



Published in final edited form as:

Int J Obes (Lond). 2021 November ; 45(11): 2335–2346. doi:10.1038/s41366-021-00909-z.

Errors in the implementation, analysis, and reporting of randomization within obesity and nutrition research: a guide to their avoidance

Colby J. Vorland¹, Andrew W. Brown¹, John A. Dawson², Stephanie L. Dickinson³, Lilian Golzarri-Arroyo³, Bridget A. Hannon⁴, Moonseong Heo⁵, Steven B. Heymsfield⁶, Wasantha P. Jayawardene¹, Chanaka N. Kahathuduwa⁷, Scott W. Keith⁸, J. Michael Oakes⁹, Carmen D. Tekwe³, Lehana Thabane¹⁰, David B. Allison³

¹Department of Applied Health Science, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA.

²Department of Nutritional Sciences, Texas Tech University, Lubbock, TX, USA.

³Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, IN, USA.

⁴Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

⁵Department of Public Health Sciences, Clemson University, Clemson, SC, USA.

Reprints and permission information is available at <http://www.nature.com/reprints> This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Correspondence and requests for materials should be addressed to C.J.V. or D.B.A. cvorland@iu.edu; allison@iu.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, D.B.A.; writing - original draft preparation & review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

COMPETING INTERESTS

In the 36 months prior to the initial submission, DBA has received personal payments or promises for the same from: American Society for Nutrition; Alkermes, Inc.; American Statistical Association; Biofortis; California Walnut Commission; Clark Hill PLC; Columbia University; Fish & Richardson, P.C.; Frontiers Publishing; Gelesis; Henry Stewart Talks; IKEA; Indiana University; Arnold Ventures (formerly the Laura and John Arnold Foundation); Johns Hopkins University; Law Offices of Ronald Marron; MD Anderson Cancer Center; Medical College of Wisconsin; National Institutes of Health (NIH); Medpace; National Academies of Science; Sage Publishing; The Obesity Society; Sports Research Corp.; The Elements Agency, LLC; Tomasik, Kotin & Kasserman LLC; University of Alabama at Birmingham; University of Miami; Nestle; WW (formerly Weight Watchers International, LLC). Donations to a foundation have been made on his behalf by the Northarvest Bean Growers Association. DBA was previously a member (unpaid) of the International Life Sciences Institute North America Board of Trustees. In the last 36 months prior to the initial submission, AWB has received travel expenses from University of Louisville; speaking fees from Kentuckiana Health Collaborative, Purdue University, and Rippe Lifestyle Institute, Inc.; consulting fees from Epigeum (Oxford University Press), LA NORC, and Pennington Biomedical Research Center. The institution of DBA, AWB, CJV, SLD, and LG-A, Indiana University, has received funds to support their research or educational activities from: NIH; USDA; Soleno Therapeutics; National Cattlemen's Beef Association; Eli Lilly and Co.; Reckitt Benckiser Group PLC; Alliance for Potato Research and Education; American Federation for Aging Research; Dairy Management Inc; Arnold Ventures; the Gordon and Betty Moore Foundation; the Alfred P. Sloan Foundation; Indiana CTSI, and numerous other for-profit and non-profit organizations to support the work of the School of Public Health and the university more broadly. BAH is an employee of Abbott Nutrition in Columbus, OH. SBH is on the Medical Advisory Board of Medifast Corporation. Other authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41366-021-00909-z>.

⁶Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA, USA.

⁷Department of Psychiatry, School of Medicine, Texas Tech University Health Sciences Center, Lubbock, TX, USA.

⁸Department of Pharmacology and Experimental Therapeutics, Division of Biostatistics, Thomas Jefferson University, Philadelphia, PA, USA.

⁹Department of Epidemiology, School of Public Health, University of Minnesota, Minneapolis, MN, USA.

¹⁰Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada.

Abstract

Randomization is an important tool used to establish causal inferences in studies designed to further our understanding of questions related to obesity and nutrition. To take advantage of the inferences afforded by randomization, scientific standards must be upheld during the planning, execution, analysis, and reporting of such studies. We discuss ten errors in randomized experiments from real-world examples from the literature and outline best practices for their avoidance. These ten errors include: representing nonrandom allocation as random, failing to adequately conceal allocation, not accounting for changing allocation ratios, replacing subjects in nonrandom ways, failing to account for non-independence, drawing inferences by comparing statistical significance from within-group comparisons instead of between-groups, pooling data and breaking the randomized design, failing to account for missing data, failing to report sufficient information to understand study methods, and failing to frame the causal question as testing the randomized assignment per se. We hope that these examples will aid researchers, reviewers, journal editors, and other readers to endeavor to a high standard of scientific rigor in randomized experiments within obesity and nutrition research.

INTRODUCTION

Randomization in scientific experiments bolsters causal inference. Determining a true causal effect would require observing the difference between two outcomes within a single unit (e.g., person, animal) in one case after exogenous manipulation (e.g., “treatment”) and in another case without the manipulation, with all else, including the time of observation, held constant [1]. However, this true causal effect would require parallel universes in which the same unit at the same time undergoes manipulation in one universe but does not in the other. In the absence of parallel universes, we can estimate average causal effects by balancing all differences between multiple units, such that one group looks as similar as possible to the other group. In practice, however, balancing all variables is likely impossible. For practical application, randomization is an alternative because the selection process is independent of the individual’s pre-randomization (observed and unobserved) characteristics that could confound the outcome, and also balances in the long run the distributions of variables that would otherwise be potential confounders, thereby providing unbiased estimation of treatment effects [2]. Randomization and exogenous treatment allow

inferential statistics to create unbiased effect estimates [3]. Departures from randomization may increase uncertainty and yield bias.

Randomization is a seemingly simple concept: just assign people (or more generically, “units” [e.g., mice, rats, flies, classrooms, clinics, families]) randomly to one treatment or intervention versus another. The importance of randomization may have been first recognized at the end of the nineteenth century, and formalized in the 1920s [4]. Yet since its inception there have been errors in the implementation or interpretation of randomized experiments. In 1930, the Lanarkshire Milk investigation tested whether raw or pasteurized milk altered weight and height vs. a control condition in 20,000 schoolchildren [5]. After publication of the experiment, William Gosset (writing as “Student” of “Student’s *t*-test” fame) critiqued the study [6], noting that while there was some random selection of students, a subset of the children were selected on the basis of being either “well fed or ill nourished,” which favored more of the smaller and lighter children being selected, rather than randomized, to the milk groups. Thus, the greater growth in individuals assigned to the milk groups could have been from receiving the milk intervention, or the result of selection bias, an invalidating design flaw. This violates the assumption that the intervention is independent of pre-randomization characteristics of the person being assigned.

Methodologists continue to improve our understanding of the implications of effective randomization, including random sequence generation, implementation (like allocation concealment and blinding), special randomization situations (e.g., randomizing groups of individuals), analysis (e.g., how to analyze an experiment with missing data), and reporting (e.g., how to describe the randomization procedures). Herein, we identify recent publications within obesity and nutrition literature that contain errors in these aspects (see Supplementary Table 1 for a structured list). These examples largely focus on errors arising in the context of null hypothesis significance testing; while there are misconceptions associated with the understanding of *p* values per se [7, 8], it is the framework by which authors typically draw conclusions. The examples span randomized experiments and trials, without or with control groups (i.e., randomized controlled trials [RCTs]). We use these examples to discuss how errors can bias study findings and fail to meet best practices for performing and reporting randomized studies. We clarify that the examples represent a convenience sample, and we make no claims about the frequency of these errors other than that they are frequent enough to have caught our attention. Our categories of errors are neither exhaustive nor in any rank order of severity. Furthermore, we make no assumptions about the circumstances that led to the errors. Rather, we share these examples in the spirit of Gosset who wrote in 1931 on the Lanarkshire Milk experiment, “...but what follows is written not so much in criticism of what was done...as in the hope that in any further work full advantage may be taken of the light which may be thrown on the best methods of arrangement by the defects as well as by the merits” [6].

ERRORS IN IMPLEMENTING GROUP ALLOCATION

1. Error: representing nonrandom allocation methods as random

Description.—Participants are allocated into treatment groups by use of methods that are not random, but the study is labeled as randomized.

Explanation.—Allocation refers to the assignment of subjects into experimental groups. The use of random methods gives each study participant a known probability of being assigned to any experimental group. When any nonrandom allocation is used, studies should not be labeled as randomized.

Examples.—Authors of studies published in a sample of Chinese journals that were labeled as randomized were interviewed about their methods, and in only ~7% was randomization determined to be properly implemented [9]. Improperly labeling studies as randomized is not uncommon in both human and animal research on topics of nutrition and obesity, and can occur in different ways.

In one instance, a vitamin D supplementation trial used a nonrandomized convenience sample from a different hospital as a control group, yet labeled the trial as randomized [10]. In a reply [11], the authors suggested that no selection bias occurred during the allocation because they detected no significant differences between groups on measured covariates. However, this assumption is unjustified because (a) unobserved or mismeasured covariates can potentially introduce bias, or measurement of a covariate may be imperfect, (b) the inferential validity of randomization rests on the assumption that the distributions of all pre-randomization variables are the same in the long run across levels of the treatment groups, not that the distributions are the same across groups in any one sample, and (c) concluding that groups are identical at baseline because no significant differences were detected entails fallaciously “accepting the null.” Regardless of the lack of observed statistical differences between groups, treatment allocation was not randomized and should not be labeled as such.

In another example, researchers first allocated all participants to the intervention to ensure a sufficient sample size and then randomized future participants [12]. This violates the assumption that every subject has some probability of being assigned to every group [13]; the participants first allocated had no probability of being in the control group. In addition, those in the initial allocation wave may have had different characteristics from those with later enrollment.

If units are not all concurrently randomized (e.g., one group is enrolled at a different time), there is also a time-associated confound [14]. This is exemplified by a study of the effects of a nutraceutical formulation on hair growth that was labeled as randomized [15]. Participants were randomized to one of two treatment groups, and then each group underwent placebo and treatment sequentially (essentially a pretest-posttest design). The sequential order suggested a hair growth-by-time confound, with hair growth differing by season [16].

Nonrandom allocation can leave a signature in baseline between-group differences. With randomization, on average, the p values of baseline group comparisons will be uniform for independent measurements. While there are limitations to applying this principle broadly to assessing literature [17-19], in some cases it has proved useful as a prompt for more information about how and whether randomization was actually employed. An analysis by Carlisle of baseline p value distributions in over 5000 trials flagged apparent deviations from this expectation [20], suggesting that many studies labeled as randomized may not

be. One trial flagged [21] was the Primary Prevention of Cardiovascular Disease with a Mediterranean Diet (PREDIMED) trial, which highlighted the significant impact of advice to consume a Mediterranean-style diet coupled with additional intake of extra-virgin olive oil or mixed nuts on risk for cardiovascular disease, compared with advice to consume a low-fat diet [22]. An audit by the PREDIMED authors discovered that members of some of the households were nonrandomly assigned to the same group as the randomized member. Furthermore, one intervention site switched from individuals to clinics as the randomization unit [23] (see section 5, “Error: failing to account for non-independence” for discussion of non-independence). Thus, the original analysis at the individual level was inappropriate for these participants because some did not have a known probability of being assigned to one of the treatment groups or the control. A retraction and reanalysis did not change the main results or conclusions [23], although causal language in the article was tempered. Conclusions from secondary analyses were affected, however, such as the 5-year change in body weight and waist circumference, which changed statistical significance for the olive oil group [24]. Use of statistical principles to examine the likelihood that randomization was properly implemented has flagged other studies related to nutrition and obesity, too [25-28]. In at least four cases, publications were retracted [22, 26, 29, 30].

Best practices.—Where randomization is impossible, methods should be clearly stated so that there is no conflation of nonrandomized with randomized experiments. Investigators should establish procedures a priori to monitor how randomization is implemented. Furthermore, although a given randomized sample may not appear balanced on all measurable baseline variables, by definition those imbalances have occurred by chance. Altering the allocation process to enforce balance with the use of nonrandom methods may introduce bias. Importantly, use of nonrandom methods may warrant changing how study results are communicated. At a practical level, most methodologists and statisticians would agree that if an RCT is properly randomized, it is reasonable to make causal claims about intervention assignment and outcomes. Whereas the purpose of most research is to seek causal effects [31], errors discussed herein break randomization, and thereby introduce additional concerns that must be satisfied to increase the confidence in unbiased estimates. While a nuanced discussion of the use of causal language is outside the scope of this review, from a purist perspective, the description of relationships as causal from nonrandom methods is inappropriate [32].

Where important pre-randomization factors are identified that could influence results if they are imbalanced (such as animal body weight), forms of restricted randomization exist to maintain the benefits of randomization with control over such factors, instead of using haphazard methods that may introduce bias. These include blocking and stratification [33, 34], which necessitate additional consideration at the analysis stage beyond a simple randomization scheme (see section 5, “Error: failing to account for non-independence”).

2. Error: failing to adequately conceal allocation from investigators

Description.—Investigators who assign treatments, and the participants receiving them, are inadequately concealed from knowing what condition was assigned.

Explanation.—Allocation concealment, when implemented properly, prevents researchers from foreknowing the allocation of the next participant. Furthermore, it prevents participants from foreknowing their assignment, who may choose to dropout if they do not receive a preferred treatment. Thus, concealment prevents selection bias and confounding [35-37]. Whereas randomization is a method to create unbiased estimates of effect, allocation concealment is necessary to remove the human element of decisions (whether conscious or unconscious) when participants are assigned to groups, and both are important for a rigorous trial. When concealment is broken, sample estimates can become biased in different ways.

Examples.—Even with the use of random allocation methods, the failure to conceal allocation means that the researchers, and sometimes participants, will know upcoming assignments. The audit of PREDIMED, as discussed in section 1, “Error: representing nonrandom allocation methods as random,” also clarified that allocation was not concealed [23], despite using computergenerated randomization tables. In the case of the Lanarkshire study as described above [5, 6], the failure to conceal allocation led to conscious bias in how schoolchildren were assigned to the interventions. In other cases, researchers may unconsciously bias allocations if they have any involvement in the allocation. For example, if the researcher who is doing the allocation is using a physical method of randomization such as rolling a die or flipping a coin in the presence of the subject, their perception of how the die or coin is rolled or flipped, or how it falls, leaves room to redo it in ways that may select for certain subjects being allocated to particular assignments.

Nonrandom allocation also may make concealment impossible; examples and explanations are presented in Table 1.

Best practices.—Appropriate concealment strategies may vary by study, but it is ideal that concealment be implemented. The random generation and storage of allocation codes is essential to allocation concealment, using generic numerals or letters unknown to the investigator. Electronic generation and storage of allocations in a protected centralized database is sometimes preferred [33, 38] to opaque sealed envelopes [39, 40], which is not completely immune to breach and can bias the results if poorly carried out or intentionally compromised [41-43]. Furthermore, if feasible, real-time generation may be favored over pregenerated allocations [44]. Regardless of physical or electronic concealment, the allocation codes and other important information about the assignment scheme, such as block size in permuted block randomization [45], should remain concealed from all research staff and participants. Initial allocation concealment can still be implemented and would improve the rigor of trials even if blinding (i.e., preventing post-randomization knowledge of group assignments) throughout the trial cannot be maintained.

3. Error: not accounting for changes in allocation ratios

Description.—The allocation ratio or number of treatment groups is changed partway through a study, but the change is not accounted for in the statistical analysis.

Explanation.—Over the course of a study, researchers may intentionally change treatment group allocations, such as adding, dropping, or combining treatment arms, for various reasons. When researchers change allocation ratios mid-study, this must be taken into account during statistical analysis [46]. Allocation ratios also change in “adaptive trials,” which have specific methods and concerns beyond what we can cover here (see [47] for more information).

Examples.—A study evaluating effects of weight loss on telomere length performed one phase by randomizing participants to three treatment groups (in-person counseling, telephone counseling, and usual care) with 1:1:1 allocation. After no significant difference was found between in-person and telephone counseling, participants in the next phase of the study were randomized with 1:1 allocation into a combined intervention of in-person and telephone counseling or usual care [48]. In addition to the authors’ choice of analyzing interim results before starting another phase (which risks increasing false-positive findings and should be accounted for in statistical analysis [49]), the analysis combined these two phases, effectively analyzing 2:1 and 1:1 allocations together [50]. Another study of low-calorie sweeteners and sucrose and weight-related outcomes [51] started by randomly allocating participants evenly to five treatment groups with 1:1:1:1:1 allocation, but changed to 2:1:1:1:1 midway through after one group had a higher attrition rate. Neither of these two studies reported accounting for these different phases of study in the statistical analysis. Using different allocation ratios for different groups can bias study results [46, 50]. This is because differences may exist between the different periods of recruitment in participant characteristics, such as baseline BMI [46, 50]. Thus, baseline differences in the wave of participants allocated at the 2:1 ratio, when pooled with the ratio of those allocated at the 1:1 ratio, would exaggerate the differences when analyzed as though all participants were allocated at the same time.

Best practices.—When allocation ratios change within studies or between randomized experiments that are pooled, caution should be used in combining data. Changes in allocation ratios must be properly taken into account in statistical analysis (see section 7, “Error: improper pooling of data”).

4. Error: replacements are not randomly selected

Description.—Participants who dropout are replaced in ways that are nonrandom, for instance, by allocating individuals to a single treatment that experienced a high percentage of participant dropout.

Explanation.—Nonrandom replacement of dropouts is another example of changing allocation ratios. Dropout is common in real-world studies and often leads to missing data, bias, and potentially the loss of power. A meta-analysis of pharmaceutical trials for obesity estimated an average 1-year dropout rate of 37% [52]. Similarly, a secondary analysis of a diet intervention estimated that the probability of completing the trial was only 60% after just 12 weeks [53]. Analytical approaches like intention-to-treat [ITT] analysis and imputation of data (described in the Errors in analysis section below) may obviate the need to consider replacing subjects after the initial randomization [52, 54]. Yet replacement is

sometimes observed in the literature and failing to use random methods to do so introduces another source of potential bias.

In a properly implemented simple RCT, every subject will have the same a priori probability of belonging to any group as any other subject. When a subject who has dropped out is replaced with the next person enrolled instead of by using randomization for assignment, the new participant did not have the same chances as the other subjects in the study of being allocated to that group. This corrupts the process of randomization, potentially introducing bias, and compromises causal inference. Furthermore, allocating participants this way makes allocation concealment impossible.

It is vital to account for dropout in the calculation of sample size and allocation ratios when designing the study. Nevertheless, if dropout was not accounted for a priori, one option is that for the number of dropouts encountered, new participants are enrolled, but each new participant is randomly assigned to groups with the same allocation ratios as the originals [55]. Note that if dropouts are higher from a particular group and if completers only are analyzed, this may result in an imbalance in the final sample group allocation, but this is not an issue if the ITT principle is adhered to (see section 8, “Error: failing to account for missing data”).

Examples.—Often, studies do not specify the methods used to replace subjects and use nondescript sentences similar to “subjects who dropped out were replaced” [56-59]. As discussed in regard to a trial on green tea ointment and pain and wound healing [60], such vagueness might suggest introduction of bias and lead to questionable conclusions.

Best practices.—Although replacing subjects may indeed help with the problem of power, the consequences can be detrimental if not properly implemented. Therefore, the decision to replace participants should be thoroughly considered, preplanned if at all possible, and performed by using correct methods, if found to be necessary.

ERRORS IN THE ANALYSIS OF RANDOMIZED EXPERIMENTS

5. Error: failing to account for non-independence

Description.—Groups of subjects (e.g., classrooms, schools, cages of animals) are randomly assigned to experimental conditions together but the data are analyzed as if they were randomized individually, or repeated within-subject measures are treated as independent. Or, measures are treated as independent when subjects individually randomized have repeated within-subject measures or are treated in groups.

Explanation.—The use of cluster randomized trial (cRCT) designs is increasing in nutrition and obesity studies, particularly for the study of school-based interventions, and in contexts where participants are exposed to the other group(s) and as such there is a lack of independence. Similarly, animals are commonly housed together (e.g., in cages, tanks) or grouped by litter. If investigators randomize all animals to treatments by groups instead of individually, this correlation must be addressed in the analysis, but is often unrecognized or ignored. These concerns also exist in cell culture experiments, for example, if treatments

are randomized to an entire plate instead of individual wells. In cluster designs, the unit of randomization is the cluster, and not the individual. A frequent error in such interventions is to power and analyze the study at the individual (e.g., person, animal) level instead of the cluster level. Failing to account for within-cluster correlation (often measured by the intraclass correlation coefficient) and cluster-level impacts during study planning leads to an overestimation of statistical power [61] and typically leads to p values and associated confidence intervals that are artificially small [62, 63].

If cRCTs are implemented incorrectly to start, valid inferential analysis for treatment effects is not possible without untestable assumptions [61]. For instance, randomly assigning one school to an intervention and one to a control yields no degrees of freedom, akin to randomizing one individual to treatment and one to control and treating multiple measurements on each of the two individuals as though those measurements were independent [61].

Studies that randomize at the individual level may also have correlated observations that should be considered in the analysis, and so it is important to identify potential sources of clustering. For example, outcome measures may be correlated when animals are individually randomized but then group housed for treatment. Likewise, participants individually randomized may be treated in group sessions (such as classes related to the intervention), or may be grouped together within surgeons that do not equally operate in all study arms. These types of scenarios require consideration in statistical analysis [64]. When repeated measurements are taken on subjects, they similarly must account for within-subject correlation. Taking multiple measurements within individuals (e.g., measuring eyesight in the left and right eye or longitudinal data within person over time) and treating them as independent will lead to invalid inferences [64].

A distinct issue exists when using forms of restricted randomization (e.g., stratification, blocking, minimization) that are employed to have tighter control over particular factors of interest. In such situations, it is important to include the factors on which randomization restrictions occur as covariates in the statistical model to account for the added correlation between groups [65, 66]. Not doing so can result in p values and associated confidence intervals that are artificially large and reduced statistical power. On the other hand, given that one is likely employing restricted randomization because of a small number of units of randomization, losing even a few “denominator” degrees of freedom due to the inclusion of additional covariates in the model may also adversely affect power [67, 68].

Examples.—Failing to account for clustering is one of the most pervasive errors in nutrition and obesity studies that we observe [6, 61, 69-79]. A review of school-based randomized trials with weight-related outcomes found that only 21.5% of studies used intraclass correlation coefficients in their power analysis, and only 68.6% applied multilevel models to account for clustering [80]. In the most severe cases that we observe, a failure to appropriately focus on the cluster as the unit of randomization invalidates any hope of deriving causal inferences [70, 75, 81]. For additional discussion of errors in implementation and reporting in cRCTs, see ref. [61].

In an example of clustering within participants, a study of vitamin E on diabetic neuropathy randomized participants to the intervention or placebo, but for outcomes related to nerve conduction, the authors conducted measurements in limbs, stating that “left and right sides were treated independently” [82]. Because these measures were taken within the same participants, within-subject correlations must be taken into account in statistical analyses. Treating non-independent measurements as independent in statistical analysis is sometimes called “pseudoreplication” and is also a common error in animal and cell culture experiments [83].

Best practices.—When planning cRCTs, it is critical to perform a power calculation that incorporates the number of clusters in the design [61]. Moreover, analyses of such designs, as well as individually randomized designs, need to include the correlations from clustering for proper treatment inferences, just as repeated measurements of outcomes within subjects must be treated as non-independent.

6. Error: basing conclusions on within-group statistical tests instead of between-groups tests

Description.—Experimental groups are analyzed separately for significant differences in the change from baseline and a difference is concluded if one is significant and the other(s) not, instead of comparing directly between groups.

Explanation.—The probative comparison for RCTs is between groups. Sometimes, however, researchers use pre-post within-group tests and draw conclusions based on whether the within-group significance is different, for example, significant in one group but not the other (the so-called “Difference in Nominal Significance” or DINS error [84]). Using these within-group tests to imply differences between groups increases the false-positive rate of 5% for equal group sizes to up to 50% (and higher for unequal groups) [85] and is therefore invalid.

Examples.—The DINS error was identified in an RCT testing isomaltulose vs. sucrose in the context of effects of an energy-reduced diet on weight and fat mass, where some conclusions, such as the outcome of fat mass, were drawn from within-group comparisons but the between-group comparison was not statistically different [86]. We observe this error frequently in nutrition and obesity research [87-103]. Sometimes using this logic still reaches the correct conclusions (i.e., the between-group and within-group comparisons are both statistically significant or not), but often it does not, and therefore it is an unreliable approach for inferences.

Best practices.—For proper analysis of RCTs, within-group testing should not be represented as the comparison of interest [71, 84, 85, 87, 102]. Journal editors, reviewers, and readers should request that conclusions be drawn from between-group comparisons.

7. Error: improper pooling of data

Description.—Data for a single RCT are pooled without maintaining the randomized design, or data from multiple RCTs are pooled (i.e., meta-analysis) without accounting for study in statistical analysis.

Explanation.—Data for statistical analysis can be pooled either within one or multiple RCTs, but errors can arise when the random elements of assignment are disregarded. Pooling within one study refers to the process of combining data across different groups, subgroups, or sites to include in a single analysis. When a single RCT is performed across multiple sites or subgroups and the same allocation ratio is not used across all sites or subgroups, or the randomization allocation to study arms changes during the course of an RCT, these different sites, subgroups, or phases of the study need to be taken into account during data analysis. This is because assignment probability is confounded with subset. If data are pooled simply with no account for subsets, any differences between subsets can bias effect estimation [50].

When combining multiple RCTs, individual participant data (IPD) can be used (i.e., IPD meta-analysis). However, if they are treated as though they came from a single RCT without accounting for site, at best it will increase the residual variance and make the analysis inefficient, and at worst will confound the results and make the effect estimates biased [104]. Another error in IPD meta-analyses is the use of data pooled across trials to compare intervention effects in one subgroup of participants with another (e.g., to test the interaction between intervention and pre-randomization subgroups) without accounting for trial in the analysis. This increases the risk of bias, owing to lack of knowledge of individual within- and across-trial interaction effects and inability to separate them, as well as inappropriate standard errors for the interaction effect [105]. This differs from “typical” meta-analyses because the effect estimates already account for the fact that both treatment groups existed in the same study.

Examples.—In the trial of how weight loss affects telomere length in women with breast cancer (see subsection “Examples” under section 3, “Error: not accounting for changes in allocation ratios”), data were pooled from two different phases of an RCT that had different allocation ratios, which was not taken into account in the analysis [50]. Another example is a pooling study that combined IPD from multiple RCTs to examine the effects of a school-based weight management program on summer weight gain among students but ignored “study” as a factor in the analysis [106].

Best practices.—When pooling data under the umbrella of one study (e.g., allocation ratio change during the study), statistical analysis should include variables for subgroups to prevent confounding [46]. When pooling IPD from multiple RCTs, care must be taken to include a term for “study” when group conditions or group allocation ratios are not identical across all included RCTs [106]. For additional information on methods for IPD meta-analysis, see ref. [105].

8. Error: failing to account for missing data

Description.—Missing data (due to dropouts, errors in measurement, or other reasons) are not accounted for in an RCT.

Explanation.—The integrity of the randomization of subjects must be maintained throughout a study. Any post-randomization exclusion of subjects or observations, or any instances of missingness in post-randomization measurements, violates both randomization and the ITT principle (analyzing all subjects according to their original treatment assignments) and thus potentially compromises the validity of any statistical analyses and the conclusions drawn from them. There are two main reasons for this. Whereas randomization minimizes potential confounding by providing similar distributions in baseline participant characteristics, missing data that are not completely at random breaks the randomization, introduces potential bias in various ways, and degrades the confidence that the effect (or lack thereof) is the result only of the experimental condition [107, 108]. Consider as an example reported income. If individuals with very low or very high incomes are less likely to report their incomes, then nonmissing income values and their corresponding covariate values cannot provide valid inference for individuals who did not report income, because the populations are simply not the same. Missing data are extremely common in RCTs, as discussed in section 4, “Error: replacements are not randomly selected.” Regardless of the intervention, investigators need to be prepared to handle missing data based on assumptions about how data are missing.

Examples.—One review found that only 50% of trials use adequate methods to account for missing data [109], and studies of obesity and nutrition are no exception. For example, in a trial of intermittent vs. continuous energy restriction on body composition and resting metabolic rate with a 50% dropout rate, reanalysis of all participants halved the magnitude of effect estimates compared with analyses of completers only [99]. As in this case, investigators will often report analyses performed only on participants who have completed the study, without also reporting an ITT analysis that includes all subjects who were randomized. Investigators may dismiss ITT analyses because they perceive them as “diluting” the effect of the treatment [110]. However, this presumes that there is an effect of treatment at all. Dropouts may result in an apparent effect that is actually an artifact. If dropouts are nonrandom, then groups may simply appear different because people remaining in the treatment group are different people from those who dropped out. Attempts to estimate whether those who dropped out differ from those who stayed in are often underpowered.

Furthermore, some investigators may not understand ITT and mislabel their analysis. For instance, in an RCT of a ketogenic diet in patients with breast cancer, the authors reported that “[s]tatistical analysis was carried out according to the intention-to-treat protocol” of the 80 randomized participants, yet the flow diagram and results suggest that the analyses were restricted to completers only [111]. Surveys of ITT practices suggest that there is a general lack of adequate reporting of information pertaining to how missing data is handled [112].

Best practices.—Many analyses can be conducted on randomized data including “per protocol” (removing data from noncompliant subjects) and ITT. However, simply comparing per protocol to ITT analyses as a sensitivity analysis is suboptimal; they estimate different things [113]. As such, the Food and Drug Administration has recently focused on the concept of estimands to clearly establish the question being tested [114]. ITT can estimate the effect of assignment, not treatment per se, in an unbiased manner, whereas the per protocol analysis can only estimate in a way that allows the possibility for bias.

In an oft-paraphrased maxim of Lachin [108], “the best way to deal with the problem [of missing data] is to have as little missing data as possible.” This goal may be furthered through diligent administrative follow-up and constant contact with subjects; further considerations on minimization of loss-to-follow-up and other missingness may be found elsewhere [115, 116]. However, having no missing data whatsoever is often not achievable in practice, especially for large, randomized studies. Thus, something must be done when missing data exist. In general, the simplest and best way to mitigate the problem of missing data is through the ITT principle when conducting the statistical analysis.

Statistical approaches for handling missing data require untestable assumptions, assumptions that lack face validity and hence are unfounded, or both [108]. Complete case analyses, where subjects with missing data are ignored, require assumptions that the data are missing completely at random that are not recommended [108]. Multiple imputation fills in missing data repeatedly, with relationship and predictions guided by other covariates, and is recommended under the assumption that data are missing at random (MAR); that is, the missingness or not of an observation is not directly impacted by its true value. Methods commonly used in obesity trials such as last observation carried forward (LOCF) [117] or baseline observation carried forward (BOCF) are not recommended because of the strict or unreasonable assumptions required to yield valid conclusions [108, 117, 118]. In such cases where values are missing not at random (MNAR; this set of assumptions may also be referred to as “not missing at random”, NMAR), explicit modeling for the missingness process is required [119], requiring stronger assumptions that may not be valid.

Finally, when it is apparent that data are MNAR, when the integrity of randomization is no longer intact, or both, estimates are no longer represented as a causal effect afforded by randomization and care should be taken that causal language is tempered. Even in cases where the assumptions are violated, however, ignoring the missingness (e.g., completers only analyses) is generally not recommended.

In summary, minimizing missing data should be a key goal in any randomized study. But when data are missing, thoughtful approaches are necessary to respect the ITT principle and produce unbiased effect estimates. Additional discussion about best practices to handle missing data in the nutrition context is available at ref. [107].

ERRORS IN THE REPORTING OF RANDOMIZATION

9. Error: failing to fully describe randomization

Description.—Published reports fail to provide sufficient information so that readers can assess the methods used for randomization.

Explanation.—Studies cannot be adequately evaluated unless methods used for randomization are reported in sufficient detail. Indeed, many examples described herein were obscured by poor reporting until we or others were able to gain clarification from the study authors through personal communication or post-publication discourse. Accepted guidelines that define the standards of reporting the results of clinical trials (i.e., Consolidated Standards of Reporting Trials for human trials (CONSORT) [120]), animal research (i.e., Animal Research: Reporting of In Vivo Experiments (ARRIVE) [121]), and others [122] have emphasized the importance of adequate reporting of randomization methods. Researchers should, to the fullest extent possible, report according to accepted guidelines as part of responsible research conduct [123].

Examples.—Most authors (including historically us), however, do not report adequately, and this includes randomization sequence generation and allocation concealment in human and animal research [124, 125]. We have noted specific examples of a failure to include sufficient details about the method of randomization and allocation ratio in a study of dairy- and berry-based snacks on nutritional status and grip strength [126], which were clarified in a reply [127]. In a personal communication regarding another trial of a nutritional intervention on outcomes in individuals with autism spectrum disorder, we learned that the authors had used additional blocking factors, and randomized some siblings as pairs, neither of which were reported in the paper nor accounted for in the statistical analysis [128]. In another study that pooled RCTs of school-based weight management programs, the reported number of participants of the included studies was inconsistent with the original publications [106]. In other cases, the methods used to account for clustering may not be appropriately described for readers to assess them [129, 130]. In one case, the authors reported randomizing in pairs, yet the number randomized was an odd number and differed between groups ($n = 21$ and $n = 24$) [131], to which the authors reported a coding error [132]. Other vague language descriptions include statements such as “the samples were randomly divided into two groups” [27].

The use of non-specific language to describe allocation methods may also lead to confusion as to whether randomized methods were actually used. For example, we observed the term “semi-random” used to reflect stratified randomization [133] or minimization [134], whereas elsewhere it may describe methods that are nonrandom or not clearly stated [135].

Best practices.—Neglecting to report essential components of how randomization was implemented hinders the ability of a reader from fully evaluating the trial and hence from interpreting the validity of the reported findings. We emphasize that reporting guidelines such as CONSORT [120] should be consulted during the study planning and publication preparation stages to ensure that essential components related to randomization are reported, such as methods used to generate the allocation sequence, implement randomization, and

conceal allocation; any matching or blocking procedures used; accuracy and consistency of the numbers in flow diagrams; and reporting baseline demographic and clinical variables. With regard to the last point, a common error is to report p values of baseline statistical comparisons and conclude covariate imbalance between groups if they are <0.05 . An example of this type of thinking is as follows: “[a]s randomization was not fully successful concerning age, it was included as covariate in the main analyses.” [136], or conversely, “The similarity between the exercise plus supplement and exercise plus placebo groups for both demographic composition and pre-intervention fitness and cognitive scores provides strong evidence that participants were randomly assigned into groups” [137]. However, as discussed in section 1, “Error: representing nonrandom allocation methods as random,” the distribution of p values from baseline group comparisons is uniform in the long run with randomization and therefore we would expect on average that $1/20$ p values will be <0.05 by chance, with some caveats [17-19]. In other words, per CONSORT, “[s]uch significance tests assess the probability that observed baseline differences could have occurred by chance; however, we already know that any differences are caused by chance” [120], and should not be reported. Baseline p values do not reflect whether imbalances might affect the results; imbalanced variables that are prognostic on the outcome that are not $p < 0.05$ can still have a strong effect on the result [138, 139]. Thus, statistical tests should not be used to determine prognostic covariates; such covariates should preferably be identified and included in an analysis plan prior to executing the study [139].

10. Error: failing to properly communicate inferences from randomized studies

Description.—The causal question is not framed as testing the randomized *assignment* per se.

Explanation.—The appropriate execution and analysis of a randomized experiment tests the effect of treatment *assignment* on the outcome of interest. The causal effect being tested is what participants are assigned to, not what they actually did. That is, if some participants drop out, do not comply with the intervention, are accidentally given the wrong treatment, or in other ways do not complete the intended treatment, the proper analysis maintains the randomized *assignment* of the subjects and tests the effect of assigning subjects to the treatment, which includes factors beyond the treatment itself. Indeed, it may be that dropout or non-compliance is caused by the assignment itself. This distinction is particularly important in nutrition trials, which often suffer from poor compliance, and is discussed in part in subsection “Explanation” under section 8, “Error: failing to account for missing data” with respect to the ITT principle. For instance, researchers may be interested in discussing the effect of *eating* their diet, when in fact what was tested was being *assigned to eat* the diet.

Examples.—As discussed in section 8, “Error: failing to account for missing data,” there is often a perception among authors that including subjects that are, e.g., noncompliant or incorrectly assigned will preclude an understanding of the true effect of the intervention on the outcome(s) of interest. But the realization of unbiased effect estimates that the principles of randomization afford us is only achieved when subjects are analyzed as they are randomized. For example, the random assignment to 25% energy restriction of

participants in a 2-year trial resulted in an average reduction of about 12% (~300 kcal) [140]. The public discussion of this trial advertised that “Cutting 300 Calories a Day Shows Health Benefits” [141]. Yet it is possible that assigning participants to cut only 300 kcal would not have produced the same benefits if they once again achieved only half of that assigned. In another example, the random assignment of high phytate bread did not lead to a statistically significant difference in whole body iron status as compared to dephytinized bread when missing data was imputed, but it was significantly higher when dropouts were excluded [98, 142, 143]. A difference cannot be concluded from these data based on the causal question of the assignment of high phytate bread, particularly because dropout was significantly higher in one group, which may create an artificial effect.

Best practices.—The appropriate framing of the treatment assignment (i.e., following the ITT principle) as the causal effect of interest is important when communicating and interpreting results of RCTs. From this perspective, maximizing the validity of randomized studies from planning, execution, and analysis is a matter of maintaining the randomized assignments to the greatest extent possible. To this end, randomized studies should be communicated carefully that the causal question is assignment to treatment.

CONCLUSION

Randomization is a powerful tool to examine causal relationships in nutrition and obesity research. Empirical evidence supports the use of both randomization and allocation concealment for unbiased effect estimates. Trials with inadequate concealment are associated with larger effect estimates than are those with adequate concealment [144-147], likely reflecting bias. Despite such undesirable potential consequences, many randomized studies of humans and animals do not adequately conceal allocation [43, 124, 148]. Although more difficult to compare in human studies, the results of nonrandomized studies sometimes differ from those of randomized trials [149], while nonrandomized animal studies are associated with increased effect sizes [148]. These empirical observations are suggestive of biased estimates, and when coupled with the theoretical arguments, indicate that randomization should be implemented whenever possible. For these reasons, where randomization is implemented per the best practices described herein, the use of causal language to communicate results is appropriate. But where it is not correctly implemented or maintained, the greater potential for bias in the effect estimates and additional assumptions that need to be met to increase confidence in causal relationships invariably changes how such effects should be communicated.

Even when randomization is implemented, errors related to randomization are common, suggesting that researchers in nutrition and obesity may benefit from statistical support during the design, execution, analysis, and reporting of randomized experiments for more rigorous, reproducible, and replicable research [150]. When errors are discovered, authors and editors have a responsibility to correct the scientific record, and journals should have procedures in place to do so expeditiously [151]. The severity of the error, ranging from invalidating the conclusions [152] to simply requiring clarification, means that different considerations exist for each type of error. For example, some invalidating errors are consequent to the design and cannot be fixed, and retractions have been issued [29, 153,

154]. For other examples such as PREDIMED, for which errors in randomization required a reanalysis as a quasi-experimental design, the reanalysis, retraction, and republication serve as an important example of scientific questioning and transparency of research methods [155]. Other cases require reanalysis or reporting of the appropriate statistical analyses but are otherwise not invalidated by design flaws [88, 156]. Yet others need clarity on the methods, for instance when a study did not really use random allocation but reported as such [157].

The involvement of professional biostatisticians and others with methodological expertise from the planning stages of a study will prevent many of these errors. The use of trial and analysis plan preregistration can aid in thinking through decisions a priori while simultaneously increasing transparency and guarding against unpublished results and inflated false positives from analytic flexibility by pre-specifying outcomes and analyses [71]. Being cognizant of these errors and becoming familiar with CONSORT and other reporting guidelines enhance the value of the time, effort, and financial investment we devote to obesity and nutrition research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Zad Rafi for insightful feedback on an earlier draft and Jennifer Holmes for editing our manuscript.

FUNDING

CJV is supported in part by the Gordon and Betty Moore Foundation. DBA and AWB are supported in part by NIH grants R25HL124208 and R25DK099080. SBH is supported in part by National Institutes of Health NORC Center Grants P30DK072476, Pennington/Louisiana and P30DK040561, Harvard. CDT research is supported by National Cancer Institute Supplemental Award Number U01-CA057030-29S2. Other authors received no specific funding for this work. The opinions expressed are those of the authors and do not necessarily represent those of the NIH or any other organization.

REFERENCES

1. Imbens GW, Rubin DB. Rubin causal model. In: Durlauf SN, Blume LE, (eds.) *Microeconometrics*. London: Springer; 2010. p. 229–41 10.1057/9780230280816.
2. Senn S. Seven myths of randomisation in clinical trials. *Stat Med*. 2013;32:1439–50. [PubMed: 23255195]
3. Greenland S. Randomization, statistics, and causal inference. *Epidemiology*. 1990;1:421–9. [PubMed: 2090279]
4. CG P, Glud C. The controlled clinical trial turns 100 years: Fibiger's trial of serum treatment of diphtheria. *BMJ*. 1998;317:1243–5. [PubMed: 9794873]
5. Leighton G, McKinlay PL. Milk consumption and the growth of school children. Report on an investigation in Lanarkshire schools. Scotland, Edinburgh: H.M.S.O.; 1930. p. 20.
6. Student. The Lanarkshire Milk experiment. *Biometrika*. 1931;23:398–406.
7. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 2016;70:129–33.
8. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol*. 2016;31:337–50. [PubMed: 27209009]

9. Wu T, Li Y, Bian Z, Liu G, Moher D. Randomized trials published in some Chinese journals: how many are randomized? *Trials*. 2009;10:1–8. [PubMed: 19128475]
10. Shub A, McCarthy EA. Letter to the Editor: “Effectiveness of prenatal vitamin D deficiency screening and treatment program: a stratified randomized field trial”. *J Clin Endocrinol Metab*. 2018;104:337–8.
11. Ramezani Tehrani F, Minooe S, Rostami M, Bidhendi Yarandi R, Hosseinpanah F. Response to Letter to the Editor: “Effectiveness of prenatal vitamin D deficiency screening and treatment program: a stratified randomized field trial”. *J Clin Endocrinol Metab*. 2018;104:339–40.
12. Williams LK, Abbott G, Thornton LE, Worsley A, Ball K, Crawford D. Improving perceptions of healthy food affordability: results from a pilot intervention. *Int J Behav Nutr Phys Act*. 2014;11:33. [PubMed: 24606876]
13. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol*. 2010;171:674–7. [PubMed: 20139125]
14. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Wilmington, MA: Houghton Mifflin Company; 1963.
15. Tenore GC, Caruso D, Buonomo G, D’Avino M, Santamaria R, Irace C, et al. Annurca apple nutraceutical formulation enhances keratin expression in a human model of skin and promotes hair growth and tropism in a randomized clinical trial. *J Med Food*. 2018;21:90–103. [PubMed: 28956697]
16. Keith SW, Brown AW, Heo M, Heymsfield SB, Allison DB. Re: “Annurca apple nutraceutical formulation enhances keratin expression in a human model of skin and promotes hair growth and tropism in a randomized clinical trial” by Tenore et al. (*J Med Food* 2018;21:90–103). *J Med Food*. 2019;22:1301–2. [PubMed: 31718391]
17. Bolland MJ, Gamble GD, Avenell A, Grey A. Rounding, but not randomization method, non-normality, or correlation, affected baseline P-value distributions in randomized trials. *J Clin Epidemiol*. 2019;110:50–62. [PubMed: 30858019]
18. Bolland MJ, Gamble GD, Avenell A, Grey A, Lumley T. Baseline P value distributions in randomized trials were uniform for continuous but not categorical variables. *J Clin Epidemiol*. 2019;112:67–76. [PubMed: 31125614]
19. Mascha EJ, Vetter TR, Pittet J-F. An appraisal of the Carlisle-Stouffer-Fisher method for assessing study data integrity and fraud. *Anesth Analg*. 2017;125:1381–5. [PubMed: 28786843]
20. Carlisle JB. Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia*. 2017;72:944–52. [PubMed: 28580651]
21. The Editors of the *Lancet Diabetes & Endocrinology*. Retraction and republication—effect of a high-fat Mediterranean diet on bodyweight and waist circumference: a prespecified secondary outcomes analysis of the PREDIMED randomised controlled trial. *Lancet Diabetes Endocrinol*. 2019;7:334 [PubMed: 31003619]
22. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med*. 2013;368:1279–90. [PubMed: 23432189]
23. Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Aros F, 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:e34. [PubMed: 29897866]
24. Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, Fitó M, Chiva-Blanch G, et al. Effect of a high-fat Mediterranean diet on bodyweight and waist circumference: a prespecified secondary outcomes analysis of the PREDIMED randomised controlled trial. *Lancet Diabetes Endocrinol*. 2019;7:e6–17. [PubMed: 31003626]
25. Mestre LM, Dickinson SL, Golzarri-Arroyo L, Brown AW, Allison DB. Data anomalies and apparent reporting errors in ‘Randomized controlled trial testing weight loss and abdominal obesity outcomes of moxibustion’. *Biomed Eng Online*. 2020;19:1–3. [PubMed: 31915014]
26. Abou-Raya A, Abou-Raya S, Helmii M. The effect of vitamin D supplementation on inflammatory and hemostatic markers and disease activity in patients with systemic lupus erythematosus: a randomized placebo-controlled trial. *J Rheumatol*. 2013;40:265–72. [PubMed: 23204220]

27. George BJ, Brown AW, Allison DB. Errors in statistical analysis and questionable randomization lead to unreliable conclusions. *J Paramed Sci.* 2015;6:153–4. [PubMed: 26949506]
28. Bolland M, Gamble GD, Grey A, Avenell A. Empirically generated reference proportions for baseline p values from rounded summary statistics. *Anaesthesia.* 2020;75:1685–7. [PubMed: 32524590]
29. Hsieh C-H, Tseng C-C, Shen J-Y, Chuang P-Y. Retraction Note to: randomized controlled trial testing weight loss and abdominal obesity outcomes of moxibustion. *Biomed Eng Online.* 2020;19:1. [PubMed: 31915014]
30. Hosseini R, Mirghotbi M, Pourvali K, Kimiagar SM, Rashidkhani B, Mirghotbi T. The effect of food service system modifications on staff body mass index in an industrial organization. *J Paramed Sci.* 2015;6:2008–4978.
31. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health.* 2018;108:616–9. [PubMed: 29565659]
32. Lazarus C, Haneef R, Ravaud P, Boutron I. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol.* 2015;15:85. [PubMed: 26462565]
33. Moher D, Hopewell S, Schulz KF, Montori V, Gotsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c869. [PubMed: 20332511]
34. Altman DG, Bland JM. How to randomise. *BMJ.* 1999;319:703–4. [PubMed: 10480833]
35. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. *Trials.* 2015;16:405. [PubMed: 26357929]
36. McKenzie JE. Randomisation is more than a coin toss: the role of allocation concealment. *BJOG.* 2019;126:1288. [PubMed: 31222903]
37. Chalmers I. Why transition from alternation to randomisation in clinical trials was made. *BMJ.* 1999;319:1372. [PubMed: 10567161]
38. Torgerson DJ, Roberts C. Understanding controlled trials. Randomisation methods: concealment. *BMJ.* 1999;319:375–6. [PubMed: 10435967]
39. Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. *J Crit Care.* 2005;20:187–91. discussion 91-3. [PubMed: 16139163]
40. Swingler GH, Zwarenstein M. An effectiveness trial of a diagnostic test in a busy outpatients department in a developing country: issues around allocation concealment and envelope randomization. *J Clin Epidemiol.* 2000;53:702–6. [PubMed: 10941947]
41. Altman DG, Schulz KF. Statistics notes: concealing treatment allocation in randomised trials. *BMJ.* 2001;323:446–7. [PubMed: 11520850]
42. Kennedy ADM, Torgerson DJ, Campbell MK, Grant AM. Subversion of allocation concealment in a randomised controlled trial: a historical case study. *Trials.* 2017;18:204. [PubMed: 28464922]
43. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: are we getting better? *BMJ.* 2016;355:i5663. [PubMed: 27856428]
44. Zhao W. Selection bias, allocation concealment and randomization design in clinical trials. *Contemp Clin Trials.* 2013;36:263–5. [PubMed: 23871796]
45. Broglio K. Randomization in clinical trials: permuted blocks and stratification. *JAMA.* 2018;319:2223–4. [PubMed: 29872845]
46. Altman DG. Avoiding bias in trials in which allocation ratio is varied. *J R Soc Med.* 2018;111:143–4. [PubMed: 29648508]
47. Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med.* 2018;16:1–15.
48. Sanft T, Usiskin I, Harrigan M, Cartmel B, Lu L, Li F-Y, et al. Randomized controlled trial of weight loss versus usual care on telomere length in women with breast cancer: the lifestyle, exercise, and nutrition (LEAN) study. *Breast Cancer Res Treat.* 2018;172:105–12. [PubMed: 30062572]

49. Demets DL, Lan KG. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13:1341–52. [PubMed: 7973215]
50. Dickinson SL, Golzarri-Arroyo L, Brown AW, McComb B, Kahathuduwa CN, Allison DB. Change in study randomization allocation needs to be included in statistical analysis: comment on ‘Randomized controlled trial of weight loss versus usual care on telomere length in women with breast cancer: the lifestyle, exercise, and nutrition (LEAN) study’. *Breast Cancer Res Treat*. 2019;175:263–4. [PubMed: 30721442]
51. Higgins KA, Mattes RD. A randomized controlled trial contrasting the effects of 4 low-calorie sweeteners and sucrose on body weight in adults with overweight or obesity. *Am J Clin Nutr*. 2019;109:1288–301. [PubMed: 30997499]
52. Elobeid MA, Padilla MA, McVie T, Thomas O, Brock DW, Musser B, et al. Missing data in randomized clinical trials for weight loss: scope of the problem, state of the field, and performance of statistical methods. *PLoS One*. 2009;4:e6624. [PubMed: 19675667]
53. Landers PS, Landers TL. Survival analysis of dropout patterns in dieting clinical trials. *J Am Diet Assoc*. 2004;104:1586–8. [PubMed: 15389418]
54. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials*. 2004;1:368–76. [PubMed: 16279275]
55. Biswal S, Jose VM. An overview of clinical trial operation: fundamentals of clinical trial planning and management in drug development. 2nd ed. 2018.
56. Lichtenstein AH, Jalbert SM, Adlercreutz H, Goldin BR, Rasmussen H, Schaefer EJ, et al. Lipoprotein response to diets high in soy or animal protein with and without isoflavones in moderately hypercholesterolemic subjects. *Arterioscler Thromb Vasc Biol*. 2002;22:1852–8. [PubMed: 12426215]
57. Shahrahmani H, Kariman N, Jannesari S, Rafieian-Kopaei M, Mirzaei M, Ghalandari S, et al. The effect of green tea ointment on episiotomy pain and wound healing in primiparous women: a randomized, double-blind, placebo-controlled clinical trial. *Phytother Res*. 2018;32:522–30. [PubMed: 29235159]
58. Draijer R, de Graaf Y, Slettenaar M, de Groot E, Wright C. Consumption of a polyphenol-rich grape-wine extract lowers ambulatory blood pressure in mildly hypertensive subjects. *Nutrients*. 2015;7:3138–53. [PubMed: 25942487]
59. de Clercq NC, van den Ende T, Prodan A, Hemke R, Davids M, Pedersen HK, et al. Fecal microbiota transplantation from overweight or obese donors in cachectic patients with advanced gastroesophageal cancer: a randomized, double-blind, placebo-controlled, phase II study. *Clin Cancer Res*. 2021;27:3784–92. [PubMed: 33883174]
60. Golzarri-Arroyo L, Dickinson SL, Allison DB. Replacement of dropouts may bias results: Comment on “The effect of green tea ointment on episiotomy pain and wound healing in primiparous women: A randomized, double-blind, placebo-controlled clinical trial”. *Phytother Res*. 2019;33:1955–6. [PubMed: 31192504]
61. Brown AW, Li P, Bohan Brown MM, Kaiser KA, Keith SW, Oakes JM, et al. Best (but oft-forgotten) practices: designing, analyzing, and reporting cluster randomized controlled trials. *Am J Clin Nutr*. 2015;102:241–8. [PubMed: 26016864]
62. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health*. 2004;94:416–22. [PubMed: 14998805]
63. Campbell M, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Stat Med*. 2007;26:2–19. [PubMed: 17136746]
64. Kahan BC, Morris TP. Assessing potential sources of clustering in individually randomised trials. *BMC Med Res Methodol*. 2013;13:1–9. [PubMed: 23297754]
65. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ*. 2012;345:e5840. [PubMed: 22983531]
66. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med*. 2012;31:328–40. [PubMed: 22139891]
67. Allison DR. When is it worth measuring a covariate in a randomized clinical trial? *J Consult Clin Psychol*. 1995;63:339. [PubMed: 7608345]

68. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. 2014;15:1–7. [PubMed: 24382030]
69. Vorland CJ, Brown AW, Dickinson SL, Gelman A, Allison DB. Comment on: Comprehensive nutritional and dietary intervention for autism spectrum disorder—a randomized, controlled 12-month trial, *Nutrients* 2018, 10, 369. *Nutrients*. 2019;11:1126.
70. Koretz RL. JPEN Journal Club 45. Cluster randomization. *JPEN J Parenter Enter Nutr*. 2019;43:941–3.
71. Brown AW, Altman DG, Baranowski T, Bland JM, Dawson JA, Dhurandhar NV, et al. Childhood obesity intervention studies: a narrative review and guide for investigators, authors, editors, reviewers, journalists, and readers to guard against exaggerated effectiveness claims. *Obes Rev*. 2019;20:1523–41. [PubMed: 31426126]
72. Golzarri-Arroyo L, Oakes JM, Brown AW, Allison DB. Incorrect analyses of cluster-randomized trials that do not take clustering and nesting into account likely lead to p-values that are too small. *Child Obes*. 2020;16:65–6. [PubMed: 32078372]
73. Li P, Brown AW, Oakes JM, Allison DB. Comment on “Intervention effects of a school-based health promotion programme on obesity related behavioural outcomes”. *J Obes*. 2015;2015:708181. [PubMed: 26101665]
74. Vorland CJ, Brown AW, Kahathuduwa CN, Dawson JA, Gletsu-Miller N, Kyle TK, et al. Questions on ‘Intervention effects of a kindergarten-based health promotion programme on obesity related behavioural outcomes and BMI percentiles’. *Prev Med Rep*. 2019;17:101022. [PubMed: 32154090]
75. Golzarri-Arroyo L, Vorland CJ, Thabane L, Oakes JM, Hunt ET, Brown AW, et al. Incorrect design and analysis render conclusion unsubstantiated: comment on “A digital movement in the world of inactive children: favourable outcomes of playing active video games in a pilot randomized trial”. *Eur J Pediatr*. 2020;179:1487–8. [PubMed: 32078037]
76. Golzarri-Arroyo L, Chen X, Dickinson SL, Short KR, Thompson DM, Allison DB. Corrected analysis of ‘Using financial incentives to promote physical activity in American Indian adolescents: a randomized controlled trial’ confirms conclusions. *PLoS One*. 2020;15:e0233273. [PubMed: 32492041]
77. Li P, Brown AW, Oakes JM, Allison DB. Comment on “School-based obesity prevention intervention in Chilean children: effective in controlling, but not reducing obesity”. *J Obes*. 2015;2015:183528. [PubMed: 26171245]
78. Wood AC, Brown AW, Li P, Oakes JM, Pavea G, Thomas DM, et al. A Comment on Scherr et al “A multicomponent, school-based intervention, the shaping healthy choices program, improves nutrition-related outcomes”. *J Nutr Educ Behav*. 2018;50:324–5. [PubMed: 29524984]
79. Mietus-Snyder M, Narayanan N, Krauss RM, Laine-Graves K, McCann JC, Shigenaga MK, et al. Randomized nutrient bar supplementation improves exercise-associated changes in plasma metabolome in adolescents and adult family members at cardiometabolic risk. *PLoS One*. 2020;15:e0240437. [PubMed: 33079935]
80. Heo M, Nair SR, Wylie-Rosett J, Faith MS, Pietrobelli A, Glassman NR, et al. Trial characteristics and appropriateness of statistical methods applied for design and analysis of randomized school-based studies addressing weight-related issues: a literature review. *J Obes*. 2018;2018:8767315. [PubMed: 30046468]
81. Meurer ST, Lopes ACS, Almeida FA, Mendonça RdD, Benedetti TRB. Effectiveness of the VAMOS strategy for increasing physical activity and healthy dietary habits: a randomized controlled community trial. *Health Educ Behav*. 2019;46:406–16. [PubMed: 30636448]
82. Ng YT, Phang SCW, Tan GCJ, Ng EY, Botross Henien NP, Palanisamy UDM, et al. The effects of tocotrienol-rich vitamin E (Tocovid) on diabetic neuropathy: a phase II randomized controlled trial. *Nutrients*. 2020;12:1522
83. Lazic SE, Clarke-Williams CJ, Munafò MR. What exactly is ‘N’ in cell culture and animal experiments? *PLoS Biol*. 2018;16:e2005282. [PubMed: 29617358]

84. George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J, et al. Common scientific and statistical errors in obesity research. *Obesity (Silver Spring)*. 2016;24:781–90. [PubMed: 27028280]
85. Bland JM, Altman DG. Comparisons against baseline within randomised groups are often used and can be highly misleading. *Trials*. 2011;12:264. [PubMed: 22192231]
86. Vorland CJ, Kyle TK, Brown AW. Comparisons of within-group instead of between-group affect the conclusions. Comment on: “Changes in weight and substrate oxidation in overweight adults following isomaltulose intake during a 12-week weight loss intervention: a randomized, double-blind, controlled trial”. *Nutrients* 2019, 11 (10), 2367. *Nutrients*. 2020;12:2335.
87. Bland JM, Altman DG. Best (but oft forgotten) practices: testing for treatment effects in randomized trials by separate analyses of changes from baseline in each group is a misleading approach. *Am J Clin Nutr*. 2015;102:991–4. [PubMed: 26354536]
88. Kroeger CM, Brown AW, Allison DB. Differences in Nominal Significance (DINS) Error leads to invalid conclusions: Letter regarding, “Diet enriched with fresh coconut decreases blood glucose levels and body weight in normal adults”. *J Complement Integr Med*. 2019;16:2.
89. Koretz RL. JPEN Journal Club 40. Differences in nominal significance. *JPEN J Parenter Enter Nutr*. 2019;43:311.
90. Dickinson SL, Brown AW, Mehta T, Heymsfield SB, Ebbeling CB, Ludwig DS, et al. Incorrect analyses were used in “Different enteral nutrition formulas have no effect on glucose homeostasis but on diet-induced thermogenesis in critically ill medical patients: a randomized controlled trial” and corrected analyses are requested. *Eur J Clin Nutr*. 2019;73:152–3. [PubMed: 29946116]
91. Brown AW, Allison DB. Letter to the Editor and Response Letter to the Editor and Author Response of assessment of a health promotion model on obese turkish children. *The Journal of Nursing Research*, 25, 436-446. *J Nurs Res*. 2018;26:373–4. [PubMed: 30192339]
92. Kaiser KA, George BJ, Allison DB. Re: Errors in Zhao et al (2015), Impact of enteral nutrition on energy metabolism in patients with Crohn’s disease. *World J Gastroenterol*. 2016;22:2867. [PubMed: 26973426]
93. Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: a tragedy of errors. *Nature*. 2016;530:27–9. [PubMed: 26842041]
94. Dawson JA, Brown AW, Allison DB. The stated conclusions are contradicted by the data, based on inappropriate statistics, and should be corrected: Comment on “Intervention for childhood obesity based on parents only or parents and child compared with follow-up alone”. *Pediatr Obes*. 2018;13:656. [PubMed: 30092611]
95. Allison D. The conclusions are unsupported by the data, are based on invalid analyses, are incorrect, and should be corrected: Letter regarding “Sleep quality and body composition variations in obese male adults after 14 weeks of yoga intervention: a randomized controlled trial”. *Int J Yoga*. 2018;11:83–4. [PubMed: 29343936]
96. Dimova RB, Allison DB. Inappropriate statistical method in a parallel-group randomized controlled trial results in unsubstantiated conclusions. *Nutr J*. 2015;15:58.
97. Dickinson SL, Foote G, Allison DB. Commentary: studying a possible placebo effect of an imaginary low-calorie diet. *Front Psychiatry*. 2020;11:329. [PubMed: 32411028]
98. Vorland CJ, Mestre LM, Mendis SS, Brown AW. Within-group comparisons led to unsubstantiated conclusions in “Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention study”. *Eur J Nutr*. 2020;59:2813–4. [PubMed: 32661682]
99. Peos J, Brown AW, Vorland CJ, Allison DB, Sainsbury A. Contrary to the conclusions stated in the paper, only dry fat-free mass was different between groups upon reanalysis. Comment on: “Intermittent energy restriction attenuates the loss of fat-free mass in resistance trained individuals. a randomized controlled trial”. *J Funct Morphol Kinesiol*. 2020;5:85.
100. Eckert I. Letter to the editor: Inadequate statistical inferences in the randomized clinical trial by Canheta et al. *Clin Nutr*. 2021;40:338. [PubMed: 33243460]
101. Vorland CJ, Foote G, Dickinson SL, Mayo-Wilson E, Allison DB, Brown AW. Letter to the Editor *Medicine Correspondence*. Blog2020. <https://journals.lww.com/md-journal/Blog/MedicineCorrespondenceBlog/pages/post.aspx?PostID=126>.

102. Sainani K. Misleading comparisons: the fallacy of comparing statistical significance. *PM R*. 2010;2:559–62. [PubMed: 20630442]
103. Allison DB, Antoine LH, George BJ. Incorrect statistical method in parallel-groups RCT led to unsubstantiated conclusions. *Lipids Health Dis*. 2016;15:1–5. [PubMed: 26728949]
104. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med*. 2015;12:e1001855. [PubMed: 26196287]
105. Fisher D, Copas A, Tierney J, Parmar M. A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *J Clin Epidemiol*. 2011;64:949–67. [PubMed: 21411280]
106. Jayawardene WP, Brown AW, Dawson JA, Kahathuduwa CN, McComb B, Allison DB. Conditioning on “study” is essential for valid inference when combining individual data from multiple randomized controlled trials: a comment on Reesor et al’s School-based weight management program curbs summer weight gain among low-income Hispanic middle school students. *J Sch Health*. 2019;89(1):59–67. *J Sch Health*. 2019;89:515–8. [PubMed: 31049962]
107. Li P, Stuart EA. Best (but oft-forgotten) practices: missing data methods in randomized controlled nutrition trials. *Am J Clin Nutr*. 2019;109:504–8. [PubMed: 30793174]
108. Lachin JM. Statistical considerations in the intent-to-treat principle. *Control Clin Trials*. 2000;21:167–89. [PubMed: 10822117]
109. Powney M, Williamson P, Kirkham J, Kolamunnage-Dona R. A review of the handling of missing longitudinal outcome data in clinical trials. *Trials*. 2014;15:237. [PubMed: 24947664]
110. Hoppe M, Ross AB, Svelander C, Sandberg AS, Hulthen L. Reply to the comments by Vorland et al. on our paper: “low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention study”. *Eur J Nutr*. 2020;59:2815–7. [PubMed: 32648021]
111. Khodabakhshi A, Akbari ME, Mirzaei HR, Seyfried TN, Kalamian M, Davoodi SH. Effects of Ketogenic metabolic therapy on patients with breast cancer: a randomized controlled clinical trial. *Clin Nutr*. 2021;40:751–8. [PubMed: 32703721]
112. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999;319:670–4. [PubMed: 10480822]
113. Morris TP, Kahan BC, White IR. Choosing sensitivity analyses for randomised trials: principles. *BMC Med Res Methodol*. 2014;14:1–5. [PubMed: 24383436]
114. ICH Expert Working Group. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials; E9(R1) 2019. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.
115. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2:109. [PubMed: 21897887]
116. Lichtenstein AH, Petersen K, Barger K, Hansen KE, Anderson CA, Baer DJ, et al. Perspective: design and conduct of human nutrition randomized controlled trials. *Adv Nutr*. 2021;12:4–20. [PubMed: 33200182]
117. Gadbury G, Coffey C, Allison D. Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obes Rev*. 2003;4:175–84. [PubMed: 12916818]
118. Veberke G, Molenberghs G, Bijnens L, Shaw D. *Linear mixed models in practice*. New York: Springer; 1997.
119. Linero AR, Daniels MJ. Bayesian approaches for missing not at random outcome data: the role of identifying restrictions. *Statist Sci*. 2018;33:198.
120. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18. [PubMed: 20334633]
121. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *J Cereb Blood Flow Metab*. 2020;40:1769–77. [PubMed: 32663096]
122. Enhancing the QUALity and Transparency Of health Research. <https://www.equator-network.org/>.

123. Altman DG, Simera I. Responsible reporting of health research studies: transparent, complete, accurate and timely. *J Antimicrob Chemother.* 2010;65:1–3. [PubMed: 19900949]
124. Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I, et al. Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ.* 2017;357:j2490. [PubMed: 28596181]
125. Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, et al. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One.* 2009;4:e7824. [PubMed: 19956596]
126. Kahathuduwa CN, Allison DB. Letter to the editor: Insufficient reporting of randomization procedures and unexplained unequal allocation: a commentary on “Dairy-based and energy-enriched berry-based snacks improve or maintain nutritional and functional status in older people in home care”. *J Nutr Health Aging.* 2019;23:396. [PubMed: 30932141]
127. Nykänen I. Insufficient reporting of randomization procedures and unexplained unequal allocation: a commentary on “Dairy-based and energy-enriched berry-based snacks improve or maintain nutritional and functional status in older people in home care”. *J Nutr Health Aging.* 2019;23:397. [PubMed: 30932142]
128. Vorland CJ, Brown AW, Dickinson SL, Gelman A, Allison DB. The implementation of randomization requires corrected analyses. Comment on “Comprehensive nutritional and dietary intervention for autism spectrum disorder—a randomized, controlled 12-month trial, *Nutrients* 2018, 10, 369”. *Nutrients.* 2019;11:1126.
129. Tekwe CD, Allison DB. Randomization by cluster, but analysis by individual without accommodating clustering in the analysis is incorrect: comment. *Ann Behav Med.* 2020;54:139. [PubMed: 31880299]
130. Morgan PJ, Young MD, Barnes AT, Eather N, Pollock ER, Lubans DR. Correction that the analyses were adjusted for clustering: a response to Tekwe et al. *Ann Behav Med.* 2020;54:140. [PubMed: 31880300]
131. Barnard ND, Levin SM, Gloede L, Flores R. Turning the waiting room into a classroom: weekly classes using a vegan or a portion-controlled eating plan improve diabetes control in a randomized translational study. *J Acad Nutr Diet.* 2018;118:1072–9. [PubMed: 29398571]
132. Erratum. *J Acad Nutr Diet.* 2019;119:1391–3. [PubMed: 31353013]
133. Douglas SM, Byers AW, Leidy HJ. Habitual breakfast patterns do not influence appetite and satiety responses in normal vs. high-protein breakfasts in overweight adolescent girls. *Nutrients.* 2019;11:1223.
134. Dalenberg JR, Patel BP, Denis R, Veldhuizen MG, Nakamura Y, Vinke PC, et al. Short-term consumption of sucralose with, but not without, carbohydrate impairs neural and metabolic sensitivity to sugar in humans. *Cell Metab.* 2020;31:493–502 e7. [PubMed: 32130881]
135. Quin C, Erland BM, Loeppky JL, Gibson DL. Omega-3 polyunsaturated fatty acid supplementation during the pre and post-natal period: a meta-analysis and systematic review of randomized and semi-randomized controlled trials. *J Nutr Intermed Metab.* 2016;5:34–54.
136. Folkvord F, Anshütz D, Geurts M. Watching TV cooking programs: effects on actual food intake among children. *J Nutr Educ Behav.* 2020;52:3–9. [PubMed: 31706794]
137. Zwilling CE, Strang A, Anderson E, Jurcsisn J, Johnson E, Das T, et al. Enhanced physical and cognitive performance in active duty Airmen: evidence from a randomized multimodal physical fitness and nutritional intervention. *Sci Rep.* 2020;10:1–13. [PubMed: 31913322]
138. Altman DG. Comparability of randomised groups. *J R Stat Soc Series D.* 1985;34:125–36.
139. Senn S. Testing for baseline balance in clinical trials. *Stat Med.* 1994;13:1715–26. [PubMed: 7997705]
140. Kraus WE, Bhapkar M, Huffman KM, Pieper CF, Das SK, Redman LM, et al. 2 years of calorie restriction and cardiometabolic risk (CALERIE): exploratory outcomes of a multicentre, phase 2, randomised controlled trial. *Lancet Diabetes Endocrinol.* 2019;7:673–83. [PubMed: 31303390]
141. O’Connor A. Cutting 300 calories a day shows health benefits. 2019. <https://www.nytimes.com/2019/07/16/well/eat/cutting-300-calories-a-day-shows-health-benefits.html>.
142. Hoppe M, Ross AB, Svelander C, Sandberg AS, Hulthen L. Correction to: Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve

- iron status of healthy Swedish females: a 12 week, randomized, parallel-design intervention study. *Eur J Nutr.* 2020;59:2819–20. [PubMed: 32648019]
143. Hoppe M, Ross AB, Svelander C, Sandberg A-S, Hulthén L. Low-phytate wholegrain bread instead of high-phytate wholegrain bread in a total diet context did not improve iron status of healthy Swedish females: a 12-week, randomized, parallel-design intervention study. *Eur J Nutr.* 2019;58:853–64. [PubMed: 29796932]
 144. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408–12. [PubMed: 7823387]
 145. Hewitt C, Hahn S, Torgerson DJ, Watson J, Bland JM. Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. *BMJ.* 2005;330:1057–8. [PubMed: 15760970]
 146. Savovi J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med.* 2012;157:429–38. [PubMed: 22945832]
 147. Page MJ, Higgins JP, Clayton G, Sterne JA, Hróbjartsson A, Savovi J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. *PLoS One.* 2016;11:e0159267. [PubMed: 27398997]
 148. Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, et al. The need for randomization in animal trials: an overview of systematic reviews. *PLoS One.* 2014;9:e98856. [PubMed: 24906117]
 149. Peinemann F, Tushabe DA, Kleijnen J. Using multiple types of studies in systematic reviews of health care interventions—a systematic review. *PLoS One.* 2013;8:e85035. [PubMed: 24416098]
 150. National Academies of Sciences, Engineering, and Medicine. Reproducibility and replicability in science. Washington, DC: The National Academies Press; 2019. p. 218.
 151. Vorland CJ, Brown AW, Ejima K, Mayo-Wilson E, Valdez D, Allison DB. Toward fulfilling the aspirational goal of science as self-correcting: a call for editorial courage and diligence for error correction. *Eur J Clin Invest.* 2020;50:e13190. [PubMed: 31788792]
 152. Brown AW, Kaiser KA, Allison DB. Issues with data and analyses: errors, underlying themes, and potential solutions. *Proc Natl Acad Sci USA.* 2018;115:2563–70. [PubMed: 29531079]
 153. Retraction Statement. LA sprouts randomized controlled nutrition, cooking and gardening program reduces obesity and metabolic risk in Latino youth. *Obesity (Silver Spring).* 2015;23:2522. [PubMed: 26524103]
 154. The effect of vitamin D supplementation on inflammatory and hemostatic markers and disease activity in patients with systemic lupus erythematosus: a randomized placebo-controlled trial. *J Rheumatol.* 2018;45:1713.
 155. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, et al. Retraction and republication: primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med.* 2013;368:1279–90. [PubMed: 23432189]
 156. Kroeger CM, Brown AW, Allison DB. Unsubstantiated conclusions in randomized controlled trial of binge-eating program due to Differences in Nominal Significance (DINS) Error. <https://pubpeer.com/publications/3596ABE0460E074A8FA5063606FFAB>.
 157. Zhang J, Wei Y, Allison DB. Comment on: "Chronic exposure to air pollution particles increases the risk of obesity and metabolic syndrome: findings from a natural experiment in Beijing". <https://hypothes.is/a/AQKsEg11EeiymitN4n0bQQ>.
 158. Hannon BA, Oakes JM, Allison DB. Alternating assignment was incorrectly labeled as randomization. *J Alzheimers Dis.* 2019;62:1767–75.
 159. Ito N, Saito H, Seki S, Ueda F, Asada T. Effects of composite supplement containing astaxanthin and sesamin on cognitive functions in people with mild cognitive impairment: a randomized, double-blind, placebo-controlled trial. *J Alzheimers Dis.* 2018;62:1767–75. [PubMed: 29614679]
 160. Rae P, Robb P. Megaloblastic anaemia of pregnancy: a clinical and laboratory study with particular reference to the total and labile serum folate levels. *J Clin Pathol.* 1970;23:379–91. [PubMed: 5476865]

161. Griffen WO Jr, Young VL, Stevenson CC. A prospective comparison of gastric and jejunoileal bypass procedures for morbid obesity. *Surg Obes Relat Dis.* 2005;1:163–72. [PubMed: 16925236]
162. Lang TA, Secic M. *How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers.* Philadelphia, PA: ACP Press; 2006.
163. Altman DG. Randomisation. *BMJ.* 1991;302:1481. [PubMed: 1855013]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Examples of why certain allocation methods are not random and how they may break concealment.

Table 1.

Example of nonrandom allocation by:	How it may break randomization	How it may break concealment
Allocation of participants first to only one treatment group until desired sample size, then randomization of the rest among treatment groups (e.g., [12])	Nonrandomized participants do not have a known probability of being in the other group(s)	The researcher knows the assignments of the participants enrolled without randomization
Alternating, such as allocating every other individual (e.g., [158, 159])	Participants may enroll in groups in nonrandom ways, and with small numbers of groups this can create imbalances	The researcher knows the next group assignment
Day of the week or time of day of enrollment [160]	Participants with certain characteristics may be more likely to be available for enrollment based on the day of the week or time of day	The researcher knows the group assignment
Patient chart number [161]	Chart numbers may be associated with known or unknown patient characteristics	If chart numbers are not randomly assigned, the researcher may be able to predict the next assignment
Participant characteristics (such as in the Lanarkshire Milk study [5, 6], or matching [162, 163])	Characteristics may not be evenly distributed (i.e., confounding may occur)	The researcher may be able to predict assignments based on characteristics