, Subar AF, Tinker LF, Vitolins M, Neuhouser ML. Comparing indices of diet quality with chronic disease mortality risk in postmenopausal women in the Women's Health Initiative Observational Study: evidence to inform national dietary guidance. Am J Epidemiol 2014;180(6):616–25.
56. Harmon BE, Boushey CJ, Shvetsov YB, Ettienne R, Reedy J, Wilkens LR, Le Marchand L, Henderson BE, Kolonel LN. Associations of key diet-quality indexes with mortality in the Multiethnic cohort: the Dietary Patterns Methods Project. Am J Clin Nutr 2015;101(3):587–97.
57. Reedy J, Krebs-Smith SM, Miller PE, Liese AD, Kahle LL, Park Y, Subar AF. Higher diet quality is associated with decreased risk of all-cause, cardiovascular disease, and cancer mortality among older adults. J Nutr 2014;144(6):881–9.
58. Akbaraly TN, Ferrie JE, Berr C, Brunner EJ, Head J, Marmot MG, Singh-Manoux A, Ritchie K, Shipley MJ, Kivimaki M. Alternative Healthy Eating Index and mortality over 18 y of follow-up: results from the Whitehall II cohort. Am J Clin Nutr 2011;94(1):247–53.
59. Buckland G, Agudo A, Travier N, Huerta JM, Cirera L, Tormo MJ, Navarro C, Chirlaque MD, Moreno-Iribas C, Ardanaz E, et al. Adherence to the Mediterranean diet reduces mortality in the Spanish cohort of the European Prospective Investigation into Cancer and Nutrition (EPIC-Spain). Br J Nutr 2011;106(10):1581–91.
60. Sotos-Prieto M, Bhupathiraju SN, Mattei J, Fung TT, Li Y, Pan A, Willett WC, Rimm EB, Hu FB. Association of changes in diet quality with total and cause-specific mortality. N Engl J Med 2017;377(2):143–53.
61. Hernan MA. The hazards of hazard ratios. Epidemiology 2010;21(1):13–5.
62. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernan MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. Stat Biosci 2011;3(1):119–43.
63. Thomson CA, Giuliano A, Rock CL, Ritenbaugh CK, Flatt SW, Faerber S, Newman V, Caan B, Graver E, Hartz V, et al. Measuring dietary change in a diet intervention trial: comparing food frequency questionnaire and dietary recalls. Am J Epidemiol 2003;157(8):754–62.
64. Lajous M, Hernan MA, Robins JM. Changes in fish and red meat intake over 20 years after cardiometabolic diseases in men and women: do people change their diets? Presented at EPI/NPAM Scientific Sessions, Atlanta, Georgia, March 22–25, 2011.
65. Schroder H, Fito M, Estruch R, Martinez-Gonzalez MA, Corella D, Salas-Salvado J, Lamuela-Raventos R, Ros E, Salaverria I, Fiol M, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011;141(6):1140–5.