

Target trial framework for determining the effect of changes in training load on injury risk using observational data: a methodological commentary

Chinchin Wang ^{1,2}, Jay S Kaufman,¹ Russell J Steele,³ Ian Shrier ²

To cite: Wang C, Kaufman JS, Steele RJ, *et al*. Target trial framework for determining the effect of changes in training load on injury risk using observational data: a methodological commentary. *BMJ Open Sport & Exercise Medicine* 2024;**10**:e002037. doi:10.1136/bmjsem-2024-002037

Accepted 25 June 2024

ABSTRACT

In recent years, a large focus has been placed on managing training load for injury prevention. To minimise injuries, training recommendations should be based on research that examines causal relationships between load and injury risk. While observational studies can be used to estimate causal effects, conventional methods to study the relationship between load and injury are prone to bias. The target trial framework is a valuable tool that requires researchers to emulate a hypothetical randomised trial using observational data. This framework helps to explicitly define research questions and design studies in a way that estimates causal effects. This article provides an overview of the components of the target trial framework as applied to studies on load and injury and describes various considerations that should be made in study design and analyses to minimise bias.

INTRODUCTION

Avoiding injury is an important goal for athletes of all sports and levels. Training load (also referred to as ‘load’ or ‘workload’) is considered an important risk factor for injury.^{1 2} Training load refers generally to a broad range of exposure variables related to sport or physical activity that can be manipulated to elicit a physiological response.^{3–5} For simplicity, the term ‘load’ will be used to refer to this concept henceforth. It is generally accepted that larger absolute loads are associated with higher injury risks.^{1 2} Mechanistically, this may occur through increased mechanical stress on tissues, increased fatigue affecting decision-making, coordination and/or neuromuscular control⁶ and increased exposure time at risk.⁷

In recent years, a large focus has been placed on the relationship between changes in load and injury. Gabbett and colleagues proposed an ‘acute-chronic workload ratio’ (ACWR) model to relate changes in load to injury based on Banister and colleagues’ fitness and fatigue performance model.^{6 8 9} In this model, athletes with similar acute loads

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ There is a large interest among athletes, coaches and clinicians in managing changes in training load for injury prevention.
- ⇒ Longitudinal data from observational studies or load monitoring programmes can provide valuable insights into the causal effect of changing load on injury risk, but data must be analysed appropriately.
- ⇒ The target trial framework is a tool for designing and analysing observational studies in a way that emulates a randomised controlled trial and minimises bias.

WHAT THIS STUDY ADDS

- ⇒ This review discusses considerations for applying the target trial framework to studies examining the causal effects of changes in load on injury risk.
- ⇒ We provide guidance for defining the research question, eligibility criteria, treatment strategies, and outcomes, and for conducting appropriate analyses.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ The application of the target trial framework in research can be used to generate valid recommendations to minimise injuries.
- ⇒ The insights outlined in this review can aid researchers in designing rigorous observational studies that estimate the causal effects of changing load on injury risk.

(causing fatigue) and chronic loads (proxy for fitness) are thought to be performing activity at a level that they are well prepared for, minimising injury risk, whereas athletes with high acute loads and low chronic loads are generally exceeding what they are prepared for, increasing injury risk.^{6 9} Athletes with low acute loads and high chronic loads are also thought to be at increased injury risk.^{6 10} Although no biological explanations were initially provided, it was later suggested that one’s past (chronic) load may promote physical adaptations (eg, tissue strengthening) that protect against injury.^{6 10} However, one’s



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

²Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Montreal, Canada

³Department of Mathematics and Statistics, McGill University, Montreal, Canada

Correspondence to

Dr Ian Shrier;
ian.shrier@mcgill.ca

recent (acute) load may cause fatigue and decrease tissue strength and mechanical stress capacity, increasing risk of injury.⁶ No biological explanations have been provided for the increased injury risk associated with low acute loads and high chronic loads (excluding a decrease in technical skill following rest periods for sports requiring high precision such as gymnastics),¹¹ and this finding is likely due to methodological flaws.^{12 13}

The monitoring of load to inform training decisions with the goal of reducing injury is now done across a variety of sport types and levels.¹⁴ Training recommendations largely depend on existing models resulting from observational studies.^{6 14} While randomised controlled trials (RCTs) are considered the gold standard for identifying causal relationships and evidence-based decision-making, they are not often feasible. RCTs generally require a large sample size and long follow-up, which is often impractical, especially in elite settings.¹⁵ As such, researchers often rely on observational data. However, none of the observational studies reported in existing systematic reviews have explicitly estimated a causal effect of changes in load on injury risk.^{1 2 16–18} Further, conventional methods used to study this relationship are prone to bias and are unlikely to correspond to true causal effects.¹²

Observational data can be used to estimate causal effects only if certain assumptions hold. Meaningful differences have been observed between results from observational studies with traditional study designs and those from RCTs, leading to concerns about their validity. However, some authors have shown that if the observational study design and analysis emulates a hypothetical randomised trial (called a ‘target trial’),¹⁹ the results are generally consistent with those from RCTs,^{20–24} although this is not always the case.^{25 26} We propose that this framework be applied to studies of load and injury risk to generate higher-quality evidence regarding their relationship.

The objective of this review is to describe the components of the target trial framework as applied to studies of load and injury risk, including potential biases and other challenges as well as strategies to address them.

TARGET TRIAL FRAMEWORK COMPONENTS

The target trial framework requires researchers to define their research question and study protocol in a way that mimics a hypothetical RCT and conduct their study analyses using observational data in a way that emulates that protocol.¹⁹ This process minimises errors and resulting biases that are common in observational analyses.

The major components of a target trial protocol are (1) eligibility criteria (population); (2) treatment strategies (intervention and comparison); (3) assignment procedures; (4) outcome; (5) follow-up period; (6) causal contrasts of interest and (7) analysis plan. These components have been described in further detail elsewhere.^{20 27} In this section, we outline these components and discuss specific considerations for studies of changes in load and injury risk.

Eligibility criteria

In an RCT, we would start by identifying our population of interest and specifying inclusion and exclusion criteria to determine eligible individuals. The same criteria should be used for an observational study. Eligibility should be determined at ‘time zero’, or the start of follow-up, and only using baseline information prior to the follow-up period.²⁸ If there are missing data on important baseline variables, results may not be meaningful given the potential for bias.

Defining the population of interest

In both RCTs and target trial emulations, the population should be defined by who we are interested in intervening. This may be a specific athletic population (eg, elite soccer players) or a general population (eg, youth). When studying general populations, we note that an intervention of a ‘change in load’ is likely to have different effects on different participants (effect heterogeneity). For instance, the same increase in load is expected to affect inactive individuals and regularly active individuals differently. This can promote generalisability; however, if we are interested in a specific subset of the population (eg, regularly active individuals) it may be appropriate to restrict our study population to those with certain baseline levels of activity measured over a run-in period, with participants only eligible for analyses following this period. Otherwise, we may explore heterogeneity using stratification or an interaction term between baseline activity and the intervention. Any subgroup analyses of primary interest (ie, not exploratory or hypothesis generating) should be considered in the sample size calculation.

We must also consider how our outcome of injury informs our population of interest. Previous injury is considered an important risk factor for new injuries.²⁹ In an RCT, we might restrict to healthy individuals (eg, those who are not currently injured or recovering from injury). We would include the same restrictions in an observational study. Data from a participant who is eligible at baseline are included until injury. Once recovered, data from the same participant would only be included once they are again eligible for the study, after several (eg, 4 or 5) weeks without injury.

Selection bias affecting internal validity

Individuals should not be included or excluded from analyses based on information gathered during follow-up. The selection of individuals based on factors that result from their intervention and outcome may cause bias through several mechanisms, affecting the internal validity of findings.³⁰ One example might be analyses that are restricted to those who attended a certain number of training sessions over the follow-up period. Participants who experience health problems (eg, illness, pain, mental health conditions) are less likely to participate in training,³¹ and health problems may be a consequence of changes in load and injury. Excluding participants based

on training participation during follow-up may, therefore, create bias in both RCTs and observational studies. Rather, alternative methods exist that address adherence/non-adherence to planned activity within RCTs (see discussion of per-protocol (PP) effects below).³² The same principles should be applied to observational studies.

Drop-outs and censoring

Drop-outs affect both RCTs and observational studies. Individuals who drop out or are lost to follow-up are considered censored, as their outcome (and potentially their intervention) is not observed.³³ Excluding censored individuals from analyses will result in selection bias when the reason for drop-out/loss to follow-up is associated with the intervention and outcome.³³ For instance, individuals who are less accustomed to activity and experience higher levels of discomfort or soreness from small increases in activity may be less motivated to remain in a study. Instead, censoring can be accounted for by imputing missing data,³⁴ or using inverse probability weighting, assuming that data are available on the covariates associated with drop-out.³³

Cross-over (nested) target trials

It is often the case with observational data that one individual will meet eligibility criteria at multiple time points. While we might allow an individual to be eligible only at one time point (the first time point or a random one), this does not make use of all available data. We can increase statistical efficiency and the effective sample size by allowing individuals to be eligible at multiple time points, creating a set of repeated cross-over trials (called 'nested' trials in the target trial literature).²⁰ These cross-over trials can overlap within individuals. For instance, if an individual were to be eligible on Monday, they would be followed up until the following Sunday. If they also met eligibility criteria on Tuesday, they would be followed up separately until the following Monday. In such a scenario, the measurement of load and injury (covered in subsequent sections) is not restricted to occurring within a Monday to Sunday calendar week. Employing a nested target trial approach requires accounting for repeated measures, as discussed in 'Analysis plan' section.

Treatment strategies

Most analyses of RCTs and observational studies compare two treatment strategies: an intervention and a comparison or control. In our context, the intervention is a change in load. Load has been operationalised in numerous ways³ and over various time frames.¹⁶ The optimal measure of load depends on the research context and available data. However, the same principles apply to defining treatment strategies regardless of the load metric.

Defining changes in load

The target trial framework prompts researchers to define treatment strategies that are relevant to stakeholders (eg, athletes, coaches, policy-makers) within the specific

sporting context. When defining a 'change in load', we must consider the baseline load, whether change is expressed as an absolute versus relative amount, and whether change is measured at a single time point or as a continuous intervention.

Measurements of change require a baseline or reference value. A simple measure of change in load might be a weekly change, or the change in load during the follow-up week (beginning at time zero) relative to the previous week. In this case, the baseline load would be the absolute load in the previous week. Other options might be an unweighted average load over multiple weeks (akin to chronic load within the ACWR framework), a weighted average or a cumulative measure. When deciding on a baseline load, researchers should consider any theories underlying the relevance of previous loads in affecting current injury risk, as well as utility for athletes, clinicians and other stakeholders. For instance, whereas large increases in load may increase susceptibility to injury, these increases are common after recovery or taper weeks which are thought to reduce injury risk.³⁵

We must also consider whether to express change as an absolute amount (eg, an hour more of training this week) or a relative amount (eg, 10% increase in running distance this week). We will distinguish between individual and policy-level interventions to illustrate this decision. Individuals are typically interested in how their injury risk may differ under different behaviours or patterns to inform their training decisions. For instance, a runner might ask questions like 'What is my injury risk if I increase my total distance covered by 5 km this week?' (absolute change) or 'What is my injury risk if I increase my total distance covered by 10% this week?' (relative change). The impact of changes in load on injury risk on an individual is expected to differ by their baseline fitness.³⁶ For instance, a 5 km increase in running distance is likely to result in a much greater injury risk for someone who regularly runs 5 km per week vs 50 km per week. Similarly, a 10% increase in distance may also result in differing injury risks for these two individuals, but perhaps not to the same extent as the absolute change. Policy-makers are interested in improving the health of an entire population. Policies are generally on an absolute scale, such as one where children within a school are mandated to take at least one physical education class,³⁷ or where youth community rugby players are allowed a maximum of 90 min of playing time per day.³⁸

Furthermore, we must decide whether we are interested in change at a single time point or as a continuous intervention. While a soccer team might be interested in increasing their practices by 1 hour in a single week (single time point), individuals training for a marathon might be interested in gradually increasing their running distance relative to their previous distance over several weeks (continuous intervention). Within continuous interventions, changes in load are not limited to an increase or a decrease. An intervention to decrease injury risk might incorporate maintenance weeks and recovery weeks (eg,



taper) where the load is unchanged or decreased. These weeks are not easily studied under single time point interventions, particularly when baseline load is measured as an average over several weeks such as in the ACWR framework.¹² Under a continuous framework, we might compare: (1) a tapering programme with a 10% increase in activity for 3 weeks followed by a 20km decrease in activity for 1 week prior to competition versus (2) a 10% increase in activity for 4 weeks prior to competition. Note that a continuous intervention can incorporate both absolute and relative changes in load.

Defining the comparison strategy

The comparison of two treatment strategies should reflect meaningful real-world decisions, such as a reasonable alternative behaviour/pattern/policy or one that is currently in place. For instance, a suitable comparison for a runner interested in increasing their total distance by 20% each week might be an increase in total distance by 10% each week, until a maximal distance is reached, while a suitable comparison for a soccer team wanting to include an extra hour of training moving forwards might be maintaining their current training schedule. A comparison for a policy mandating at least one physical education class per week might be to not have this mandate in place, allowing the population to participate in physical education as they choose.

To determine causal effects, ideally, all aspects of training would be maintained between the treatment strategy and comparator except for the aspect that is being intervened on. For instance, if we were interested in increasing training volume (eg, distance run), we would want to keep intensity (eg, pace) constant. This may not be feasible using observational data, and we may instead be limited to assessing the impact of increasing training volume on injury risk regardless of intensity. This is a limitation of using observational data compared with RCTs. At the same time, it is a strength of the target trial emulation approach because it makes these challenges more transparent compared with traditional observational approaches.

Thus far, we have only considered comparisons between two treatment strategies, with specific yet arbitrary values for changes in load. In practice, researchers may choose to dichotomise or categorise changes in load when defining their treatment strategies (eg, increase distance by 5–10km vs increase distance by 0–4km). These categorisations should be done in a way that reflects realistic training practices, rather than arbitrarily. Determining the effect of a continuous range of changes in load on injury risk is analogous to determining a dose-response curve. The development of a dose-response curve requires a single RCT with many arms or multiple RCTs. This remains true with the target trial emulation approach, and therefore, requires defining multiple comparison strategies and a more complex analytical strategy (covered in more detail in ‘Analysis plan’ section).

Consistency and positivity

Positivity and consistency are two conditions necessary for causal inference (along with exchangeability, covered in the following section).³⁹ Under positivity, each individual should theoretically have a positive probability of receiving each level of exposure for every combination of covariates.^{39–40} As such, each individual should be theoretically capable of changing their load by a specified amount, which may not be the case for large relative increases in load (eg, tripling training time in a day when someone is currently training for 8 hours per day). The treatment strategies should be realistic given the eligibility criteria for a study.

Briefly, consistency requires that treatments be defined unambiguously so that there cannot be two versions of a single treatment that would result in the same individual having different outcomes.^{39–41} In our context, this involves specifically defining what a change in load represents, including the type of activity, frequency, intensity and/or duration. Further, there cannot be interference, where an individual’s outcome depends on another individual’s treatment. In our context, one individual’s load should not affect another individual’s injury risk. Consistency is likely to be violated when there is a broad intervention, such as an increase in activity duration that does not account for intensity over a variety of sports. While this can be avoided by having more specific research questions, in reality, stakeholders may be interested in general recommendations. Researchers should aim to strike a balance between defining clear treatment strategies and generalisability.

Assignment procedures

Controlling for baseline confounders

In an RCT, treatments are assigned at random at baseline. This achieves exchangeability, one of the necessary conditions for causal inference, given a large enough sample size and perfect adherence to the assigned treatment strategy.³⁹ Simply, exchangeability means that there is no inherent difference in the risk of injury between treatment and control groups, and that any observed differences are due to the treatment itself. Under full exchangeability, the outcomes for the intervention group are the same as the outcomes for the control group had the control group received the intervention, and the outcomes for the control group are the same as the outcomes for the intervention group had the intervention group not received intervention, all else being equal.⁴²

Training decisions are rarely random in observational data. An individual’s magnitude of change in load may be influenced by factors such as sex, age, experience, baseline activity levels, planned strength and conditioning training, recent recovery or taper weeks, or previous injuries. These factors may also influence injury risk, and therefore, act as confounders. As full exchangeability requires that there be no unmeasured confounding,⁴² confounders must be adjusted for in observational

analyses through methods such as the inverse probability of treatment weighting,⁴³ multivariable regression or both (doubly robust estimation).⁴⁴

For a treatment strategy that occurs at a single time point (eg, increase in load in a single week), adjustment must only be done for factors measured at baseline. Adjustment for factors measured during follow-up affected by the treatment or outcome (eg, illness) may result in bias^{30 45} and decrease precision.⁴⁵ For treatment strategies that occur over a period of time (eg, consistently increasing load by 10% each week), there may be time-varying confounders that affect injury risk and subsequent changes in loads. One example is fatigue or soreness causing one to decrease their load. Time-varying confounding must be handled using specialised methods developed by Robins.³⁹

Timing of treatment assignment and immortal time bias

Treatment assignment, or the observational analogue of defining an individual's exposure, must be done at baseline to properly emulate a target trial. However, observational studies of changes in load and injury often only measure acute load at the end of the follow-up period. As such, any injury occurring during follow-up can affect one's measured load and cause a bias akin to immortal time bias in other fields of epidemiology.^{28 46} For instance, load may be measured as one's activity performed over a week. Athletes who get injured earlier in the calendar week will not be able to perform their planned activity for the rest of the week and will have systematically lower loads than athletes who get injured later in the week (figure 1) or complete the week without injury.¹² The same principles apply when daily averages are used to calculate loads, but with reduced bias.¹²

Researchers sometimes impose an injury lag period, in which only injuries occurring in a specified time window (eg, 1 week) subsequent to the load window will be

attributed to that load.¹⁶ In this setting, treatment assignment would occur at the beginning of the follow-up period, defined as the week following the load window. This eliminates the bias explained above but ignores the principle that current load is the inciting factor for injury and assumes that the load between the end of the load window and the time of injury is not relevant.¹² Alternatively, researchers may use planned loads rather than observed loads to calculate changes in load and estimate an intention-to-treat (ITT) effect of changes in load on activity. ITT effects are discussed further under 'Causal contrasts of interest' section.

Outcome

A well-designed study requires a clear definition of the outcome. Injury can be defined in many ways. Common categorisations include any athlete-reported complaint, medical attention injuries and/or time-loss injuries.⁴⁷ The onset of injury might be defined at the time of first complaint, initiation of time lost from sport or at the time of medical diagnosis.

Multiple injuries

Injuries can and often do occur more than once in the same individual, and one's risk of subsequent injury may be affected by previous injuries.²⁹ Furthermore, injuries often influence one's subsequent activity patterns. As such, previous or current injuries are a confounder for the relationship between changes in load and injury and must be accounted for in study design or analyses.

Previous or current injuries at the start of follow-up can be adjusted for as baseline confounders in observational studies. These might be included as dichotomous variables (eg, yes/no injury in the previous \times months) or continuous variables (eg, number of injuries in the previous \times months). However, most RCTs would only include healthy individuals as part of their eligibility

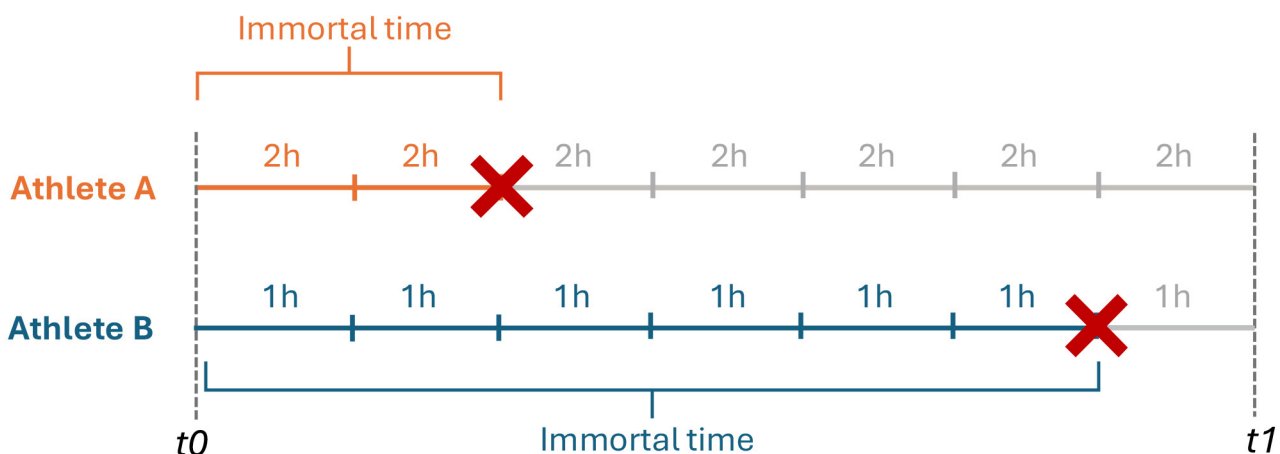


Figure 1 Immortal time bias in the measurement of load. Loads (measured as duration) are indicated for days 1–7 during the calendar week. Observed loads are indicated in orange for athlete A, and blue for athlete B. Planned loads that were not observed due to injury are indicated in grey. Injury is represented by a red X. Follow-up for injury starts at time zero (t_0 , beginning of the calendar week) and ends at t_1 (end of the calendar week). Load is assessed at t_1 . Despite having planned a larger load and having been exposed to a larger load up to the point of injury, a smaller load is observed for athlete A than athlete B who completed the week without injury. This creates a bias known as 'immortal time bias' in epidemiology.^{28 46}

criteria, excluding those who have returned to training but are not fully healed. We might emulate this criterion by only including individuals in our study up to their initial injury, after which they are no longer eligible. However, this would greatly reduce our effective sample size. Alternatively, we may believe that one's injury risk is unaffected by previous injuries after a certain time period (eg, 1 month). Similar to an RCT that might restrict to individuals who have not been injured in the past month, we can restrict our observational analyses to those who have been uninjured for 1 month prior to the start of follow-up. This is equivalent to a 'washout' period commonly employed in pharmacoepidemiology studies, where participants are observed for a period of time prior to follow-up to ensure that outcomes are not due to exposures that occurred prior to the study.^{48 49} However, if we are interested in a sustained intervention such as an increase in load over several time points, we must treat injuries occurring during follow-up as time-varying confounders and account for them using the appropriate methods.¹²

Finally, we may explore effect heterogeneity between initial and subsequent injuries through stratified analyses or by assessing interactions if relevant to our research question.

Causal contrasts of interest

Data from RCTs can be used to obtain an ITT or PP effect estimate.⁵⁰ Analogues of these effects can be estimated using observational data.²⁰

ITT versus PP effects

The ITT estimate addresses the question 'What is the effect of assigning a policy or intervention on injury?'. Participants are analysed in the group that they were assigned during randomisation, irrespective of the treatment they actually received, non-adherence or drop-out. This maintains exchangeability between groups assuming no drop-outs but will generally result in conservative effect estimates for the treatment actually received due to noncompliance.^{50 51} The ITT estimate may be of interest on a policy level because not everyone is expected to comply with policies or recommendations in real life.³² For instance, coaches or clinicians may be interested in ITT estimates because they prescribe training plans rather than follow them.

The PP estimate addresses the question 'What is the effect of a policy or intervention on injury if everyone adhered to the policy or intervention?'.³⁹ Traditional methods to estimate the PP effect include 'as-treated' analyses which compare participants based on the treatment they actually took, or 'naïve per protocol'/'on-treatment' analyses that are restricted to participants who followed their assigned treatment.⁵² These analyses are essentially observational, as individuals are able to choose their intervention. To properly estimate the PP effect, more sophisticated analyses with additional assumptions are required to adjust for confounding and non-adherence

to avoid bias, even in an RCT setting.^{39 52} For example, although the objective of a recent RCT was to estimate the ITT effect of providing a load management software programme on injury risk, the conclusion referred to 'managing training loads' (a PP effect).⁵³ Such a conclusion would require more assumptions, different analyses and higher-quality data. The PP estimate is generally of greater interest to individuals for informing decisions (eg, athletes trying to minimise injury risk).³⁹

The ITT and PP estimates will differ when there is non-adherence to treatment assignment. Non-adherence to a training plan or pattern may occur due to reasons such as injuries at baseline, fatigue, soreness, illness or motivation. Importantly, individuals who get injured during follow-up and stop training should be considered as having adhered to their treatment assignment so long as they were following their strategy up to the point of injury, as we would not expect injured participants to continue their regular training.

Estimating ITT effects using observational data

To estimate ITT effects using observational data, we must determine an individual's treatment assignment using their planned loads at baseline and adjust for baseline confounders related to their planned training. This is only feasible if planned loads such as a weekly training programme are available. We recommend that planned training schedules be collected in observational studies to allow ITT analyses to be conducted and to avoid immortal time bias as discussed above. Within team sports, participants generally have the same training schedule and planned loads. However, their baseline loads may differ due to absences, non-adherence, etc. The planned 'changes in load' should be based on each individual's observed baseline load. Further, because participants on the same team may have similar training schedules and propensities for injury, clustering by team should be accounted for in analyses.

Estimating PP effects in observational data

To estimate PP effects using observational data, we must compare individuals based on their actual activity patterns, as opposed to their planned training.

Above, we discussed how immortal time bias can occur if acute load is measured at the end of the follow-up period. This creates difficulties in estimating PP effects, as we are unable to obtain an unbiased measure of an individual's observed exposure or training. To estimate PP effects for a specific change in load at a single time point, we must impose an injury lag period and follow-up for injuries after the acute load is measured. For instance, we could define the outcome as injuries occurring within a day after the current week and assign treatments based on an individual's change in load for that week compared with the previous week. In a nested target trial approach, the follow-up period for injury would be on the following Monday for the trial where load was measured from Monday to Sunday, the

following Tuesday for the trial where load was measured from Tuesday to Monday, and so forth. However, this approach would ignore variations in load in the current day that might affect injury risk. Despite the advantages of target trial emulation, it does not solve the challenge of estimating PP effects for specific single time point interventions which may be of interest to athletes, coaches and clinicians.

PP effects can be estimated using a cloning and censoring approach when the treatment is categorical.²⁰ For instance, we might be interested in whether an increase in distance by 5–10 km increases injury risk compared with an increase in distance by 0–4 km. Under this approach, we would clone each individual in our analyses, assigning each clone to a different treatment at time zero (5–10 km vs 0–4 km). We would follow up each clone for injury, and censor the clone at the point that their observed load is no longer consistent with their treatment assignment. Such an analysis is only feasible for dichotomous treatments or treatments with few categories, as assessing injury risk for a continuous range of changes in load would result in an infinite number of clones.

For sustained interventions such as a consistent increase in load over several weeks, we must adjust for time-varying confounders related to non-adherence and injury using methods such as inverse probability weighting or g-estimation.⁵² Important confounders include fatigue and soreness, and this information should be collected in load and injury surveillance or studies to determine causal effects. Alternatively, if training schedules are available, planned training can be used as an instrumental variable to estimate the effect of changing load on injury in the presence of unmeasured confounding, providing the underlying assumptions are likely to hold true.^{12 52}

Analysis plan

Generally, the study analysis requires creating a statistical model that reflects the relationship between the exposure and outcome and estimating the effect of interest.⁵⁴

Pooling multiple trials

In RCTs, eligible participants are typically identified and randomised into one of two treatment groups at baseline or 'time zero'. In observational data, an individual may meet eligibility criteria at multiple time points. To increase the number of observations and effective sample size, we might allow individuals to contribute multiple trials or follow-up periods, given they meet eligibility requirements.²⁵ This is analogous to a repeated measures design in an RCT, where individuals participate in a trial multiple times.^{39 55} In both a repeated measures RCT and observational study, we would have to account for repeated measures in the analyses (eg, through cluster bootstrapping,⁵⁶ mixed models,^{57 58} or generalised estimating equations⁵⁹).

Estimating effects using observed versus predicted data

The majority of studies employing the target trial framework assign individuals to a treatment group consistent with their observed data. For instance, if we were interested in the PP effect on injury risk for an increase in load by twofold or more versus less than twofold, we would categorise each individual into a group based on their observed increase in load assessed at time zero. If an individual's observed exposure was compatible with multiple treatments at time zero, we could employ a cloning and censoring approach to minimise bias.²⁰

Treatment assignment using observed data becomes inefficient for treatments that are continuous variables. For instance, we may be interested in comparing injury risk for an increase in load by twofold compared with onefold. Any individual with an increase in load by a value other than twofold or onefold would be excluded from analyses, drastically reducing the sample size. Instead, we can employ marginal standardisation.^{60 61} Briefly, we create a model reflecting the relationship between continuous increases in load and injury risk (appropriately accounting for confounding, loss to follow-up, etc), and predict each individual's outcomes under different hypothetical treatments. In this scenario, we could include all eligible individuals in our predictive model, and predict whether or not they would become injured under either treatment (twofold increase vs onefold increase). We can then use these results to estimate the average treatment effect across the different treatments, with bootstrapping to calculate SEs and CIs.^{60 61}

CONCLUSION

To inform training recommendations and prevent injuries among athletes, we require evidence of the relationships between changes in load and injury. While observational data are often used in studying the relationship between changes in load and injury risk, conventional analytical approaches are prone to bias. The target trial framework is a valuable and simple tool to explicitly define causal questions and design studies to estimate causal effects using observational data. By applying this framework, we can strengthen the validity of future research in the sport medicine field. Although the target trial framework solves some of the challenges compared with current approaches, other challenges remain including isolating the effects of a single aspect of load, implementing ITT or instrumental variable analyses when planned loads are not available and limitations in estimating PP effects.

Contributors CW conceptualised and drafted the initial manuscript. JSK, RJS and IS critically reviewed and revised the manuscript. All authors provided final approval for submission. CW is responsible for the overall content as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Chinchin Wang <http://orcid.org/0000-0002-8472-9936>

Ian Shrier <http://orcid.org/0000-0001-9914-3498>

REFERENCES

- Drew MK, Finch CF. The relationship between training load and injury, illness and soreness: a systematic and literature review. *Sports Med* 2016;46:861–83.
- Eckard TG, Padua DA, Hearn DW, et al. The relationship between training load and injury in athletes: a systematic review. *Sports Med* 2018;48:1929–61.
- Impellizzeri FM, Marcora SM, Coutts AJ. Internal and external training load: 15 years on. *Int J Sports Physiol Perform* 2019;14:270–3.
- Lambert MI, Borresen J. Measuring training load in sports. *Int J Sports Physiol Perform* 2010;5:406–11.
- Coutts AJ, Crowcroft S, Kempton T. Developing athlete monitoring systems: theoretical basis and practical applications. In: *Sport, Recovery, and Performance*. Routledge, 2017.
- Soligard T, Schwelnus M, Alonso J-M, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med* 2016;50:1030–41.
- Kalkhoven JT, Watsford ML, Coutts AJ, et al. Training load and injury: causal pathways and future directions. *Sports Med* 2021;51:1137–50.
- Calvert TW, Banister EW, Savage MV, et al. A systems model of the effects of training on physical performance. *IEEE Trans Syst Man Cybern* 1976;SMC-6:94–102.
- Hulin BT, Gabbett TJ, Blanch P, et al. Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *Br J Sports Med* 2014;48:708–12.
- Gabbett TJ. The training–injury prevention paradox: should athletes be training smarter and harder. *Br J Sports Med* 2016;50:273–80.
- Orlando C, Levitan EB, Mittleman MA, et al. The effect of rest days on injury rates. *Scand J Med Sci Sports* 2011;21:e64–71.
- Wang C, Vargas JT, Stokes T, et al. Analyzing activity and injury: lessons learned from the acute:chronic workload ratio. *Sports Med* 2020;50:1243–54.
- Impellizzeri FM, Tenan MS, Kempton T, et al. Acute:chronic workload ratio: conceptual issues and fundamental pitfalls. *Int J Sports Physiol Perform* 2020;15:907–13.
- Bourdon PC, Cardinale M, Murray A, et al. Monitoring athlete training loads: consensus statement. *Int J Sports Physiol Perform* 2017;12:S2161–70.
- Wang C, Stokes T, Vargas JT, et al. Predicting injury risk over changes in physical activity in children using the acute:chronic workload ratio. *Am J Epidemiol* 2022;191:665–73.
- Andrade R, Wik EH, Rebelo-Marques A, et al. Is the acute:chronic workload ratio (ACWR) associated with risk of time-loss injury in professional team sports? A systematic review of methodology, variables and injury risk in practical situations. *Sports Med* 2020;50:1613–35.
- Jones CM, Griffiths PC, Mellalieu SD. Training load and fatigue marker associations with injury and illness: a systematic review of longitudinal studies. *Sports Med* 2017;47:943–74.
- Griffin A, Kenny IC, Comyns TM, et al. The association between the acute:chronic workload ratio and injury and its application in team sports: a systematic review. *Sports Med* 2020;50:561–80.
- Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA* 2022;328:2446–7.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
- Kutcher SA, Brophy JM, Banack HR, et al. Emulating a randomised controlled trial with observational data: an introduction to the target trial framework. *Can J Cardiol* 2021;37:1365–77.
- Boyne DJ, Cheung WY, Hilsden RJ, et al. Association of a shortened duration of adjuvant chemotherapy with overall survival among individuals with stage III colon cancer. *JAMA Netw Open* 2021;4:e213587.
- Cain LE, Saag MS, Petersen M, et al. Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. *Int J Epidemiol* 2016;45:2038–49.
- Wang SV, Schneeweiss S, RCT-DUPLICATE Initiative, et al. RCT-DUPLICATE initiative. emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA* 2023;329:1376–85.
- Danaei G, García Rodríguez LA, Cantero OF, et al. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol* 2018;96:12–22.
- Pearce N, Vandenbroucke JP. Are target trial emulations the gold standard for observational studies. *Epidemiology* 2023;34:614–8.
- Matthews AA, Danaei G, Islam N, et al. Target trial emulation: applying principles of randomised trials to observational studies. *BMJ* 2022;378:e071108.
- Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016;79:70–5.
- Fulton J, Wright K, Kelly M, et al. Injury risk is altered by previous injury: a systematic review of the literature and presentation of causative neuromuscular factors. *Int J Sports Phys Ther* 2014;9:583–95.
- Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol* 2010;39:417–20.
- Clarsen B, Bahr R, Myklebust G, et al. Improved reporting of overuse injuries and health problems in sport: an update of the Oslo sport trauma research center questionnaires. *Br J Sports Med* 2020;54:390–6.
- Shrier I, Steele RJ, Verhagen E, et al. Beyond intention to treat: what is the right question. *Clin Trials* 2014;11:28–37.
- Boston 677 Huntington Avenue, Ma 02115 +1495-1000. Causal Inference: What If (the book). Miguel Hernan's Faculty Website, 2012. Available: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/> [Accessed 6 Mar 2023].
- Altman DG. Missing outcomes in randomized trials: addressing the dilemma. *Open Med* 2009;3:e51–3.
- Mujika I, Padilla S. Scientific bases for precompetition tapering strategies. *Med Sci Sports Exerc* 2003;35:1182–7.
- Mohseni MM, Filmlater SE, Taylor WC, et al. Factors associated with half- and full-Marathon race-related injuries: a 3-year review. *Clin J Sports Med* 2021;31:e277–86.
- The Ontario curriculum, grades 9–12: health and physical education. 2015. Available: <https://www.publications.gov.on.ca/browse-catalogues/health-physical-education-curriculum/secondary/the-ontario-curriculum-grades-9-12-health-and-physical-education-2015> [Accessed 17 May 2024].
- 2023 Age grade law variations introduced across Canada. Rugby Canada; 2023. Available: <https://rugby.ca/en/news/2023/04/2023-age-grade-law-variations-introduced-across-canada> [Accessed 17 May 2024].
- Hernán M, Robins J. Causal inference: what if. Chapman & Hall/CRC; 2020. Available: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/> [Accessed 6 Oct 2021].
- Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology (Sunnyvale)* 2009;20:3–5.
- Rehkopf DH, Glymour MM, Osypuk TL. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Curr Epidemiol Rep* 2016;3:63–71.
- Sarvet AL, Wanis KN, Stensrud MJ, et al. A graphical description of partial exchangeability. *Epidemiology (Sunnyvale)* 2020;31:365–8.
- Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J* 2022;15:14–20.
- Funk MJ, Westreich D, Wiesen C, et al. Doubly robust estimation of causal effects. *Am J Epidemiol* 2011;173:761–7.
- Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology (Sunnyvale)* 2009;20:488–95.
- Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol* 2008;167:492–9.
- Bahr R, Clarsen B, Derman W, et al. International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sports 2020 (including the STROBE extension for sports injury and illness surveillance (STROBE-SIIS)). *Orthop J Sports Med* 2020;8:2325967120902908.
- Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Lippincott Williams & Wilkins, 2008.

- 49 Harvey RD, Mileham KF, Bhatnagar V, *et al.* Modernizing clinical trial eligibility criteria: recommendations of the ASCO-friends of cancer research Washout period and concomitant medication work group. *Clin Cancer Res* 2021;27:2400–7.
- 50 Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: intention-to-treat versus per-protocol analysis. *Perspect Clin Res* 2016;7:144–6.
- 51 Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res* 2011;2:109–12.
- 52 Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. *Clin Trials* 2012;9:48–55.
- 53 Dalen-Lorentsen T, Bjørneboe J, Clarsen B, *et al.* Does load management using the acute:chronic workload ratio prevent health problems? A cluster randomised trial of 482 elite youth Footballers of both sexes. *Br J Sports Med* 2021;55:108–14.
- 54 Petersen ML, van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology* 2014;25:418–26.
- 55 Singh V, Rana RK, Singhal R. Analysis of repeated measurement data in the clinical trials. *J Ayurveda Integr Med* 2013;4:77–81.
- 56 Huang FL. Using cluster Bootstrapping to analyze nested data with a few clusters. *Educ Psychol Meas* 2018;78:297–318.
- 57 Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990;46:673–87.
- 58 Wallace D, Green SB. Analysis of repeated measures designs with linear mixed models. In: *Modeling intraindividual variability with repeated measures data*. Psychology Press, 2001.
- 59 Ballinger GA. Using generalized estimating equations for longitudinal data analysis. *Organ Res Methods* 2004;7:127–50.
- 60 Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol* 2014;43:962–70.
- 61 Naimi AI, Whitcomb BW. Estimating risk ratios and risk differences using regression. *Am J Epidemiol* 2020;189:508–10.