This work is a preprint and
has not been peer-reviewed
doi: 10.51224/SRXIV.379

Laboratory and Field-Based Data Collection (Quantitative)

Bryan Saunders[1,2,3], Felipe Marticorena[1], Philip Hurst[4], Lewis Gough[5]

1 – Applied Physiology and Nutrition Research Group – School of Physical Education and Sport and Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, Brazil.
2 – Center of Lifestyle Medicine, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, Brazil
3 – Nutrology Academy, Rio de Janeiro, Brazil
4 – School of Psychology and Life Sciences, Canterbury Christ Church University, Canterbury, United Kingdom.
5 – Human Performance and Health Research Group, Centre for Life and Sport Sciences (CLaSS), Department of Sport and Exercise, Birmingham City University, Birmingham, United Kingdom.

This is a preprint of a chapter forthcoming in the following edited textbook: Eimear Dolan & James Steele (Eds). Research Methods in Sport and Exercise Science. An Open-Access Primer. Published by the Society for Transparency, Openness and Replication in Kinesiology.

Please cite as: Saunders, B., Marticorena, F., Hurst, P. and Gough, L. (2024). Preprint version - Laboratory and Field-Based Data Collection (Quantitative). In: Eimear Dolan and James Steele (Eds). Research Methods in Sport and Exercise Science. An Open-Access Primer. Published by the Society for Transparency, Openness and Replication in Kinesiology.
Preprint DOI: 10.51224/SRXIV.379

doi: 10.51224/SRXIV.379

Feedback:

If you would like to provide feedback on this chapter, please email me directly (drbryansaunders@outlook.com) or use the comments section on the https://web.hypothes.is/. For the latter you can create a free account in about a minute. Then download the browser extension or bookmarklet, download this pdf locally but open it in your browser, activate the hypothes.is extension/bookmarklet, and then you can add feedback directly via annotations. Afterwards, email or share the pdf file and I will be able to see this feedback via the hypothes.is fingerprint on the file. See this video for more information https://web.hypothes.is/annotating-pdfs-tutorial/

Correspondence:

Dr Bryan Saunders
Applied Physiology & Nutrition Research Group,
Faculty of Medicine FMUSP,
Av. Dr. Arnaldo, 455 - Cerqueira César - CEP: 01246903
University of São Paulo,
São Paulo, SP, Brazil.
E-mail: drbryansaunders@outlook.com
Phone: +55 11 97776-6047

 doi: 10.51224/SRXIV.379

Table of Contents

 doi: 10.51224/SRXIV.379

1. Introduction

Rigorous assessment of sport and exercise measures is a requirement for any scientist aiming to answer a research question. Sport and exercise scientists may strive to answer questions such as, "Does caffeine improve an athlete's performance?", "What are the physiological determinants of endurance running?" and "When can an athlete return to training after injury?". Researchers aim to answer these questions through data collection in experimental studies that are designed to test a hypothesis and provide robust evidence on a topic. This is pertinent as the replicability of findings in sport and exercise research has been questioned (Mesquida et al., 2022). By prioritizing methodological quality in research, researchers can enhance the credibility and trustworthiness of their findings and, in turn, promote the replicability of research findings in the field of sport and exercise science. To help researchers design their studies, there are several guidelines that offer recommendations on appropriate reporting (Consolidated Standards Of Reporting Trials, CONSORT) with some more specific to exercise nutrition (Proper Reporting of Evidence in Sport and Exercise Nutrition Trials, PRESENT) (Betts et al., 2020). While these provide excellent considerations to ensure reporting of the scientific method is complete, they can also be used as guidelines implemented prior to data collection to ensure that the study results are robust.

Quantitative data collection in sport and exercise research can include different methods including surveys and questionnaires, biomechanical and physiological measures and exercise capacity and performance measures. These data can be obtained in controlled laboratory environments or in an applied setting (*e.g.*, during a race) depending upon the specific research question. Here we aim to focus primarily on practical data collection, such as obtaining measures of physiological responses and exercise performance. Furthermore, fundamental to this is the use of randomised controlled trials, which are often regarded as the cornerstone of any data collection researchers conduct across the field of science. Below, we provide an overview of the essential components that researchers should consider both in the laboratory and field, with emphasis given to collecting data during randomised controlled trials.

2. Ethical considerations

1. Institutional review or ethics committee

Prior to initiating data collection, researchers are required to submit their project to their Institutional Review Committee or Institutional Ethics Committee which is formally

 doi: 10.51224/SRXIV.379

designated to review and approve research involving human participants according to ethical principles such as the Declaration of Helsinki developed by the World Medical Association (https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/). Specifically, the primary role of the ethics committee is to safeguard the rights, welfare, and privacy of those participating in research studies. Researchers are required to submit detailed proposals outlining their study objectives, methodology, participant recruitment procedures, and outcome measures to ensure confidentiality and informed consent. The committee then evaluates all aspects of these study proposals to ensure that potential risks to participants are minimized and that the anticipated benefits of the research justify any potential harm. The committee may request some changes to the proposal if they believe that the risk of certain procedures is too high or outweighs the societal benefits. Only once a study has been approved by the ethics committee can a study initiate participant recruitment, following obtention of informed consent (see Section 2.2 Informed consent). This ensures the ethical and responsible conduct of research while protecting the rights and well-being of the research participants.

2. Informed consent

Before the collection of any data, it is prerequisite to gain informed consent, ideally in writing, from participants in the study. Non-written consent (i.e., verbal) should be supported by witness statements or audio or video recordings to ensure all parties are covered and avoid disputes as to whether consent was given. All informed consent forms should be stored safely and confidentially (see Section 3.4.2. Data management). To ensure participant well-being, and protect them from harm, informed consent ensures that participants are aware of the aims, the method, and potential outcomes and risks associated with the study. To achieve this, researchers need to provide unbiased, up-to-date, relevant information of their decision to participate in the study and importantly, that participation is completely voluntary, for which they can choose to withdraw at any time without reason and consequence. To help participants decide whether to participate in the study, and understand potential consequences, they should be provided with an information sheet that contains brief and clear information on the essential aspects of the study. The Standards for Ethics in Sport and Exercise Research (Harriss et al., 2022) lists what need to be included in the information sheet (see Table 1).

It is important to note that any information should be written clearly and be easy to read for a layman. The use of technical and jargon should be avoided, but if required, should be first

 doi: 10.51224/SRXIV.379

explained in a plain, accessible language. Researchers may not consider that a lot of the language used in their day-to-day work is in fact technical. Words used throughout this chapter, for example - randomisation, sample size, blinding and validity – may be complex for a participant, and as such, should be avoided to ensure they are fully aware of what is required of them. Given this, researchers should aim to provide information about the study in both written and spoken form. The former can be emailed or sent to participants prior to visiting the data collection site (*e.g.*, the laboratory), so that they have ample opportunity to read all information and be cognisant of what to expect in the study. The latter offers the opportunity to expand on technical areas and provides participants the opening to question and alleviate any concerns.

While most data collection will sample the general population, researchers may also be interested in sampling other populations that are more vulnerable, including children, the elderly, and those with intellectual impairments. Researchers will therefore need to consider additional ethical concerns and be aware that it may not be possible to gain consent or that they need more time. Passive assent, which can involve a parent or guardian, should be avoided where possible, and every effort should be made to involve the participant in the informed consent process. Explaining the details for informed consent for vulnerable groups are outside the scope of this chapter, but readers are directed towards the UK Research and Innovation guidance (UKRI, 2023).

Table. 1 Brief outline of information required for a participant information sheet

#	Information given to participant	Elaboration
1	Researcher details	Names and institutional affiliations
2	The aims of the research	Why is the work being undertaken?
3	Methods of the study	What will participants be asked to do?
4	Sources of funding	Has an organisation funded the study?
5	Conflicts of interest	Would financial or personal consideration compromise the research?
6	Anticipated benefits	What benefits can participants receive?
7	Potential risks	What harms of consequences come from participation?
8	Right to decline	Participants do not have to take part and can do so without consequence

doi: 10.51224/SRXIV.379

9	Right to withdraw	Participants can withdraw consent without consequence
10	Handling of data	Where will data be stored, shared and accessed
11	Retention of data	How long will data be stored?
12	Contact details	Who can participant's contact if they have questions or complaints?

Note: Content is adapted from Harriss et al. (2022)

3. Experimental design

The extent to which the observed results of an experimental study represent the true effect of the intervention depends on the rigour of the methodology. Internal validity is the term used to describe whether the methodology was conducted adequately to answer the research question without substantial bias (Andrade, 2018; Halperin et al., 2015). There is an extensive list of confounding factors which could potentially influence exercise performance (e.g., diet, sleep, fatigue) and, thus, should be considered and/or controlled to various extents depending upon the research question being asked and how they might impact upon the data. External validity relates to how generalisable the current data are to other contexts (Andrade, 2018). For example, a study looking at the effects of a training intervention in elderly individuals with type II diabetes will likely not be entirely generalisable to a young and athletic population. Ecological validity is a sub-section of external validity applied to the real-world, specifically, whether the study can be generalisable to everyday life. For example, a study showing the side-effects of caffeine (*e.g.*, anxiety) on participants in a resting and relaxed state in a seamlessly controlled laboratory may have high internal validity but is in stark contrast to the high-pressure environment of competitive sport, and results may therefore not be directly applicable. Understanding of internal and external validity is vital to design and conduct studies and to understand the limitations of that research. The following sections aims to critically discuss their importance in relation to data collection.

1. Laboratory and field-based research

Most research questions are focused on determining the mechanistic characteristics (*e.g.*, physiological, psychological, biomechanical, sociological) of sport or the effectiveness of sport and exercise science interventions, both in the field (applied) and laboratory. The advancement and development of cutting edge and portable technologies means that researchers have a plethora of methods through which to answer their research questions in both the laboratory

 doi: 10.51224/SRXIV.379

and field. While the laboratory is often the preferred choice, given its high reliability, sensitivity, and ability in which to control several variables, such as temperature and humidity, researchers can conduct research within the field, which offers more ecological validity that can help translate findings into real world scenarios. Nevertheless, both have their own limitations.

1. Laboratory-based research

The primary benefit of laboratory research is that more extraneous factors can be controlled compared to field or remote data collection, including the ability to control factors such as the environment (humidity, temperature) and using ‘gold standard’ laboratory equipment to enhance internal validity (in most cases). These added layers of control allow the researcher to be confident that performance measures are not a result of extraneous factor(s). In most laboratories, temperature is controllable through air conditioning systems, and in most cases, humidity will also be constant. It is advised to keep this consistent during data collection both between and within participant procedures, with records being kept for each experimental session. The main drawback of laboratory research is that the environment is largely artificial, especially in sport where athletes often compete in an environment that is constantly changing (e.g., weather, temperature, typography, anxiety from high-pressured environments). As a result, the findings in rigorously controlled laboratories lose generalisability to sport practitioners (*i.e.*, ecological validity). Moreover, demand characteristics could impact the findings whereby participants may behave differently when being observed (Nichols & Maner, 2008).

2. Field-based research

Field-based research has become a more common approach within sport and exercise sciences due to the ability to increase the ecological validity of the findings. An area of concern with field-based research is selecting an exercise protocol or using equipment that are valid against laboratory or ‘gold standard’ measurements (Halperin et al., 2018). Exercise protocols such as the multistage 20-m shuttle run test (more commonly known as the “bleep test”) have been shown to correlate to a good level with maximal rate of oxygen output ($\dot{V}O_{2max}$) (Léger & Lambert, 1982; Paliczka et al., 1987; Ramsbottom et al., 1988) making it an appropriate surrogate in the field. Considering physiological measures and blood lactate as an example, analysers were traditionally a large benchtop equipment that was not readily portable. The

doi: 10.51224/SRXIV.379

development of a portable handheld device such as the Lactate Pro 2 (Arkray, Japan) has overcome such issues, and research has shown it to be useable in the field, and importantly, valid (Bonaventura et al., 2015) and reliable (Tanner et al., 2010) against 'gold standard' laboratory analysers. As a result, the findings in studies using field-based measurements and techniques can then make valid inferences to guide practitioners. Despite some successes, in sport sciences, this is perhaps not implemented as often as it should. One example includes electromyography (EMG), which is commonly used to infer muscle hypertrophy with higher versus lower amplitudes, however, it is unknown if this is a causal relationship (Halperin et al., 2018). The use of valid techniques is an area that sport science could improve to help practitioners make informed decisions with participants from a sports performance, but also a health perspective (Abt et al., 2022).

Another factor to consider in the field is the lack of control versus laboratory settings, such as weather, temperature, and aerodynamics. This is particularly common if data collection is ongoing during a competition. While these extraneous factors could influence the results and compromise internal validity, particularly if the study design is a crossover design and researchers are attempting to determine changes from multiple different treatments (*e.g.*, a supplement study to assess the impact on exercise performance), it can be minimised by conducting the test at the same time of day, season (*i.e.*, summer vs. winter) and in similar air density (*e.g.*, indoor track cycling). The best approach for this type of research is to measure and describe as much as possible so that the reader can interpret the extraneous factors that might have influenced results. Furthermore, the authors themselves may use the measured variables to apply a correction factor to standardise conditions for test performed on different days. For example, one study investigating caffeine on 100-m running performance measured temperature, humidity, atmospheric pressure and wind speed to standardise measurements (Matsumura et al., 2022). While it may reduce the generalisability of the research, if all can be accounted for, the benefit of field-based studies is the increased level of ecological validity, which in turn, usually means greater impact within the given sport of focus.

3. Remote data collection

Since the COVID-19 pandemic in 2020, a contemporary approach has been to collect data remotely due to the obvious constraints on face-to-face contact (Souza et al., 2022). This is unique compared to field-based testing as it requires no observer (i.e., researcher) of the data

 doi: 10.51224/SRXIV.379

collection process. In the context of sport and exercise research, this could increase the inclusivity and reduce the carbon footprint of research, as well as opening opportunities for multicentre experiments. For example, one study collected 165 data sessions on a cycle ergometer remotely over a 2-month period using the commercially available software TrainingPeaks™ (Bennett et al., 2021). Given that small sample sizes are common in sport science and can cause issues with power (Abt et al., 2020) and difficulties in translating to real-world settings, remote data collection provides an opportunity to sport and exercise researchers in recruiting larger homogeneous and heterogenous samples. For this to grow and become the norm within the discipline, however, attempts to maintain the reliability and validity must be factored into the study design. Like field-based studies, this includes using consistent methodologies and equipment across participants (*e.g.*, software and equipment), and visual inspections of data collection where possible (*i.e.*, raw data checks, virtual observation of experimental trials taking place). An example of this approach was shown by Matta and colleagues (2022) whereby the reproducibility of a 20-min cycling time-trial was assessed using a home-based protocol. Participants completed two exercise trials using their own home setup on a commercially available software platform (Zwift™) and cycle ergometer (and power meter). This type of approach could be adopted for similar studies, except it would be encouraged that the researcher could watch experimental trials being performed virtually using software (*e.g.*, Microsoft Teams™, Zoom™), which the researchers opted against in their study design (Matta et al., 2022). With this addition, there would be little difference between this approach and both laboratory and field study designs providing no complex data collection is required (*e.g.*, blood sampling, physiological measures).

2. *Randomisation*

Randomisation is considered a critical component of an experimental study that ensures each participant has an equal chance of being assigned to a specific treatment group (in a parallel group study; *e.g.*, 4-week of either beta-alanine or placebo supplementation) or intervention order (in a crossover study; *e.g.*, receiving caffeine first then placebo, or placebo first then caffeine). In performing this allocation entirely randomly, we avoid distorting results due to non-random allocation, which could lead to group differences due to baseline characteristics or identical treatment orders for all participants that, in turn, could bias outcomes.

 doi: 10.51224/SRXIV.379

Where possible, simple randomisation methods should be preferred. Randomisation is as simple as allocating participants to a treatment group (*e.g.*, beta-alanine or placebo) or order (*e.g.*, "caffeine – placebo" or "placebo – caffeine") using a coin flip or throwing a dice (Schulz & Grimes, 2002). There is a limitation with simple randomisation in that small sample sizes (<200; (Schulz & Grimes, 2002)), which are common in sport and exercise research, may lead to an uneven number of participants allocated to a particular order, or uneven group sizes. Nonetheless, with increasing sample sizes, this chance is diminished. Block randomisation is also often employed, whereby participants are allocated in an equal ratio (*e.g.*, 1:1 or 2:2) to a treatment group or order. Additionally, studies in sport and exercise science often wish to avoid baseline differences in fitness or performance of participants between groups, and can use stratified randomization to do so (Kang et al., 2008). For example, in a study examining the effects of different training protocols (*e.g.*, high-intensity intermittent exercise vs. continuous exercise) on changes in $\dot{V}O_{2max}$, it would be undesirable for the two training groups to differ significantly in their baseline $\dot{V}O_{2max}$ since those with lower baseline values are likely more susceptible to greater training responses (Støren et al., 2017), regardless of the specific training protocol. Thus, participants could be stratified according to groups based upon their baseline $\dot{V}O_{2max}$. One way this could be achieved is allocating participants to chosen groups of baseline $\dot{V}O_{2max}$ (*e.g.*, 45-50; >50-55; >55-60;>60-65 $mL \cdot kg^{-1} \cdot min^{-1}$) and within each group, an equal number of participants are randomly allocated to each training condition. This helps ensure that baseline $\dot{V}O_{2max}$ does not differ between groups. An obvious limitation is that the researchers are reliant on equal numbers of participants in each sub-group, and that drop-outs may occur more so in one group that another, which is something that cannot be predicted, and may lead to significant baseline differences. If this occurs, researchers should report the differences in baseline and/or number of dropouts for each condition, and exercise caution in their conclusions.

The randomisation procedure should be performed by somebody not involved in data collection so that there is no knowledge from the participant or researcher about the intervention being administered, a concept termed allocation concealment. This maintains the blinding of the study should it be necessary (see Section 4. Blinding) and minimises the chance of selection bias, an error that occurs if proper randomisation is not performed resulting in skewed or unrepresentative samples. The person undertaking the randomisation may wish to use free online tools such as <u>Randomization Plans: Never the same thing twice!</u>

 doi: 10.51224/SRXIV.379

(jerrydallal.com) or Research Randomizer. As much information as possible as to how the randomisation was performed should be included in any subsequent publication to allow readers to evaluate whether proper randomisation was implemented, of whether possible bias has occurred due to improper randomisation which may occur unwittingly (Schulz & Grimes, 2002). We direct the reader towards further reading to gain a more in-depth overview of the methods and techniques for randomisation (Kang et al., 2008; Suresh, 2011).

Personal view: In our study on caffeine supplementation and exercise performance (Saunders, de Oliveira, et al., 2017), block randomisation was performed by someone not involved in data collection so that all possible orders in which participants could receive the supplements (6 different orders to receive three treatments; caffeine, placebo and control) were balanced across 42 participants.

3. Blinding

Participant and researcher expectations about the intervention can significantly affect outcomes during data collection. As a result, within randomised controlled trials, a fundamental decision is to consider whether participants, and those conducting data collection, are blinded to the intervention (*i.e.*, they do not know what interventions are being provided). For example, imagine a research study examining whether caffeine improves 5000-m running time compared to placebo. If a participant is aware they received caffeine, and expect it improves performance, they may change how they perform the trial than when they receive placebo (see for example (Hurst, Schiphof-Godart, et al., 2020)). Researchers would therefore be unable to determine if it was caffeine that improved performance or the change in behaviour. Similarly, if a researcher is aware they are administering caffeine to participants, they may change their behaviour during the trial, such as their body language, words used during administering the caffeine, and type of encouragement given during the trial. As a result, even if the participant is unaware they received caffeine, they may perform the trial differently based on the behaviour of the researcher.

Blinding in research studies generally takes three forms. First, researchers can use a single-blind design, which involves ensuring only participants do not know which intervention has been administered. This will most likely occur when resources are limited and the person conducting the data collection also needs to administer the intervention. Second, a double-blind

doi: 10.51224/SRXIV.379

intervention can be conducted, in which both the participant and researcher administering the intervention are unaware of what has been administered. In this design, a third-party not involved in data collection disguises both the intervention and placebo so that they are identical in appearance. The researcher would then administer the intervention or placebo to the participant, and both would be unaware what had been administered. Finally, in a triple-blind study, to remove any biases relating to how the data is analysed, the person analysing the data following the completion of data collection can also be unaware of which data is related to the intervention or placebo.

Blinding is more than just keeping the name of the intervention hidden. Blinding relates to the entirety of the study. This includes, but is not limited to, researchers developing the blinding, witnessing other participants receiving the intervention, perceptual cues of the interventions (*e.g.*, taste, colour, smell) and even physiological responses. The latter can be inherently difficult to blind, especially for some interventions that have noticeable physiological responses, such as sodium bicarbonate that can cause gastrointestinal symptoms (McNaughton, 1992; Saunders et al., 2014). If a participant experiences such effects, then the blinding has failed, and any further data collected is likely biased. It is generally considered that successful blinding ensures the results of the study are not subject to bias. Nonetheless, it is possible that participants experience side-effects related to the active ingredient despite having received a placebo, which may be intrinsically linked to expectation and the information provided regarding the intervention. Blinding success can be assessed by directly asking participants which intervention they think was administered and this data can then be analysed using a number of different tools, such as the Bang's Blinding Index (Bang et al., 2004), which can be used to evaluate the blinding of each intervention (e.g., in a caffeine vs. placebo study, you can determine whether blinding was successful both within the caffeine visit and the placebo visit). Bang's Blinding Index provides a value between -1 and 1, with successful blinding considered between -0.30 to 0.30 (Bang et al., 2010). If blinding was unsuccessful, then blinding may have been compromised and influenced the result of the study, something which researchers may wish to consider upon interpretation of the data.

It can sometimes be impractical or unfeasible to blind participants to an intervention. It would be impossible to blind a participant to, for example, Normatec compression therapy, physiotherapy, or high-intensity interval training (HIIT) since participants know when they are

 doi: 10.51224/SRXIV.379

receiving these interventions. As a result, in such studies, it would be necessary for the researcher to understand participant expectations of the intervention and whether they believed it influenced outcomes. This can be achieved via a questionnaire prior to or post study (*e.g.*, asking participants on a Likert-type scale from 1-5 how much they expect it to affect outcomes), or through post-study interviews, and assessing how much they expected the intervention to influence outcomes (Gurton et al., 2022). The results of this data should be considered during the main analyses and can help determine if they influenced outcomes of the intervention.

While blinding is often regarded as the gold standard during experimental data collection, sometimes researchers may be interested in understanding the effects of an intervention that has already been shown to be beneficial in blinded studies. This design is called open-label, and is arguably best conducted within the field, where outcomes are of interest under real-world conditions. Given that caffeine has shown to be efficacious during double-blind randomised controlled trials (Grgic et al., 2020), it would be useful to understand if these effects are translated to the field, when participants are aware they have received caffeine. There would be no need to blind participants to what they received, and researchers can understand if caffeine improves performance when given openly.

Personal view: We conducted a double-blind, randomised controlled trial to determine if an acute dose of dietary nitrate improved 5-km running performance (Hurst, Saunders, et al., 2020). We purchased the placebo from the supplier "Beet-IT", who developed a placebo product identical in taste, smell and appearance (Gilchrist et al., 2014). To ensure we administered the correct intervention to participants, we asked another person to label one "X" and another "Y". During data collection, participants received the X or Y intervention, and we were unaware of whether it was the dietary nitrate or placebo until after the study had been completed.

4. Confounding Variables

1. Observers and researchers

One factor that could impact research is the number and/or sex of observers present at data collection, which could enhance or hinder the participant's performance. Winchester et al. (2012) reported that ratings of perceived exertion, a subjective measure of how hard the participant believes the exercise is, was reduced with both female and male observers when

 doi: 10.51224/SRXIV.379

men were completing a run at ~60% peak running speed. This seemed to be due to the changes in affect scores as these were significantly higher compared to a control trial. In another study, van der Meij et al. (2008) reported that testosterone increased in men by 8% when a woman was introduced to the experimental trials versus a 0.5% change when this was a man. Similarly, 24 young male handball players' performance was improved in the presence of female versus male observers. In contrast, the exercise performance of women when in the presence of observers appears hampered, although in some cases it was unchanged. Based on this evidence, researchers should be aware of these potential issues and ensure their research environment limits these impacts. This can be achieved using private research spaces or the use of screens to block the viewing of external individuals within open laboratory spaces.

The number of people observing exercise can influence exercise performance, whether indirectly or directly observing in the environment (Halperin et al., 2015). One study showed that an audience of fifteen individuals directly watching participants perform a 1-RM bench press improved performance compared to either a passive audience of co-actors (not directly watching; 12.9% increase) or a competitive scenario (fewer direct observers; 2% increase). A factor that might mitigate or enhance these responses is whether the observer is known to the participant, where it has been shown that if this is the case, performance may not change, whereas if the additional person is unknown, a reduction in performance may be found (Guerin, 1986). It is worth noting that this change is more likely to be seen for complex tasks (*e.g.*, team sport actions) than simple ones (*e.g.*, capacity or stamina tests). This impact is related to the work of Guerin (1983) who suggested only if the additional audience are not known to the participant would this cause uncertainty and the performance might reduce. This contrasts an early theory such as the generalised drive hypothesis (Zajonc, 1965), whereby a participant's performance will be improved simply through the presence of others. Whilst such theories have since been criticised and many are not discussed in this chapter (for full review see (Strauss, 2002)), it is worthwhile for a researcher to consider this within their laboratory research project to reduce the interference of observers in the results. Our recommendation would be that researchers standardise the number and sex of the researchers who will be present at all main data collection sessions throughout a project.

 doi: 10.51224/SRXIV.379

2. *Verbal encouragement*

Verbal encouragement is often seen as a key factor to help participants produce their best effort. However, approximately one third of participants can experience a neutral or negative response to verbal encouragement (Midgley et al., 2018). Unfortunately, there is little evidence to guide recommendations with limited literature to date (Midgley et al., 2018). Of the available evidence, Andreacci et al. (2002) reported that verbal encouragement every 20-s and 60-s improved running performance, whilst no effect was found with encouragement every 180-s. Therefore, for maximal efforts verbal encouragement in a frequency of every 20-60-s could assist participants performance. During resistance training, verbal encouragement can improve performance, as Weakley et al. (2020) reported improvements in weight lifted during barbell back squats within a group of 12 semi-professional rugby players. Binboğa et al. (2013) reported that those with low conscientiousness significantly improved their maximal voluntary contraction of the *triceps surae*, but reported no improvements in those with high conscientiousness (9.7% vs. 2.4%). Reasons for discrepancies between Weakley et al. (2020) and Binboğa et al. (2013) may be the sample size (n = 12 vs n = 83) and the different exercise tests (barbell back squat vs. maximal voluntary contractions). Nonetheless, this suggests that for resistance type exercise, verbal encouragement may be beneficial to produce a best effort performance, however, this might be dependent on the level of conscientiousness within individuals.

Although most research has focused on positive feedback, there is a small body of research examining negative feedback. Instead of stating “great effort”, “excellent values” and “looking strong”, when researchers state “you're not trying”, “low values” or “you can do better”, this may improve performance (Halperin et al., 2020). This was hypothesised to be due to participants experiencing some level of anger and exerting greater effort due to the suggestion that their initial efforts were lacking. However, caution is advised since negative feedback might not elicit positive effects in the long-term due to effects on motivation and self-efficacy. Since positive feedback improved performance over no feedback (Halperin et al., 2020), this type of feedback should be preferred. Based on this evidence, it is reasonable to suggest that if verbal encouragement is to be offered it should, at the very least, be standardised. Preferably, the level and frequency of encouragement would also match the type of exercise to achieve the desired effect. Equally, it may be intuitive to match the encouragement based on the level of conscientiousness of participants where possible.

 doi: 10.51224/SRXIV.379

Personal view: In our studies (Gough et al., 2018; Gough, Rimmer, et al., 2019), we have used multiple approaches for verbal encouragement; however, all have been standardised either to time or distance based (for both time-to-exhaustion and time-trial tests) at approximately 60-s intervals. In one study, the encouragement was provided every 500 m across a 4-km time-trial. We also attempt to standardise the phrases used throughout (e.g., 'good work, keep going') by using a phrase bank for encouragement.

3. *Familiarisation or habituation to the exercise protocol*

A key element of rigorous study control is whether participants within a research study are familiarised to the exercise protocol. Familiarisation sessions are usually included in experimental designs to reduce the effect of learning. This is especially important when utilizing untrained samples and participants not familiar to the exercise protocol. Participants in research are often unfamiliar with the exact demands of the exercise task being undertaken (*e.g.*, many cyclists may not be familiar with a 4-km cycling time-trial). Including a session whereby participants perform the exercise task to become familiar with it, researchers can reduce the coefficient of variation and increase test-retest reliability between exercise sessions (Stevens & Dascombe, 2015) which avoids confounding the effect of the intervention with learning. The importance of this is highlighted by the work of Stein and colleagues, who first published their study showing that caffeine improved performance, but were forced to retract their article after discovering results were due to data tabulation error (discussed below) and that the effect was due to a lack of a familiarisation to the exercise protocol and a learning effect (Stein et al., 2020a; Stein et al., 2020b).

While many researchers perform a solitary familiarisation session in which participants are made familiar with the exercise task, this should not be confused with habituation of a participant to an exercise task. That is, a familiarisation offers participants to become familiar with the exercise protocol, whereas habituation is determined when performance does not change after subsequent visits to the laboratory and can be determined via statistical assessment (*e.g.*, the difference between consecutive tests is very small). It is unclear how many familiarisation sessions are required to attain habituation to an exercise protocol, and will be protocol and participant specific, but this would substantially increase study costs and the number of laboratory visits required for the participant. Nonetheless, we consider it essential that at least one familiarisation is performed prior to initiating the main interventions. There

doi: 10.51224/SRXIV.379

are exceptions wherein it may be appropriate not to include a specific familiarisation protocol. This would be specific to when then the sample population being studied is already familiar with the exercise being undertaken. For example, it is common for rowers to perform regular 2000 m rowing tests on a rowing ergometer. Similarly, professional football players will likely perform several YoYo Intermittent Recovery Tests throughout a season to determine exercise capacity. In these situations, it would be appropriate to forgo a specific familiarisation session and simply report that the athletes are well acquainted with the exercise test undertaken.

Personal view: Our (BS, FM) research laboratory generally aims to include two familiarisation sessions to any exercise protocol to ensure participants are well familiarised to the exercise protocol. This is what was required of participants in our study on caffeine supplementation and exercise performance (Saunders, de Oliveira, et al., 2017), in which trained cyclists performed two familiarisation trials of a simulated time-trial before the main intervention session.

4. *Time of day*

Several aspects of exercise performance appear to be influenced by the time of day at which they are measured, including strength (Grgic et al., 2019) and endurance (Küüsmaa et al., 2016) exercise, with afternoon and evening performance generally superior to that in the morning. Since the time of day at which individuals exercise can influence exercise performance, when participants attend the laboratory for data collection, researchers should strive to ensure that tests are performed at the same time of day for each participant. Although it may be desirable for all participants to perform exercise when performance appears to be optimised, it is highly improbable that all studies can perform data collection during this very limited late afternoon/early evening timeframe. As such, while it appears unnecessary to require all participants in a study to perform exercise at the same time of day (unless this is a specific aim of the study), each participant should attend the laboratory for data collection within a study at the same time according to their own schedule. Once a participant has performed their first visit, all subsequent visits should then be performed at the same or a similar time to avoid potential influence of circadian variation on the outcome measures.

Personal view: In our laboratory, we aim to have participants attend the lab at the best time of day that suits them. This might be early morning for some, or late evening for others. For

 doi: 10.51224/SRXIV.379

example, in our study on caffeine and exercise performance (Saunders, de Oliveira, et al., 2017), most participants favoured a morning (06:00 – 08:00) or evening (18:00 – 20:00) start due to their working day; this also coincided with their usual training hours. All visits were subsequently performed within a ±1 h period of the initial visit for each participant, since it was impossible to always begin at exactly the same time.

5. Dietary control

A person's diet strongly influences their health (Willett, 1994) and exercise performance (Burke & Hawley, 2018). Therefore, it is crucial to monitor or control dietary intake of participants enrolled to the study. While it is common to criticize the lack of control over participants' diet or the way in which such control was carried out, generic criticisms stem from the false belief that all studies should approach dietary control in the same way. To reflect on this, the researcher should not assume that dietary control must be done, but rather evaluate whether there is a need for it and, if so, how to implement it. To develop a good experimental design, there must be clarity with respect to the main research question, namely what will be evaluated, and what the primary outcome (dependent variable *e.g.*, $\dot{V}O_{2max}$, power output, force) is.

Once researchers have determined if monitoring or controlling diet in the study is truly necessary, the next step is to determine how to do it. It is crucial that the way dietary data is collected and evaluated is valid and appropriate for the study aims. Many options exist including the duplicate diet approach, food consumption recording, 24-h dietary recall, dietary record, dietary history and food frequency questionnaires. Detailing each of these is beyond the scope of the current chapter but those wishing to obtain more specific information about each of these dietary assessment methods are directed towards further reading (Shim et al., 2014; Thompson & Subar, 2017). Where possible, dietary assessment should be performed by the same experienced nutritionist to minimise errors and variation, although some errors between actual and estimated/calculated dietary intake are always likely (Stables et al., 2021). From this point, the researcher should aim to determine whether diet should be monitored, replicated or intervened.

a) Monitoring: In this situation, the researcher does not control the participant's diet in any way, but simply measures it via one of several methods available to monitor the quality,

doi: 10.51224/SRXIV.379

composition, or a specific bioactive compound. For example, a study that aims to evaluate carbohydrate consumption in the week leading up to a sports competition may ask a volunteer to record their food consumption via daily food diaries. Or a study that aims to determine the dietary habits and nutritional status of a distinct group of athletes (*e.g.*, endurance runners or CrossFit® athletes). A consideration here is the observer effect; participants may actively make different choices throughout the study to appear healthier or to be more knowledgeable about food choices, meaning the data may not be an accurate representation of their true diet.

b) Replication: Participants should be requested to maintain their normal dietary intake and avoid major changes throughout their participation in a study. In situations where changes in diet may cause unwanted changes in the primary outcome, participants should be requested to replicate their diet for a period of between 24-72 h. For example, during a crossover study aiming to determine whether sodium bicarbonate supplementation is ergogenic during a 100-km time-trial on a cycle ergometer, it is possible that carbohydrate intake (and other nutrients) impacts performance, which is the primary outcome for the study. As a result, dietary replication may be advisable before every visit so that this does not influence performance and differences can be attributed to the intervention and not to differences in diet. Replication could occur via one of two ways. Firstly, participants could record their dietary intake during the prespecified period (*e.g.*, 24-72 h pre-test) prior to the first main test, and then be asked to repeat this as closely as possible prior to each subsequent visit. The second option would be to provide participants with pre-prepared food prior to each main test. The former option may be more favourable for studies that do not have funds for food purchases but is reliant on participants repeating their food choices closely which may not always be done. The second option certainly provides more study control since the participants are instructed to eat the food provided by the researchers.

c) Intervention: This related to when the diet is the independent variable, meaning it is the intervention itself. For example, a study that aims to investigate whether a ketogenic diet impacts the performance of rowers in a 2000-m rowing test compared to a carbohydrate-rich diet. Ideally, since the diet is the intervention, strict control over the diet is desired and all food is provided to the participants. Unfortunately, we do not live in an ideal world and many studies would not have the resources to provide this, and thus dietary advice would likely be provided to participants while dietary monitoring would occur throughout the study to ensure

 doi: 10.51224/SRXIV.379

participants are adhering to their respective diets. The frequency and method (see below) via which this information is obtained will depend upon the researchers. Some studies may be more mechanistic and acute, for example investigating whether carbohydrate ingestion alongside beta-alanine supplementation aids in the entry of beta-alanine into the muscle. In this case, participants can be provided with a standardised carbohydrate-rich meal with and without beta-alanine on separate occasions to determine whether there are differences in muscle levels of beta-alanine. In this context, it is necessary that the provided meal is standardised according to carbohydrate (and other nutrients) content.

Sport science studies often prohibit certain foods and drinks in the day(s) prior to exercise tests, including alcohol and caffeine, to avoid any influence on exercise performance. Alcohol can negatively impact performance (Shirreffs & Maughan, 2006), and while caffeine can positively influence exercise performance (Guest et al., 2021), the quantities found in coffee can vary up to 100% even when the same quantities and brewing methods are applied (Desbrow et al., 2012; Desbrow et al., 2007; McCusker et al., 2003). Therefore, it makes sense to ensure participants do not ingest these prior to their laboratory visits as they may interfere with the outcomes of the study. Since carbohydrate intake is known to impact endurance performance (Bergström et al., 1967; Jensen et al., 2020), it may be desirable to monitor or control for this in the lead up to an exercise task. since it is known that this can impact upon endurance performance. Similarly, a debated topic is whether research participants should perform exercise in a fasted or fed state. As with most of these factors, the choice should depend upon the primary aims of the study. If the aim of a study is to determine whether nitrate could be a useful pre-exercise supplement to improve 16-km cycling time-trial performance in competition, then it makes sense to have participants consume a pre-exercise diet that the participant would regularly have. However, if the study is mechanistic in nature, such as whether nitrate supplementation increases the rate of oxygen consumption during 16-km time-trial cycling, then researchers may wish to have participants exercise in a fasted state as an easy method of dietary control. Nonetheless, results of such a study may not be entirely applicable to a real-world scenario where athletes are likely to ingest a pre-exercise meal.

Personal view: In our laboratory, we perform studies with dietary supplements to determine their influence on exercise performance. In these studies, we try to be as applicable to the real-world as possible, and generally simply ask volunteers to maintain their normal dietary

 doi: 10.51224/SRXIV.379

patterns throughout their participation in the study. Since diet can influence exercise performance, we request that participants record their dietary intake in the 24-h prior to the first intervention session and ask them to replicate this as closely as possible prior to the subsequent sessions. The participants are still required to perform 24-h dietary records prior to these subsequent sessions so that we can analyse how closely these were followed.

5. Exercise control

1. *Prior to main sessions*

A key component to a sport and exercise research study is the control of exercise prior to experimental trials, which is important since this may have negative or positive effects on the outcome of the experimental trial. Specifically, exercise close to an experimental trial may lead to carry-over fatigue, which could impact exercise performance when a best performance is required. This is common when the participants studied, for example, are triathletes, who are reported to train at least once per day (Korkia et al., 1994). The solution would be to allow exercise prior to an experimental trial, however, ensure that this is standardised and recorded. As previously discussed, (see Section 3.1.2. Field-based research), commercial software can assist with checking adherence to this approach (*e.g.*, Strava^{TM}). This would be stronger than attempting to make certain populations refrain from exercise 24-48 hours prior to a trial when this is highly unlikely in practice. Monitoring exercise can also allow the researcher to prescribe the exercise, such as the intensity and volume that would minimise the impact on the experimental trial. For example, if the aim of a study was to investigate the changes in muscle glycogen during a 3-hour simulated time-trial, the researcher could instruct participants to only complete exercise that will not deplete glycogen stores in the 24-48 hours prior to the experimental trial which will help minimise the impact of this on the 3-hour simulated time-trial.

2. *Throughout short-term studies*

The longer the duration of involvement in a study, the longer biological variability might influence outcome measures. Biological variability is defined as "non-intervention related processes that cause true scores to change" (Swinton et al., 2018). Parallel group designs somewhat account for this, whereby a separate group of participants complete the trial under control conditions (*i.e.*, without the intervention). However, crossover designs, such as those using acute supplements such as caffeine, do not. In this instance, it is recommended that

 doi: 10.51224/SRXIV.379

participants complete all main trials in as short a time period as is feasible to avoid substantial changes in biological variability. In previous work (Gough, Deb, et al., 2019; Gough et al., 2018), participants completed the study within a three-week window to minimise the impact of training adaptations, which are typically studied (or periodised) over an 8–12-week period (Solli et al., 2019). Using a short time frame of approximately 2-4 weeks should allow for the influence of training adaptations to be minimal. Of course, this approach also needs to be balanced with the time frame between each experimental trial. Generally, a time frame of between 2-3 days between experiments trials has been used (Gough, Deb, et al., 2019; Gough et al., 2018) in dietary supplement studies, and this ensures that sufficient recovery is provided for the physiological systems to reach homeostasis bearing in mind both the influence of the exercise and the supplement (Siegler et al., 2012; Stanley et al., 2013). A caveat to this would be the exercise task employed in the study. If the study involves longer duration exercise, such as running a half marathon or full marathon, then a longer period of recovery may be required. However, for longer duration exercise a parallel-group design is usually preferred when there are either carryover effects or repeated bout effects (Bacchieri & Della Cioppa, 2007). Additionally, it is important to note, albeit anecdotally, that participants consenting to research can often see the research study as a chance to change other elements of their behaviour such as nutrition and training (*i.e.*, to begin a health kick). It could also have the opposite effect, whereby participants feel because they are being healthy in the study they can be unhealthy outside of it (*i.e.*, a licensing effect) (Chiou et al., 2011). This makes it vital at the outset to explain to participants that the intervention is not intended to support this and that other than what the intervention intends to change, all else should remain consistent (other than typical daily variation).

3. *Throughout longer-term studies*

With advances in technology, it is now possible to monitor factors such as physical activity, sleep, and training, whereby the latter can even be controlled (or prescribed) for long-term intervention studies. In the example of a 12-week training study, training monitoring can be completed using applications such as Strava™ and TrainingPeaks™. Due to the autonomous nature of commercial applications, there is no longer a need to rely on written logs that can also increase the level or error compared to commercial applications that track work completed through global positioning satellites (GPS), power meters or heart rate, although these can still have small error themselves (Rampinini et al., 2015). These platforms, however, only cover a

 doi: 10.51224/SRXIV.379

few sports such as running and cycling and rely on expensive equipment (*e.g.*, power meter). In other sports, written training diaries might be a more practical method through which to monitor external influences over a long-term study due to the incompatibility of commercial applications (*e.g.*, swimming). The use of written logs may be a benefit to the study to help reduce participant attrition as reflection can lead to better adherence of the experimental procedures (Pirotta et al., 2019), although the opposite might also be expected due to more time being dedicated to the study. In respect of that point, strategies to reduce the amount of participant attrition is vital in research, as the procedures are usually logistically difficult and time consuming. Equally, it can lead to issues of internal and external validity through those dropping out from the research would change the outcome of the study (*i.e.*, negative response), yet would not be included in analysis (dropouts are typically excluded) (Barry, 2005). There is a statistical concept called intention-to-treat analysis that suggests including every participant that was randomised to a treatment group or order in the analysis, regardless of incomplete data, and more reading on this can be done elsewhere (Gupta, 2011). To counter the problem of dropouts, researchers may also wish to consider financial incentives and/or frequent reporting points to complete studies that are long term as this has been shown to increase participant adherence (Pirotta et al., 2019). Researchers should report, as a minimum, how many participants were initially recruited and how many dropped out, and best practice would be to attempt to identify why the participants dropped out. If the dropout was due to the intervention than this should be discussed and interpreted to reduce internal validity issues.

Personal view: In a study investigating 4-weeks of beta-alanine supplementation on cycling performance in trained cyclists (Perim et al., 2022), we wanted to ensure that potential changes in training did not influence our results. To do this, we monitored participant's training volumes for 4 weeks prior to supplementation, and during the 4 weeks of supplementation, and compared the two to ensure there were no differences. This was done using the participant's own GPS of preference, the data from which was uploaded to Strava™ from where we could have access to all the information regarding training.

 doi: 10.51224/SRXIV.379

4. DATA COLLECTION

1. Equipment

Equipment used for data collection should be calibrated according to standards or manufacturer recommendations prior to every use. It is recommended that researchers understand what "normal" values are expected for whatever measurement they are making so that they can immediately identify whether an equipment reading is off. It is always worth keeping records of calibration values as these can be a good way to check if the equipment is working correctly and provides an audit trail for accreditation purposes (*e.g.*, BASES laboratory accreditation). It is important to note that researchers should aim to use the same exercise equipment, not just the same make or model, during repeat testing as there may be subtle variability in outcomes. From our own experience, we found that two different exercise ergometers of the same make and model reported differences of~3%, which is large enough to mask any changes after administering an intervention. This applies to field-based research as well as the laboratory. For example, if a running test is performed on a grass surface, ensure all subsequent tests are performed on the same surface so that changes in performance are not influenced by different floor surfaces.

2. Exercise protocol

1. Exercise protocol validity

The type of exercise protocol that is chosen in a research study is important and can depend upon the specific aims of the study. Sometimes the choice is straightforward, for example, if the aim is to determine the efficacy of caffeine on 100-m sprint performance, then the exercise test should be a 100-m sprint (Matsumura et al., 2022). However, this choice is not always as easy, for example, if the aim is to investigate the effect of beta-alanine on football (soccer) performance as performance during such activities are numerous and difficult to measure (*e.g.*, it can be difficult to determine what a performance improvement in soccer is). Often, researchers will develop a test that replicates the demands of the activity, which in the case of football is the YoYo Intermittent Recovery Test (Krustrup et al., 2003), a running test consisting of 2 x 20 m runs which 10 s active recovery until exhaustion. Such a protocol should resemble performance during the activity that it is attempting to simulate as closely as possible, an aspect called validity, though there are many types of validity with further reading suggested (Currell & Jeukendrup, 2008). YoYo Intermittent Recovery Test performance is strongly correlated to running trends during match play (Krustrup et al., 2003; Krustrup et al., 2006)

 doi: 10.51224/SRXIV.379

making it a good surrogate for match performance. Since Saunders et al. (2012) showed a positive effect of beta-alanine supplementation on YoYo Intermittent Recovery test performance, this can then be extrapolated to suggest that beta-alanine may be effective for in-match football performance.

Sometimes the choice of an exercise test is to determine the underpinning mechanisms of an intervention. For example, Hill et al. (2007) developed a high intensity cycling capacity test that is performed until exhaustion and limited by muscle acidosis. This makes it an excellent model to determine whether increased muscle buffering capacity (which delays acidosis), achieved via beta-alanine supplementation, can improve performance during exercise limited by acidosis. They showed not that beta-alanine is effective for a sport-specific exercise, but that it can improve performance during exercise limited by muscle acidosis.

Researchers using exercise capacity tests, in which participants perform exercise at a fixed intensity until no longer tolerable (also called a time-to-exhaustion protocol), are often criticised for not considering the ecologically validity of the test (*i.e.*, they do not necessarily replicate a real-world situation). This is particularly true for supramaximal intensities in which the participant is instructed to exercise at an intensity well above their usual maximum, meaning that they will fatigue rapidly. Nonetheless, in addition to providing potential mechanistic insights, for many athletes trying to maintain race pace with the leader, this is a true reflection in an applied setting. For example, in road cycling, an end sprint on a climb would likely be supramaximal and close to a time-to-exhaustion test since the athlete will aim to be exert themselves maximally and aim to be completely depleted by the finish line. Thus, knowing how long they could realistically maintain such a high intensity, and how this might be improved, could provide valuable information.

Some studies in sport science evaluate measures of performance or fatigue during exercise to determine how this differs between, for example, sex or ability (McKay et al., 2022). This could be achieved by using specific exercise protocols replicating real-world competition such as a 100-m running sprint or a 4-km cycling time-trial. It is natural to question whether laboratory measurement of a particular sporting activity truly represents the physiological demands of competition, but studies do exist showing that they may not be different. One study showed that physiological responses to a 5-km cycling time-trial were not different when

doi: 10.51224/SRXIV.379

measured in the laboratory or during a competition (Foster et al., 1993). Some exercise protocols have been developed to measure a specific component of exercise capacity. For example, the 30-s cycling Wingate test, in which participants cycle maximally (all-out) against a fixed resistance for 30 s, was developed to measure muscular power and anaerobic exercise capacity (Bar-Or, 1987; Bar-Or et al., 1977). This test can then subsequently be used to determine differences in anaerobic capacity between athletic groups (*e.g.*, endurance vs. sprint cyclists) or whether a nutritional intervention can improve anaerobic capacity (*e.g.*, sodium bicarbonate supplementation).

Personal view: In a study performed by our laboratory, we supplemented participants with beta-alanine for 24-weeks to see how much muscle carnosine could be increased and whether improvements in exercise performance followed suit (Saunders, Painelli, et al., 2017). We used the high intensity cycling capacity test employed by Hill et al. (2007) because they had previously shown it to be limited by muscle acidosis and improved by 4 weeks of beta-alanine supplementation making it an appropriate model for our study. The aim was not to determine whether beta-alanine improved a specific sport, but how closely performance improvements mimicked muscle changes.

2. *Exercise protocol reliability*

The reliability of an exercise protocol is an important consideration, particularly when considering that many intervention effects may be small. For example, supplementation effects are generally 1-3% (Carr et al., 2011; Hobson et al., 2012). It is, therefore, key that the day-to-day variability in performance during the exercise test is minimal, as it may render the test unable to detect intervention changes. Test-retest studies typically have participants perform the same exercise test on two separate occasions, usually following at least one familiarisation, and under the same strict controlled conditions. The performance difference between sessions is then calculated using metrics such as the coefficient of variation (CV), Pearson’s correlation, intraclass correlation 95% limits of agreement or typical error (for more reading see (Currell & Jeukendrup, 2008; Hopkins, 2000; Swinton et al., 2018)). The CV is considered an appropriate statistic, easy to interpret as it is expressed as a percentage since it uses the standard deviation as a percentage of the mean, and allows easy comparison between different exercise protocols (Currell & Jeukendrup, 2008). The higher the CV value, the greater the variation between one visit and the next, which is undesirable. Though there is no specific cut-off limit,

 doi: 10.51224/SRXIV.379

CVs above 10% are often considered too high rendering the test inadequate. Such high CVs are generally seen in time-to-exhaustion exercise capacity tests performed at low intensities (Currell & Jeukendrup, 2008; Jeukendrup et al., 1996), though high-intensity capacity tests often show more suitable CVs below 10% (Higgins et al., 2014; Saunders et al., 2013). Time-trial tests generally show excellent reliability (<5%) (Currell & Jeukendrup, 2008; Jeukendrup et al., 1996) meaning they are often the preferred choice for intervention studies. Researchers should also be aware that training status positively influences test-retest reliability (Benton et al., 2013), meaning that less trained participants may exhibit higher variability than is desired. Clinical populations may also show different consistency in performance dependent upon their disorder and the exercise test being employed. Anyone initiating data collection should be aware of the reliability of the exercise protocol being used and the expected changes with the intervention under investigation so appropriate decision-making can be made. Furthermore, protocols with large variability may explain equivocal results in some intervention studies, for example, in which the variation of the exercise protocol will likely have masked the small effect of a dietary supplement.

Personal view: We previously sought to employ a time-to-exhaustion cycling protocol performed at 75% of peak power output to determine the effects of caffeine supplementation on performance. Pilot testing with a handful of cyclists revealed a day-to-day variation of approximately 30%, similar to that shown by Jeukendrup et al. (1996), which led us to choose a time-trial protocol with a smaller variation of ~3% (Oliveira et al., 2017).

3. Blood sampling

Blood samples are often taken in sport science studies to determine a plethora of measures depending upon the aims of the study. Some blood analytes can be measured almost immediately using standard laboratory equipment, such as blood lactate concentration or pH. Other compounds, such as markers of muscle damage or stress (*e.g.*, creatine kinase or lactate dehydrogenase) may be more complex and require blood samples to be collected and adequately stored (see Sample management below) for posterior analysis using intricate analytical techniques and equipment. As with most factors, there are a number of considerations to be addressed regarding blood sampling, the most important being which blood parameters are being analysed as this will affect where blood will be sampled from (*e.g.*, the arm, finger, earlobe), the type needed (venous vs. arterial vs. arterialised) and the amount

doi: 10.51224/SRXIV.379

required. Sampling at different sites may lead to different values for certain measures. For example, many studies may choose to measure blood lactate from the ear (for example, during rowing exercise), but researchers should be aware that results are not directly comparable to those obtained from the fingertip (Feliu et al., 1999). Participant posture can also influence the measurement of many clinical blood measurements depending on whether the participant is in a seated vs. standing position (Lima-Oliveira et al., 2017; Lippi et al., 2015). The type of blood collected may also modify the measure in question, for example, venous blood provides lower glucagon-like peptide-1 concentrations than arterialized blood in the postprandial (*i.e.*, fed) state (Chen et al., 2018). Nonetheless, since many sport science studies are unlikely to take place in a hospital, venous or capillary blood samples are usually preferable. This may not always be an issue, as in the example of blood pH and bicarbonate, which shows high levels of agreement whether sampled from venous or arterial blood (Ayaz et al., 2021; Kelly et al., 2004), meaning venous blood is an acceptable substitute for arterial for these measurements. To avoid any unwanted variability in blood sampling, researchers should aim to always take blood samples from the same site (which should be researched and chosen based upon the study aims and accessibility) with the participants in the same position (standing, seated or supine). Anecdotally, researchers may wish to familiarize their participants to blood sampling since a fear of needles may artificially increase blood lactate or glucose levels, though this fear is likely to subside after multiple exposures.

Personal view: In our studies, we often take venous blood samples with participants seated on a bike. To ensure sampling differences are not encountered due to postural differences, despite cannulation occurring in a supine position, we then sample blood with participants in a seated position.

4. Sample and data management

Participant information in research should be confidential, to ensure that the identities and their associated information is protected. Researchers must follow ethical guidelines to ensure that the data collected is handled appropriately and with respect for participants' privacy so it cannot be linked to specific participants. One way to maintain confidentiality is by assigning a unique identifier to each participant, as opposed to directly using their name or other identifying information. This unique identifier is then used in all future data, samples or notes relating to that particular participant, while the identifying information linking the

 doi: 10.51224/SRXIV.379

participant to the unique identifier should be kept separate and safe. Some studies may require complete anonymity to protect participant privacy, particularly when sensitive topics are being studied. This may involve removing any identifying information from participant records and using coding systems or anonymous questionnaires to collect data.

1. *Sample storage*

Biological samples such as blood, muscle, sweat, saliva, or other such samples are sensitive materials with potential risk of and for contamination, meaning they need to be handled and stored with the utmost care. Many countries may have governing policies on this type of collection with specific regulations that researchers must adhere to. One example is the United Kingdom with the Human Tissue Act (2004) (https://www.hta.gov.uk/guidance-professionals/hta-legislation/human-tissue-act-2004). Samples should be put into appropriate containers and properly labelled with information containing the unique identifier of the participant, the specific moment of collection and potentially the study to which they belong (e.g., CAF001BS, V1A; this might refer to a specific caffeine study [CAF], participant 001 with initials BS, Visit 1 [V1] and the first timepoint of data collection [A]). Samples can then be organised into larger airtight containers such as freezer boxes or plastic bags which are subsequently stored at the appropriate temperature other specific conditions to prevent degradation or contamination. The ideal storage conditions will vary depending on the type of sample and analysis to be performed but are often stored at -20°C or -80°C for long-term storage. Organisation of samples within a freezer or similar (*e.g.*, liquid nitrogen) should be detailed in an inventory using a computerized tracking system or manual logbook, and access should be restricted to authorized personnel only. Samples should be stored until analysed and then disposed of correctly (*i.e.*, according to university or company guidelines regarding disposal of contaminated samples).

2. *Data management*

Data management should be considered a critical component of research as it ensures that the information collected is accurate, reliable, and easily accessible. Laboratory books are an essential tool for researchers to document their experimental methods, observations, and results. Tabulation of data is an important step following data collection as it allows researchers to organize and analyse their data more effectively. It is recommended that researchers tabulate their data immediately (into Excel, for example) following a collection session to avoid losing

 doi: 10.51224/SRXIV.379

data. This can also help the researchers to evaluate whether there is any issue in data collection by visually inspecting whether data appear normal. Data can then be backed up to secure online storage networks or to portable drives to ensure that it is saved to multiple locations in the (hopefully unlikely) event that a laboratory book is lost, file becomes corrupted, or somebody steals your computer (a favourite excuse of a final year undergraduate student to gain more time). Online storage networks, such as OneDrive or DropBox, may be particularly favourable since they allow remote access from any device. Researchers should also take care to store participant data securely, such as in a locked cabinet or password-protected computer file. Storage and maintenance of data for an appropriate period are necessary to ensure that the data can be accessed and reviewed for future research or audits. The length of this retention period varies depending on the type of data, funding requirements, and the research area but is often considered to be 5 years for sport and exercise science.

Personal view: Each student in our laboratory has their own laboratory book in which they are to write down all their results and are strongly encouraged to extract any data file immediately and back it up, tabulate all data as soon as possible and back it up to an online server. The laboratory book should also be stored in a secure location.

doi: 10.51224/SRXIV.379

References

Abt, G., Boreham, C., Davison, G., Jackson, R., Nevill, A., Wallace, E., & Williams, M. (2020). Power, precision, and sample size estimation in sport and exercise science research. *J Sports Sci*, *38*(17), 1933-1935. https://doi.org/10.1080/02640414.2020.1776002

Abt, G., Jobson, S., Morin, J. B., Passfield, L., Sampaio, J., Sunderland, C., & Twist, C. (2022). Raising the bar in sports performance research. *J Sports Sci*, *40*(2), 125-129. https://doi.org/10.1080/02640414.2021.2024334

Andrade, C. (2018). Internal, External, and Ecological Validity in Research Design, Conduct, and Evaluation. *Indian J Psychol Med*, *40*(5), 498-499. https://doi.org/10.4103/ijpsym.Ijpsym_334_18

Andreacci, J. L., Lemura, L. M., Cohen, S. L., Urbansky, E. A., Chelland, S. A., & Duvillard, S. P. v. (2002). The effects of frequency of encouragement on performance during maximal exercise testing. *Journal of Sports Sciences*, *20*(4), 345-352. https://doi.org/10.1080/026404102753576125

Ayaz, F., Furrukh, M., Arif, T., Ur Rahman, F., & Ambreen, S. (2021). Correlation of Arterial and Venous pH and Bicarbonate in Patients With Renal Failure. *Cureus*, *13*(11), e19519. https://doi.org/10.7759/cureus.19519

Bacchieri, A., & Della Cioppa, G. (2007). Experimental Design: Fallacy of "Before-After" Comparisons in Uncontrolled Studies. In A. Bacchieri & G. Della Cioppa (Eds.), *Fundamentals of Clinical Research: Bridging Medicine, Statistics and Operations* (pp. 183-199). Springer Milan. https://doi.org/10.1007/978-88-470-0492-4_8

Bang, H., Flaherty, S. P., Kolahi, J., & Park, J. (2010). Blinding assessment in clinical trials: A review of statistical methods and a proposal of blinding assessment protocol. *Clinical Research and Regulatory Affairs*, *27*(2), 42-51. https://doi.org/10.3109/10601331003777444

Bang, H., Ni, L., & Davis, C. E. (2004). Assessment of blinding in clinical trials. *Control Clin Trials*, *25*(2), 143-156. https://doi.org/10.1016/j.cct.2003.10.016

Bar-Or, O. (1987). The Wingate anaerobic test. An update on methodology, reliability and validity. *Sports Medicine*, *4*(6), 381-394. https://doi.org/10.2165/00007256-198704060-00001

 doi: 10.51224/SRXIV.379

Bar-Or, O., Dotan, R., & Inbar, O. (1977). A 30-second all-out ergometric test - its reliability and validity for anaerobic capacity. *A 30-second all-out ergometric test - its reliability and validity for anaerobic capacity.*, *13*, 326.

Barry, A. E. (2005). How attrition impacts the internal and external validity of longitudinal research. *J Sch Health*, *75*(7), 267-270. https://doi.org/10.1111/j.1746-1561.2005.00035.x

Bennett, S., Tiollier, E., Brocherie, F., Owens, D. J., Morton, J. P., & Louis, J. (2021). Three weeks of a home-based “sleep low-train low” intervention improves functional threshold power in trained cyclists: A feasibility study. *Plos One*, *16*(12), e0260959. https://doi.org/10.1371/journal.pone.0260959

Benton, M. J., Raab, S., & Waggener, G. T. (2013). Effect of training status on reliability of one repetition maximum testing in women. *J Strength Cond Res*, *27*(7), 1885-1890. https://doi.org/10.1519/JSC.0b013e3182752d4a

Bergström, J., Hermansen, L., Hultman, E., & Saltin, B. (1967). Diet, muscle glycogen and physical performance. *Acta Physiol Scand*, *71*(2), 140-150. https://doi.org/10.1111/j.1748-1716.1967.tb03720.x

Betts, J. A., Gonzalez, J. T., Burke, L. M., Close, G. L., Garthe, I., James, L. J., Jeukendrup, A. E., Morton, J. P., Nieman, D. C., Peeling, P., Phillips, S. M., Stellingwerff, T., van Loon, L. J. C., Williams, C., Woolf, K., Maughan, R., & Atkinson, G. (2020). PRESENT 2020: Text Expanding on the Checklist for Proper Reporting of Evidence in Sport and Exercise Nutrition Trials. *International Journal of Sport Nutrition and Exercise Metabolism*, *30*(1), 2-13. https://doi.org/10.1123/ijsnem.2019-0326

Binboğa, E., Tok, S., Catikkas, F., Guven, S., & Dane, S. (2013). The effects of verbal encouragement and conscientiousness on maximal voluntary contraction of the triceps surae muscle in elite athletes. *J Sports Sci*, *31*(9), 982-988. https://doi.org/10.1080/02640414.2012.758869

Bonaventura, J. M., Sharpe, K., Knight, E., Fuller, K. L., Tanner, R. K., & Gore, C. J. (2015). Reliability and accuracy of six hand-held blood lactate analysers. *J Sports Sci Med*, *14*(1), 203-214.

Burke, L. M., & Hawley, J. A. (2018). Swifter, higher, stronger: What's on the menu? *Science*, *362*(6416), 781-787. https://doi.org/10.1126/science.aau2093

Carr, A. J., Hopkins, W. G., & Gore, C. J. (2011). Effects of acute alkalosis and acidosis on performance: a meta-analysis. *Sports Medicine*, *41*(10), 801-814. https://doi.org/10.2165/11591440-000000000-00000

Chen, Y.-C., Edinburgh, R. M., Hengist, A., Smith, H. A., Walhin, J.-P., Betts, J. A., Thompson, D., & Gonzalez, J. T. (2018). Venous blood provides lower glucagon-like peptide-1 concentrations than arterialized blood in the postprandial but not the fasted state: Consequences of sampling methods. *Experimental Physiology*, *103*(9), 1200-1205. https://doi.org/https://doi.org/10.1113/EP087118

Chiou, W.-B., Yang, C.-C., & Wan, C.-S. (2011). Ironic Effects of Dietary Supplementation:Illusory Invulnerability Created by Taking Dietary Supplements Licenses Health-Risk Behaviors. *Psychological Science*, *22*(8), 1081-1086. https://doi.org/10.1177/0956797611416253

Currell, K., & Jeukendrup, A. E. (2008). Validity, Reliability and Sensitivity of Measures of Sporting Performance. *Sports Medicine*, *38*(4), 297-316. https://doi.org/10.2165/00007256-200838040-00003

Desbrow, B., Henry, M., & Scheelings, P. (2012). An examination of consumer exposure to caffeine from commercial coffee and coffee-flavoured milk. *Journal of Food Composition and Analysis*, *28*(2), 114-118. https://doi.org/https://doi.org/10.1016/j.jfca.2012.09.001

Desbrow, B., Hughes, R., Leveritt, M., & Scheelings, P. (2007). An examination of consumer exposure to caffeine from retail coffee outlets. *Food Chem Toxicol*, *45*(9), 1588-1592. https://doi.org/10.1016/j.fct.2007.02.020

Feliu, J., Ventura, J. L., Segura, R., Rodas, G., Riera, J., Estruch, A., Zamora, A., & Capdevila, L. (1999). Differences between lactate concentration of samples from ear lobe and the finger tip. *J Physiol Biochem*, *55*(4), 333-339.

Foster, C., Green, M. A., Snyder, A. C., & Thompson, N. N. (1993). Physiological responses during simulated competition. *Med Sci Sports Exerc*, *25*(7), 877-882. https://doi.org/10.1249/00005768-199307000-00018

Gilchrist, M., Winyard, P. G., Fulford, J., Anning, C., Shore, A. C., & Benjamin, N. (2014). Dietary nitrate supplementation improves reaction time in type 2 diabetes: development and application of a novel nitrate-depleted beetroot juice placebo. *Nitric Oxide*, *40*, 67-74.

doi: 10.51224/SRXIV.379

Gough, L. A., Deb, S. K., Brown, D., Sparks, S. A., & McNaughton, L. R. (2019). The effects of sodium bicarbonate ingestion on cycling performance and acid base balance recovery in acute normobaric hypoxia. *J Sports Sci*, *37*(13), 1464-1471. https://doi.org/10.1080/02640414.2019.1568173

Gough, L. A., Deb, S. K., Sparks, S. A., & McNaughton, L. R. (2018). Sodium bicarbonate improves 4 km time trial cycling performance when individualised to time to peak blood bicarbonate in trained male cyclists. *J Sports Sci*, *36*(15), 1705-1712. https://doi.org/10.1080/02640414.2017.1410875

Gough, L. A., Rimmer, S., Sparks, S. A., McNaughton, L. R., & Higgins, M. F. (2019). Post-exercise Supplementation of Sodium Bicarbonate Improves Acid Base Balance Recovery and Subsequent High-Intensity Boxing Specific Performance. *Frontiers in Nutrition*, *6*, 155. https://doi.org/10.3389/fnut.2019.00155

Grgic, J., Grgic, I., Pickering, C., Schoenfeld, B. J., Bishop, D. J., & Pedisic, Z. (2020). Wake up and smell the coffee: caffeine supplementation and exercise performance—an umbrella review of 21 published meta-analyses. *British Journal of Sports Medicine*, *54*(11), 681-688.

Grgic, J., Lazinica, B., Garofolini, A., Schoenfeld, B. J., Saner, N. J., & Mikulic, P. (2019). The effects of time of day-specific resistance training on adaptations in skeletal muscle hypertrophy and muscle strength: A systematic review and meta-analysis. *Chronobiology International*, *36*(4), 449-460. https://doi.org/10.1080/07420528.2019.1567524

Guerin, B. (1986). Mere presence effects in humans: A review. *Journal of Experimental Social Psychology*, *22*(1), 38-77. https://doi.org/https://doi.org/10.1016/0022-1031(86)90040-5

Guest, N. S., VanDusseldorp, T. A., Nelson, M. T., Grgic, J., Schoenfeld, B. J., Jenkins, N. D. M., Arent, S. M., Antonio, J., Stout, J. R., Trexler, E. T., Smith-Ryan, A. E., Goldstein, E. R., Kalman, D. S., & Campbell, B. I. (2021). International society of sports nutrition position stand: caffeine and exercise performance. *J Int Soc Sports Nutr*, *18*(1), 1. https://doi.org/10.1186/s12970-020-00383-4

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspect Clin Res*, *2*(3), 109-112. https://doi.org/10.4103/2229-3485.83221

doi: 10.51224/SRXIV.379

Gurton, W. H., Matta, G. G., Gough, L. A., & Hurst, P. (2022). Efficacy of sodium bicarbonate ingestion strategies for protecting blinding. *European Journal of Applied Physiology*, *122*(12), 2555-2563.

Halperin, I., Pyne, D. B., & Martin, D. T. (2015). Threats to internal validity in exercise science: a review of overlooked confounding variables. *Int J Sports Physiol Perform*, *10*(7), 823-829. https://doi.org/10.1123/ijspp.2014-0566

Halperin, I., Ramsay, E., Philpott, B., Obolski, U., & Behm, D. G. (2020). The effects of positive and negative verbal feedback on repeated force production. *Physiol Behav*, *225*, 113086. https://doi.org/10.1016/j.physbeh.2020.113086

Halperin, I., Vigotsky, A. D., Foster, C., & Pyne, D. B. (2018). Strengthening the Practice of Exercise and Sport-Science Research. *Int J Sports Physiol Perform*, *13*(2), 127-134. https://doi.org/10.1123/ijspp.2017-0322

Harriss, D., Jones, C., & MacSween, A. (2022). Ethical standards in sport and exercise science research: 2022 update. *International Journal of Sports Medicine*, *43*(13), 1065-1070.

Higgins, M. F., James, R. S., & Price, M. J. (2014). Familiarisation to and reproducibility of cycling at 110% peak power output. *J Sports Med Phys Fitness*, *54*(2), 139-146.

Hill, C. A., Harris, R. C., Kim, H. J., Harris, B. D., Sale, C., Boobis, L. H., Kim, C. K., & Wise, J. A. (2007). Influence of beta-alanine supplementation on skeletal muscle carnosine concentrations and high intensity cycling capacity. *Amino Acids*, *32*(2), 225-233. https://doi.org/10.1007/s00726-006-0364-4

Hobson, R. M., Saunders, B., Ball, G., Harris, R., & Sale, C. (2012). Effects of β-alanine supplementation on exercise performance: a meta-analysis. *Amino Acids*, *43*(1), 25-37. https://doi.org/10.1007/s00726-011-1200-z

Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*(1), 1-15. https://doi.org/10.2165/00007256-200030010-00001

Hurst, P., Saunders, S., & Coleman, D. (2020). No differences between beetroot juice and placebo on competitive 5-km running performance: A double-blind, placebo-controlled trial. *International Journal of Sport Nutrition and Exercise Metabolism*, *30*(4), 295-300.

Hurst, P., Schiphof-Godart, L., Hettinga, F., Roelands, B., & Beedie, C. (2020). Improved 1000-m running performance and pacing strategy with caffeine and placebo: a balanced placebo design study. *International Journal of Sports Physiology and Performance*.

Jensen, R., Ørtenblad, N., Stausholm, M. H., Skjaerbaek, M. C., Larsen, D. N., Hansen, M., Holmberg, H. C., Plomgaard, P., & Nielsen, J. (2020). Heterogeneity in subcellular muscle glycogen utilisation during exercise impacts endurance capacity in men. *The Journal of Physiology*, *598*(19), 4271-4292. https://doi.org/10.1113/jp280247

Jeukendrup, A., Saris, W. H., Brouns, F., & Kester, A. D. (1996). A new validated endurance performance test. *Med Sci Sports Exerc*, *28*(2), 266-270. https://doi.org/10.1097/00005768-199602000-00017

Kang, M., Ragan, B. G., & Park, J. H. (2008). Issues in outcomes research: an overview of randomization techniques for clinical trials. *J Athl Train*, *43*(2), 215-221. https://doi.org/10.4085/1062-6050-43.2.215

Kelly, A. M., McAlpine, R., & Kyle, E. (2004). Agreement between bicarbonate measured on arterial and venous blood gases. *Emerg Med Australas*, *16*(5-6), 407-409. https://doi.org/10.1111/j.1742-6723.2004.00642.x

Korkia, P. K., Tunstall-Pedoe, D. S., & Maffulli, N. (1994). An epidemiological investigation of training and injury patterns in British triathletes. *Br J Sports Med*, *28*(3), 191-196. https://doi.org/10.1136/bjsm.28.3.191

Krustrup, P., Mohr, M., Amstrup, T., Rysgaard, T., Johansen, J., Steensberg, A., Pedersen, P. K., & Bangsbo, J. (2003). The yo-yo intermittent recovery test: physiological response, reliability, and validity. *Med Sci Sports Exerc*, *35*(4), 697-705. https://doi.org/10.1249/01.Mss.0000058441.94520.32

Krustrup, P., Mohr, M., Nybo, L., Jensen, J. M., Nielsen, J. J., & Bangsbo, J. (2006). The Yo-Yo IR2 test: physiological response, reliability, and application to elite soccer. *Med Sci Sports Exerc*, *38*(9), 1666-1673. https://doi.org/10.1249/01.mss.0000227538.20799.08

Küüsmaa, M., Schumann, M., Sedliak, M., Kraemer, W. J., Newton, R. U., Malinen, J.-P., Nyman, K., Häkkinen, A., & Häkkinen, K. (2016). Effects of morning versus evening combined strength and endurance training on physical performance, muscle hypertrophy, and serum hormone concentrations. *Applied Physiology, Nutrition, and Metabolism*, *41*(12), 1285-1294. https://doi.org/10.1139/apnm-2016-0271 %M 27863207

doi: 10.51224/SRXIV.379

Léger, L. A., & Lambert, J. (1982). A maximal multistage 20-m shuttle run test to predict VO2 max. *Eur J Appl Physiol Occup Physiol*, *49*(1), 1-12. https://doi.org/10.1007/bf00428958

Lima-Oliveira, G., Guidi, G. C., Salvagno, G. L., Danese, E., Montagnana, M., & Lippi, G. (2017). Patient posture for blood collection by venipuncture: recall for standardization after 28 years. *Rev Bras Hematol Hemoter*, *39*(2), 127-132. https://doi.org/10.1016/j.bjhh.2017.01.004

Lippi, G., Salvagno, G. L., Lima-Oliveira, G., Brocco, G., Danese, E., & Guidi, G. C. (2015). Postural change during venous blood collection is a major source of bias in clinical chemistry testing. *Clinica Chimica Acta*, *440*, 164-168. https://doi.org/https://doi.org/10.1016/j.cca.2014.11.024

Matsumura, T., Tomoo, K., Sugimoto, T., Tsukamoto, H., Shinohara, Y., Otsuka, M., & Hashimoto, T. (2022). Acute Effect of Caffeine Supplementation on 100-m Sprint Running Performance: A Field Test. *Med Sci Sports Exerc*. https://doi.org/10.1249/mss.0000000000003057

Matta, G., Edwards, A., Roelands, B., Hettinga, F., & Hurst, P. (2022). Reproducibility of 20-min Time-trial Performance on a Virtual Cycling Platform. *Int J Sports Med*, *43*(14), 1190-1195. https://doi.org/10.1055/a-1848-8478

McCusker, R. R., Goldberger, B. A., & Cone, E. J. (2003). Caffeine content of specialty coffees. *J Anal Toxicol*, *27*(7), 520-522. https://doi.org/10.1093/jat/27.7.520

McKay, A. K. A., Stellingwerff, T., Smith, E. S., Martin, D. T., Mujika, I., Goosey-Tolfrey, V. L., Sheppard, J., & Burke, L. M. (2022). Defining Training and Performance Caliber: A Participant Classification Framework. *Int J Sports Physiol Perform*, *17*(2), 317-331. https://doi.org/10.1123/ijspp.2021-0451

McNaughton, L. R. (1992). Bicarbonate ingestion: Effects of dosage on 60 s cycle ergometry. *Journal of Sports Sciences*, *10*(5), 415-423. https://doi.org/10.1080/02640419208729940

Mesquida, C., Murphy, J., Lakens, D., & Warne, J. (2022). Replication concerns in sports and exercise science: a narrative review of selected methodological issues in the field. *Royal Society Open Science*, *9*(12), 220946. https://doi.org/doi:10.1098/rsos.220946

Midgley, A. W., Marchant, D. C., & Levy, A. R. (2018). A call to action towards an evidence-based approach to using verbal encouragement during maximal exercise

doi: 10.51224/SRXIV.379

testing. *Clinical Physiology and Functional Imaging*, *38*(4), 547-553. https://doi.org/https://doi.org/10.1111/cpf.12454

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: investigating participant demand characteristics. *J Gen Psychol*, *135*(2), 151-165. https://doi.org/10.3200/genp.135.2.151-166

Oliveira, L. F. d., Yamaguchi, G., Painelli, V. S. d., Silva, R. P. d., Gonçalves, L. S., Gualano, B., & Saunders, B. (2017). Comprehensive reliability analysis of a 16 km simulated cycling time-trial in well-trained individuals. *Journal of Science and Cycling*, *6*(1), 11-17.

Paliczka, V. J., Nichols, A. K., & Boreham, C. A. (1987). A multi-stage shuttle run as a predictor of running performance and maximal oxygen uptake in adults. *Br J Sports Med*, *21*(4), 163-165. https://doi.org/10.1136/bjsm.21.4.163

Perim, P., Gobbi, N., Duarte, B., Oliveira, L. F. d., Costa, L. A. R., Sale, C., Gualano, B., Dolan, E., & Saunders, B. (2022). Beta-alanine did not improve high-intensity performance throughout simulated road cycling. *European Journal of Sport Science*, *22*(8), 1240-1249. https://doi.org/10.1080/17461391.2021.1940304

Pirotta, S., Joham, A., Hochberg, L., Moran, L., Lim, S., Hindle, A., & Brennan, L. (2019). Strategies to reduce attrition in weight loss interventions: A systematic review and meta-analysis. *Obes Rev*, *20*(10), 1400-1412. https://doi.org/10.1111/obr.12914

Rampinini, E., Alberti, G., Fiorenza, M., Riggio, M., Sassi, R., Borges, T. O., & Coutts, A. J. (2015). Accuracy of GPS devices for measuring high-intensity running in field-based team sports. *Int J Sports Med*, *36*(1), 49-53. https://doi.org/10.1055/s-0034-1385866

Ramsbottom, R., Brewer, J., & Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *Br J Sports Med*, *22*(4), 141-144. https://doi.org/10.1136/bjsm.22.4.141

Saunders, B., de Oliveira, L. F., da Silva, R. P., de Salles Painelli, V., Goncalves, L. S., Yamaguchi, G., Mutti, T., Maciel, E., Roschel, H., Artioli, G. G., & Gualano, B. (2017). Placebo in sports nutrition: a proof-of-principle study involving caffeine supplementation. *Scand J Med Sci Sports*, *27*(11), 1240-1247. https://doi.org/10.1111/sms.12793

Saunders, B., Painelli, V. S., LF, D. E. O., V, D. A. E. S., RP, D. A. S., Riani, L., Franchi, M., Goncalves, L. S., Harris, R. C., Roschel, H., Artioli, G. G., Sale, C., & Gualano, B. (2017). Twenty-four Weeks of beta-Alanine Supplementation on Carnosine

doi: 10.51224/SRXIV.379

Content, Related Genes, and Exercise. *Med Sci Sports Exerc*, *49*(5), 896-906. https://doi.org/10.1249/MSS.0000000000001173

Saunders, B., Sale, C., Harris, R. C., Morris, J. G., & Sunderland, C. (2013). Reliability of a high-intensity cycling capacity test. *J Sci Med Sport*, *16*(3), 286-289. https://doi.org/10.1016/j.jsams.2012.07.004

Saunders, B., Sale, C., Harris, R. C., & Sunderland, C. (2014). Sodium bicarbonate and high-intensity-cycling capacity: variability in responses. *Int J Sports Physiol Perform*, *9*(4), 627-632. https://doi.org/10.1123/ijspp.2013-0295

Saunders, B., Sunderland, C., Harris, R. C., & Sale, C. (2012). β-alanine supplementation improves YoYo intermittent recovery test performance. *Journal of the International Society of Sports Nutrition*, *9*(1), 39. https://doi.org/10.1186/1550-2783-9-39

Schulz, K. F., & Grimes, D. A. (2002). Generation of allocation sequences in randomised trials: chance, not choice. *The Lancet*, *359*(9305), 515-519. https://doi.org/10.1016/S0140-6736(02)07683-3

Shim, J. S., Oh, K., & Kim, H. C. (2014). Dietary assessment methods in epidemiologic studies. *Epidemiol Health*, *36*, e2014009. https://doi.org/10.4178/epih/e2014009

Shirreffs, S. M., & Maughan, R. J. (2006). The effect of alcohol on athletic performance. *Curr Sports Med Rep*, *5*(4), 192-196. https://doi.org/10.1007/s11932-006-0046-8

Siegler, J. C., Marshall, P. W., Bray, J., & Towlson, C. (2012). Sodium bicarbonate supplementation and ingestion timing: does it matter? *J Strength Cond Res*, *26*(7), 1953-1958. https://doi.org/10.1519/JSC.0b013e3182392960

Solli, G. S., Tønnessen, E., & Sandbakk, Ø. (2019). Block vs. Traditional Periodization of HIT: Two Different Paths to Success for the World's Best Cross-Country Skier. *Front Physiol*, *10*, 375. https://doi.org/10.3389/fphys.2019.00375

Souza, H. L. R., Bernardes, B. P., Prazeres, E. O. d., Arriel, R. A., Meireles, A., Camilo, G. B., Mota, G. R., & Marocolo, M. (2022). Hoping for the best, prepared for the worst: can we perform remote data collection in sport sciences? *Journal of Applied Physiology*, *133*(6), 1430-1432. https://doi.org/10.1152/japplphysiol.00196.2022

Stables, R. G., Kasper, A. M., Sparks, S. A., Morton, J. P., & Close, G. L. (2021). An Assessment of the Validity of the Remote Food Photography Method (Termed Snap-N-Send) in Experienced and Inexperienced Sport Nutritionists. *Int J Sport Nutr Exerc Metab*, *31*(2), 125-134. https://doi.org/10.1123/ijsnem.2020-0216

doi: 10.51224/SRXIV.379

Stanley, J., Peake, J. M., & Buchheit, M. (2013). Cardiac parasympathetic reactivation following exercise: implications for training prescription. *Sports Medicine*, *43*(12), 1259-1277. https://doi.org/10.1007/s40279-013-0083-4

Stein, J. A., Ramirez, M., & Heinrich, K. M. (2020a). Acute Caffeine Supplementation Does Not Improve Performance in Trained CrossFit((R)) Athletes. *Sports (Basel)*, *8*(4). https://doi.org/10.3390/sports8040054

Stein, J. A., Ramirez, M., & Heinrich, K. M. (2020b). Retraction: Stein, J.A. et al. The Effects of Acute Caffeine Supplementation on Performance in Trained CrossFit Athletes. Sports 2019, 7, 95. *Sports (Basel)*, *8*(2), 24. https://www.mdpi.com/2075-4663/8/2/24

Stevens, C. J., & Dascombe, B. J. (2015). The reliability and validity of protocols for the assessment of endurance sports performance: an updated review. *Measurement in Physical Education and Exercise Science*, *19*(4), 177-185.

Støren, Ø., Helgerud, J., Sæbø, M., Støa, E. M., Bratland-Sanda, S., Unhjem, R. J., Hoff, J., & Wang, E. (2017). The Effect of Age on the V˙O2max Response to High-Intensity Interval Training. *Med Sci Sports Exerc*, *49*(1), 78-85. https://doi.org/10.1249/mss.0000000000001070

Strauss, B. (2002). Social facilitation in motor tasks: a review of research and theory. *Psychology of Sport and Exercise*, *3*(3), 237-256. https://doi.org/https://doi.org/10.1016/S1469-0292(01)00019-X

Suresh, K. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*, *4*(1), 8-11. https://doi.org/10.4103/0974-1208.82352

Swinton, P. A., Hemingway, B. S., Saunders, B., Gualano, B., & Dolan, E. (2018). A Statistical Framework to Interpret Individual Response to Intervention: Paving the Way for Personalized Nutrition and Exercise Prescription. *Frontiers in Nutrition*, *5*, 41. https://doi.org/10.3389/fnut.2018.00041

Tanner, R. K., Fuller, K. L., & Ross, M. L. (2010). Evaluation of three portable blood lactate analysers: Lactate Pro, Lactate Scout and Lactate Plus. *European journal of applied physiology*, *109*(3), 551-559. https://doi.org/10.1007/s00421-010-1379-9

Thompson, F. E., & Subar, A. F. (2017). Chapter 1 - Dietary Assessment Methodology. In A. M. Coulston, C. J. Boushey, M. G. Ferruzzi, & L. M. Delahanty (Eds.), *Nutrition in*

 doi: 10.51224/SRXIV.379

the Prevention and Treatment of Disease (Fourth Edition) (pp. 5-48). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-802928-2.00001-1

UKRI. (2023). *Research with potentially vulnerable people*. https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/research-with-potentially-vulnerable-people/

van der Meij, L., Buunk, A. P., van de Sande, J. P., & Salvador, A. (2008). The presence of a woman increases testosterone in aggressive dominant men. *Horm Behav*, *54*(5), 640-644. https://doi.org/10.1016/j.yhbeh.2008.07.001

Weakley, J., Wilson, K., Till, K., Banyard, H., Dyson, J., Phibbs, P., Read, D., & Jones, B. (2020). Show Me, Tell Me, Encourage Me: The Effect of Different Forms of Feedback on Resistance Training Performance. *J Strength Cond Res*, *34*(11), 3157-3163. https://doi.org/10.1519/jsc.0000000000002887

Willett, W. C. (1994). Diet and Health: What Should We Eat? *Science*, *264*(5158), 532-537. https://doi.org/doi:10.1126/science.8160011

Winchester, R., Turner, L. A., Thomas, K., Ansley, L., Thompson, K. G., Micklewright, D., & St Clair Gibson, A. (2012). Observer effects on the rating of perceived exertion and affect during exercise in recreationally active males. *Percept Mot Skills*, *115*(1), 213-227. https://doi.org/10.2466/25.07.05.Pms.115.4.213-227