Loughborough University

This item was submitted to Loughborough's Research Repository by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Present 2020: Text expanding on the checklist for proper reporting of evidence in sport and exercise nutrition trials

PLEASE CITE THE PUBLISHED VERSION

https://doi.org/10.1123/ijsnem.2019-0326

PUBLISHER

Human Kinetics

VERSION

VoR (Version of Record)

PUBLISHER STATEMENT

This is an Open Access Article. It is published by Human Kinetics under the Creative Commons Attribution 4.0 Unported Licence (CC BY). Full details of this licence are available at:
http://creativecommons.org/licenses/by/4.0/

LICENCE

CC BY 4.0

REPOSITORY RECORD

Betts, James A., Javier T. Gonzalez, Louise M. Burke, Graeme L. Close, Ina Garthe, Lewis James, Asker E. Jeukendrup, et al. 2020. “Present 2020: Text Expanding on the Checklist for Proper Reporting of Evidence in Sport and Exercise Nutrition Trials”. Loughborough University. https://hdl.handle.net/2134/11871762.v1.

International Journal of Sport Nutrition and Exercise Metabolism, 2020, 30, 2-13
https://doi.org/10.1123/ijsnem.2019-0326
© 2020 Human Kinetics, Inc.

Human Kinetics
SCHOLARLY REVIEW

PRESENT 2020: Text Expanding on the Checklist for Proper Reporting of Evidence in Sport and Exercise Nutrition Trials

James A. Betts and Javier T. Gonzalez
University of Bath

Louise M. Burke
Australian Institute of Sport

Graeme L. Close
Liverpool John Moores University

Ina Garthe
Norwegian Olympic and Paralympic Committee and Confederation of Sport

Lewis J. James and Asker E. Jeukendrup
Loughborough University

James P. Morton
Liverpool John Moores University

David C. Nieman
Appalachian State University

Peter Peeling
The University of Western Australia

Stuart M. Phillips
McMaster University

Trent Stellingwerff
Canadian Sport Institute-Pacific and University of Victoria

Luc J.C. van Loon
Maastricht University Medical Centre+

Clyde Williams
Loughborough University

Kathleen Woolf
NYU Steinhardt

Ron Maughan
St Andrews University

Greg Atkinson
Teesside University

Keywords: exercise performance, exercise physiology, sport physiology

The CONSORT (CONsolidated Standards Of Reporting Trials) 2010 guidelines (http://www.consort-statement.org/consort-2010) were developed to improve the reporting of parallel-group randomized controlled trials, whereby compliance with established standards can be demonstrated via completion of the CONSORT 2010 checklist (Schulz et al., 2010). Leading medical journals have endorsed this initiative, which has undoubtedly improved the conduct and reporting of clinical and health care research. Research in the field of sports nutrition and exercise metabolism stands to benefit from similar standards, but it commonly involves research designs other than parallel-group trials, such as cross-over experiments.

A CONSORT extension covering randomized cross-over trials has now been published (Dwan et al., 2019), with a revised checklist that focuses on issues of primary relevance to clinical trials involving medicine or health care outcomes. However, such issues specific to clinical trials may have different relevance when considered relative to the tightly controlled, laboratory-based, mechanistic experiments that are common in exercise science. For example, cross-over designs may involve order effects between assessments; in medical trials, this tends to occur most commonly due to the treatment or intervention itself, requiring a sufficient wash-out interval before repeated assessments. By contrast, the carry-over effect in sports nutrition research is commonly related to the assessment itself, which often tends to be more invasive or demanding for the participant than a snapshot of health status. Indeed, exercise tests of human

Betts and Gonzalez are with the Department for Health, University of Bath, Bath, United Kingdom. Burke is with the Australian Institute of Sport, Belconnen, ACT, Australia. Close and Morton are with the Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom. Garthe is with the Norwegian Olympic and Paralympic Committee and Confederation of Sport, Oslo, Norway. James, Jeukendrup, and Williams are with the School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, United Kingdom. Nieman is with the Human Performance Laboratory, Appalachian State University, North Carolina Research Campus, Kannapolis, NC, USA. Peeling is with the School of Human Sciences (Exercise and Sport Science), The University of Western Australia, Crawley, WA, Australia. Phillips is with the Department of Kinesiology, McMaster University, Hamilton, ON, Canada. Stellingwerff is with Canadian Sport Institute-Pacific, Victoria, BC, Canada; and the Department of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada. van Loon is with the Department of Human Biology, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands. Woolf is with the Department of Nutrition and Food Studies, NYU Steinhardt, New York, NY, USA. Maughan is with the School of Medicine, St Andrews University, St Andrews, United Kingdom. Atkinson is with the School of Health and Social Care, Teesside University, Middlesbrough, United Kingdom. Betts (J.Betts@bath.ac.uk) is corresponding author.

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

performance are particularly prone to learning or fatigue effects and even physical adaptations that can persist for weeks or months after the first test. For example, the so-called "repeated bout effect," which describes how a single exposure to unaccustomed physical exercise that induces muscle damage can impart profound and lasting protection from similar exercise in the future (Byrnes et al., 1985; McHugh et al., 1999). Participants in exercise trials may also be elite athletes whose habitual levels of physical activity (and diet) may show profound variation over time (i.e., periodization), thus further complicating the interpretation of longitudinal studies.

The PRESENT (Proper Reporting of Evidence in Sport and Exercise Nutrition Trials) 2020 checklist (see Appendix) has therefore been adapted from the CONSORT guidelines to specifically address the unique combination of challenges and opportunities facing researchers within the broad fields of sports nutrition and exercise metabolism. This current paper complements and expands upon the CONSORT checklist by providing emphasis and examples that are commonplace or of greatest relevance to research in this subject area. The PRESENT 2020 checklist was designed with consideration of the need to minimize the burden on submitting authors while ensuring that standards for reporting research are met; it should allow researchers to quickly determine whether all relevant information is included in their manuscript. Of course, it is possible to meet all the factors on the checklist despite either having conducted a poor study and/or having reported a good study poorly, whereas some items on the checklist may not be applicable even for rigorously conducted research. Nonetheless, consideration and discussion of the factors identified in the checklist should improve the reporting of exercise- and nutrition-related research in the immediate future and has the potential to enhance the design and conduct of trials in the long term.

The following sections expand on and justify the rationale for each of the items included in the associated submission checklist.

Title

1a The title should accurately reflect the primary findings of the study, preferably via an informative statement (e.g., "Caffeine improves 200 m swimming time in elite swimmers" rather than "The effects of caffeine ingestion on swimming"). Correlation does not infer causality, so causal language should be reserved for the title of experimental research only (e.g., terms such as *impaired*, *resulted*, or *improved*), whereas observational research should employ appropriate noncausal statements (e.g., terms such as *related*, *correlated*, or *associated*). Titles should specifically identify the measured variable(s) rather than proxy or indirect measures (e.g., "blood lactate production" or "blood glucose clearance" should not be stated when only blood concentrations of the respective metabolites have been assessed).

1b The title should identify the study population if characteristics are directly relevant to the study design (e.g., sex and/or training status). If using nonhuman models, the species should be stated.

Abstract

Some readers may not have access to the full paper, so a properly formatted and well-written abstract is imperative. Authors should give priority to information about the current study rather than using the abstract for an extensive background or rationale.

2a Methods: Key information regarding the study design, methods, and population should be summarized to enable broad understanding of the study from the abstract.

2b Results: Readers are interested in extracting key data that reflect the main findings of the study. The abstract should present data (e.g., the absolute magnitude of values and the size/precision of effects—specifying which measures of central tendency and variability are stated) rather than simply stating the presence, absence, or direction of effects. The presentation of p values or similar inferential statistics is no substitute for reporting actual data (Maughan, 2004).

2c Conclusion: Priority should be given to the reporting of results as per the previous section, with only a brief concluding statement thereafter. A concise conclusion based on what was actually measured in the study is preferred to speculative interpretations, with cautious use of language to avoid hyperbole or improper inference of causality (Brown et al., 2013). It is not appropriate or necessary to identify further research priorities here.

Introduction

3a This section need not provide a comprehensive review of the subject area, as relevant review articles can be referenced when briefly introducing underlying theory, processes, and/or mechanisms. Priority of publication should be recognized—recent publications that simply confirm earlier findings should not be cited. Therefore, the primary focus should be on what has been done already, particularly the most relevant previous studies, ensuring a fair balance of different perspectives wherever current evidence is equivocal. It should be clear to the reader *both* what new information the study aims to provide *and* why that information is important. The former most commonly means identifying novelty (i.e., what will this study show that has not been shown before) but could also mean justifying why replication of a past study is necessary. If the novel element of a study relies on a long list of qualifiers (e.g., the effect of A on B has already been established but not within population C and context D), then the "Introduction" should justify why those qualifiers are interesting and warrant further investigation (i.e., What is important about population C and/or context D?). Thus, the reader should have a clear understanding of the rationale for the work.

3b A formal hypothesis may not always be appropriate for work of an exploratory or qualitative nature, but the "Introduction" should conclude with some form of clear aim, objective, or research question (i.e., sufficient to express the central question and present the primary variables under investigation).

Methods

4 Ethics

The details regarding ethics approval should provide adequate reference to the approving body to enable verification of approval if needed. Authors are encouraged to report the ethics committee approval number (if available) and date of approval. Informed consent should be explicitly noted for human trials where relevant—and carefully justified for any trials where deception or testing of participants without prior consent was warranted (Harriss et al., 2017). In the case of participants below the legal age of

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

consent, their assent and the consent of their parent or guardian should be obtained. Pretrial registration of studies that qualify for recognition as "Clinical Trials" is strongly encouraged, and details of access to the Clinical Trials Registry should be provided. Where a paper reports data that have been collected during routine monitoring of individuals as a condition of their employment or participation (e.g., athletes or military populations), a clear statement about the source of data and the reasons for the absence of prior approval by an appropriate body, such as their requirement as a condition of employment, must be made explicitly (Winter & Maughan, 2009). In such situations, it may be appropriate to introduce some degree of separation into the process, such as a gatekeeper to maintain the anonymity and/or confidentiality of data.

5 Design

Full details of the research design should be summarized early in the "Methods" section to establish context for the reader, using common terminology and nomenclature (e.g., parallel trial/crossover, randomized, counterbalanced, blinding, observational, etc.) both to provide a top-line overview from the outset and to facilitate accurate data extraction for those conducting future systematic reviews and meta-analyses. All descriptive work, including surveys and case studies, should be identified as observational research. If the data presented in the paper are from a secondary analysis of a wider/previous study rather than from primary research, this should also be clearly identified in the introductory information about the study design, with commentary around whether the study was specifically or adequately designed for the purposes reported in the present paper and whether any data are replicated from an earlier publication.

6 Sampling

6a Recruitment methods, eligibility (inclusion/exclusion) criteria, and sampling methods should be explained, ideally providing a rationale for the target population. In particular, authors of case studies must justify the interest in, or choice of, their unique participant(s). If the study authors intentionally recruit an unrepresentative proportion of male and female participants, this should be justified.

6b To provide full transparency around the population to which the study findings may be generalized, a thorough account of participant characteristics is essential. All data should be reported with appropriate precision relative to the measurement tool and units in question (Kordi et al., 2011); for example, it is unnecessary to record or express age in decimal years or with greater precision than the integer (e.g., it is appropriate to report an age in years as 25 rather than 25.24). The choice of which demographic, anthropometric, or lifestyle variables to report should be considered in terms of relevance to trial interpretation and generalization within this field (e.g., training status, competitive level, habituation to diet). The term "elite" is commonly misused when describing athletes: Descriptive terms should be used appropriately and qualified by information that is objective (e.g., personal best for a relevant event, level of international/national representation, ranking, or point score where it has global meaning, current scores from a quality assured test battery, etc.).

6c The study setting should be described both in terms of laboratory- versus field-based testing and the region/site(s) of data collection. If data are collected across multiple independent laboratories, details of which measurements were made at each location should be provided, with relevant reliability data specific to each site. Start and end dates of testing are also required. Appropriate environmental data (e.g., temperature, relative humidity, altitude) should be stated where these are relevant to interpretation of the study.

6d Justification should be provided for the size of the sample that was recruited and tested. The pragmatic nature of labor/time-intensive, and often invasive, experiments in this field is such that sample size is sometimes dictated by the availability of human volunteers and/or resources. It is also the case, however, that it can be unethical to include more participants than are necessary to test the hypothesis. An estimation of the statistical power/precision should therefore be provided based on the minimal worthwhile effect and available sample size (Batterham & Atkinson, 2005), and all components of the calculations should be clearly justified and reported in sufficient detail to enable replication (i.e., alpha, beta, minimal worthwhile effect, variance estimate—e.g., the *SD* of pre-to-post change). Post hoc power estimations (based on the effect size observed in the study itself) should be avoided. It should also be clearly expressed whether the hypothesis is of a "superiority" nature, where one treatment/intervention is hypothesized to differ from the control or another intervention, or whether the study is more akin to an "equivalence trial" in which one treatment/intervention is deemed to be of similar effectiveness to another intervention—but perhaps simpler, less costly or less invasive in some way. In the latter situation, the null hypothesis testing procedure is not appropriate—$p > .05$ should never be used to infer that there is "no difference" between two or more sample estimates (Assel et al., 2019) nor should p values marginally greater than .05 be interpreted as a "trend."

Sample size estimations may be based on null hypothesis testing or the desired width of a confidence interval, but the choice should be justified. The "Discussion" section may provide authors with further opportunity to frame their statistical inferences alongside recent concerns about replication of study results. For example, the notion that an observed effect size is large merely due to the sampling error associated with a relatively small sample—an issue known as the "winner's curse"—may be assisted by the reporting of statistics such as the "Surprisal (S) value" (Greenland, 2019). Importantly, inadequate sample sizes are not necessarily justified merely because the population in question are rare or difficult to recruit (e.g., elite athletes) or that similar sample sizes have been used in previous publications—such instances may be better presented as pilot work or case studies, or alternatively the sample size could be supplemented by including data from a broader population (e.g., subelite athletes), unless there is reason to expect the primary outcome to respond distinctly in the more focused sample.

7 Interventions

The intervention/treatment is arguably the single-most important element of methods reporting; even seemingly trivial details of the independent variable may be critical to a proper understanding of the responses attributed to the groups/conditions. For nutritional compounds with complex ingredients, a side-by-side comparison (possibly tabulated) is often advisable for clear and comprehensive reporting. Critically, it is wholly inadequate to simply provide a manufacturer/product name for a scientific report because the actual composition of such commercially available goods may be proprietary, reformulated, discontinued, and/or different from the reported composition of each batch produced; the prospective

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

scientific value of the results depends entirely on the reader being able to establish what the intervention involved. For this reason, there should be careful consideration of whether the nutritional compound under investigation requires analysis to verify both that the reported ingredients are present in the quantities stated and that undisclosed ingredients or contaminants are absent (in some cases, it may be that certain physiological measurements can be used to verify the efficacy of the supplementation protocol and so may be capable of verifying the mere presence or absence of key ingredients between treatments). Simple products that are formulated in-house (e.g., glucose solutions prepared according to a clear protocol) may not require such verification, but transparent and credible investigation of commercially available products often requires such checks of the specific batch used in the research project. Lastly, where particular nutritional compounds or blends have been selected for examination, it can be informative to explain the rationale for decisions regarding the choice of intervention relative to outcomes (e.g., see *Approach to the Research Question* subsection in Bailey et al., 2011).

8 Measurements

8a The primary outcome measures intended to answer the stated research questions should be clearly specified. The eventual conclusions and interpretation should then be based on these prestated primary outcomes (especially in the case of null results), rather than focusing on outcomes according to which were most responsive or were consistent with the hypothesis. The categorization of variables that are measured on a continuous scale ("dichotomania") should be avoided unless there is a robust reason to do so (Senn & Julious, 2009).

8b A rationale for the selected measurement tools should be provided. Metrics pertaining to dietary analysis and/or exercise testing are frequently involved in sports nutrition research. Regarding the former, the method of dietary assessment should be precisely described (e.g., self-report vs. direct measure, prospective vs. retrospective collection, weighed vs. estimation from household measures quantification), and the limitations should be considered, along with full details of the name and version of any dietary software used. Regarding the latter, information of interest includes the nature of the test (e.g., exercise capacity/time-to-exhaustion vs. exercise performance/time trial), the familiarity/familiarization of participants with the testing protocol and whether intensity is absolute or relative to another parameter (e.g., $\%\dot{V}O_2max$). Protocols that aim to measure sports performance should be valid, reliable, and sensitive (Currell & Jeukendrup, 2008). In all cases, the degree of measurement error associated with each outcome should be expressed using relevant reliability statistics (Atkinson & Nevill, 1998). For biochemical assays, reliability data should ideally be derived from in-house analysis rather than using the intra- or inter-assay reliability reported by the manufacturer. Researchers should note whether clinical chemistry analyzers intended for diagnostic use provide the precision necessary for research purposes.

8c Authors should clearly identify and justify the smallest difference between treatments that is deemed to be meaningful, which should be consistent with the sample size estimates presented under Item 6d and be taken into account when interpreting data (i.e., an effect smaller than that deemed meaningful should not then be overinterpreted, even if it happens to be statistically significant). It is important to note the distinction between the smallest worthwhile effect or association and the smallest detectable effect (de Vet et al., 2006). The former value, often termed the "target difference" or "minimal clinically important difference" (MCID), is the value of change or association that is deemed to be important for participants, athletes, or patients. Ideally this threshold would be based on known relationships between change in the outcome of interest and change in real athletic performance, physiological adaptation, or morbidity/mortality in the context of clinical trials. Nevertheless, such knowledge can be difficult to derive. We refer readers to the DELTA and DELTA2 publications for a full treatment of this topic, including all the approaches for arriving at a MCID (Cook et al., 2014, 2018).

The smallest detectable change is a threshold based on the likelihood of a change in an individual athlete or patient being due to random within-subjects (test–retest) error or not, with a certain coverage probability. Although this may be useful for monitoring individuals, it is important to note that the smallest detectable change and the MCID might be different in magnitude. For research purposes, the MCID and the approaches for deriving this value are most important. Similarly, with regard to correlations, interpretation should always consider that the strength, direction, and form of a relationship; for example, a weak, inverse, linear relationship (e.g., $r = -.1$) may be of limited meaning or utility irrespective of whether it is statistically significant.

9 Randomization

Randomization in a trial can be a critical component of experimental validity, yet it is one where important details are often unreported. Full details should be provided to identify precisely how the random allocation sequence was generated and implemented and by whom (Schulz & Grimes, 2002a, 2002c, 2002e). For example, such details could include the type of random allocation sequence, the allocation ratio, and details of any minimization or restriction (e.g., stratification, blocking, and block size), with clear identification of the individuals responsible for generating/managing the allocation sequence, recruiting/enrolling volunteers, assigning treatments, conducting data collection, and analyzing samples.

10 Blinding

Although experiments are commonly designed and instigated on the basis of a working hypothesis regarding the anticipated answer to a prestated research question, it has been recognized for over 150 years that measurements should be made without any preconceived ideas or a priori assumptions regarding the expected results (Bernard, 1865). This reasoning forms the basis for the modern concept of experimental blinding, which aims to control the conscious or subconscious biases of experimenters, participants, and/or outcome assessors by keeping them unaware of the expected response to a given stimulus (Schulz & Grimes, 2002b). Blinding is a technique that allows researchers to conduct the study without preconceived expectations of the results, despite the existence of a working hypothesis. This ideal can be achieved by concealing the treatment allocation and/or the purpose of the experiment from anyone involved who can conceivably influence the measurements being made.

Although blinding is important at various levels, when outcome measures are subject to conscious or subconscious control, the blinding of participants becomes a critical issue. Examples of such measures in the field of sport nutrition and exercise metabolism

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

include metrics around physical performance (e.g., maximal *voluntary* muscle contraction, time to fatigue in an exercise task) or subjective perceptions (e.g., rating of perceived exertion or soreness). These outcome measures might be compared with objective physiological variables that are generally less prone to expectancy/placebo effects (e.g., changes in systemic metabolite concentrations or substrate utilization during exercise). However, some interventions are difficult or essentially impossible to truly blind (e.g., cryotherapy or exercise) and in some cases the participants' awareness of the intervention is pivotal to the proposed mechanism of action (e.g., psychological interventions). In such situations where blinding of treatment allocation is impossible or undesirable, it may be justified to blind participants to the purpose of the experiment by assessing the primary outcome measure covertly within the context of the wider study. For example, participants may be offered a meal upon completion of testing, after which consent can be sought post hoc to determine natural ad libitum food intake based on any left-over/uneaten food. Examples of the various levels of blinding are provided in Table 1.

When performing a blinded study, it is especially important to assess the success of blinding. For example, a treatment can be implemented in a blinded manner whereby participants are not informed of treatment allocation, and every effort can be made to conceal treatment allocation using placebo/control conditions that are taste and texture-matched, using opaque containers, anonymized labels, sham treatments, or even novel methods of administration (e.g., nasogastric delivery of nutrients; Funnell et al., 2019). However, the trial is only truly blinded in the sense that conscious/subconscious bias is controlled if participants are unaware of their treatment allocation. Therefore, beyond the above-mentioned methods intended to conceal treatment allocation, the success of blinding can easily be assessed by way of a formal exit questionnaire. For example, for a cross-over design in which each participant has the opportunity to experience and compare conditions, this can be achieved via three straightforward binary "yes/no" responses: (a) did the participant (or experimenter/assessor) detect any difference between treatments; if yes, (b) did the they feel able to identify their treatment allocation; if yes, (c) did they correctly identify their treatment allocation? For interventions where participants may conceivably detect differences between treatments, it may also be informative to survey their prior knowledge, preconceptions, and beliefs about the expected effects of treatment.

Ultimately, research should only be reported as blinded if the researcher is confident that participants were not aware of treatment allocation; if it transpires that a proportion of the overall cohort were able to distinguish treatments, then it may be informative to report the extent to which observed responses may have been influenced by the success of blinding. Although blinding can reduce bias in experiments, it does not overcome a lack of randomization, which is a common misconception (Schulz & Grimes, 2002b). In addition, if complete blinding to a treatment is difficult or impossible and exercise performance is the primary outcome, then exercise testing should follow recommendations to withhold relevant performance feedback during the exercise tests (Currell & Jeukendrup, 2008).

11 Standardization

Standardization of participants' behavior and environment ahead of testing and during data capture can reduce the variability in baseline measurements both between and within individuals. This can increase the ability to detect true effects with relatively small sample sizes, effectively increasing the signal-to-noise ratio. Key considerations include the length of time of pretest controls and whether standardization is performed across different participants (i.e., controlling interindividual variation) or only within-participants (i.e., controlling intraindividual variation), which largely depends on whether the research involves a parallel or cross-over design. Common parameters around participant behavior/characteristics to control within the field of sport nutrition and exercise metabolism include physical activity, diet/hydration, medication/supplement use, and menstrual cycle. It should be clearly stated whether this control was directly facilitated by the investigators (e.g., providing meals) or, if not, whether any objective assessments were employed to verify successful replication (e.g., heart rate monitoring to confirm abstinence from intense physical exertion prior to testing). If such monitoring has been completed, it is often informative to report the data to provide future context when contrasting the

Table 1 Summary of Various Forms of Experimental Blinding

Type of blinding	Description	Rationale
Open label	All categories of individuals (experimenters, participants, and outcome assessors) are aware of who has received which intervention throughout the trial.	Some interventions are impossible to truly blind (e.g., exercise) or participants' awareness of the intervention is inherent to the research question.
Single blind	One category of individuals (normally the participants, but potentially the experimenters or outcome assessors) is unaware of who has received which intervention throughout the trial.	When participants are blinded: • Less likely to have biased psychological or physical responses to the intervention. • Less likely to seek additional adjunct interventions. • More likely to comply with intervention. • Less likely to leave the trial without providing data.
Double blind	Experimenters, participants, and outcome assessors are all unaware of the intervention assignments during the trial.	When investigators/assessors are blinded: • Less likely to transfer inclinations or attitudes to participants. • Less likely to have biases affect outcome assessments, especially with subjective outcomes.
Triple blind	Experimenters, participants, and outcome assessors are all unaware of the intervention assignments during the trial and during data analysis.	Less likely to have biases affect statistical analyses and interpretation.

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

results with other trials. For example, even if diet was closely matched between conditions, readers may benefit from knowing whether nutritional interventions were contrasted against a background diet that is relatively high or low in energy or other nutrients. A more detailed summary of methods to implement and report dietary standardization techniques is provided elsewhere (Jeacocke & Burke, 2010).

Variance in the presence and stages of menstrual cycle should be considered where relevant: There are numerous approaches to control for this factor, each potentially justified based on a balance of internal and external validity. A rationale should be provided to explain whether menstrual phase was controlled for within or between participants and which phase of the cycle was selected to minimize confounding influences and/or to be most representative. If a specific phase of the menstrual cycle is stated, it should be made clear whether this was based on self-reported dates or measurements of sex hormones.

For further information about the benefits and techniques for the pre- and during-trial standardization of factors such as recent exercise, acclimation, noise/distractions, encouragement, and the awareness of the passage of time or previous results, the reader is referred to other reviews (Burke & Peeling, 2018; Currell & Jeukendrup, 2008).

12 Order Effects

The potential for order effects between serial trials of the same participant should be assessed and reported, particularly during studies using a repeated-measures crossover design. Critically, if an initial assessment under one condition exerts lasting effects or adaptations that exceed the wash-out interval before the repeat assessment under the other condition, then randomization and counterbalancing of treatment order do nothing to prevent the problem of order/period/sequence effects. This scenario is particularly problematic if there is any interaction between the effect of the intervention and the magnitude of the carry-over effects between measurements; indeed, if order effects are apparent, then no antidote can rectify the problem and isolate the effects of the intervention. Examples of preemptive measures to minimize the likelihood of order effects include the following: allowing adequate washout between assessments; providing adequate prior familiarization with testing; and/or sampling a population already accustomed to the protocol and thus unlikely to exhibit marked learning or fatigue effects between exposures (e.g., using trained athletes for exercise tests). However, more extended wash-out periods between assessments (i.e., weeks to months) could induce training/detraining effects and/or introduce seasonal variance as confounders.

Experiments where order effects are apparent are not necessarily unfit for publication, but the order effect must be clearly referenced when reporting outcomes. Exploring data for order effects may take the form of applying similar statistical techniques as used for the primary analysis between conditions but with reference to the sequential assessments (Wellek & Blettner, 2012), with due consideration that effects of the intervention and treatment order may interact (e.g., the placebo may be consistently inferior only when applied in the first trial). However, given the complexities of such statistical analyses and that the absence of detectable differences between sequential tests does not necessarily infer that they were equivalent, it may be worthwhile inspecting individual data for evidence of any systematic changes according to trial order.

13 Statistics

13a The comparisons that will inform primary inferences about the study data should be identified, with statistical approaches described in adequate detail. In particular, there should be consistency between the proposed approach, the rationale presented in the Introduction, and the data reported. For example, if the study has been designed to contrast responses over time to an intervention versus a placebo, the analysis should target the differences between those conditions rather than the relative presence or absence of changes from baseline within each condition (Bland & Altman, 2015).

13b Any unplanned or exploratory analyses should be clearly distinguished from the primary purpose and analysis, with justification provided for any interim analysis or potential conditions under which a trial might stop, including consideration of intention-to-treat versus per-protocol analyses. In particular, be cautious of subgroup analysis with stratification at baseline because regression to the mean can compromise inferences (Thomas et al., 2019). Similarly, stratifying on the study outcome for responder/nonresponder analysis is fraught with difficulties and pitfalls (Atkinson & Batterham, 2015). These and other common statistical pitfalls are listed along with relevant examples, further explanation, and possible solutions in Table 2.

13c Researchers should be aware of the assumptions underlying each selected statistical approach and report clearly how any departures from these assumptions were handled. The most important data to check for approximate "normality" are typically the residuals of the selected statistical model. In the context of a paired *t* test, these data are the individually paired change scores or differences. Furthermore, parametric analyses may often be robust, even when raw data are nonnormally distributed. Any data transformation should be clearly justified (Bland & Altman, 1996). Blanket logarithmic transformation of data without checking first whether the data require or benefit from this transformation is not good practice. Researchers should also be aware that statistical tests of skewness (where the null hypothesis is that data are "normally" distributed) are prone to problems when sample sizes are small. Similar to the time-series design described under item 13a, in which there are repeated measures over serial time points within each condition, it may also be necessary to adjust for baseline differences in the primary outcome measure (i.e., analysis of covariance; Vickers, 2005). In this respect, authors should interpret any differences in pre-to-post change carefully when there are meaningful differences between groups or conditions in the study outcome(s) at baseline—such a problem may not be rectified by calculating percentage changes from baseline or by interpreting a Group/Condition × Time interaction term (Vickers, 2001).

Results

14 Participant Flow

14a Depending on the complexity and nature of the experimental design (e.g., multiple follow-ups with poor adherence, many withdrawals and exclusions), it may be appropriate to present a flowchart of the numbers of participants who were recruited, eligible, enrolled, randomized, assigned, received treatment, and analyzed at each time point. For more straightforward designs, it may be possible to simply communicate this information within text, but it remains essential to clearly identify the number of participants (denominator) included in each analysis (i.e., specific to each time point and outcome measure as required). In particular,

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

Table 2 Common Statistical Pitfalls in Sport Nutrition and Exercise Metabolism Trials

Pitfall	Example	Explanation	Solutions
Pooling data from multiple groups or conditions and analyzing as a single combined sample. See Bland and Altman (1995a; 1995b)	Pooling data from 10 participants across three diet conditions and examining the correlation between two measured variables, as if $n = 30$.	This is "pseudoreplication," whereby sample size is erroneously inflated and there is obfuscation between within- and between-subjects inferences.	Any grouping factor should be present in the statistical model. Replicates may be averaged within each participant to quantify between-subjects correlations while retaining the correct degrees of freedom. Similarly, all "clusters" (e.g., laboratory site) should be included in the statistical model.
Comparing unadjusted (raw) changes/responses between different trial arms or subgroups. See Vickers (2001, 2005)	Comparing pre–post changes between two different age groups or men/women because these groups may differ at baseline.	Regression to the mean may influence inferences when subgroups are formed on the basis of baseline status or other factors that may lead to baseline imbalance for the study measured variable.	An analysis of covariance model is appropriate whereby the baseline values for each group are entered into the statistical model as a covariate. Analysis of covariance has been reported to be superior to comparing percentage changes between groups or calculating a Group × Time interaction term.
Forming groups or "conditioning" on the basis of the outcome variable and quantifying group differences in baseline values. See Atkinson and Batterham (2015), Atkinson et al. (2019)	Comparing baseline values between groups of "responders" and "nonresponders."	This creates a similar problem of regression to the mean as in the case above, rendering the responders as lower at baseline (and vice versa) due to this statistical artifact.	There are many pitfalls in "responder counting" and analysis of factors that might influence individual response. Conditioning on the magnitude of response is rarely appropriate.
Correlating proposed baseline status or other predictors with treatment response only in the intervention group. See Hecksteden et al. (2015), Lonergan et al. (2017)	Correlating baseline $\dot{V}O_2$max with the baseline to follow-up change in $\dot{V}O_2$max in response to training.	Regression to the mean could again influence inferences here.	Any predictor of response should be modeled as a Predictor × Trial arm interaction with covariate adjustment for baseline values of the outcome (because the predictor and outcome may be already correlated at baseline).
Correlating variables that are already mathematically related. See Bland and Altman (1994), Tu and Gilthorpe (2007)	Correlating total triathlon time with cycling phase time, or correlating oxygen uptake ($ml \cdot kg^{-1} \cdot min^{-1}$) with body mass index ($kg \cdot m^{-2}$).	This is "mathematical coupling" or "relating a part to a whole," whereby one variable is a component in the calculation of another variable already. Such correlations can be spurious.	This practice should be avoided. Sometimes partial correlations and multivariable-adjusted associations can help "adjust" the correlation between two variables when there is a "third variable" deemed to influence both variables, but this third variable should not be inherent in the maths of the other two variables.
Reporting individual differences in response only in the intervention/treatment group data. See Atkinson and Batterham (2015), Atkinson et al. (2019)	Plotting the response of $\dot{V}O_2$max to training for each individual in the intervention group, or reporting the *SD* of response for intervention group only.	This practice does not acknowledge that similar individual differences in change can occur in the control group in the context of the "counterfactual" in randomized controlled trials.	Control group change variance should be included in the analysis approach for quantifying any hypothesized individual response heterogeneity.
Arriving at conclusions solely on the basis of statistically significant (or not significant) *p* values. See Greenland et al. (2016)	Reporting that an improvement in performance is statistically significant but by a margin that is not worthwhile for athletes, or vice versa.	Here, clinical or practical relevance of the magnitude of effects is not being considered appropriately.	A minimum important effect should be rationalized a priori and compared with the observed effect size. 95% confidence intervals should be reported routinely. A nonsignificant *p* value should never be used to conclude "no effect" or "no difference," as the null hypothesis testing process is not designed to make inferences about equivalence.

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

clear reasons should be provided for any exclusions from the final data set, with changes from original treatment assignment acknowledged and justified (Schulz & Grimes, 2002d).

14b Researchers should avoid pooling data for participants over multiple experimental conditions or groups and analyzing the data without modeling these conditions and groups, a problem known as "pseudoreplication" (Lazic, 2010). Examples of such transgressions include the pooling of data for 10 participants across three conditions and then calculating correlation coefficients between measured variables as $n = 30$ rather than $n = 10$. This approach breaks the assumption of independence of cases, inflates degrees of freedom, and can be misleading: Indeed, within-subjects correlations between two variables measured over time may be different from between-subjects correlations between two variables measured at the same point in time.

15 Outcomes

15a Data should be reported according to the International System of Units (i.e., SI units) or accepted derivatives. An exception, as commonly applied in nutrition science, are units for measuring energy intake or expenditure, where either kilocalorie or kilojoule may be used. The measures of central tendency and variability used to describe each data set should be clear, as well as and how effects and estimated precision are expressed (e.g., 95% confidence interval). If p values are to be reported, then it should be recognized that those close to but higher than .05 do not indicate a "trend" or an approach to significance, whereas a value equal to 0 is not possible (Assel et al., 2019). Units of measurement should not be used in place of variables when describing data: for example, data should be referred to as energy intake rather than kilojoule (or calorie) intake.

15b Wherever reasonable, there should be an effort to present the full range of observed data (i.e., individual measurements or responses) to illustrate the consistency of effects rather than just group summary statistics. Any normalization of measurements according to another variable or time point (e.g., % change or g/kg) requires careful consideration and justification, ideally complemented by some reporting of the original data so the reader can interpret findings within the context of the unadjusted/absolute values. Nevertheless, participants who produce a result that differs in the direction or magnitude of the group mean outcomes should not be labeled as "nonresponders" to the treatment, based solely on this solitary observation. Where the likelihood of individual variability to a treatment is anticipated, the study design could include features such as repeated testing of the same treatment or the collection of mechanistic data that might corroborate true differences in physiological/psychological responses (i.e., biological samples that may verify or explain the efficacy of effect). The possibility that day-to-day variability in the primary outcome measure explains divergent results should be considered.

15c In the interest of balanced reporting, any unforeseen negative consequences, harms, or unintended consequences must be fully disclosed alongside the primary results (rather than noted in "Discussion" section). For example, unexpected gastrointestinal issues associated with a nutritional supplement should be reported in full, irrespective of whether primary outcomes, such as metabolic and ergogenic responses, were negatively affected.

Discussion

16a Interpretation—The Discussion should maintain focus on the new data generated by the research rather than extensively reviewing the wider literature, although it should be made clear how the novel insight provided by the research complements and advances existing evidence. More speculative suggestions that go beyond the data may be acceptable in this section, but must be clearly identified as such. As per the Title and Introduction, special care should again be taken in this section to ensure the choice of language avoids hyperbole or improper inference of causality.

16b Generalization—Consideration should be given to how well the outcomes of the research project can be generalized beyond the population and context in which they were measured. The fact that the research involved a relatively homogenous or heterogeneous sample is neither a strength nor a limitation in itself, but it has implications for how broadly or specifically findings may translate to others.

16c Strengths and limitations—A succinct and truthful summary of the strengths of the design, conduct, and outcomes of the study should be integrated throughout the Discussion. Not only does this provide the reader (and reviewer) with an opportunity to reflect on the confidence they place in the study findings but, by drawing attention to the careful methodological and reporting features of the study, other researchers will be encouraged to adopt such measures in their future work. Potential sources of error or confounding variables should also be acknowledged where relevant throughout the Discussion, allowing conclusions to be tempered accordingly. A more integrated reference to limitations is usually preferred to a stand-alone paragraph toward the end of the Discussion that lists a selection of weaknesses, which are not reflected in the overall interpretation. In any case, the final conclusions need to reflect a balance of all characteristics noted in this section.

Other

17 Disclosures

The onus is on the authors to determine which relationships they believe others might conceivably consider to represent conflicts of interest or that should be disclosed for any other reason. This could include, but is not limited to, industry relationships, employment/shares, provision of nutritional supplements or consumables, book sales, or any other personal biases.

18 Protocol

Any publicly registered or published protocol should be referenced, and deviations from any prior records should be listed and explained.

References

Assel, M., Sjoberg, D., Elders, A., Wang, X., Huo, D., Botchway, A., … Vickers, A.J. (2019). Guidelines for Reporting of Statistics for Clinical Research in Urology. *The Journal of Urology, 201,* 595–604. PubMed ID: 30633111 doi:10.1097/JU.0000000000000001

Atkinson, G., & Batterham, A.M. (2015). True and false interindividual differences in the physiological response to an intervention.

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

Experimental Physiology, 100, 577–588. PubMed ID: 25823596 doi:10.1113/EP085070

Atkinson, G., & Nevill, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine, 26,* 217–238. PubMed ID: 9820922 doi:10.2165/00007256-199826040-00002

Atkinson, G., Williamson, P., & Batterham, A.M. (2019). Issues in the determination of 'responders' and 'non-responders' in physiological research. *Experimental Physiology, 104,* 1215–1225. PubMed ID: 31116468 doi:10.1113/EP087712

Bailey, D.M., Williams, C., Betts, J.A., Thompson, D., & Hurst, T.L. (2011). Oxidative stress, inflammation and recovery of muscle function after damaging exercise: Effect of 6-week mixed antioxidant supplementation. *European Journal of Applied Physiology and Occupational Physiology, 111,* 925–936. PubMed ID: 21069377 doi:10.1007/s00421-010-1718-x

Batterham, A.M., & Atkinson, G. (2005). How big does my sample size need to be? A primer on the murky world of sample size estimation. *Physical Therapy in Sport, 6,* 153–163. doi:10.1016/j.ptsp.2005.05.004

Bernard, C. (1865). *An introduction to teh study of experimental medicine*. London, UK: Courier Corporation.

Bland, J.M., & Altman, D.G. (1994). Some examples of regression towards the mean. *BMJ, 309,* 780. PubMed ID: 7950567 doi:10.1136/bmj.309.6957.780

Bland, J.M., & Altman, D.G. (1995a). Calculating correlation coefficients with repeated observations: Part 1–Correlation within subjects. *BMJ, 310,* 446. doi:10.1136/bmj.310.6977.446

Bland, J.M., & Altman, D.G. (1995b). Calculating correlation coefficients with repeated observations: Part 2–Correlation between subjects. *BMJ, 310,* 633. doi:10.1136/bmj.310.6980.633

Bland, J.M., & Altman, D.G. (1996). Transforming data. *BMJ, 312,* 770. doi:10.1136/bmj.312.7033.770

Bland, J.M., & Altman, D.G. (2015). 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. *The American Journal of Clinical Nutrition, 102,* 991–994. PubMed ID: 26354536 doi:10.3945/ajcn.115.119768

Brown, A.W., Bohan Brown, M.M., & Allison, D.B. (2013). Belief beyond the evidence: Using the proposed effect of breakfast on obesity to show 2 practices that distort scientific evidence. *The American Journal of Clinical Nutrition, 98,* 1298–1308. doi:10.3945/ajcn.113.064410

Burke, L.M., & Peeling, P. (2018). Methodologies for Investigating Performance Changes With Supplement Use. *International Journal of Sport Nutrition and Exercise Metabolism, 28,* 159–169. doi:10.1123/ijsnem.2017-0325

Byrnes, W.C., Clarkson, P.M., White, J.S., Hsieh, S.S., Frykman, P.N., & Maughan, R.J. (1985). Delayed onset muscle soreness following repeated bouts of downhill running. *Journal of Applied Physiology, 59,* 710–715. PubMed ID: 4055561 doi:10.1152/jappl.1985.59.3.710

Cook, J.A., Hislop, J., Adewuyi, T.E., Harrild, K., Altman, D.G., Ramsay, C.R., … Vale, L.D. (2014). Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference ELicitation in TriAls) review. *Health Technology Assessment, 18,* v–vi, 1–175. PubMed ID: 24806703 doi:10.3310/hta18280

Cook, J.A., Julious, S.A., Sones, W., Hampson, L.V., Hewitt, C., Berlin, J.A., … Vale, L.D. (2018). DELTA(2) guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ, 363,* k3750. PubMed ID: 30560792 doi:10.1136/bmj.k3750

Currell, K., & Jeukendrup, A.E. (2008). Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine, 38,* 297–316. PubMed ID: 18348590 doi:10.2165/00007256-200838040-00003

de Vet, H.C., Terwee, C.B., Ostelo, R.W., Beckerman, H., Knol, D.L., & Bouter, L.M. (2006). Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health and Quality of Life Outcomes, 4,* 54. PubMed ID: 16925807 doi:10.1186/1477-7525-4-54

Dwan, K., Li, T., Altman, D.G., & Elbourne, D. (2019). CONSORT 2010 statement: Extension to randomised crossover trials. *BMJ, 366,* l4378. PubMed ID: 31366597 doi:10.1136/bmj.l4378

Funnell, M.P., Mears, S.A., Bergin-Taylor, K., & James, L.J. (2019). Blinded and unblinded hypohydration similarly impair cycling time trial performance in the heat in trained cyclists. *Journal of Applied Physiology, 126,* 870–879. doi:10.1152/japplphysiol.01026.2018

Greenland, S. (2019). Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolution with s-values. *The American Staistician, 73,* 106–114. doi:10.1080/00031305.2018.1529625

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N., & Altman, D.G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31,* 337–350. PubMed ID: 27209009 doi:10.1007/s10654-016-0149-3

Harriss, D.J., Macsween, A., & Atkinson, G. (2017). Standards for ethics in sport and exercise science research: 2018 update. *International Journal of Sports Medicine, 38,* 1126–1131. PubMed ID: 29258155 doi:10.1055/s-0043-124001

Hecksteden, A., Kraushaar, J., Scharhag-Rosenberger, F., Theisen, D., Senn, S., & Meyer, T. (2015). Individual response to exercise training—A statistical perspective. *Journal of Applied Physiology, 118,* 1450–1459. PubMed ID: 25663672 doi:10.1152/japplphysiol.00714.2014

Jeacocke, N.A., & Burke, L.M. (2010). Methods to standardize dietary intake before performance testing. *International Journal of Sport Nutrition and Exercise Metabolism, 20,* 87–103. doi:10.1123/ijsnem.20.2.87

Kordi, R., Mansournia, M.A., Rostami, M., & Maffulli, N. (2011). Troublesome decimals; a hidden problem in the sports medicine literature. *Scand J Med Sci Sports, 21,* 335–336. PubMed ID: 21564305 doi:10.1111/j.1600-0838.2011.01312.x

Lazic, S.E. (2010). The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? *BMC Neuroscience, 11,* 5. doi:10.1186/1471-2202-11-5

Lonergan, M., Senn, S.J., McNamee, C., Daly, A.K., Sutton, R., Hattersley, A., … Pirmohamed, M. (2017). Defining drug response for stratified medicine. *Drug Discovery Today, 22*(1), 173–179. PubMed ID: 27818254 doi:10.1016/j.drudis.2016.10.016

Maughan, R.J. (2004). Returning to the writing of abstracts. *Journal of Sports Sciences, 22,* 603. doi:10.1080/02640410410001724987

McHugh, M.P., Connolly, D.A.J., Eston, R.G., & Gleim, G.W. (1999). Exercise-induced muscle damage and potential mechanisms for the repeated bout effect. *Sports Medicine, 27,* 157–170. PubMed ID: 10222539 doi:10.2165/00007256-199927030-00002

Schulz, K.F., Altman, D.G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Trials, 11,* 32. PubMed ID: 20334632 doi:10.1186/1745-6215-11-32

Schulz, K.F., & Grimes, D.A. (2002a). Allocation concealment in randomised trials: Defending against deciphering. *The Lancet, 359,* 614–618. doi:10.1016/S0140-6736(02)07750-4

Schulz, K.F., & Grimes, D.A. (2002b). Blinding in randomised trials: Hiding who got what. *The Lancet, 359,* 696–700. doi:10.1016/S0140-6736(02)07816-9

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

Schulz, K.F., & Grimes, D.A. (2002c). Generation of allocation sequences in randomised trials: Chance, not choice. *The Lancet, 359,* 515–519. doi:10.1016/S0140-6736(02)07683-3

Schulz, K.F., & Grimes, D.A. (2002d). Sample size slippages in randomised trials: Exclusions and the lost and wayward. *The Lancet, 359,* 781–785. doi:10.1016/S0140-6736(02)07882-0

Schulz, K.F., & Grimes, D.A. (2002e). Unequal group sizes in randomised trials: Guarding against guessing. *The Lancet, 359,* 966–970. doi:10.1016/S0140-6736(02)08029-7

Senn, S., & Julious, S. (2009). Measurement in clinical trials: A neglected issue for statisticians? *Statistics in Medicine, 28,* 3189–3209. PubMed ID: 19455540 doi:10.1002/sim.3603

Thomas, D.M., Clark, N., Turner, D., Siu, C., Halliday, T.M., Hannon, B.A., … Allison, D.B. (2019). Best (but oft-forgotten) practices: Identifying and accounting for regression to the mean in nutrition and obesity research. *The American Journal of Clinical Nutrition*. Advance online publication. doi:10.1093/ajcn/nqz196

Tu, Y.K., & Gilthorpe, M.S. (2007). Revisiting the relation between change and initial value: A review and evaluation. *Statistics in Medicine, 26,* 443–457. PubMed ID: 16526009 doi:10.1002/sim.2538

Vickers, A.J. (2001). The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Medical Research Methodology, 1,* 6. PubMed ID: 11459516 doi:10.1186/1471-2288-1-6

Vickers, A.J. (2005). Analysis of variance is easily misapplied in the analysis of randomized trials: A critique and discussion of alternative statistical approaches. *Psychosomatic Medicine, 67,* 652–655. PubMed ID: 16046383 doi:10.1097/01.psy.0000172624.52957.a8

Wellek, S., & Blettner, M. (2012). On the proper use of the crossover design in clinical trials: Part 18 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International, 109,* 276–281. PubMed ID: 31765436

Winter, E.M., & Maughan, R.J. (2009). Requirements for ethics approvals. *Journal of Sports Sciences, 27,* 985. doi:10.1080/02640410903178344

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

Appendix

PRESENT 2020 (PROPER REPORTING OF EVIDENCE IN SPORT AND EXERCISE NUTRITION TRIALS)

CHECKLIST OF INFORMATION TO INCLUDE WHEN REPORTING RESEARCH IN SPORT NUTRITION AND EXERCISE METABOLISM[a]

Section	Item	Checklist	Page/Line Number (if applicable)
Title			
	1a	State the independent (groups/conditions) and dependent (outcome) variables	_____
	1b	Identify the study population or case	_____
Abstract			
	2a	Specify the research design, methods, and characteristics of study population	_____
	2b	Report a balanced account of the results and cite actual data	_____
	2c	Restrict conclusions to measured variables, without speculation or unsupported recommendations	_____
Introduction			
	3a	Present a scientific rationale based on an objective review of available evidence	_____
	3b	State the aims, objectives, research questions, and/or hypotheses	_____
Methods			
Ethics	4	Provide details of ethical approval (citing conduct of human research in accordance with the Declaration of Helsinki)	_____
Design	5	Summarize the research design (e.g., parallel trial/cross-over, randomized, counterbalanced, blinding, observational)	_____
Sampling	6a	List the eligibility (inclusion/exclusion) criteria and sampling method	_____
	6b	Characterize the study sample (e.g., demographics, anthropometry, lifestyle)	_____
	6c	Report the setting/location and periods of recruitment and data collection	_____
	6d	Justify the sample size (presenting the selected target effect size and error variances to replicate sample size estimates)	_____
Interventions[b]	7	Detail all aspects of the groups/conditions (considering the need to verify the composition of ingested substances)	_____
Measurements	8a	Define the pre-specified primary, secondary and/or mechanistic outcome variables	_____
	8b	Rationalize the selection of test protocols, considering validity and reliability (e.g., coefficient of variation, familiarization)	_____
	8c	Justify the smallest worthwhile effect or minimal clinically important difference	_____
Randomization	9	Detail the exact mechanisms of generating and concealing the random allocation sequence	_____
Blinding[b]	10	Document whether participants and/or researchers were aware of allocation (e.g., exit questionnaire)	_____
Standardization	11	Describe within- and between-participant controls (e.g., replication/reporting of diet, physical activity, sleep, menstrual cycle)	_____
Order Effects	12	Detail control of systematic influences of serial measurements (e.g., sequence effect in analysis model, wash-out interval)	_____

(continued)

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC

PRESENT 2020 Checklist *(continued)*

Section	Item	Checklist	Page/Line Number (if applicable)
Statistics	13a	Specify the contrast for primary inferences (i.e., relative to the appropriate control, not changes from baseline in each group/condition)	_____
	13b	Clearly distinguish and fully justify any unplanned, interim or exploratory subgroup analyses	_____
	13c	Describe any adjustments for violated statistical assumptions and for relevant covariates (e.g., baseline measures)	_____
Results			
Participant Flow	14a	Report the sample size at each phase from recruitment to analysis (with reasons for losses and exclusions)	_____
	14b	Ensure data analysis matches research design, avoiding data pooling across groups/conditions (i.e., pseudoreplication)	_____
Outcomes	15a	Report SI units and report measures of central tendency, variability, and effect size/precision (confidence intervals)	_____
	15b	Report individual data/responses (e.g., draw figures showing the raw data in each group/condition)	_____
	15c	Document all relevant harms and unintended consequences observed	_____
Discussion			
	16a	Present an objective and balanced interpretation of the observed data within the context of existing evidence	_____
	16b	Consider the applicability and/or practical relevance of the research findings (e.g., external validity)	_____
	16c	Acknowledge strengths and limitations of the research relevant to accurate interpretation (e.g., internal validity)	_____
Other			
Disclosures	17	State any relevant relationships (e.g., financial, technical, material support)	_____
Protocol	18	Identify any publicly registered or published protocol (explaining any deviations)	_____

[a]Adapted from the 2010 CONSORT (CONsolidated Standards Of Reporting Trials) checklist for reporting randomized controlled trials and can be used in conjunction with the associated paper that expands on each item.
[b]Items 7 (Interventions) and 10 (Blinding) are relevant for experimental research, including single-/double-blind contrasts of nutritional supplements.

Brought to you by PILKINGTON LIB GBR | Downloaded 02/12/20 09:44 AM UTC