

BMJ Open The Garmin-RUNSAFE Running Health Study on the aetiology of running-related injuries: rationale and design of an 18-month prospective cohort study including runners worldwide

Rasmus Østergaard Nielsen,¹ Michael Lejbach Bertelsen,¹ Daniel Ramskov,¹ Camma Damsted,¹ René Korsgaard Brund,² Erik Thorlund Parner,³ Henrik Sørensen,⁴ Sten Rasmussen,^{5,6} Søren Kjærgaard⁷

To cite: Nielsen RØ, Bertelsen ML, Ramskov D, *et al.* The Garmin-RUNSAFE Running Health Study on the aetiology of running-related injuries: rationale and design of an 18-month prospective cohort study including runners worldwide. *BMJ Open* 2019;9:e032627. doi:10.1136/bmjopen-2019-032627

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-032627>).

Received 27 June 2019
Revised 02 August 2019
Accepted 08 August 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Rasmus Østergaard Nielsen; roen@ph.au.dk

ABSTRACT

Introduction Running injuries affect millions of persons every year and have become a substantial public health issue owing to the popularity of running. To ensure adherence to running, it is important to prevent injuries and to have an in-depth understanding of the aetiology of running injuries. The main purpose of the present paper was to describe the design of a future prospective cohort study exploring if a dose–response relationship exists between changes in training load and running injury occurrence, and how this association is modified by other variables.

Methods and analysis In this protocol, the design of an 18-month observational prospective cohort study is described that will include a minimum of 20 000 consenting runners who upload their running data to Garmin Connect and volunteer to be a part of the study. The primary outcome is running-related injuries categorised into the following states: (1) no injury; (2) a problem; and (3) injury. The primary exposure is change in training load (eg, running distance and the cumulative training load based on the number of strides, ground contact time, vertical oscillation and body weight). The change in training load is a time-dependent exposure in the sense that progression or regression can change many times during follow-up. Effect-measure modifiers include, but is not limited to, other types of sports activity, activity of daily living and demographics, and are assessed through questionnaires and/or by Garmin devices.

Ethics and dissemination The study design, procedures and informed consent have been evaluated by the Ethics Committee of the Central Denmark Region (Request number: 227/2016 – Record number: 1-10-72-189-16).

INTRODUCTION

Physical activity should be taken seriously and performed regularly to decrease the risk of lifestyle diseases and to diminish the increased mortality risks associated with prolonged sitting.^{1–3} The scientific literature offers abundant justification that promotion

Strengths and limitations of this study

- Monitoring of the running activities and the running injury status of more than 20 000 runners for up to 18 months, which allows for in-depth analyses on the dose–response relationship between change in training load and injury occurrence using advanced time-to-event analyses on time-dependent exposures and outcomes.
- Use of a web-based solution that conforms to the General Data Protection Regulation regulative that allows for access to data from a very large cohort of runners.
- Self-reporting of injury status without a standardised examination procedure may lead to information problems, which could lead to overestimation or underestimation of the effect size.
- The design of the study will not allow all variables playing a role in the causal mechanism to injury to be quantified. The totality of relevant associations between exposures associated with injury risk will remain unexplored in the present study.

of regular physical activity should be prioritised worldwide as part of a comprehensive strategy to reduce non-communicable diseases.⁴ The popularity of running as a physical activity has risen much in the past 20 years, which is demonstrated, among others, in the growth in road-race events and the increased prevalence proportion of the population level indicating that they run on a regular basis.⁵ For novices, recreational and competitive runners alike, running is a time-efficient, easily accessed and relatively inexpensive activity.⁶ From a health-related perspective, running provides beneficial effects on body mass, body fat, resting heart rate, VO₂max, triglycerides and the cholesterol level.⁷ The longer the training period,

the larger the achieved health benefits.⁷ Based on the considerable interest in running as a preferred type of physical activity, it seems crucial to ensure adherence to running by shedding light on the barriers to continued running.

Busyness, childcare, illness and running-related injuries are well-known barriers that may lead to a temporary or permanent cessation of running activities.⁸ Of these barriers, running-related injuries have become a considerable public health issue because such injuries affect millions of persons every year.^{9 10} Although some minor running injuries pass quickly, others are associated with a long rehabilitation period exceeding 12 months.¹¹ This is problematic, since running injuries were the most common reason for permanently dropping out of a running regime among men, and the third most common reason among women according to a 10-year prospective cohort study.⁸ To ensure adherence to running, attention to and prevention of running injuries is highly important.

Prevention of running injuries has been extensively studied and discussed in the scientific literature.^{12–15} To effectively prevent injuries, an in-depth understanding of the aetiology of running injury is needed.¹⁶ For decades, running injury researchers have sought to identify risk factors for injury, and the number of publications on this topic has risen steeply in the past 10 years.¹⁷ Despite the great research efforts made in previous studies,¹⁸ it is currently impossible to draw any definitive conclusions on the aetiology of running injuries, and hence to guide runners to structure their running activity to minimise their injury risk. Study-specific drawbacks relating to previous studies include: (1) limited sample sizes,¹⁹ (2) retrospective study designs,²⁰ (3) a lack of objective measures of running activity²¹ and (4) a limited number of variables included in the statistical analyses.²² In addition, a shift from traditional risk factor identification towards identifying the strength and the proximity of risk factors within a causal framework for injury aetiology has been proposed.^{23 24}

Causal frameworks have previously been used in a variety of research contexts to guide the data collection strategies and the subsequent analytical approach,^{25–27} and to demonstrate why prospective cohort studies may plausibly assume that dose–response relationships exist between change in training load and injury occurrence.²⁸ However, in running injury research, questionable mathematically driven stepwise selection procedures have been used to compute a list of variables significantly associated with injury.^{29–34} This approach may lead to bias in a small sample setting.³⁵ Moreover and more importantly, it prevents the researcher from explaining precisely how and why exposures are interrelated in a causal perspective.²³ In addition, simple crude associations have been presented that enable identification of predictors of injury potentially playing a role in the causal mechanisms of injury occurrence.^{30 36 37} Taking body mass index (BMI) as an example, traditional risk factor studies have been able to identify that runners with a high BMI are at increased

risk of sustaining a running-related injury.^{17 30 38} Although this approach identifies a subpopulation at increased risk of injury, it fails to identify the amount of running associated with a decreased or increased injury risk among runners with a high BMI.³⁹ Accordingly, the evidence and knowledge about injury aetiology produced from studies based on traditional epidemiological approaches have not always been translatable into advice for runners on how different ways to schedule their running activities may reduce or increase their risk of injury. This is unfortunate since runners themselves identify factors such as ‘*excess of training*’ and ‘*exceeding the body’s limitations*’ as major risk factors for running injuries.⁴⁰ These opinions are also strongly rooted in the scientific literature since ‘*Running too much too soon*’ or sudden changes in training load seem to be a key factor in running injury development.^{22 41 42} The underlying assumption here is that runners are unable to sustain a running-related injury without running too much—or in other words, that the runners’ capacity to withstand load is exceeded.^{28 42}

In this light, a plethora of studies examining the role of sudden changes in training load on injury occurrence should exist. However, even today the scientific attempts to examine the role of sudden changes in training load on injury occurrence are limited.^{22 43–45} The reasons why the ‘*Too much, too soon*’ theory is under-researched in a running injury setting may be twofold: (1) that only limited attention has been given to the advanced concept of exposures and outcomes that change status over time,⁴⁴ and (2) that many previous scientific studies have not quantified running activities, such as number of steps per running session and vertical oscillation per step, prospectively during follow-up.²²

Monitoring running activities prospectively using commercially available devices^{21 46 47} is necessary to elucidate the ‘*too much, too soon*’ theory.^{22 28} Recently, commercially available devices have made it possible to quantify distance, time-spent-running and running pace.²¹ Likewise, more sophisticated measures of running dynamics, such as number of strides, cadence, stride length, ground contact time, ground contact time balance, vertical oscillation and vertical ratio, can be quantified in large-scale epidemiological studies.⁴⁶ These data allow researchers to explore the relationship between sudden changes in training load and injury occurrence in different types of runners with different anthropometrics, demographic characteristics and different sports gear. Given the presence of multiple predisposing exposures that influence how much running a runner can tolerate, future research questions should address how much running the musculoskeletal system can tolerate *before* injury is sustained. Using this approach may help explain precisely *why* and *how* risk factors interplay, and how much running is acceptable for runners of different sizes and shapes.^{15 39} To better understand the influence of sudden changes in training load on running injury occurrence, more studies are needed. The main purpose of the present paper was to describe the design of a future prospective cohort study

exploring if a dose–response relationship exists between changes in training load and injury occurrence, and how this association is modified by other variables.

Hypotheses

In this study, the main hypotheses to be examined will be:

H1: A dose–response relationship between change in training load and injury occurrence exists in the sense that more runners will sustain injury following a large change (progression) in training load compared with a lower change.

H2: Runners with a higher BMI and age, runners with a history of previous injury and runners with a lower activity of daily living sustain more running-related injury compared with lighter and younger runners with no history of previous injury who comparably change their training load.

H3: A biological interaction (absolute excess risk due to interaction on an additive scale) exists when examining the synergy between changes in training load and changes in other variables.

METHODS AND ANALYSIS

This study protocol describes the design of the observational prospective cohort study entitled ‘The Garmin-RUNSAFE Running Health Study’. Owing to consecutive inclusion starting in the summer of 2019 combined with a fixed end of follow-up in December 2020, the length of follow-up period can vary up to 18 months between participants. Prior to inclusion, all runners will be required to sign an online informed consent form. The runners will be free to discontinue participation at any time with no obligation to provide their reason why. The study is observational and therefore needs no permission from the system of research ethics committees according to the Danish Act on Research of Health Projects, Section 14, no. 2 (see letter from the Local Ethics Committee Central Denmark Region in online supplementary material S1). The Danish Data Protection Agency has approved the study, including the data collection procedures and data storage (see online supplementary material S2).

Study population

The study’s base population will comprise consenting English-speaking runners who own a Garmin watch that supports tracking of running activities and who upload their data from running sessions to Garmin Connect, which is a worldwide web-based training diary (<https://connect.garmin.com/>). All types of runners (ie, elite, recreational, novice) with varying years of running experience are eligible for inclusion.

Patient and public involvement

Some of the researchers (RØN, DAR, CD, SR) involved in the present study have worked in clinical practice dealing with injured runners. As they have shared their stories underpinning injury occurrence, these runners

indirectly assisted in the hypothesis-making process for the current study. In addition, the RUNSAFE research group have conducted several research studies in which runners attended a clinical examination in the case they were injured. These studies also aided the rationale. No runners were, however, invited to take active part in the design of the current study via, for example, knowledge-transfer schemes.

Recruitment strategy

Following four pilot studies conducted in the spring of 2019, runners will be recruited in July, August, September, October and November 2019, pending appropriate data flow and logistics. However, the inclusion will be open-ended in the sense that the active recruitment phase allows runners to join at a later stage. Users of Garmin Connect will be contacted by email by Garmin. The email provides a short description of the study including a link to a RUNSAFE recruitment page (<https://garmin-runsafe.com/>) where further information about the study will be provided. Besides distributing a link to a recruitment page through emails, runners will be recruited through notifications on Garmin social media pages (such as Facebook, Twitter), Garmin blog post and through a media release which can be picked up by third-party online sites. In addition, RUNSAFE may recruit runners through the social media and via contact to running clubs. Since the recruitment material, informed consent information and follow-up questionnaires will be distributed in English, runners will be recruited predominantly from English-speaking countries (such as England, USA, Australia, New Zealand) and countries in which a majority of the population are familiar with written English (ie, Denmark, Norway, Sweden, Germany, the Netherlands and Belgium). Runners who will be eligible for inclusion include those:

- ▶ Willing to provide informed content.
- ▶ Willing to provide the research team with access to their Garmin Connect activities via Garmin’s Health Application Programming Interface (API).
- ▶ Who are familiar with the English language.
- ▶ Willing to respond to an enrolment questionnaire and a one baseline questionnaire in English.
- ▶ Willing to respond to scheduled questionnaires in English on a weekly, monthly and quarterly basis until end of follow-up at 31 December 2020 or until they leave the study, whichever comes first.
- ▶ Regularly uploading their running sessions to Garmin Connect (more than 90% of their running sessions).
- ▶ Who are above the age of 18 years.

Runners will be excluded if their Garmin Connect account is used by more than one person or if more than one person uses the same watch. In cases where runners regularly wear more than one watch, they will be eligible for inclusion provided they do not wear two watches at the same time or upload two or more data sets from similar time spans to Garmin Connect. For instance, it will be acceptable if runners use a Forerunner watch

Table 1 Overview of data collection of exposure-related variables and outcome

	Quantitative data from Garmin Connect		Subjective data from questionnaires	
	Health API	Baseline	Weekly	Biweekly
Running activity data (<i>req. info</i>) For example, strides, distance, time spent running, vertical oscillation, cadence, heart rate, pace, stride length	√			
Running injury data (<i>req. info</i>) Symptoms, pain and consequences (ie, medical consultations, time loss, stride-to-stride changes)			√	
Activities of daily living All-day step count, all-day walk/run distance, all-day heart rate	√	√		
Well-being index				√
Other sports activities Cycling, swimming, strength training	√			
Demographics (<i>req. info</i>) Gender, date of birth, height, weight, email address		√		

Information related to baseline is included either in the recruitment questionnaire (online supplementary material S3) or in the baseline questionnaire (online supplementary material S4). The weekly and biweekly questionnaires are provided in online supplementary material S5 and S6, respectively.

API, Application Programming Interface; Req. info, this information is required from study participants.

while running and another type of watch during activities of daily living.

Data collection

Data will be collected objectively through watches and running accessories such as heart-rate monitors; and subjectively through one enrolment and one baseline questionnaire; and scheduled emails sent out on a weekly basis. If a runner is followed during an 18-month follow-up period, he/she will be exposed to 72 weekly questionnaires. In addition, biweekly questionnaires are available in each of the runner's personal page. Table 1 presents an overview of the variables assessed. Personalised (first name, last name, email address), subjective and objective data are stored at password-protected servers located at Aarhus University, Denmark. The software (eg, Microsoft ASP.Net Core programme) will be continuously updated to ensure that the system conforms to the rules and regulations of the Danish Data Protection Agency and the provisions outlined by the European General Data Protection Regulation (GDPR). No person-identifiable information will be published or shared. Person-specific label numbers (Tokens) will be used to merge Garmin Connect data with injury data from questionnaires. These person-specific label numbers will be used during all sharing and processing of data within the RUNSAFE research group at Aarhus University.

Quantitative data measured by Garmin devices

When a runner is included in the study, Aarhus University will be granted access to Garmin Connect data for this particular runner. Note that Garmin does not grant

access to any user data without explicit consent by users. The access will be restricted to the study-specific follow-up period and to activity files going back 30 days from the date of the first activity file upload after consent has been approved. The Garmin Health API (<http://developer.garmin.com/garmin-connect-api/overview/>) will be used to access the following data specific to running activity from included runners: number of strides, time spent running, distance, pace, cadence, stride length, ground contact time, ground contact time/balance, vertical ratio, vertical oscillation and heart rate. These variables will be measured using (1) the Global Positioning System (GPS) in the watch; and/or (2) the accelerometer in the watch or in running accessories such as a heart-rate strap; and/or (3) heart rate monitor in a heart-rate strap or in the watch. The metrics will be computed based on motion measured by a GPS and/or three-dimensional accelerometer using Garmin proprietary algorithms in Garmin watches. Quantitative data on running activities will be needed since runners are unable to self-report their running data, such as distance, in a reliable manner,⁴⁷ whereas the validity of Garmin Forerunner watches have been found to be acceptable for use in scientific studies when measuring running distance and speed,²¹ as well as cadence, ground contact time and vertical oscillation.⁴⁶

In addition to the running-activity-related data, data from other types of sports activities such as cycling, swimming, walking and strength training will be assessed through the Health API. Furthermore, the Health API will be used to assess information on all-day step count, all-day calorie count, all-day distance, sleep duration, all-day

heart rate and index scale information (such as weight, body fat, bone mass). Importantly, such information will only be assessable if the included participants wear their watch throughout the day. As a consequence, information related to activities of daily living is missing for some participants in the study since it is not an inclusion criterion to wear the watch all day.

Self-reported subjective data

Questionnaires filled in on enrolment, at baseline and at follow-up will be employed to capture information about the runners' injury status or non-running-related variables (such as body mass, shoe use, previous injuries). At the recruitment page, runners will be required to respond to a short enrolment questionnaire to provide their name, last name, email address and demographic characteristics (online supplementary material S3). After signing the informed consent form, an email will be sent to welcome the runner to the study. In addition, a link to a baseline questionnaire will be provided in the email and the runners are kindly invited, when their time allows and in quiet surroundings, to complete the baseline questionnaire no later than 1 week after inclusion. The baseline questionnaire consists of easily read sections/tabs addressing running participation and running shoes, previous or current running-related injuries, well-being using the WHO-5 well-being index^{48 49} (see online supplementary material S6), and health-related diseases and conditions (see online supplementary material S4).

Participants will be prompted to respond to follow-up questionnaires online through a unique, participant-specific link included in an email. They have provided their email addresses in the recruitment/enrolment questionnaire to RUNSAFE for this purpose. The weekly questionnaire consists of the following sections/tabs: Introduction, Running-related problems in the past week and Equipment (see online supplementary material S5). The time needed to respond to the weekly questionnaire will be 20 s in case of no injury and approximately 4 min when injured once familiarised with the questionnaire.

All emails containing links to questionnaires will be distributed using an Aarhus University-based system, which is specifically developed for questionnaires sent to participants included in research studies.^{50 51} This email distribution system conforms to the rules and regulations of the Danish Data Protection Agency.

Outcome

Running-related injury is the primary outcome. Information about injury status will be assessed in the weekly questionnaire. In the question '*In the past week, have you had a musculoskeletal injury or have you experienced a problem to muscles, tendons or bones that is fully or partly caused by running?*', the runners are able to classify themselves as injury-free, as uninjured but with problems (new problems or same problems as reported last week) or as injured (a new injury or the same injury as reported last week). Altogether, answers provided to this question will

allow us to determine on a weekly basis which injury state each runner belongs to: (1) injury-free, (2) uninjured, but with a problem or (3) injured. If a participant reports a problem or an injury, he or she will be led to additional questions inquiring information about the injury or problem's location, pain level and origin of the injury (see the weekly questionnaire in online supplementary material S5). The runners are informed that a problem is less severe than an injury and a problem is something that can be painful and irritating; however, running activity continues in full. In addition, they are informed that an injury is more severe than a problem and that an injury is something that is painful and irritating leading to a reduction in running activity (eg, volume, intensity, frequency). This approach was introduced based on experiences using the Yamato *et al*⁵² consensus definition and the Oslo Trauma Research Center questionnaire.⁵³ Importantly, the injury outcome should be considered as a time-varying covariate as described in Nielsen *et al*.⁵⁴

Exposure

Change in training load is the primary exposure. Training load is defined using different variables such as running distance or cumulative training load calculated based on stepwise loads based on ground contact time, vertical oscillation and body weight. As other ways to define and calculate training load can be used, the definition of change in training load is not fixed and predefined. Likewise, the change in training load is defined using various equations. For instance, to calculate a weekly change, the ratio between two weekly loads expressed as a percentage of change as calculated in Nielsen *et al*⁴³ could be used. The ratio will be calculated in the following manner: after each running session during the study period, the total training load (ie, distance, strides, time spent running) from that session will be added to the total training load covered in a 6-day period prior to that session. Accordingly, the training load over a 1-week period (week 1) will be determined. Then, the training load from day 7 to day 13 prior to the training session of interest (week 0) will be calculated. Based on these two weekly training loads (week 1 and week 0), the progression (or increase) or regression (or decrease) between these two periods will be calculated by dividing the two weekly loads and then multiplying the result by 100 (ratio between weekly distances = (total training load week 1)/(total training load week 0)×100). After calculating the ratio after each running session, the change in weekly training load (progression or regression) will be categorised into 1 of the following 10 exposure states: (1) regression below -50%; (2) regression between -50% and -20%; (3) regression between -20% and -10%; (4) regression between -10% and 0%; (5) progression between 0% and 10%; (6) progression between 10% and 20%; (7) progression between 20% and 30%; (8) progression between 30% and 40%; (9) progression between 40% and 50%; and (10) progression greater than 50%. The 10% cut-off was chosen based on the general belief that a graded

training programme could become injurious at a progression in weekly distance exceeding 10%.⁵⁵ The 30% cut-off was chosen based on the findings from a 1-year prospective cohort study.⁴³ Because most participants presumably vary in their running routines, each runner will be able to move/transition between the 10 exposure states every time that runner completes a new running session during the study period. Statistically, such movement between exposure groups is known as a multistate transition.⁵⁶ If a runner does not run in week 0, it is not possible to calculate a ratio between the weekly training load in week 1 and week 0 because the denominator is zero. In such cases, participants will be categorised into a 'not available' group. To summarise, after each running session, participants are continuously categorised into 1 of the 11 exposure states using 0% to 10% progression as the reference group.

Other ways of calculating changes in activity have been used in the literature. For instance, a workload ratio has been used to describe the acute training load (ie, the training load of the past week) to the chronic load (ie, the 4-week rolling average of load) under the assumption that if an acute load exceeds the chronic load, then the athlete is considered underprepared and likely to be facing an increased risk of injury.^{43 57–60} Importantly, many additional ways of calculating change beyond weekly changes exist including, but not limited to, session-specific, monthly or bimonthly changes.

The ratio between the training loads is a time-dependent exposure in the sense that progression (a positive ratio) or regression (a negative ratio) can change many times during the study period (effectively after each running session). This allows for data analysis of a time-dependent exposure variable that allows the participants to move into other exposure states (or stay in the same exposure state) each time they run. Importantly, this approach is much different from the average change

in load (ie, running distance), which is not time-dependent and has been used as the exposure of interest in previously published studies.^{22 44}

Effect-measure modifiers

To explain precisely *why* and *how* risk factors interact, one needs to consider the runners' ability to tolerate the amount of running activity to which they expose themselves. A plethora of non-running-related variables determine, for better or worse, the change in training load a runner can tolerate without sustaining an injury. To name a few, the number of days/hours between running sessions,⁴¹ running experience (ie, total years and months of running practice),³³ running ability (ie, current performance level),³⁷ running shoe use,³² hours of sleep,⁶¹ nutrition intake,⁶¹ musculoskeletal problems sustained in other activities than running,³⁷ activity level in other sports and activities of daily living.¹⁷ Therefore, information on other types of sports activities, activities of daily living, demographics and anthropometrics need to be quantified and/or assessed in order to examine if the risk of sustaining an injury after different changes in running activity differs across strata of other variables. In the epidemiological literature, such variables are known as effect-measure modifiers,⁶² which should be considered vastly different from confounders.⁶³ Also, it will be investigated if the relative excess risk due to interaction is larger than the expected values if the change in training load and/or the change in load tolerance are severe (table 2).

Power calculation

The power calculation was based on a superiority study. Based on comparisons from previous studies^{42 43 45 64–66} and including experiences from clinical practice, a 1-year cumulative injury incidence proportion of 20% is expected in the reference state (equivalent to 0%–10% change in running activity). In the '20% to 30%' state,

Table 2 The hypothesised interaction between change in training load and change in load tolerance on risk of running-related injury

		Change in load tolerance		
		Increase	Same level	Decrease
Change in running activity	Below -50%	Green	Green	Green
	-50% to -20%	Green	Green	Green
	-20% to -10%	Green	Green	Green
	-10% to 0	Green	Green	Green
	0 to 10%	Green	Green	Green
	10% to 20%	Green	Yellow	Red
	20% to 30%	Yellow	Red	Red
	30% to 40%	Red	Red	Red
	40% to 50%	Red	Red	Red
	Above 50%	Red	Red	Black

Green is associated with minor risk, yellow with moderate risk, red with high risk and black with extremely high risk. We expect the relative excess risk due to interaction to be considerably higher than the expected risk in the black and red areas.

an injury incidence proportion is hypothesised to reach 24%, whereas a 28%, 35% and 50% incidence proportion target the '30% to 40%', '40% to 50%' and 'above 50%' states, respectively.

To be able to show a minimum difference in injury risk between the reference state and the 20% to 30% state of 1% (corresponding to a number needed to treat of 100 runners assuming a causal relationship), a sample of 4500 runners are required in the reference group and 2200 in the '20%–30%' state to reach a power of 79%. An accommodation to a potential loss to follow-up is necessary to include in determining the number of participants needed. In prospective studies with a self-structured running regime and a follow-up ≥ 6 months, the loss to follow-up has been reported to be approximately 22%–30%.^{30 67} In case of a 30% dropout in both states, a required sample size of 9571 runners will be required.

Statistics

Time-to-event models are used to estimate the cumulative risk difference between the different progression states using a generalised linear model (pseudo-observation method) using the normal distribution. We compute CI and p values using robust variance estimation to account for non-normality of the pseudo-observations and we will use the id- and log-link function to model cumulative risk differences and risk ratios, respectively.^{56 68–70} Recently, the pseudo-observation method was updated allowing for delayed entry in case the data is assumed to be subject to right-censoring.⁷¹ Consequently, the inclusion of time-dependent exposures and/or outcomes into the analysis has become possible. Time-dependent exposure enables each participant to move continuously between exposure states (after each running session) using multistate time-to-event models.⁵⁶ In these analyses, cumulative risk difference will be used as a measure of association. To comply with the assumptions behind the statistical model, at least 10 injuries per explanatory variable included in the analysis are needed.⁴⁴ The unit of analysis will be each runner (or each leg if data allow for calculation of stepwise loads rather than stride-wise loads). The following time scales will be used: calendar days, strides and/or kilometres. In addition, participants will be censored in case of disease, lack of motivation, no uploaded data to Garmin Connect during a 6-month period, unwillingness to continue in the study regardless of the reason or end of follow-up by 31 December 2020, whichever comes first.

In the case analyses are performed on running injuries occurring in a specific anatomical location (ie, the knee or the foot), we will analyse cause-specific hazards of the instantaneous risk of injury from a specific injury category using a competing risk model. To avoid violating of the assumption regarding right-censored data, hazard rate ratio will be used as measure of association in analyses on location-specific injury rate.

Group-based differences in rates or risks are used in the analyses described above. In addition, the development in individual incidence rates will be calculated.

The purpose of an individualised calculation of development in rate or injury development is to correct for the healthier runner selection, which presumably occurs during follow-up. Most likely, the more vulnerable runners sustain injury in the first part of the follow-up, while the less injury-prone runners continue throughout. In order to take into account this 'healthy runner effect', Cox regression or Poisson regression with shared frailty (or via robust variance estimation) on the recurrent injuries will be used.^{72 73} This approach will allow us to calculate a frailty factor for each individual. A 'frailty factor' is a number stating a participant's frailty, or likelihood of sustaining an injury. The 'frailty factor' will be high for participants sustaining multiple injuries during a short period of time and low if the participants remain uninjured throughout the follow-up period. The 'frailty factor' is then maintained. This will make it possible to evaluate if *each* subject's injury incidence rate changes over time. It should be noted that at this stage, researchers must assume that the frailty factor is constant over time.

Results are considered statistically significant at $p \leq 0.05$. In addition, for proper interpretation of study results, estimated effect size and estimated precision (95% confidence limits) will be calculated.⁷⁴ In case the synergy between two or more exposures is explored, the absolute excess risk due to interaction on an additive scale, which is equivalent to biological interaction, is used rather than relative excess risk due to interaction.³⁹ All analyses will be performed using Stata V.14 or greater (StataCorp LP, College Station, Texas, USA).

DISCUSSION

The Garmin-RUNSAFE Running Health Study will be the first prospective cohort study to include a large group of runners and to quantify running dynamics such as number of strides, cadence, stride length, ground contact time, ground contact time balance, vertical oscillation and vertical ratio,⁴⁶ as well as injury status during a long-term follow-up. The goal of the study is to provide evidence explaining precisely *why* and *how* risk factors interact in order to identify progression schemata in running practice associated with minimised injury risk in different types of runners. Ultimately, this will allow for identification of modifiable factors that are likely to be suitable targets for prevention and intervention strategies.

A major strength of the present study is the quantification of running activity data from runners who are followed prospectively over time. Such data allow for examining if a dose–response relationship between change in training load and injury occurrence exists using advanced time-to-event analyses of time-dependent exposures and outcomes.⁴⁴ Based on this, it may become possible to determine to which degree one or more running sessions are excessive in terms of injury risk and when runners are at a high risk of exceeding their body's limitations, taking into account various effect-measure modifiers. In the scientific literature, this phenomenon

has been described as ‘*running too much, too soon*’,^{22 41} which across many sporting activities, has lately become a hot topic in many sports science communities.^{28 75}

Another major strength is the sample size counting more than 20 000 runners. Previously, prospective cohort studies and trials have included 100–2000 runners.^{55 76–79} Although these studies have been considered large scale in terms of their sample size,⁸⁰ the statistical analyses have been restricted to inclusion of three to seven, often time-fixed, exposure variables in order to achieve robust analyses without violating the assumptions regarding events per variable.⁴⁴ For instance, using hazard rate ratio as a measure of association requires 10 injuries per variable included in the analyses.^{81 82} Taking change in running distance, categorised into five groups, as an example, at least $10 \times (5-1) = 40$ injuries would be required only for this variable in case it is analysed using states. If one were to examine the transitions⁴⁴ between the four states, 16 transitions are possible. This would require a minimum number of injuries of $10 \times (16-1) = 150$ only for this single variable. Based on this, analyses of the synergy between a changing training load variable and multiple other exposures have been impossible in the previous studies based on sample sizes of 100–2000 runners and, consequently, between 20 and 250 injuries per study.⁴⁴ Therefore, new large-scale prospective data collections as the one described in the present article are necessary to reach a sufficient number of events per variable. Only such studies will allow for robust statistical analysis on the synergy between change in training load and non-running-activity-related variables on running-injury occurrence.

A third strength is the information technology (IT) support provided by a commercial collaborator. In running research, an underdiscussed challenge is the difficulties related to obtaining the funding needed to ensure that IT systems and gadgets are up-to-speed with commercially available devices. Clearly, prospective monitoring of individual load using objective measures, as promoted in a recent consensus-based statement by Soligard *et al*,²⁸ is crucial for better understanding the aetiology behind running injuries. In this light, researchers need the objective data collected by commercially available devices. Researchers in the running injury community previously developed web-based data infrastructures to gather data on runners included in their studies.^{21 64 65 83} Although the efforts have been outstanding, the amount of funding needed to keep the web-based data collection system up-to-date with the recent development in GPS watches commonly used by the runners is considerable and beyond the reach of most research studies. Since it is a challenge for researchers to locate funding sources, alternative options for data collection should be considered. Manufactures of commercial devices often have web-based solutions that allow runners to upload their data to a web-based diary easily and efficiently. A strength of such a web-based solutions is the continuous software updates provided by the manufacturer to keep the system up-to-speed with the most recent hardware

developments. Another strength is the access to the large cohort of runners who use these web-based solutions. In research, the recruitment of runners is a time-consuming process. Drawing on the manufacturer’s assistance in the recruitment phase allows us to include more than 20 000 runners into our prospective cohort; a number, which all things considered, must be considered extremely large scale. The question remains if results from studies on users of an internet-based training diary, like Garmin Connect, are generalisable to all types of runners. Most likely, the users of Garmin Connect are rather experienced. This affects the external validity of the study since the aetiology of running injury development presumably differs between novice runners and more experienced runners.¹⁷ To investigate the comparability across study populations, the demographic characteristics of the participants in the Garmin-RUNSAFE Running Health Study will be compared with those of novice runners in the DANO-RUN study,⁶⁴ with the recreational runners in the RUNCLEVER trial,⁸⁴ and with the half-marathon runners in the ProjectRun21.⁸⁵

A major limitation in the present study is the use of self-reporting of injury status. Ideally, in the present study like in previous studies, all injured runners should attend a clinical examination to validate the injury diagnoses.^{11 84} Although the sample size in the present study is a strength, it is also a weakness in the sense that it is impossible to establish a setup with possibilities for clinical examination in all the countries from which the runners are recruited. In addition to the limitation regarding the need for clinical examinations, the choice of injury definition could be discussed since it has been a topic of much debate in the scientific literature.^{9 52 53 86} As a further limitation, it is important to stress that continuous measurement of factors such as muscle flexibility, strength deficits, increased or decreased range of motion, to name a few, is needed to fully grasp the mechanisms behind running-related injuries.^{23 87} Since it is impossible to quantify all relevant data from the included runners, it is unlikely that we will be able to identify all aetiological mechanisms leading to injury based on the data set collected.

ETHICS AND DISSEMINATION

The study design, its procedures and informed consent form were presented to the Ethics Committee Central Denmark Region (Request number: 227/2016 – Record number: 1-10-72-189-16). The Committee waived its right to consider the study, which is not necessary under Danish law owing to the observational nature of the study. The Danish Data Protection Agency approved the study (the Danish Data Protection Agency’s record number: 2015-57-0002; Aarhus University’s record number: 62908, serial number 309). All included participants will provide informed written consent prior to inclusion.

The publication of research results is of great importance for scientific qualification. Considering this and in

appreciation of academic freedom, Garmin has agreed that the authors are entitled to publish, present or by other means make public any findings or results generated by the applicants in accordance with good international standards for publication of research results (such as the Vancouver Guidelines), provided that participants have provided explicit consent for the study and that no personal data (name, email address) are included in any such publication or presentation.

The authors aim to publish their findings as peer-reviewed articles in international journals, mainly journals listed in the top 30% in the 'Sports Science', 'Public Health' or 'Statistics & Probability' category at ISI Web of Science/Journal Citation Reports. Examples of such journals include British Medical Journal, Sports Medicine, British Journal of Sports Medicine, American Journal of Sports Medicine and International Journal of Epidemiology. The results are to be reported according to the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for prospective cohort studies. We expect to publish several publications based on data collected in collaboration with Garmin before 1 January 2021. All results (negative, positive and inconclusive findings) will be disseminated and published.

Author affiliations

¹Section for Sports Science, Department of Public Health, Aarhus Universitet, Aarhus, Denmark

²Department of Affective Disorders, Aarhus University Hospital, Aarhus, Denmark

³Department of Public Health, Aarhus University, Section for Biostatistics, Aarhus, Denmark

⁴Section of Sports, Department of Public Health, Aarhus University, Aarhus, Denmark

⁵Orthopaedic Surgery Research Unit, Aarhus University Hospital – Aalborg Hospital, Aalborg, Denmark

⁶Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

⁷Section for Environment, Occupation and Health, Department of Public Health, Aarhus Universitet, Aarhus, Denmark

Acknowledgements Garmin is greatly acknowledged for recruiting study participants and providing assistance in understanding some of Garmin's running metrics. Dr Adam Hulme is greatly acknowledged for providing constructive feedback on the content of the questions regarding injury in the weekly questionnaire.

Contributors All authors (RØN, MLB, CD, RKB, DR, HS, ETP, SR, SK) conceived the idea behind the study and provided advice on the study design. RØN, SR and SK drafted a working agreement with Garmin. SK, RØN, MLB, CD, RKB and DR obtained funding for the study. RØN, MLB, CD, RKB, DR, ETP and SR are responsible for data acquisition, data management and statistical analyses. RØN developed the enrolment questionnaire. RKB and MLB developed the baseline and weekly questionnaire with feedback from RØN, CD, SK and DR. RØN is the main investigator. All authors are entitled to explore the data set and publish on prespecified hypotheses. RØN drafted the article, while all other authors revised the article for important intellectual content. All authors read and approved the final manuscript.

Funding Aarhus University and the Aarhus University Research Fund (grant number: AUFF-E-2015-FLS-9-9) provided funding for this project.

Competing interests None declared.

Patient consent for publication Not required.

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/>.

REFERENCES

- Andersen LB, Mota J, Di Pietro L. Update on the global pandemic of physical inactivity. *The Lancet* 2016;388:1255–6.
- Ekelund U, Steene-Johannessen J, Brown WJ, *et al*. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. *The Lancet* 2016;388:1302–10.
- Das P, Horton R. Physical activity—time to take it seriously and regularly. *The Lancet* 2016;388:1254–5.
- Ding D, Lawson KD, Kolbe-Alexander TL, *et al*. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet* 2016;388:1311–24.
- Pilgaard M, Rask S. *Danskernes motions- OG sportsvaner 2016 (in Danish)*. Copenhagen, Denmark: Danish Institute of Sports Studies, 2016.
- Pilgaard M. *Danskernes motions- OG sportsvaner 2007 (in Danish)*. Copenhagen, Denmark: Danish Institute of Sports Studies, 2008.
- Hespanhol Junior LC, Pillay JD, van Mechelen W, *et al*. Meta-Analyses of the effects of habitual running on indices of health in physically inactive adults. *Sports Med* 2015;45:1455–68.
- Koplan JP, Rothenberg RB, Jones EL. The natural history of exercise: a 10-yr follow-up of a cohort of runners. *Med Sci Sports Exerc* 1995;27:1180–4.
- Kluitenberg B, van Middelkoop M, Diercks R, *et al*. What are the differences in injury proportions between different populations of runners? A systematic review and meta-analysis. *Sports Med* 2015;45:1143–61.
- Videbæk S, Bueno AM, Nielsen RO, *et al*. Incidence of Running-Related injuries per 1000 H of running in different types of runners: a systematic review and meta-analysis. *Sports Med* 2015;45:1017–26.
- Nielsen RO, Rønnow L, Rasmussen S, *et al*. A prospective study on time to recovery in 254 injured novice runners. *PLoS One* 2014;9:e99877.
- Yeung SS, Yeung EW, Gillespie LD. Interventions for preventing lower limb soft-tissue running injuries. *Cochrane Database Syst Rev* 2011;7:CD001256.
- Yeung EW, Yeung SS. Interventions for preventing lower limb soft-tissue injuries in runners. *Cochrane Database Syst Rev* 2001;(3):CD001256.
- Fields KB, Sykes JC, Walker KM, *et al*. Prevention of running injuries. *Curr Sports Med Rep* 2010;9:176–82.
- Gingrich S, Harrast M. Injury prevention in novice runners: an evidence-based approach and literature review. *Curr Phys Med Rehabil Rep* 2015;3:18–24.
- Finch C. A new framework for research leading to sports injury prevention. *J Sci Med Sport* 2006;9:3–9.
- Hulme A, Nielsen RO, Timpka T, *et al*. Risk and protective factors for middle- and long-distance Running-Related injury. *Sports Med* 2017;47:869–86.
- Hulme A, Finch CF. From monocausality to systems thinking: a complementary and alternative conceptual approach for better understanding the development and prevention of sports injury. *Inj Epidemiol* 2015;2.
- Van Gent RN, Siem D, Van Middelkoop M, *et al*. Incidence and determinants of lower extremity running injuries in long distance runners: a systematic review. *Sport & Geneeskunde* 2007;40:16–29.
- Wen DY. Risk factors for overuse injuries in runners. *Curr Sports Med Rep* 2007;6:307–13.
- Nielsen RO, Cederholm P, Buist I, *et al*. Can GPs be used to detect deleterious progression in training volume among runners? *J Strength Cond Res* 2013;27:1471–8.
- Nielsen RO, Buist I, Sorensen H, *et al*. Training errors and running related injuries: a systematic review. *Int J Sports Phys Ther* 2012;7:58–75.
- Bertelsen ML, Hulme A, Petersen J, *et al*. A framework for the etiology of running-related injuries. *Scand J Med Sci Sports* 2017;27:1170–80.
- Nielsen RO, Bertelsen ML, Møller M, *et al*. Training load and structure-specific load: applications for sport injury causality and data analyses. *Br J Sports Med* 2018;52:1016–7.
- Shrier I, Steele RJ, Zhao M, *et al*. A multistate framework for the analysis of subsequent injury in sport (M-FASIS). *Scand J Med Sci Sports* 2016;26:128–39.

26. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol* 2008;8:70:2288-8-70.
27. Hernán MA, Taubman SL. Does obesity shorten life? the importance of well-defined interventions to answer causal questions. *Int J Obes* 2008;32(S3):S8-S14.
28. Soligard T, Swelling 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.
29. Buist I, Bredeweg SW, Bessem B, et al. Incidence and risk factors of running-related injuries during preparation for a 4-mile recreational running event. *Br J Sports Med* 2010;44:598-604.
30. Nielsen RO, Buist I, Parner ET, et al. Predictors of Running-Related injuries among 930 novice runners: a 1-year prospective follow-up study. *Orthopaedic Journal of Sports Medicine* 2013;1:1-7.
31. Van Middelkoop M, Kolkman J, Van Ochten J, et al. Risk factors for lower extremity injuries among male marathon runners. *Scand J Med Sci Sports* 2008;18:691-7.
32. Malisoux L, Ramesh J, Mann R, et al. Can parallel use of different running shoes decrease running-related injury risk? *Scand J Med Sci Sports* 2015;25:110-5.
33. Malisoux L, Chambon N, Delattre N, et al. Injury risk in runners using standard or motion control shoes: a randomised controlled trial with participant and assessor blinding. *Br J Sports Med* 2016;50:481-7.
34. Malisoux L, Chambon N, Urhausen A, et al. Influence of the Heel-to-Toe drop of standard Cushioned running shoes on injury risk in leisure-time runners: a randomized controlled trial with 6-month follow-up. *Am J Sports Med* 2016;44:2933-40.
35. Steyerberg E, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets a simulation study of bias in logistic regression analysis. *J Clin Epidemiol* 1999;52:935-42. 10.1016/S0895-4356(99)00103-1
36. Rasmussen CH, Nielsen RO, Juul MS, et al. Weekly running volume and risk of running-related injuries among marathon runners. *Int J Sports Phys Ther* 2013;8:111-20.
37. Buist I, Bredeweg SW, Lemmink KAPM, et al. Predictors of running-related injuries in novice runners enrolled in a systematic training program: a prospective cohort study. *Am J Sports Med* 2010;38:273-80.
38. Buist I, Bredeweg SW. Higher risk of injury in overweight novice runners. *Br J Sports Med* 2011;45:338. 10.1136/bjism.2011.084038.79
39. Nielsen RO, Bertelsen ML, Parner ET, et al. Running more than three kilometers during the first week of a running regimen may be associated with increased risk of injury in obese novice runners. *Int J Sports Phys Ther* 2014;9:338-45.
40. Saragiotto BT, Yamato TP, Lopes AD. What do recreational runners think about risk factors for running injuries? A descriptive study of their beliefs and opinions. *J Orthop Sports Phys Ther* 2014;44:733-8.
41. Etiology HA. Prevention, and early intervention of overuse injuries in runners: a biomechanical perspective. *Phys Med Rehabil Clin N Am* 2005;16:651-67.
42. Malisoux L, Nielsen RO, Urhausen A, et al. A step towards understanding the mechanisms of running-related injuries. *J Sci Med Sport* 2015;18:523-8.
43. Nielsen Rasmus Østergaard, Parner ET, Nohr EA, et al. Excessive progression in Weekly running distance and risk of running-related injuries: an association which varies according to type of injury. *J Orthop Sports Phys Ther* 2014;44:739-47.
44. Nielsen Rasmus Østergaard, Malisoux L, Møller M, et al. Shedding light on the etiology of sports injuries: a look behind the scenes of time-to-event analyses. *J Orthop Sports Phys Ther* 2016;46:300-11.
45. Kluitenberg B, van der Worp H, Huisstede BMA, et al. The NLstart2run study: Training-related factors associated with running-related injuries in novice runners. *Journal of Science and Medicine in Sport* 2016;19:642-6.
46. Adams D, Pozzi F, Carroll A, et al. Validity and reliability of a commercial fitness Watch for measuring running dynamics. *J Orthop Sports Phys Ther* 2016;46:471-6.
47. Dideriksen M, Soegaard C, Nielsen RO. Validity of self-reported running distance. *J Strength Cond Res* 2016;30:1592-6.
48. Topp CW, Østergaard SD, Søndergaard S, et al. The WHO-5 well-being index: a systematic review of the literature. *Psychother Psychosom* 2015;84:167-76.
49. Ellervik C, Kvetny J, Christensen KS, et al. Prevalence of depression, quality of life and antidepressant treatment in the Danish General suburban population study. *Nord J Psychiatry* 2014;68:507-12.
50. Obling KH, Overgaard K, Juul L, et al. The mile study: a motivational, individual and locally anchored exercise intervention among 30-49 year-olds with low levels of cardiorespiratory fitness: a randomised controlled study in primary care. *BMC Public Health* 2013;13:1224,2458-13-1224.
51. Obling KH, Hansen A-LS, Overgaard K, et al. Association between self-reported and objectively measured physical fitness level in a middle-aged population in primary care. *Prev Med Rep* 2015;2:462-6.
52. Yamato TP, Saragiotto BT, Lopes AD. A consensus definition of running-related injury in recreational runners: a modified Delphi approach. *J Orthop Sports Phys Ther* 2015;45:375-80.
53. Clarsen B, Myklebust G, Bahr R. Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo sports trauma research centre (OSTRC) overuse injury questionnaire. *Br J Sports Med* 2013;47:495-502.
54. Nielsen RO, Bertelsen ML, Ramskov D, et al. Time-to-event analysis for sports injury research Part 2: time-varying outcomes. *Br J Sports Med* 2019;53:70-8.
55. Buist I, Bredeweg SW, Lemmink KAPM, et al. The GRONORUN study: is a graded training program for novice runners effective in preventing running related injuries? design of a randomized controlled trial. *BMC Musculoskelet Disord* 2007;8:24.
56. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med* 2007;26:2389-430.
57. Gabbett TJ, Hulin BT, Blanch P, et al. High training workloads alone do not cause sports injuries: how you get there is the real issue. *Br J Sports Med* 2016;50:444-5.
58. Gabbett TJ, Ullah S, Finch CF. Identifying risk factors for contact injury in professional rugby League players – application of a frailty model for recurrent injury. *J Sci Med Sport* 2012;15:496-504.
59. Hulin BT, Gabbett TJ, Lawson DW, et al. The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med* 2016;50:231-6.
60. Gabbett TJ. The training – injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med* 2016;50:273-80.
61. von Rosen P, Frohm A, Kottorp A, et al. Too little sleep and an unhealthy diet could increase the risk of sustaining a new injury in adolescent elite athletes. *Scand J Med Sci Sports* 2017;27:1364-71. 10.1111/sms.12735
62. Knol MJ, VanderWee TJ. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol* 2012;41:514-20.
63. Meeuwisse WH. Athletic injury etiology: distinguishing between interaction and confounding. *Clin J Sport Med* 1994;4:171-5.
64. Nielsen RO, Buist I, Parner ET, et al. Foot pronation is not associated with increased injury risk in novice runners wearing a neutral shoe: a 1-year prospective cohort study. *Br J Sports Med* 2014;48:440-7.
65. Bredeweg SW, Zijlstra S, Bessem B, et al. The effectiveness of a pre-conditioning programme on preventing running-related injuries in novice runners: a randomised controlled trial. *Br J Sports Med* 2012;46:865-70.
66. Buist I, Bredeweg SW, van Mechelen W, et al. No effect of a graded training program on the number of Running-Related injuries in novice runners. *Am J Sports Med* 2008;36:33-9.
67. Lun V, Meeuwisse WH, Stergiou P, et al. Relation between running injury and static lower limb alignment in recreational runners. *Br J Sports Med* 2004;38:576-80.
68. Parner ET, Andersen PK. Regression analysis of censored data using pseudo-observations. *Stata J* 2010;10:408-22.
69. Klein JP, Logan B, Harhoff M, et al. Analyzing survival curves at a fixed point in time. *Stat Med* 2007;26:4505-19.
70. Andersen PK, Hansen MG, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Anal* 2004;10:335-50.
71. Overgaard M, Andersen PK, Parner ET. Regression analysis of censored data using pseudo-observations: an update. *Stata J* 2015;15:809-21.
72. Ullah S, Finch CF. Applications of functional data analysis: a systematic review. *BMC Med Res Methodol* 2013;13:43:2288-13-43.
73. Finch CF, Cook J. Categorising sports injuries in epidemiological studies: the subsequent injury categorisation (sic) model to address multiple, recurrent and exacerbation of injuries. *Br J Sports Med* 2014;48:1276-80.
74. Stang A, Poole C, Kuss O. The ongoing tyranny of statistical significance testing in biomedical research. *Eur J Epidemiol* 2010;25:225-30.
75. Swelling M, Soligard T, Alonso J-M, et al. How much is too much? (Part 2) International Olympic Committee consensus statement on load in sport and risk of illness. *Br J Sports Med* 2016;50:1043-52.
76. Kluitenberg B, van Middelkoop M, Diercks RL, et al. The NLstart2run study: health effects of a running promotion program in novice runners, design of a prospective cohort study. *BMC Public Health* 2013;13:685.

77. Nielsen R, Ramskov D, Sørensen H, *et al.* Protocol for the dano-run study: a 1-year observational follow up study on running related injuries in 1000 novice runners. *Br J Sports Med* 2011;45:365–6.
78. Bredeweg SW, Zijlstra S, Buist I. The GRONORUN 2 study: effectiveness of a preconditioning program on preventing running related injuries in novice runners. The design of a randomized controlled trial. *BMC Musculoskelet Disord* 2010;11:196.
79. Baltich J, Emery CA, Whittaker JL, *et al.* Running injuries in novice runners enrolled in different training interventions: a pilot randomized controlled trial. *Scand J Med Sci Sports* 2017;27:1372–83.
80. Nigg BM, Baltich J, Hoerzer S, *et al.* Running shoes and running injuries: mythbusting and a proposal for two new paradigms: 'preferred movement path' and 'comfort filter'. *Br J Sports Med* 2015;49:1290–4.
81. Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
82. Peduzzi P, Concato J, Feinstein AR, *et al.* Importance of events per independent variable in proportional hazards regression analysis. II. accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503–10.
83. Malisoux L, Frisch A, Urhausen A, *et al.* Monitoring of sport participation and injury risk in young athletes. *J Sci Med Sport* 2013;16:504–8.
84. Ramskov D, Nielsen RO, Sørensen H, *et al.* The design of the run clever randomized trial: running volume, -intensity and running-related injuries. *BMC Musculoskelet Disord* 2016;17:177.
85. Damsted C, Parner ET, Sørensen H, *et al.* The association between changes in Weekly running distance and Running-Related injury: preparing for a half marathon. *J Orthop Sports Phys Ther* 2019;49:230–8.
86. Timpka T, Jacobsson J, Bickenbach J, *et al.* What is a sports injury? *Sports Med* 2014;44:423–8.
87. Nielsen RO, Bertelsen ML, Ramskov D, *et al.* Time-to-event analysis for sports injury research Part 1: time-varying exposures. *Br J Sports Med* 2019;53:61–8.