Statistical Commentary

Best (but oft-forgotten) practices: sensitivity analyses in randomized controlled trials[1,2]

*Russell J de Souza,[3–5] Rebecca B Eisen,[3] Stefan Perera,[3] Bianca Bantoto,[3] Monica Bawor,[3] Brittany B Dennis,[3] Zainab Samaan,[3,6,10] and Lehana Thabane[3,7–9]**

[3]Department of Clinical Epidemiology & Biostatistics, Faculty of Health Sciences, [4]Department of Nutritional Sciences, Faculty of Medicine, University of Toronto, Toronto, Canada; [5]Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada; [6]Department of Psychiatry and Behavioral Neurosciences, [7]Center for Evaluation of Medicine, [8]Population Health Research Institute, Hamilton Health Sciences, and [9]System-Linked Research Unit, McMaster University, Hamilton, Canada; and [10]St. Joseph's Healthcare Hamilton and Hamilton Health Sciences Mood Disorders Program, Hamilton, Canada

ABSTRACT

A randomized controlled clinical trial is the best way to minimize bias in ascertaining treatment effects, but the credibility of the results of a trial depends on the validity of the methods used to analyze the data, and the conditions under which such methods produce valid answers. A sensitivity analysis is a method to determine the robustness of trial findings by examining the extent to which results are affected by changes in methods, models, values of unmeasured variables, or assumptions. The goal of a sensitivity analysis is to identify results that are most dependent on questionable or unsupported assumptions. In this article, we briefly review the current use of sensitivity analyses in a random sample of published nutrition literature and provide a guide on the use of sensitivity analyses in randomized trials as to when to consider them, what to consider when planning them, and different methods of implementing them. We propose an 8-step strategy for improving the approach to conducting and reporting sensitivity analyses in nutrition-based trials. *Am J Clin Nutr* 2016;103:5–17.

Keywords: analysis, commentary, sensitivity, statistics, tutorial

INTRODUCTION

Well-designed and -conducted randomized controlled trials of interventions represent the highest level of evidence of treatment efficacy. The randomized controlled clinical trial, if ethical and possible, is the best way to minimize bias in ascertaining treatment effects. Through random assignment, groups of patients are established with similar distributions of characteristics that could determine their responses to the intervention. If prognostic factors are balanced in the 2 (or more) groups, any differences between the groups in the outcome that are observed can be confidently attributed to the experimental intervention. For this reason, randomized controlled trials are the cornerstone of evidence-based medicine, and the results guide evidence-based health care practices and therapeutic choices to improve patient outcomes.

The credibility of the results of a trial depends on the validity of the methods used to analyze the data and the assumptions under which such methods produce valid answers (1). If these conditions are not met, the confidence in the results of a trial may be lowered. Examples of factors that may affect confidence in the results of a trial include the method of statistical analysis (2), deviations from the protocol (i.e., nonadherence) (3, 4), accounting for the correlation of repeated measures (5, 6), alternative outcome or treatment definitions (7), the presence of outliers (8, 9), high rates of dropout or other forms of missing data (10, 11), and the potential identification of subgroup effects (12–14). An analytic approach to assessing the impact of these factors in the context of randomized controlled trials is known as a sensitivity analysis (SA)[11] (1).

An SA is "a method used to determine the robustness of an assessment by examining the extent to which results are affected by changes in methods, models, values of unmeasured variables, or assumptions" (15). SAs play a crucial role in assessing the robustness of the findings or conclusions on the basis of primary analyses of data in clinical trials, which is a critical way to

[1] The authors reported no funding received for this study.

[2] Supplemental Material is available from the "Online Supporting Material" link in the online posting of the article and from the same link in the online table of contents at http://ajcn.nutrition.org.

*To whom correspondence should be addressed. E-mail: thabanl@mcmaster.ca.

[11] Abbreviations used: ITT, intention to treat; MAR missing at random; MCAR, missing completely at random; MNAR, missing not at random; POUNDS LOST, Preventing Overweight Using Novel Dietary Strategies; PREDIMED, Prevención con Dieta Mediterránea; SA, sensitivity analysis.

Received August 18, 2015. Accepted for publication November 4, 2015.

First published online December 16, 2015; doi: 10.3945/ajcn.115.121848.

assess the impact, effect, or influence of key assumptions or variations (such as different methods of analysis, definitions of outcomes, protocol deviations, missing data, and outliers) on the overall conclusions of a study. If the findings of the SA are consistent with those of the primary analysis, this indicates that the findings are robust even if ideal experimental and analytic conditions have not been met, thereby increasing confidence in the results.

In this article, we briefly review the current use of SAs in a random sample of published nutrition literature and provide a guide on the use of SAs in randomized trials as to when to consider them, what to consider when planning them, and different methods of implementing them.

WHAT IS AN SA?

An SA is defined as "a method to determine the robustness of an assessment by examining the extent to which results are affected by changes in methods, models, values of unmeasured variables, or assumptions" (16). The aim of SAs is to identify "results that are most dependent on questionable or unsupported assumptions" (16).

HOW TO BUILD AN SA INTO THE PROTOCOL

The United States Food and Drug Administration offers guidance on statistical approaches to clinical trials and emphasizes that "it is important to evaluate the robustness of the results and primary conclusions of the trial" (17). Therefore, it is wise, at the protocol stage, to think carefully about the primary analytic approach as well as alternative approaches to the data analysis that might reasonably be used and assumptions under which the validity of the primary analysis plan may be questionable. For example, an SA to confirm the impact of modeling assumptions should be considered at this stage and thoughtfully planned out.

Some considerations that could be addressed when planning SAs include whether there are viable alternative ways of measuring the outcome; whether there are multiple statistical analytic approaches that could be used in a valid manner to address the same research question; what is the anticipated missing data mechanism, and what would be the best way to account for missing data in the trial; how influential are minor deviations from the protocol; and the distribution of the variables and the possibility of outliers. Appropriate consideration should be given to collecting sufficient data on each of these factors to facilitate informative SAs that assess the influence of the variation in each factor on the robustness of the main analysis.

For example, the PREDIMED (Prevención con Dieta Mediterránea) trial was a parallel-group, multicenter, randomized trial of a Mediterranean diet for the primary prevention of cardiovascular disease. Under ideal conditions, the effect of the diet on the rate of occurrence of events would *1*) be constant over time, *2*) be similar for all cardiovascular outcomes, *3*) not differ according to study center, and *4*) have no participants lost to follow-up. However, such assumptions are reasonably questionable; therefore, the authors presented SAs for different ways of handling missing data points (complete case, single imputation, and multiple imputation), different outcome definitions (including angina plus revascularization as a composite endpoint), variations in results by recruitment site (excluding centers with atypically low or high event rates), and the time course of the effect of the intervention effect (by excluding events occurring early or late) (18).

HOW OFTEN ARE SAs REPORTED IN PRACTICE?

To determine the use and reporting of SAs in the field of nutrition research, we reviewed the tables of contents of the following 8 leading journals in the field of nutrition and dietetics from 1 January 2013 through 20 March 2015: *The American Journal of Clinical Nutrition*, the *Journal of the American College of Nutrition*, the *Journal of the Academy of Nutrition and Dietetics*, the *Journal of Nutrition*, the *Journal of Parenteral and Enteral Nutrition*, *Appetite*, the E*uropean Journal of Clinical Nutrition*, and the *Journal of Nutritional Biochemistry.* We identified 184 reports of randomized controlled trials of which we reviewed 100 reports in detail that were selected at random (**Supplemental Material**). Of the 100 included articles, 40 articles were published in 2013, 49 articles were published in 2014, and 11 articles were published in 2015. The trials were conducted in 29 countries; studies conducted within the United States contributed the most trials of any single country (n = 29 articles). The majority of the trials were published in *The American Journal of Clinical Nutrition* (n = 56) followed by those published in the *Journal of the American College of Nutrition* (n = 17), the *Journal of the Academy of Nutrition and Dietetics* (n = 10), the *Journal of Nutrition* (n = 5), the *Journal of Parenteral and Enteral Nutrition* (n = 4), *Appetite* (n = 3), the *European Journal of Clinical Nutrition* (n = 3), and the *Journal of Nutritional Biochemistry* (n = 2).

Of the 100 randomly selected studies, 18 studies reported having conducted SAs (**Table 1**). The sensitivity approaches documented included the impact of noncompliance or protocol deviations (n = 11), the impact of missing data (n = 4), the impact of competing risks in a trial of composite outcomes (n = 1), the impact of a baseline imbalance (n = 1), the impact of prognostic factors (n = 1), and the impact of different assumptions underlying the statistical model (n = 1) (**Table 2**). Eighteen studies reported having performed one or more SAs. Of these 18 studies, 5 studies reported having planned the SAs a priori, and 6 studies reported post hoc SAs. In 7 of the studies, it was unclear whether the SA was conducted a priori or post hoc.

Fifteen of the 18 studies (83%) reported the final results on the basis of the original findings, whereas 3 studies (17%) reported results on the basis of the SAs (**Table 3**). Eleven studies (61%) showed significant results [i.e., the between-treatment difference in response was declared of statistical significance at the specified α level (e.g., $P < 0.05$)] in the primary analyses, and 7 studies did not (39%) report significant results in the primary analyses. In 2 of the 11 studies (18%) that reported significant results from the primary analyses, the SAs were nonsignificant, and in one of the 7 studies (14%) that did not report significant results in the primary analyses, the SAs were significant. Six of 18 studies that performed SAs (33%) mentioned the SAs in the final conclusion.

AN OVERVIEW OF COMMON SAs

We describe some of the major SAs encountered in nutrition trials with reference to examples drawn from the nutrition literature. Examples of these SAs from the literature are listed in **Table 4**.

TABLE 1
Characteristics of 18 articles reporting sensitivity analyses in nutrition journals since 2013 (of 100 randomly selected articles)

Domain	*n* (%)
Articles reviewed	100 (100)
Articles reporting sensitivity analyses	18 (18)
Journal	
The American Journal of Clinical Nutrition	14 (77.8)
European Journal of Clinical Nutrition	1 (5.6)
Journal of Parenteral and Enteral Nutrition	1 (5.6)
Journal of the American College of Nutrition	1 (5.6)
Journal of the Academy of Nutrition and Dietetics	1 (5.6)
Country	
United States	5 (27.8)
France	2 (11.1)
Australia	2 (11.1)
Netherlands	1 (5.6)
Haiti	1 (5.6)
United Arab Emirates	1 (5.6)
China	1 (5.6)
Gambia	1 (5.6)
Ireland	1 (5.6)
Spain	1 (5.6)
Germany	1 (5.6)
Multiple (Belgium, Germany, Italy, Poland, and Spain)	1 (5.6)
Randomized trial design	
Parallel	14 (77.8)
Factorial	2 (11.1)
Crossover	1 (5.6)
Cluster	1 (5.6)
Type and intensity of intervention	
Outpatient, supplement study	9 (50)
Controlled feeding	4 (22.2)
Outpatient, advice only	3 (16.7)
Acute feeding	1 (5.6)
Metabolic ward	1 (5.6)
Referenced a published protocol	
Yes	3 (16.7)
No	15 (83.3)
Listed in a trial registry	
Yes	16 (88.9)
No	2 (11.1)
Centers, *n*	
Single	12 (66.7)
Multiple	6 (33.3)
Funding source	
Nonprofit or government	10 (55.6)
Industry	3 (16.7)
Unclear	5 (27.8)

Impact of nonadherence or protocol deviations

In nutrition trials, interventions may be delivered in several ways, ranging in intensity from dietary advice that participants follow in a free-living setting to supermarket models in which participants receive all foods free of charge for 3–6 mo from a hospital- or university-based supermarket, and the diet composition is controlled by the bar codes and a computer program, to controlled-feeding studies in which all study food is provided and eating occasions are monitored, and to metabolic ward studies involving admission to a unit for observation during which time all food intake and waste output is carefully measured. Adherence is often suboptimal in advice-only trials (19, 41) and much better in supermarket studies (42), controlled-feeding studies (43, 44), or metabolic ward studies (45, 46). Incomplete adherence to the assigned intervention poses several challenges for interpretation and may leave a trial underpowered to detect the planned effect size.

In most randomized controlled trials, the preferred approach to the primary analysis is to follow the intention-to-treat (ITT) principle in which all participants who are randomly assigned are analyzed according to their assigned study group (the maxim "as randomized, so analyzed") ignoring the degree of adherence, deviations from protocol, or the fact that some participants are lost during follow-up (47). This analysis plan preserves the unique advantage of the randomized trial design (i.e., the balancing of known and unknown confounders) and avoids presenting biased or optimistic estimates of treatment effects on the basis of only those individuals who are highly adherent to the study protocol. It is important to distinguish high trial retention from high dietary adherence. The former refers to participants attending a high proportion of study measurement visits, whereas the latter refers to participants who actually consume a high proportion of the food they are assigned to consume. One does not necessarily imply the other; e.g., in the Preventing Overweight Using Novel Dietary Strategies (POUNDS LOST) randomized trial, participant retention was very high (80% over 2 y), but adherence to assigned macronutrient ranges was suboptimal (19, 48).

The following 2 SAs are often performed in addition to the ITT analysis: a per protocol analysis, in which participants who do not adhere to the assigned intervention are excluded from the analysis set, and an as-treated analysis, in which participants are analyzed according to the treatment they actually took during the study (49).

TABLE 2
Characteristics of sensitivity analyses identified in 18 published articles in nutrition journals since 2013

Domain	*n* (%)
Sensitivity analysis plan	
A priori	5 (27.8)
Post hoc	6 (33.3)
Unclear	7 (38.9)
Reported any missing data	
Yes	18 (100)
No	0 (0)
Percentage lost during follow-up (dropouts), %	
<5	4 (22.2)
5–9.9	3 (16.7)
10–19.9	5 (27.8)
≥20	6 (33.3)
Type of sensitivity analysis performed	
Impact of noncompliance or protocol deviations	11 (61.1)
Impact of missing data	4 (22.2)
Impact of competing risks in analysis of composite outcomes	1 (5.6)
Impact of baseline imbalance	1 (5.6)
Impact of prognostic factors	1 (5.6)
Impact of different assumptions underlying statistical models	1 (5.6)

TABLE 3
Impact of Sensitivity Analyses on reporting of results and conclusions of 18 articles

Domain	*n* (%)
Analytic plan reflected in summary and conclusion?	
Primary analysis plan	15 (83.3)
Sensitivity analyses findings	3 (16.7)
Were the results of the primary analysis significant?	
Yes	11 (61.1)
No	7 (38.9)
Sensitivity analysis yielded different result(s) from primary analysis?	
Yes	4 (22.2)
No	13 (72.2)
Unclear	1 (5.6)
Sensitivity analysis mentioned in final conclusion/summary?	
Yes	6 (33.3)
No	12 (66.7)

Although standard approaches to handling missing data because of poor retention are suitable for nutrition trials, methods of approaching poor dietary adherence are less straightforward. In trials of whole diets, clear as-treated analyses are seldom feasible because participants are rarely totally nonadherent or fully adherent, but adherence to desired dietary interventions may be as low as 50–60% in advice-only, long-term studies (19) but may reach as high as >90% (50) when food is provided in shorter-term studies. A common approach to an as-treated analysis in these trials is to relate adherence to the effect size. In a randomized trial of a dietary portfolio approach to cholesterol reduction, adherence to the 4 emphasized dietary portfolio components (nuts, soy, viscous fiber, and plant sterol) was regressed on the LDL-cholesterol change and was shown to be significantly associated with the percentage reduction in LDL cholesterol in participants who completed the study (51). This sensitivity or explanatory analysis showed that the effectiveness of the intervention was directly related to the degree of adherence to the diet prescribed.

Impact of outliers

Outliers are observations that are not representative of the sample population enrolled in the study and typically present a value >2 SDs from the group mean or also on the basis of face validity [e.g., BMI (in kg/m^2) >50.0] or clinical reasons (e.g., fasting glucose concentration >7.0 mmol/L in participants without diabetes) (52). Such implausible data points may unduly influence the statistical model, particularly in studies with a small sample size. Care should be taken to ensure outlying values are true outliers and not simply a result of a clerical error. The most-comprehensive and transparent approach is to present analyses both with and without outlying values and to assess the robustness of the main analysis as to the presence or absence of these extreme cases. However, if the outliers are not clerical errors and are not from a different population than the sample, there is no compelling reason to exclude them from the analysis.

Thus, a third possibility to dealing with outlying values is robust regression, which offers a compromise between excluding these points entirely from the analysis and including all the data points and treating all of them equally in a regression model. Robust regression is an alternative to ordinary least-squares regression in the presence of outliers or influential observations, and it can also be used to detect influential observations. The idea of robust regression is to weigh the observations differently on the basis of how closely an observation fits the regression line. Simply put, values with large deviations from the line of best fit (i.e., with large residuals) are given less weight in the regression analysis. As the absolute value of the residual decreases, the weight of the observation is increased.

This approach was used in a study of weight changes over 1 and 2 y in 303 women enrolled in a low-fat dietary-intervention trial. A multivariable regression analysis was used to quantify the independent predictive values of changes in energy and the percentage of energy from fat on the change in body weight (29). The primary analysis used ordinary least-squares regression. No effect of energy reduction was seen once the effect of dietary fat was controlled for. SAs that *1*) excluded 22 women with outlying values for weight change and *2*) used robust-regression approaches that used all observations but minimized the influence of outliers by systematically downweighting them were undertaken. The magnitude of the regression coefficients were sensitive to the regression-modeling approach, but all sensitivity approaches supported the primary analysis and showed that the change in the percentage of energy from fat was strongly and significantly associated with the change in body weight.

Impact of missing outcome data

Missing data, which is loosely defined as "missing some information on the phenomena in which we are interested" (53), is a widespread problem in clinical research. Our review of 100 articles in the field of nutrition showed that 14% of articles reported >20% of missing outcome data; 34% of articles reported ≥10% of missing outcome data; and 52% of articles had ≥5% of missing data. In total, 92% of trials had some missing data on the primary outcome. This finding is compatible with a recent review of musculoskeletal randomized trials that showed 95% (of 91 trials) reported at least some missing outcome data (54).

Several mechanisms give rise to missing data, and the most appropriate analytic approach in the presence of missing data is dependent on the pattern of the missingness. The 3 common missing data mechanisms are missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). When data are MCAR, participants with missing outcome data are accepted to represent a random sample of all enrolled participants (picture the case in which a box of participant records is inadvertently shredded in confidential waste). When data are MAR, the missing data are dependent on other observed values (picture a study in which younger participants are more likely to not attend a clinic visit than older participants are; the pooled analysis of completers will be biased toward the result in older participants, but analyses stratified by age will still be valid). A more-serious problem arises when data are MNAR, the so-called nonignorable missing data; in this case, the missing data are dependent on a process that resulted in the value being missing (picture a participant who fails to provide a final vitamin D sample in a trial of vitamin D supplementation because they did not attend their final visit because of nausea from the supplement).

TABLE 4
Examples of common scenarios for sensitivity analyses in clinical nutrition trials[1]

Scenario	Option	Example
Nonadherence or protocol violation	ITT analysis	In a 4-arm, 2 × 2 factorial RCT of macronutrient composition and weight loss, an ITT analysis in which long-term weight loss for persons who withdrew from the study early (after ≥6 mo of participation) was imputed on the basis of a rate of 0.3 kg regained weight/mo and a rate of 0.3 cm regained waist circumference/mo after withdrawal. The main analysis showed no significant differences in weight changes according to the assigned macronutrient over 2 y. The magnitudes of the difference in the change in weight between high- and low-protein arms [−0.6 kg (ITT) vs. −0.9 kg (completers)] and high- and low-fat arms [0.04 kg (ITT) vs. 0.2 kg (completers)] were similar (19).
	As-treated analysis	In a 2-arm, parallel trial to determine whether calcium carbonate supplementation would reduce BP in Gambian women, the primary analysis comparing BP between treatment and placebo groups with the use of completers only was supplemented with an as-treated analysis that related the number of tablets a women consumed to her BP at week 36. In the as-treated analysis, there was no significant difference between calcium and placebo groups in the change between 20 and 36 wk of gestation in SBP (20).[2]
	Per-protocol analysis	In a 2-arm, parallel trial to determine the safety and efficacy of vitamin D to improve markers of insulin sensitivity and resistance and reduce inflammation, the primary analysis (ITT) was supplemented with an analysis with the use of only data form participants who completed all 3 expected study visits. Results were similar in magnitude and direction (21).[2] In a 3-arm, parallel trial to determine whether fortified juices increased vitamin concentrations in children, the primary analysis (with the use of all available data from each participant; the analysis sample) was supplemented with an analysis that carried forward missing values or replaced missing values with the mean of all participants (ITT). Results were not significantly different from those of the analysis sample (22).[2] In a 2-arm, parallel trial to determine whether replacing refined grains with whole grains in the context of a weight-loss intervention would improve MetS criteria in individuals with or at risk of MetS, the primary analysis (with the use of linear mixed models; all-available data) was supplemented with an analysis that removed noncompliant individuals (n = 3 of 60; 5%). Results showed a stronger effect of the intervention on glucose [difference of 4.1 mg/dL (P = 0.02) vs. 3.3 mg/dL (P = 0.03)] (23).[2] In a 2-arm, parallel trial to determine whether protecting parenteral nutrition solutions from light decreases the rate of bronchopulmonary dysplasia or death in very-low-birth-weight infants, the primary analysis did not show significant beneficial effects of photoprotection (with the use of nonlinear mixed models; all available data). A sensitivity analysis used a per-protocol analysis with the exclusion of infants in whom the intervention was not delivered as intended (n = 45; 8% missing data). Results of the per-protocol analysis were similar to those obtained with the ITT analysis (24).[2] In a 2-arm, parallel trial of probiotics in obese pregnancy to reduce maternal fasting glucose, no differences in metabolic variables were noted in an ITT analysis. No significant difference was found in any outcome between intervention and control in the main analysis. In a sensitivity analysis, noncompliers (n = 21; 15%) and antibiotic users during the study (n = 6; 4%) were excluded. The sensitivity analyses showed no differences in any outcome between intervention and control groups (25).[2] In a randomized, balanced, crossover study to test the satiating value of aerated drinks with differing gastric stabilities, 3 skim-milk–based test products were consumed and compared with respect to self-reported appetite ratings. Compared with the control, both aerated drinks increased the gastric volume and reduced hunger with the use of a mixed model with all available data at each time point (ITT). Three participants did not complete the study after starting (17%). Both ITT and per-protocol populations were analyzed, and both groups yielded similar results and conclusions. Only ITT results were presented (26).[2] In a 4-arm, parallel, factorial design, the effect of cocoa and theobromine on serum HDL-cholesterol concentrations was examined. The full-set analysis included all subjects for whom end-of-intervention values were available. A per-protocol analysis excluded data from participants who were noncompliant with test-product consumption (n = 9; 6%; defined as <80% of the total prescribed intake; included subjects who had a body weight change >2 kg or who did not follow dietary restrictions). The per-protocol analysis and full-set analysis showed similar results; only the results of the full-set analysis were presented (27).[2] In a 2-arm, 2-phase, parallel trial to determine whether vitamin D supplementation improved metabolic control in an obese type 2 Emirati population, the primary analysis was compared with the use of repeated-measures ANOVA with no imputation in the ITT

(Continued)

TABLE 4 (*Continued*)

Scenario	Option	Example
		analysis. A sensitivity per-protocol analysis of only participants who provided complete data was also conducted and yielded results with similar significance and direction (28).[2]
Outliers	Robust regression	In a study of weight changes over 1 and 2 y in 303 women enrolled in a low-fat dietary intervention trial, a multivariable regression analysis was used to quantify the independent predictive values of changes in energy and the percentage of energy from fat on the change in body weight The primary analysis used an ordinary least-squares regression. No effect of energy reduction was seen once the effect of dietary fat was controlled for. Sensitivity analyses that *1*) excluded 22 women with outlying values for weight change and *2*) used robust-regression approaches that used all observations but minimized the influence of outliers by systematically downweighting them were undertaken. The magnitudes of the regression coefficients were sensitive to the regression-modeling approach, but all sensitivity approaches supported the primary analysis and showed that the change in the percentage of energy from fat was strongly and significantly associated with the change in body weight (29).[2]
Missing data	Analyze only completers data	In a 4-arm, 2 × 2 factorial RCT of macronutrient composition and weight loss, an ITT analysis in which long-term weight loss for persons who withdrew from the study early (after ≥6 mo of participation) was imputed on the basis of a rate of 0.3 kg regained weight/mo and a rate of 0.3 cm regained waist circumference/mo after withdrawal. The main analysis showed no significant differences in weight changes according to the assigned macronutrient over 2 y. The magnitudes of the difference in the change in weight between high- and low-protein arms [−0.6 kg (ITT) vs. −0.9 kg (completers)] and high- and low-fat arms [0.04 kg (ITT) vs. 0.2 kg (completers)] were similar (19).
	Impute missing data with the use of single or multiple imputation and redo analysis	In a 3-arm, parallel RCT to determine whether manipulating the glycemic index of the diet resulted in improved weight loss, the primary analysis was by ITT, which included all randomly assigned participants with the use of the LOCF to replace missing data points (single imputation). At weeks 16, 20, and 24, decreases in BMI were higher in the low–glycemic index–diet group than in the low-fat–diet group. The per-protocol analysis excluded individuals who did not attend the final visit (n = 17 of 122; 14%). Results at 6 mo (24 wk) in the per-protocol analysis were similar to that of the ITT analysis (30).[2] In a 2-arm, double-blind, parallel RCT of perioperative probiotic treatment on serum zonulin concentrations and postoperative infectious complications after colorectal cancer surgery, the probiotics group, compared with the placebo group, had a lower postoperative serum zonulin concentration and postoperative infection rate. A per-protocol sensitivity analysis was conducted including only subjects who completed the trial (excluding n = 13; 9%). Similar results were showed in the per-protocol data (31).[2] In a 3-arm, parallel RCT of mobile technology to monitor physical activity and nutrition over 12 mo in patients with MetS, 3 approaches to handling missing data were presented. The primary analysis used the mixed-model approach with all available data at each time point. Sensitivity analyses used the following 3 methods of imputation of outcomes for dropouts (n = 43; 23%): LOCF, BOCF, and multiple imputation. Dropout rates were not equal across groups (36% in the control arm; 16% in the ABC arm; 18% in the 4S arm). The sensitivity analysis suggested that lost cases were those with the smallest weight loss (3.7 kg in control for the mixed-model approach, 3.0 kg for the LOCF, and 2.5 kg for the BOCF); results from control subjects who completed the study likely overestimated the true effects (32).[2] In a 2-arm, parallel RCT of a low–glycemic index diet vs. a low-fat diet on body composition and components of MetS in obese Hispanic youth, the primary analysis used a multiple-imputation approach of dealing with missing data with the assumption of missing-at-random data and not-at-random data. Predictors of BMI change were used to inform the imputation as follows: BMI, age, sex, maternal education, and waist circumference at baseline. A sensitivity analysis used only completers and yielded similar results (33).[2] In a cluster-randomized trial that appropriately accounted for clustering, the impact of a short term (<6 mo) telephone intervention aimed at improving children's fruit and vegetable intake on longer-term fruit and vegetable intake was examined in a cluster-randomized controlled design. The primary endpoint was fruit and vegetable intake at 18 mo after baseline. The primary analysis used generalized estimating equations to assess efficacy with the use of all available data at 18 mo. At the 12-mo follow-up, Children's Dietary Questionnaire fruit and vegetable subscale scores were significantly higher in children in the intervention group, indicating greater child fruit and vegetable intake. Sensitivity analyses to deal with missing data included *1*) baseline carried forward and *2*) subgroup analyses by parent socioeconomic status and educational attainment to assess the consistency of the main findings across subgroups. The effect remained significant when

(Continued)

TABLE 4 (*Continued*)

Scenario	Option	Example
		baseline values were carried forward, and there was no interaction between treatment and socioeconomic status or education (34).[2]
		In the Childhood Obesity Project, a 2-arm, randomized trial was conducted of higher- vs. lower-protein formulas within the first year of life on the development of obesity. The primary outcome was the BMI *z* score at 6 y, and the treatment difference was assessed with the use of a linear regression without adjustment for covariates. Several sensitivity analyses were performed as follows: *1*) Adjustment was made for potential confounders (sex, age at measurement, country, highest education level of the mother and father, smoking during pregnancy, and mother and father BMI); *2*) To test for the effect of missing data (98 children did not have measured BMI; 13%), the 6-y analysis also included a parental report of weight and height at 6 y via telephone; and *3*) The imputation of missing values was conducted with the use of chained equations. When missing values were replaced by values from telephone interviews and multiple imputations, the estimated effects decreased (35).[2]
Baseline imbalance	Analysis with and without adjustment for baseline	In a 2-arm, parallel trial of probiotics in obese pregnancy to reduce maternal fasting glucose, an imbalance of baseline BMI (in kg/m^2) was noted between the probiotic and placebo groups (32.9 ± 2.4 vs. 34.1 ± 2.7; $P = 0.007$). Thus, the authors performed "an additional analysis of primary and secondary outcomes [...] by using the general linear model with BMI as the covariate and intervention as the fixed factor," but the article and accompanying materials presented only the models adjusted for baseline BMI (25).[2]
	Stratify the analysis according to prognostic values	In a 2-arm, parallel trial of normal protein intake vs. moderate protein intake on renal function, which was measured as the change in eGFR, the primary analysis was by ITT according to assigned groups. Changes in eGFR were shown to be related to the baseline eGFR. A post hoc analysis was performed and stratified by baseline eGFR. In this sensitivity analysis, participants with stage 1, 2, or 3 renal disease (less-severe disease) had improved renal function with weight loss; those with hyperfiltration (more-severe disease) had worsened renal function with weight loss (36).[2]
	Various methods of adjustment: multivariable regression and propensity score matching (more common in observational studies)	In a double-blind, placebo-controlled trial of vitamin B-12 supplementation for neurologic and cognitive function in moderately vitamin B-12–deficient older adults, the primary analysis, which was adjusted for baseline neurologic function (with the use of an ANCOVA), showed no evidence of an effect of supplementation on the primary outcome of the posterior tibial compound muscle action potential amplitude at 12 mo. A sensitivity analysis, which used an ANCOVA to adjust for age and sex in addition to baseline neurologic function, showed virtually identical effect sizes and did not change the main analysis result (37).
Definition of outcomes	Perform analyses on outcomes of different cutoffs or definitions	In the Physicians Health Study II randomized trial of vitamin E or vitamin C supplementation and prostate cancer, 356 incident cases of prostate cancer were observed over a mean of 2.8 y after the trial ended (10 y of supplementation). In this study, the primary analysis, with the use of a Cox proportional hazards model, showed neither vitamin C nor vitamin E to be significantly better than a placebo for preventing total cancer, prostate cancer, or other site-specific cancers. Two sensitivity analyses were conducted as follows: *1*) participants were censored at the last date of follow-up and *2*) an analysis was conducted that was limited to posttrial cancers that were new (i.e., that occurred in men who did not have a cancer at the same site during the intervention period). The results did not materially change (38).[2]
		In the POUNDS LOST study, the primary outcome was the mean change in body weight; a sensitivity analysis also examined differences in the proportions of participants achieving a threshold of weight loss that was deemed to be clinically important. For example, at 2 y, 31–37% of the participants had lost $\geq$5% of their initial body weight, 14–15% of the participants in each diet group had lost $\geq$10% of their initial weight, and 2–4% of the participants had lost $\geq$20 kg (*P*-comparisons between diets > 0.20). The lack of differences between diets in the proportion of people achieving these targets was consistent with the primary analysis (19).[2]
Distributional assumptions	Parametric vs. nonparametric	We did not identify an article in the leading nutrition journals that analyzed the same data with the use of both parametric and nonparametric approaches; typically, one of the 2 approaches was used. In a study of the effectiveness of nutrition therapy in pediatric patients with chronic liver diseases awaiting liver transplantation, one group of children were assigned an intensive nutrition protocol (oral diet plus enteral nutrition), and the other group was assigned an oral diet alone. Comparisons between height *z* scores in girls were made with the use of an independent-samples *t* test and a Mann-Whitney *U* test. In the *t*-test approach, the height *z* score was 1.16 U higher (95% CI: 0.09, 2.23 U higher) in children receiving

(Continued)

TABLE 4 (*Continued*)

Scenario	Option	Example
		the intensive nutrition protocol; and the *P* value for the Mann-Whitney *U* test for the same comparison was 0.022 (39).
	Random effects vs. fixed effects	In a longitudinal study of a daily lipid-based nutrient supplement for increased linear growth in young children, healthy infants were assigned to either a control or a 3- or 6-mo supplementation period of vitamin A, vitamin B-12, iron, and zinc. The effect of the intervention was modeled with the use of a generalized-least squared model with random effects and mixed effects (random plus fixed effects). The mixed model was chosen for analysis because assumptions for random effects were not justifiable (i.e., individual effects were uncorrelated with the independent variables, and dropout was at random) (40).[2]

[1]ABC, Active Body Control Program of University of Magdeburg; BOCF, baseline-observation carried forward; BP, blood pressure; eGFR, estimated glomerular filtration rate; ITT, intention to treat; LOCF, last-observation carried forward; MetS, metabolic syndrome; POUNDS LOST, Preventing Overweight Using Novel Dietary Strategies; RCT, randomized controlled trial; SBP, systolic blood pressure; 4S, 4sigma telephone coaching.

[2]Represents an article identified in our literature review. Articles without a superscript 2 were chosen as illustrative examples by a targeted literature search.

Unfortunately, in practice, missing data rarely arise from a single mechanism, and it is impractical, if not impossible, to distinguish among the missing data mechanisms. The only true way to distinguish between MNAR and MAR is to measure some of that missing data. For example, one could contact a group who withdrew from the study and ask a few key questions about variables of interest in the study (e.g., weight or the number of fruit eaten in the past week). This method would allow for a comparison of respondents to nonrespondents on some key characteristics. If the responses between subjects who withdrew and those who remained in the study differ substantially, this difference is evidence that the data are MNAR.

An MCAR mechanism can be detected with the use of Little's test (55) or by creating dummy (indicator) variables for whether a variable is missing (1) or observed (0) at a planned measurement occasion. A chi-square test can be used to assess whether the proportion of missingness on the variable is related to observed values of other variables. For example, the differences in the percentages of missing values for the outcome of body weight between athletes and nonathletes in a study of weight-management practices in college students could be compared, and a significant chi-square test may reveal that the percentage of missing data for weight is higher in nonathletes.

The best way to handle missing data is to take all of the steps possible to avoid missing data during the study design phase by designing a protocol that is highly acceptable for enrolled participants. Little et al. (11) provided an excellent review of the prevention and treatment of missing data in clinical trials. Missing data points may also be partially recovered. For example, in a weight-loss trial, participants who drop out of a trial may be asked to return for one final weigh-in visit (56, 57) or telephoned and asked to weigh themselves on a home scale; or in a blood pressure–reduction trial, participants who drop out may be asked to share the most-recent blood pressure measurement taken by their family physician.

There are 2 general approaches to handling missing data during the analysis phase as follows: imputation-based methods (single or multiple) and likelihood-based methods. Single-imputation methods are most commonly used, such as the baseline observation carried forward where baseline values of participants replace unobserved values (58), the last-observation carried-forward method where the observation at the time point closest to the missing value replaces the unobserved value (59), mean imputation where the mean of observed values for other participants replaces the unobserved values (60), or a regression approach where the predicted value for an individual of, e.g., the same age, sex, or random assignment replaces the unobserved values. Although computationally straightforward, single-imputation methods generally lead to biased estimates of the treatment effect and underestimate the variability. The last observation carried forward may be particularly misleading when changes in the outcome measure tend to be more pronounced early in a study with a gradual regression to the mean, which is typical of weight-loss interventions (10).

Less biased but more-computationally intensive approaches to handling missing data are also available. For example, multiple imputation fills in all missing values *m* times chosen at random from a distribution of values thereby creating *m* new, complete pseudo–data sets. These *m* complete data sets are analyzed, and the results are pooled with the between-imputation and within–data set variation taken into account (61). If data are assumed to be MAR, mixed models (62), generalized estimating equations (for continuous data) (63), or generalized linear mixed models (for count or binary data) (64) can be used to generate variable estimates with the use of all available data.

For example, in the PREDIMED trial (18), all primary analyses were performed under the ITT principle. Time-to-event data were analyzed with the use of Cox models. SAs were conducted by imputing data for missing values, and participants who dropped out (523 of 7447 randomly assigned subjects; a 7% loss to follow-up). In the primary analysis, the HR for the comparison of a Mediterranean diet supplemented with extra-virgin olive oil with a control diet was 0.70 (95% CI: 0.53, 0.91), and the HR for the comparison of a Mediterranean diet supplemented with nuts with a control diet was 0.70 (95% CI: 0.53, 0.94). In SAs with the use of the complete-case analysis, the HRs became 0.69 and 0.70, respectively; with the use of single imputation, the HRs became 0.70 and 0.72, respectively; and with the use of multiple imputation, the HRs became 0.67 and 0.70, respectively.

In the POUNDS LOST trial (19), all primary analyses were performed under the ITT principle. Mean weight loss was

analyzed with the use of unpaired *t* tests between groups on the change from baseline values in body weight. In this study, long-term weight loss for persons who withdrew from the study after ≥6 mo of participation (n = 166 of 811; 20%) was imputed on the basis of a rate of 0.3 kg regained weight/mo after withdrawal. The difference in weight change over 2 y between high- and usual- protein groups was −0.6 kg (95% CI: −1.6, 0.4 kg; $P = 0.22$, $n = 811$) with the use of the ITT approach and −0.9 kg (95% CI: −2.1, 0.2 kg; $P = 0.10$, $n = 645$) with the use of the completers-only approach. In the subset of participants with a body composition analysis, the dropout rate was twice as high as reported for the main trial (n = 188 of 424; 44%) (65). In the ITT analysis, the baseline body fat percentage was applied to the imputed weight to estimate fat and lean mass at the missing time point. The difference in the change of body fat over 2 y between high- and usual-protein groups was −0.5 kg ($P = 0.30$, $n = 424$) with the use of the ITT approach and −0.1 kg ($P = 0.90$, n = 236) with the use of the completers-only approach.

These examples serve to illustrate that, when the proportion of missing data is expected to be small and nondifferential (as in the PREDIMED trial), the choice of methodology to deal with missing data may be relatively inconsequential; however, when the proportion of missing data is expected to be large (as was planned for in the POUNDS LOST trial), missing-data approaches must be carefully considered.

UNDER WHAT GENERAL CONDITIONS IS A COMPLETE-CASE ANALYSIS APPROPRIATE?

When the proportion of missingness is low and in similar proportions within each randomly assigned group, the impact of the choice of method of handling missing data (e.g., complete-case compared with imputation methods) is not likely to change the conclusions of an ITT analysis. Complete-case analyses are also useful in pilot or efficacy (does it work?) trials. In efficacy trials, the goal is to determine whether the intervention worked under ideal circumstances in an ideal resource-intensive setting and in a highly selected, homogenous population who are cared for by a highly experienced and well-trained team who strictly enforce and standardize the intervention, and there is no contamination (66).

UNDER WHAT GENERAL CONDITIONS IS A COMPLETE-CASE ANALYSIS INAPPROPRIATE?

When the proportion of missingness is high (i.e., >30%) or not equally distributed between randomly assigned groups, careful attention must be paid to the treatment of missing data. Under conditions of nonignorable missingness (MNAR), a complete-case analysis should be interpreted with a great deal of caution and preferably avoided because there is no optimal way to arrive at an unbiased answer in the presence of nonignorable missing data. When data are MCAR or MAR, multiple imputation-based methods provide efficient answers even when the degree of missingness is high (67). An ITT approach is the preferred approach for effectiveness trials (68). In effectiveness trials, the goal is to determine whether the intervention will work in a real-world setting in a typical clinic with a heterogeneous population representative of the people who will use the intervention and are cared for by representative, usual providers who allow flexibility with respect to the intervention and for concurrent interventions and possible crossover (66).

Impact of baseline imbalance

The major advantage of a randomized trial for causal inference is that random assignment balances known and unknown confounders; if the random assignment is performed correctly, treatment effects can be attributed to the intervention. However, a residual imbalance in baseline measures of the outcome or important covariate (e.g., age, BMI, or sex) may occur by chance. An ANCOVA (i.e., including the baseline value of the outcome measure or covariate) may yield more-precise estimates of the treatment effect (25, 37).

To avoid the possibility of an imbalance, block randomization is a commonly used technique in clinical trial design (69). This approach increases the probability that each arm will contain an equal number of individuals by sequencing participant assignments into small groups called blocks, and within a block, there is a balanced allocation. Although this approach is helpful to ensure a balance, the allocation process may be predictable (e.g., when the investigator is not blind, and the block size is fixed). For example, in a 2-group parallel trial with a fixed block size of 4, after 2 participants are allocated to arm 1, the investigator can predict that the next 2 assignments will be to arm 2. Selection bias may be reduced with the use of random block sizes and keeping the investigator blind to the size of each block (e.g., some of 2, some of 4, and some of 8).

Impact of outcome definition

Often an outcome is defined by achieving or not achieving a certain level or threshold of a measure. For example, in a study measuring adherence rates to medication, levels of adherence can be dichotomized as achieving or not achieving ≥80%, ≥85%, or ≥90% of pills taken. The choice of the level a participant has to achieve can affect the outcome, whereby it might be harder to achieve 90% adherence than 80% adherence. Therefore, an SA could be performed to see how redefining the threshold changes the observed effect of a given intervention.

In the POUNDS LOST study, the primary outcome was the mean change in body weight; an SA also examined differences in the proportions of participants who achieved a threshold of weight loss that was deemed to be clinically important. For example, at 2 y, within each diet group, 31–37% of participants had lost ≥5% of their initial body weight, 14–15% of participants had lost ≥10% of their initial weight, and 2–4% of participants had lost ≥20 kg (P-comparisons between diets > 0.20) (19). This recasting of the outcomes did not affect the conclusions of the primary analysis.

In the Physicians Health Study II randomized trial of vitamin E or vitamin C supplementation and prostate cancer, 356 incident cases of prostate cancer were observed over a mean of 2.8 y after the trial ended (10 y of supplementation). In this study, the primary analysis, with the use of a Cox proportional hazards model, showed neither vitamin C nor vitamin E was significantly better than a placebo for preventing total cancer, prostate cancer, or other site-specific cancers. Two SAs were conducted as follows: first, participants were censored at the last date of follow-up; and second, an analysis that was limited to posttrial cancers

that were new (i.e., that occurred in men who did not have a cancer at the same site during the intervention period). The results did not materially change (38).

Impact of distributional assumptions

Many statistical analyses depend on normally distributed continuous outcomes; binomially distributed dichotomous outcomes; or Poisson-distributed count outcomes, negative binomial distributions, or zero-inflated negative binomial distributions. Goodness-of-fit tests can help assess departures from the underlying distribution required for the chosen test (70). SAs may invoke nonparametric tests to test the robustness of the primary analyses assuming distributional assumptions are met (**Table 5** reviews some common nonparametric analogs of parametric tests; we suggest a discussion of these alternatives with an experienced statistician during the development of an analysis plan).

HOW TO INTERPRET SAs

When the primary analysis and all SAs provide similar estimates of the magnitude and direction of treatment effect, similar *P* values, and similar widths of CIs, one can be reasonably confident that the finding of the trial is robust. A brief statement to this effect may suffice.

When the primary analysis and SAs differ, the situation is more difficult. Our guidance in this situation is to remember that the goal of an SA is not to select the best results. Rather, the aim is to assess the robustness or consistency of the results under, e.g., different methods, subgroups, definitions, and assumptions. The assessment of robustness is often based on the magnitude, direction, or statistical significance of the estimates. A sensitivity analysis cannot be used to choose an alternative conclusion to a study. Authors should state the conclusion on the basis of the primary analysis plan and use the presentation of one or more sensitivity analyses as a way of assessing the robustness and validity of the primary analysis. If the SA suggests that the primary analysis is not robust, it may point to the need for future research that might address the source of the inconsistency. To answer the question of which method is best and under what conditions, simulation studies comparing the different approaches on the basis of bias, precision, coverage, or efficiency may be necessary.

A common interpretation of SAs in nutrition trials is to distinguish relative efficacy from the effectiveness of an intervention. An efficacy study aims to determine whether the intervention works under ideal circumstances; an effectiveness study aims to determine whether the intervention works in real-world practice (66). For example, an analysis of only those individuals who completed a dietary intervention trial and achieved a high level of adherence (e.g., took close to all of the prescribed diet or supplement) would provide the best estimate of an intervention's efficacy (i.e., did it work under ideal circumstances?). An ITT analysis, with the use of all

TABLE 5
Parametric and nonparametric tests[1]

Type of data	Aim	Parametric	Parametric hypotheses (2-tailed)	Nonparametric	Nonparametric hypotheses (2-tailed)
Continuous	Compare 2 independent groups	Student's *t* test of means	H_O: $\mu_1 = \mu_2$	Mann-Whitney *U* test of ranks	H_O: samples come from the same distribution (there is no difference between medians)
			H_A: $\mu_1 \neq \mu_2$		H_A: samples come from different distributions (there is a difference between medians)
Continuous	Compare 2 dependent groups or matched pairs	Paired *t* test of means	H_O: $\mu_d = 0$	Wilcoxon's signed-rank test	H_O: the median difference between pairs of observations is zero (the difference between mean absolute ranks is zero)
			H_A: $\mu_d \neq 0$		H_A: the median difference between pairs of observations is not zero (the difference between mean absolute ranks is not zero)[2]
Continuous	Compare >2 independent groups	ANOVA	H_O: $\mu_1 = \mu_2 = ... = \mu_k$	Kruskal-Wallis test	H_O: mean ranks of groups are the same
			H_A: at least one mean differs from at least one other mean		H_A: the mean rank of one group differs from the mean rank of at least one other group
Continuous	Compare ≥2 independent groups, adjusting for one or more covariates	ANCOVA	H_O: $\mu_1 = \mu_2 = ... = \mu_k$ H_A: at least one mean differs from at least one other mean (after adjustment for the influence of other variables of importance)	Nonparametric ANCOVA	H_O: $\mu_1 = \mu_2 = ... = \mu_k$ H_A: at least one mean differs from at least one other mean (after adjustment for the influence of other variables of importance)

[1] H_A, alternative hypothesis; H_O, null hypothesis; μ_d, mean difference between paired observations; μ_k, mean of the k^{th} group; μ_1, mean of group 1; μ_2, mean of group 2.

[2] If the distribution is assumed to have a finite first moment and is symmetric about the median, the 2 null hypotheses (for parametric and nonparametric tests) are the same.

randomly assigned participants (i.e., as they were randomly assigned, regardless of whether they consumed close to or all of the prescribed diet or supplement) would provide the best estimate of an intervention's effectiveness (i.e., did it work in the real world?). For example, if a completers-only analysis shows that a dietary intervention reduces serum cholesterol, but the ITT analysis finds no effect of the intervention on cholesterol, we could say that the intervention is efficacious but not effective for a population intervention.

WHEN AND HOW SHOULD I REPORT AN SA?

A priori, planned SAs should be written into the study protocol and described in the Methods section of a manuscript. Post hoc SAs should be clearly identified as such, and the rationale for such analyses should be clearly justified. For example, if the dropout in a study is large and higher than anticipated when the power and sample size for the study was determined (57), there may be a sufficient dilution of the intended treatment effect such that the ITT analysis is underpowered.

The Methods section of a manuscript should describe the conduct and rationale for all a priori SAs outlined in the statistical analysis plan as well as for all post hoc SAs conducted. The results of all a priori and post hoc SAs should be reported. It may be sufficient to state that the SAs were conducted but did not yield materially different conclusions; and these data could be presented in summary form as supplemental material. If an SA yields results that are substantially different from that of the primary analysis plan, the results of the SA should be reported in the Results section to allow a reader to more-directly compare these results with those of the primary analysis plan.

It is important that, when an SA yields substantially different conclusions than those of the primary analytic approach, the Discussion section of a manuscript should mention the SA, and the limitations and implications of the findings are discussed. The impact of the SA on the interpretation of the data should be discussed in a context that allows readers to assess whether the primary analysis or SA is more believable or applicable to their practice and how the assumptions of the SA differ from that of the primary analysis plan.

In one case (of 18 cases), a post hoc SA replaced the prespecified analysis, and the prespecified analysis was not presented in the main text or supplemental material (25). We suggest that this practice be avoided and that the primary specified analysis plan be followed.

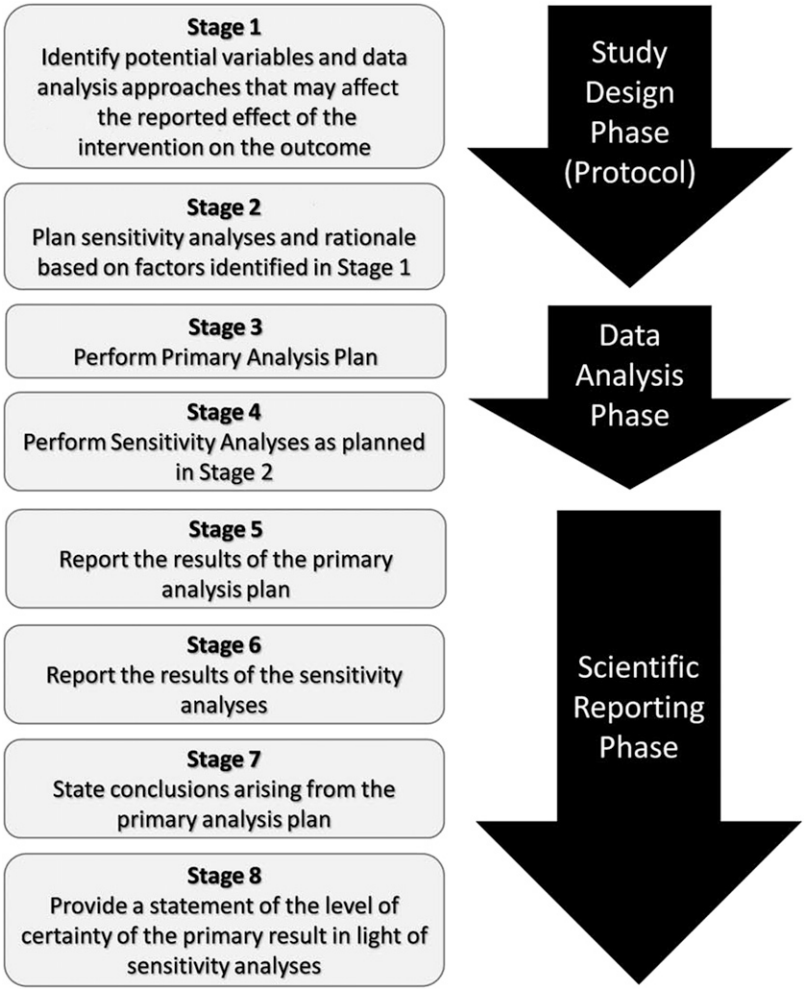

FIGURE 1 Proposed process for planning, conducting, and reporting sensitivity analyses in randomized controlled trials in the field of nutrition.

A MULTIPLE-STAGE APPROACH TO SAs

We propose an 8-stage process for planning, conducting, and reporting SAs in randomized controlled trials in the field of nutrition (**Figure 1**). At the study-design phase, a primary analysis plan should be clearly laid out with careful thought from statisticians, methodologists, and subject-matter experts given to potential variables or analytic methods that are likely to be most sensitive to changes in the methods, models, values of unmeasured variables, or assumptions (stage 1). After these have been identified, SAs should be planned that address the impact of each of the identified variables on the results provided by the primary analysis (stage 2). We believe that these 2 steps, which are performed before any data are collected, are the most crucial to ensure the highest validity of SAs. At the data-analysis phase, with a clear analysis plan prespecified and all data collected, the primary data-analysis plan should be executed (stage 3). After execution of the primary analysis plan, one or more SAs that address the alternative scenarios under which the primary analysis results may change, as prespecified in stage 1 and planned in stage 2, should be carried out (stage 4). We recommend that, in all cases, the primary analysis plan should be reported (stage 5). SAs performed should also be reported (stage 6). We recommend that the results of any preplanned SAs be reported in published articles either in the main text or as supplemental online content. The Discussion section of article should provide the context for the SAs and propose reasons for any discrepancies between the primary results and the results of the SAs (stage 7). Finally, a brief statement should be made about the impact of SAs on the conclusions of the study (stage 8).

In summary, in this article, we provide an overview of the use of SAs in randomized trials in the nutrition literature and provide examples of different types of SAs that might be considered when designing a randomized controlled trial. The reporting of SAs in nutrition is infrequent and suboptimal, and we propose an 8-step strategy for improving the approach to SAs and to facilitate transparency in reporting.

The authors' responsibilities were as follows—RJdS, ZS, and LT: study concept and design, drafting of the manuscript, and study supervision; RJdS and RBE: statistical analysis; RJdS, RBE, ZS, and LT: analysis and interpretation of data; RBE, SP, BB, MB, and BBD acquisition of the data; and ZS and LT: critical revision of the manuscript for important intellectual content. None of the authors reported a conflict of interest related to the study.

REFERENCES

1. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, Thabane M, Giangregorio L, Dennis B, Kosa D, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol 2013;13:92.
2. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. BMC Med Res Methodol 2001;1:6.
3. Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. Control Clin Trials 1987;8:12–9.
4. Bendtsen P, McCambridge J, Bendtsen M, Karlsson N, Nilsen P. Effectiveness of a proactive mail-based alcohol Internet intervention for university students: dismantling the assessment and feedback components in a randomized controlled trial. J Med Internet Res 2012;14: e142.
5. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004;61:310–7.
6. Keselman HJ, Algina J, Kowalchuk RK. The analysis of repeated measures designs: a review. Br J Math Stat Psychol 2001;54:1–20.
7. NINDS NET-PD Investigators. A randomized, double-blind, futility clinical trial of creatine and minocycline in early Parkinson disease. Neurology 2006;66:664–71.
8. Dixon WJ. Processing data for outliers. Biometrics 1953;9:74–89.
9. Andrews DF, Pregibon D. Finding the outliers that matter. J Royal Statistical Society Series B (Methodological) 1978;40:85–93.
10. Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med 2003;348:2136–7.
11. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, et al. The prevention and treatment of missing data in clinical trials. N Engl J Med 2012;367:1355–60.
12. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. JAMA 2014;311:405–11.
13. Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, Amstutz A, Bengough T, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ 2014;349:g4539. Erratum in: BMJ. 2014;349:4921.
14. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine–reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189–94.
15. Porta MS, Greenland S, Hernán M, dos Santos Silva I, Last JM. A dictionary of epidemiology. New York: Oxford University Press; 2014.
16. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. Pharmacoepidemiol Drug Saf 2006;15:291–303.
17. International conference on harmonisation; guidance on statistical principles for clinical trials; availability–FDA. Notice. Fed Regist 1998;63:49583–98.
18. 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. N Engl J Med 2013;368:1279–90.
19. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, McManus K, Champagne CM, Bishop LM, Laranjo N, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med 2009;360:859–73.
20. Goldberg GR, Jarjou LM, Cole TJ, Prentice A. Randomized, placebo-controlled, calcium supplementation trial in pregnant Gambian women accustomed to a low calcium intake: effects on maternal blood pressure and infant growth. Am J Clin Nutr 2013;98:972–82.
21. Belenchia AM, Tosh AK, Hillman LS, Peterson CA. Correcting vitamin D insufficiency improves insulin sensitivity in obese adolescents: a randomized controlled trial. Am J Clin Nutr 2013;97:774–81.
22. Economos CD, Moore CE, Hyatt RR, Kuder J, Chen T, Meydani SN, Meydani M, Klein E, Biancuzzo RM, Holick MF. Multinutrient-fortified juices improve vitamin D and vitamin E status in children: a randomized controlled trial. J Acad Nutr Diet 2014;114:709–17.
23. Harris Jackson K, West SG, Vanden Heuvel JP, Jonnalagadda SS, Ross AB, Hill AM, Grieger JA, Lemieux SK, Kris-Etherton PM. Effects of whole and refined grains in a weight-loss diet on markers of metabolic syndrome in individuals with increased waist circumference: a randomized controlled-feeding trial. Am J Clin Nutr 2014;100:577–86.
24. Laborie S, Denis A, Dassieu G, Bedu A, Tourneux P, Pinquier D, Kermorvant E, Millet V, Klosowski S, Patural H, et al. Shielding parenteral nutrition solutions from light: a randomized controlled trial. JPEN J Parenter Enteral Nutr 2015;39:729–37.
25. Lindsay KL, Kennelly M, Culliton M, Smith T, Maguire OC, Shanahan F, Brennan L, McAuliffe FM. Probiotics in obese pregnancy do not reduce maternal fasting glucose: a double-blind, placebo-controlled, randomized trial (Probiotics in Pregnancy Study). Am J Clin Nutr 2014;99:1432–9.
26. Murray K, Placidi E, Schuring EA, Hoad CL, Koppenol W, Arnaudov LN, Blom WA, Pritchard SE, Stoyanov SD, Gowland PA, et al. Aerated drinks increase gastric volume and reduce appetite as assessed by MRI: a randomized, balanced, crossover trial. Am J Clin Nutr 2015;101:270–8.
27. Neufingerl N, Zebregs YE, Schuring EA, Trautwein EA. Effect of cocoa and theobromine consumption on serum HDL-cholesterol concentrations: a randomized controlled trial. Am J Clin Nutr 2013;97:1201–9.
28. Sadiya A, Ahmed SM, Carlsson M, Tesfa Y, George M, Ali SH, Siddieg HH, Abusnana S. Vitamin D supplementation in obese type 2 diabetes subjects in Ajman, UAE: a randomized controlled double-blinded clinical trial. Eur J Clin Nutr 2015;69:707–11.
29. Sheppard L, Kristal AR, Kushi LH. Weight loss in women participating in a randomized trial of low-fat diets. Am J Clin Nutr 1991;54:821–8.

30. Juanola-Falgarona M, Salas-Salvado J, Ibarrola-Jurado N, Rabassa-Soler A, Diaz-Lopez A, Guasch-Ferre M, Hernandez-Alonso P, Balanza R, Bullo M. Effect of the glycemic index of the diet on weight loss, modulation of satiety, inflammation, and other metabolic risk factors: a randomized controlled trial. Am J Clin Nutr 2014;100:27–35.
31. Liu ZH, Huang MJ, Zhang XW, Wang L, Huang NQ, Peng H, Lan P, Peng JS, Yang Z, Xia Y, et al. The effects of perioperative probiotic treatment on serum zonulin concentration and subsequent postoperative infectious complications after colorectal cancer surgery: a double-center and double-blind randomized clinical trial. Am J Clin Nutr 2013;97: 117–26.
32. Luley C, Blaik A, Gotz A, Kicherer F, Kropf S, Isermann B, Stumm G, Westphal S. Weight loss by telemonitoring of nutrition and physical activity in patients with metabolic syndrome for 1 year. J Am Coll Nutr 2014;33:363–74.
33. Mirza NM, Palmer MG, Sinclair KB, McCarter R, He J, Ebbeling CB, Ludwig DS, Yanovski JA. Effects of a low glycemic load or a low-fat dietary intervention on body weight in obese Hispanic American children and adolescents: a randomized controlled trial. Am J Clin Nutr 2013;97:276–85.
34. Wolfenden L, Wyse R, Campbell E, Brennan L, Campbell KJ, Fletcher A, Wiggers J, Bowman J, Heard TR. Randomized controlled trial of a telephone-based intervention for child fruit and vegetable intake: long-term follow-up. Am J Clin Nutr 2014;99:543–50.
35. Weber M, Grote V, Closa-Monasterolo R, Escribano J, Langhendries JP, Dain E, Giovannini M, Verduci E, Gruszfeld D, Socha P, et al. Lower protein content in infant formula reduces BMI and obesity risk at school age: follow-up of a randomized trial. Am J Clin Nutr 2014;99:1041–51.
36. Jesudason DR, Pedersen E, Clifton PM. Weight-loss diets in people with type 2 diabetes and renal disease: a randomized controlled trial of the effect of different dietary protein amounts. Am J Clin Nutr 2013; 98:494–501.
37. Dangour AD, Allen E, Clarke R, Elbourne D, Fletcher AE, Letley L, Richards M, Whyte K, Uauy R, Mills K. Effects of vitamin B-12 supplementation on neurologic and cognitive function in older people: a randomized controlled trial. Am J Clin Nutr 2015;102:639–47.
38. Wang L, Sesso HD, Glynn RJ, Christen WG, Bubes V, Manson JE, Buring JE, Gaziano JM. Vitamin E and C supplementation and risk of cancer in men: posttrial follow-up in the Physicians' Health Study II randomized trial. Am J Clin Nutr 2014;100:915–23.
39. Giammanco M, Bavetta S, Greco R, Di Majo D, La Guardia M. Therapeutic effectiveness of nutrition therapy in pediatric patients with chronic liver diseases awaiting liver transplantation. J Food Res 2013; 2:179–90.
40. Iannotti LL, Dulience SJ, Green J, Joseph S, Francois J, Antenor ML, Lesorogol C, Mounce J, Nickerson NM. Linear growth increased in young children in an urban slum of Haiti: a randomized controlled trial of a lipid-based nutrient supplement. Am J Clin Nutr 2014;99:198–208.
41. Women's Health Initiative Study Group. Dietary adherence in the Women's Health Initiative Dietary Modification Trial. J Am Diet Assoc 2004;104:654–8.
42. Skov AR, Toubro S, Raben A, Astrup A. A method to achieve control of dietary macronutrient composition in ad libitum diets consumed by free-living subjects. Eur J Clin Nutr 1997;51:667–72.
43. Appel LJ, Sacks FM, Carey VJ, Obarzanek E, Swain JF, Miller ER 3rd, Conlin PR, Erlinger TP, Rosner BA, Laranjo NM, et al. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: results of the OmniHeart randomized trial. JAMA 2005;294:2455–64.
44. 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;336:1117–24.
45. Bray GA, Smith SR, de Jonge L, Xie H, Rood J, Martin CK, Most M, Brock C, Mancuso S, Redman LM. Effect of dietary protein content on weight gain, energy expenditure, and body composition during overeating: a randomized controlled trial. JAMA 2012;307:47–55.
46. Anderson JW, Story L, Sieling B, Chen WJ, Petro MS, Story J. Hypocholesterolemic effects of oat-bran or bean intake for hypercholesterolemic men. Am J Clin Nutr 1984;40:1146–55.
47. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869.
48. Astrup A, Brand-Miller J. Reply: totality of evidence needed for nutrition recommendations. Nat Rev Endocrinol 2014;10:310.
49. Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ 2001;165: 1339–41.
50. Jenkins DJ, Kendall CW, Marchie A, Faulkner DA, Wong JM, de Souza R, Emam A, Parker TL, Vidgen E, Lapsley KG, et al. Effects of a dietary portfolio of cholesterol-lowering foods vs lovastatin on serum lipids and C-reactive protein. JAMA 2003;290:502–10.
51. Jenkins DJ, Jones PJ, Lamarche B, Kendall CW, Faulkner D, Cermakova L, Gigleux I, Ramprasath V, de Souza R, Ireland C, et al. Effect of a dietary portfolio of cholesterol-lowering foods given at 2 levels of intensity of dietary advice on serum lipids in hyperlipidemia: a randomized controlled trial. JAMA 2011;306:831–9.
52. Thiese MS, Arnold ZC, Walker SD. The misuse and abuse of statistics in biomedical research. Biochem Med (Zagreb) 2015;25:5–11.
53. McKnight PE, McKnight KM, Sidani S, Figueredo AJ. Missing data: a gentle introduction. New York: Guildford; 2007.
54. Joseph R, Sim J, Ogollah R, Lewis M. A systematic review finds variable use of the intention-to-treat principle in musculoskeletal randomized controlled trials with missing data. J Clin Epidemiol 2015;68:15–24.
55. Little RJ. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 1988;83:1198–202.
56. Pi-Sunyer X, Astrup A, Fujioka K, Greenway F, Halpern A, Krempf M, Lau DC, le Roux CW, Violante Ortiz R, Jensen CB, et al. A Randomized, Controlled Trial of 3.0 mg of Liraglutide in Weight Management. N Engl J Med 2015;373:11–22.
57. McManus K, Antinoro L, Sacks F. A randomized controlled trial of a moderate-fat, low-energy diet compared with a low fat, low-energy diet for weight loss in overweight adults. Int J Obes Relat Metab Disord 2001;25:1503–11.
58. Chen L, Appel LJ, Loria C, Lin PH, Champagne CM, Elmer PJ, Ard JD, Mitchell D, Batch BC, Svetkey LP, et al. Reduction in consumption of sugar-sweetened beverages is associated with weight loss: the PREMIER trial. Am J Clin Nutr 2009;89:1299–306.
59. St-Onge MP, Bosarge A. Weight-loss diet that includes consumption of medium-chain triacylglycerol oil leads to a greater rate of weight and fat mass loss than does olive oil. Am J Clin Nutr 2008;87:621–6.
60. Chang KT, Lampe JW, Schwarz Y, Breymeyer KL, Noar KA, Song X, Neuhouser ML. Low glycemic load experimental diet more satiating than high glycemic load diet. Nutr Cancer 2012;64:666–73.
61. Tate DF, Turner-McGrievy G, Lyons E, Stevens J, Erickson K, Polzien K, Diamond M, Wang X, Popkin B. Replacing caloric beverages with water or diet beverages for weight loss in adults: main results of the Choose Healthy Options Consciously Everyday (CHOICE) randomized clinical trial. Am J Clin Nutr 2012;95:555–63.
62. Katcher HI, Legro RS, Kunselman AR, Gillies PJ, Demers LM, Bagshaw DM, Kris-Etherton PM. The effects of a whole grain-enriched hypocaloric diet on cardiovascular disease risk factors in men and women with metabolic syndrome. Am J Clin Nutr 2008;87:79–90.
63. Mason C, Xiao L, Imayama I, Duggan C, Wang CY, Korde L, McTiernan A. Vitamin D3 supplementation during weight loss: a double-blind randomized controlled trial. Am J Clin Nutr 2014;99:1015–25.
64. Kramer MS, Matush L, Vanilovich I, Platt RW, Bogdanovich N, Sevkovskaya Z, Dzikovich I, Shishko G, Collet JP, Martin RM, et al. Effects of prolonged and exclusive breastfeeding on child height, weight, adiposity, and blood pressure at age 6.5 y: evidence from a large randomized trial. Am J Clin Nutr 2007;86:1717–21.
65. de Souza RJ, Bray GA, Carey VJ, Hall KD, LeBoff MS, Loria CM, Laranjo NM, Sacks FM, Smith SR. Effects of 4 weight-loss diets differing in fat, protein, and carbohydrate on fat mass, lean mass, visceral adipose tissue, and hepatic fat: results from the POUNDS LOST trial. Am J Clin Nutr 2012;95:614–25.
66. Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. Clin Transl Gastroenterol 2014;5:e45.
67. Nakai M, Chen D-G, Nishimura KYM. Comparative study of four methods in missing value imputations under missing completely at random mechanism. Open J Stat 2014;4:27–37.
68. Armijo-Olivo S, Warren S, Magee DJ. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. Phys Ther Rev 2009;14:36–49.
69. Efird J. Blocked randomization with randomly selected block sizes. Int J Environ Res Public Health 2011;8:15–20.
70. D'Agostino R, Stephens MA. Goodness-of-fit techniques. New York: Marcel Dekker Inc.; 1986.