Title: Natural experiments in marathon running

Authors:

Daniel Muniz-Pumares [1]; Ed Maunder [2]; Barry Smyth [3]; Ben Hunter[4].

Affiliations

[1] School of Health, Medicine and Life Sciences, University of Hertfordshire, Hatfield, UK.

[2] Sports Performance Research Institute New Zealand, Auckland University Technology, Auckland, New Zealand

[3] Insight Centre for Data Analytics, School of Computer Science, University College Dublin, Dublin, Ireland

[4] School of Human Sciences, London Metropolitan University, London, United Kingdom

Corresponding Author:

Daniel Muniz-Pumares

School of Health, Medicine and Life Sciences

University of Hertfordshire

Hatfield

AL10 9EU

UK

D.muniz@herts.ac.uk

Telephone: (+44) 01707 284800

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

Abstract

Natural experiments are a form of passive experiments that analyse populations exposed to environmental or behavioural stimuli. Here, we review natural experiments using marathon running as an example of an unusual physiological stressor to illustrate how such approaches can generate new hypotheses or test competing ones using large datasets of training and performance data. Natural experiments have been able to demonstrate that marathons are completed close to, but below, critical speed, and that faster athletes complete marathons at a higher fraction of critical speed compared with slower runners. Environmental conditions, such as heat and altitude, as well as training and racing behaviours including the characteristics of training programmes and pacing strategies, have also been demonstrated to impact marathon performance. Findings from the natural experiments discussed herein often align with results from smaller, laboratory-controlled studies. However, the scale of these datasets has allowed for the detection of subtle trends and interactions that otherwise would remain undetected in smaller cohorts. In conclusion, natural experiments have helped advanced our understanding of marathon physiology and can effectively complement traditional laboratory-based interventions. Future research may explore hybrid approaches in which researchers conducting well-controlled laboratory experiments collaborate with running platforms to design integrated research studies that combine large sample sizes with laboratory-controlled interventions.

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

1. Introduction

Passive experiments are defined as studies with little or no active intervention from the investigator, typically on individuals exposed to unique or unusual physiological stressors.[1] Passive experiments can be further classified as experiments of nature and natural experiments. The former examines individuals with rare genetic or acquired conditions, such as McArdle disease, whilst natural experiments analyse datasets from populations that differ in their exposure to environmental or behavioural stimuli, enabling researchers to measure specific outcomes and test targeted hypotheses.[1]

Exercise, in its broad definition of any form of physical activity which aims to improve performance or health,[2] can hardly be considered as a unique physiological stressor.[3] However, under certain circumstances, exercise might be considered a unique physiological stressor and thus be used to conduct natural experiments. For instance, the physiological effects of regular exercise over relatively short time periods (weeks to months) are well documented. However, studies investigating the effects of long-term exercise interventions (years to decades) are scarce and remain unusual, and natural experiments reporting fitness data across the lifespan have been used to bridge this gap.[1]

Marathon running has experienced an increase in popularity over the last few years. Applications for the London Marathon increased from 247,069 in 2016 to nearly 1.4 million for the 2027 ballot. However, despite this growing interest, running a marathon remains a relatively rare feat. Less than ~0.005% of the global population take part in one of the seven World Marathon Majors each year, and it has been estimated that less than 1% of the population has completed a marathon in the last five years.[4] Therefore, despite humans exhibiting evolutionary traits that favour marathon running,[5] and the recent increase in popularity, running a marathon can be considered an unusual physiological stressor.

Much of the research on endurance running has aimed to understand and expand the limits of human performance.[6–8] Studies of marathon performance typically rely on laboratory tests

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

to measure physiological or biomechanical markers of performance, even though recent research is starting to incorporate field-based measures.[9,10] Laboratory testing can be expensive and time consuming, and with modest ecological validity. In addition, elite athletes are, by definition, a small population,[11] so research studies requiring athletes to modify their training programmes prior to a major competition can be challenging.[12] This is particularly the case if athletes are asked to be exposed to a condition or practice with the potential to have a deleterious effect on performance.

Endurance athletes often monitor performance and physiological responses during training and competition using devices that collect an increasingly wide range of variables, including metrics related to external workload, such as speed, altitude, or acceleration,[13] and physiological parameters including heart rate and heart rate variability,[14] breathing frequency,[15] interstitial glucose[16] and lactate concentrations,[17] or sleep.[18] These data are then uploaded to online platforms, enabling athletes and coaches to manage and design training programmes (e.g. TrainingPeaks), monitor progression, or predict future performance. In turn, these platforms collect large sets of training, physiological, and performance data from thousands of athletes. For example, >10 billion exercise activities are now uploaded per year to Strava alone.[19]

Given marathons are relatively unusual physiological stressors and performing laboratory-controlled studies is challenging, natural experiments using wearable-derived endurance training and racing data can be used to test and generate hypotheses. The aim of this review is to highlight the potential of performance and physiological data to generate new and test existing hypotheses, specifically relating to exercise physiology and marathon performance. First, we focus on how large datasets have been used to further the understanding of physiological factors underpinning marathon performance. Second, we investigate how large datasets of running data can be used to establish the effect of environmental factors, such as weather and altitude, and how behaviours such as training and pacing strategies, can affect marathon performance. The final section of the paper will outline the practical applications,

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

limitations, and future research in this field, highlighting the potential of combining passive experiments, typically retrospective analysis of existing data, with prospective experiments where certain hypotheses are tested. Whilst the focus of the review relies on marathon performance as an exemplar, evidence drawn from other sports or events have been included where appropriate.

2. Formulating and Testing Hypotheses Through Training Data

2.1. Markers of Marathon Performance from Training Data

Running a marathon requires a metabolic steady state to be maintained throughout, virtually, the entire race. If exercise is performed in the severe domain, steady state metabolism cannot be maintained,[20,21] as the resynthesis of ATP through oxidative phosphorylation of fats and carbohydrates is not sufficient to satisfy the energetic demand. Exercise in the severe domain is therefore characterised by a substantial physiological perturbation at local and systemic levels,[21] including depletion of phosphocreatine, accumulation of hydrogen ions and inorganic phosphate, and inexorable increases in oxygen uptake, culminating in exercise cessation within a few minutes (see, for a review [22]). The best performance in a marathon is, therefore, likely to be at the intensity corresponding, or very close, to the maximal metabolic steady state.

There has been substantial interest in how the maximal metabolic steady state can be determined. Despite some controversy,[23,24] the asymptote of the speed-duration relationship, critical speed (CS), is generally accepted to represent the speed corresponding to the maximal metabolic steady state.[25–27] A frequent criticism of CS is the time-intensive protocol it requires. The ‘gold-standard’ protocol to determine CS involves multiple maximal efforts at slightly different intensities within the severe exercise domain, intended to result in a range of test durations of ~2 to 15 min.[27,28] The speed and duration of these tests are then used to construct the speed-duration relationship, from which CS is derived. Several alternative approaches have been proposed to estimate CS,[29,30] including an analysis of the best performance observed over a range of distances, where the relationship between speed and duration is

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

constructed from the observed best efforts, instead of the maximal effort attained during a testing session(s).

Some studies have used training and performance data to determine CS in marathon runners. For example, Jones and Vanhatalo[31] reported that elite athletes complete the marathon at virtually the speed corresponding to their CS (~96%). Using a large dataset of training data, it was possible to test the hypothesis that training data during the weeks leading up to a major marathon could be used to estimate CS.[32] The authors reported that, in >25,000 runners, the best performance over distances from 400 m to 5000 m recorded during training was used to determine CS, using models that generally had a good fit (R^2 > 0.99), which suggests CS can be predicted with reasonable accuracy from training data. Further, CS derived from training data predicted performance with ~7.7% error.[32] On average runners completed the marathon at ~85% of their predicted CS, irrespective of their sex or age (>40 vs <40 y). However, there was substantial variation depending on the marathon speed. The fastest runners in this dataset completed the marathon at >90% CS, but the fraction of CS at which runners completed the marathon decreased in a linear fashion with marathon finishing time. This relationship persists when considering the fraction of CS which elite male and female athletes complete marathons (Figure 1).

**** Figure 1 around here ****

The finding that faster runners also sustain a higher fraction of CS throughout the marathon may be explained due to a superior capacity to preserve physiological function during exercise.[33–35] Indeed, the ability to preserve physiological function from deteriorating during prolonged exercise has been defined as 'durability' [36] or 'resilience'.[37] Whilst there are ongoing discussions to establish clear definitions of these terms, for the purpose of this article we will define durability as the deterioration in physiological characteristics over time during prolonged exercise.[38] There appears to be several methodological considerations that can influence the assessment of durability.[39] One potential method to assess durability is by monitoring the ratio between internal workload (e.g. heart rate, HR) and external workload (e.g. speed) during

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

prolonged exercise.[36,39,40] A greater magnitude of decoupling between internal and external workloads is associated with worsened durability, as assessed via laboratory-based measurements of work rates at the intensity domain transitions before and after prolonged exercise.[40] The hypothesis that decoupling of heart rate-to-speed ratio during a marathon contributes to explain performance in marathon running was tested by Smyth et al.[41] In this natural experiment, athletes who demonstrated high decoupling, defined as a heart rate-to-speed ratio greater than 1.3 at the end of the marathon, relative to the 5-10 km segment (i.e., a ≥30% increase), completed the marathon in 3 h 58 min, compared to 3 h 37 min in athletes who exhibited low decoupling (i.e., a ratio <1.1). Furthermore, the authors reported that a regression model containing CS alone could predict marathon performance with an error of ~6.5%,[41] comparable to the prediction error reported in a previous analysis (~7.7%, see reference [32]). However, incorporating the magnitude of decoupling reduced the error in the prediction of marathon performance from ~6.5 to ~5.2%.

The approach taken in these experiments assumes that the best performance over certain distances recorded over a specific period of time (e.g. 16 weeks prior to the marathon in the studies by Smyth et al.) closely match the performance observed during ‘intentional’ maximal efforts. It should be highlighted that relying on performance data such as personal bests may overestimate CS by including performance no longer feasible for an athlete, whilst using training data may underestimate CS as it assumes the best observed performance represents a maximal effort. Nonetheless, good agreement between best observed performance during training, and performance attained during intentional maximal efforts performed has been reported.[30] Therefore, whilst the gold standard for CS estimation should require an athlete to perform intentional maximal efforts,[28] the natural experiments presented herein suggest that training and performance data may provide reasonable predictions of performance, particularly for recreational runners, and further the understanding of physiology underpinning marathon performance.

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

2.2. Factors affecting Marathon Performance: Environmental conditions and running behaviours

Large datasets containing training and performance data may also be valuable tools to perform natural experiments, investigating individuals undertaking the same physiological challenge (e.g. running a marathon), but facing different challenges (e.g. environmental conditions), or to investigate how different behavioural strategies (e.g. pacing, training strategies) affect marathon performance.

2.2.1. Environmental conditions and marathon performance

Performance is affected by environmental factors. For example, as the partial pressure of oxygen decreases with altitude, performance in events that rely on oxidative phosphorylation, such as the marathon, rapidly deteriorates. This can be illustrated by the results from the 1968 Olympic Games in Mexico City, held at an altitude of 2,240 m above sea level. Mamo Wolde won the gold medal for the marathon with a time of 2:20:26, which was ~8% slower than the world record at that time, held by Derek Clayton with a time of 2:09:36. (Of note, the first women's Olympic marathon was held at the 1984 Summer Olympics in Los Angeles). This pattern was consistent across other endurance and long-distance events, and no new world records were set in any endurance discipline. In contrast, 24 athletic world records were set in the 1968 Olympic Games in Mexico City in short-distance track and field events due to the lower air resistance at altitude. Hamlin et al.[42] investigated the performance of 1889 elite athletes at 794 venues, with a total of 132,104 track and field events analysed. The results revealed that endurance performance in long-distance track and field running events (i.e. up to 10,000 m) was negatively affected even at very modest altitudes of 150 to 299 m above sea level, and this deterioration rapidly accelerates as altitude increases above 1000 m. In marathon running specifically, a study of the 1280 performances observed in the top finishers from 16 races at altitudes from sea level to 2800 m above sea level reported a negative correlation between marathon performance and altitude, though this was only evident at altitudes of over 700 meters.[43] The authors re412ported that for every 100 m above sea level,

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

marathon performance decreased by ~1%. These studies demonstrate that marathon performance is highly sensitive to altitude and suggests marathon performance will not be optimal in cities located at relatively modest elevations like Munich (~500 m), or Madrid (~650 m).

Weather conditions are also known to influence performance outcomes. Early laboratory-based studies reported an inverted U-shape relationship between ambient temperature and performance, with mild temperatures of 10-20°C suggested as optimal for endurance performance.[44] Natural experiments, however, have further advanced our understanding of the extent to which environmental conditions affect marathon performance. Mantzios et al.[45] investigated the effects of weather parameters on endurance running performance in 7,867 elite athletes competing across 1,258 races in 42 countries. The authors were able to establish that the 'optimal' conditions for peak performance occur at 7.5°C–15°C wet bulb globe temperature (or 10°C–17.5°C air temperature). In marathon running specifically, the best performance was reported at a wet bulb globe temperature of 7.5°C. These results are in agreement with previous studies looking at performance in major marathons, which demonstrate a progressive deterioration of marathon performance as wet bulb globe temperature increases from 5 to 25 °C.[46] The negative effect of increasing temperature on marathon performance appears to be consistent both across different locations and within individual races, such as the Berlin,[47] Boston[48], or New York Marathon.[49] Further to its effect on marathon performance, natural experiments have been used to establish the effect of environmental conditions on the incidence of exertional heat stroke during a marathon, a serious and potentially life-threatening medical complication. Breslow et al.[50] reported that between 2015 and 2019 there were 3.7 cases of heat stroke per 10,000 marathon starters, but younger and faster runners had higher risk, and greater increases in heat stress from start to peak during a marathon may exacerbate risk. These findings, overall, should be considered to adjust predictions of performance, but also by race management teams when establishing a 'do-not-start' criteria to safeguard athletes in endurance events like the marathon.[51]

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

Endurance events, including Olympics Games and major city marathons, are typically hosted by large cities, where levels of pollution might be high. These events are often scheduled in the morning to avoid peak heat during midday hours, yet this might coincide with rush hour and potentially highest levels of pollution (see Mougin et al.[52] for a review). This is significant, because air pollution may also affect endurance performance. Well-controlled randomised studies investigating pollution levels on performance, however, may not be feasible or ethical, due to the potentially harmful effect of pollutants on health. Fine airborne particulate matter smaller than 2.5 microns in diameter ($PM_{2.5}$) is produced during fuel combustion processes, and can cross the lung–blood barrier and enter the bloodstream, where it can trigger inflammation in various tissues.[53] Some studies have reported that high levels of $PM_{2.5}$ can impair airway function, which may in turn impact performance.[54] An analysis of the top three athletes from seven competitive US marathons (168 race-years) revealed that, in women, for every 10 $\mu g \cdot m^{-3}$ increase in PM_{10}, marathon performance was decreased by 1.4%, even when air pollutants rarely exceeded the health limits.[55] A criticism of the above studies was that air pollution was determined from monitoring stations, which may be far from the racecourses. More recently, $PM_{2.5}$ was estimated for 140 event-years (2.5 million marathon finishers) using a high-resolution machine-learning model that reconstructs $PM_{2.5}$ concentrations at specific locations such as marathon courses.[56] Greater $PM_{2.5}$ exposure was associated with slower marathon times, with an increase of 1 $\mu g \cdot m^{-3}$ associated with ~30 s slower marathon (32 s in men, 25 s in women).[56] Overall, these natural experiments offer insight into the negative effects that air pollution has on marathon performance, with a level of sensitivity that would be very difficult to achieve through traditional laboratory interventions.

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

2.2.2. Pacing strategies and marathon performance

Pacing refers to how athletes distribute energy and regulate effort and intensity to optimise performance.[57] In endurance events, particularly long-distance races, it is generally accepted that an even pacing strategy is optimal for performance. Studies investigating pacing strategies and marathon performance suggest that females and more experienced runners are more likely to adopt an even pacing strategy, compared to males and younger runners, who are more likely to adopt a positive split, i.e., starting fast in the first half to then slowing down in the second half of the marathon.[58] Furthermore, an analysis of 1.7 million marathons supported that both positive and negative pacing strategies are associated with reduced marathon performance, compared to an even pacing strategy.[59] These studies reveal a tendency for male recreational runners to adopt a positive split, which may hinder their performance. Whilst the reasons remain unclear, it may be speculated that the difference in pacing strategies may originate from psychological differences between sexes.[60]

Marathon performance is, indeed, a multifaceted phenomenon, regulated by a complex interaction of psychological, as well as physiological, factors. In an attempt to integrate both scientific disciplines and previous work, and building upon the central governor model,[61] St Clair Gibson et al.[62] proposed the Integrative Governor Theory whereby psychological and physiological drives compete and interact with each other, creating a dynamic system that regulates and controls observed behaviour. In the case of a marathon, poor pacing may result in excessive speed that triggers a feedback loop whereby perturbations within the exercising muscle and other physiological systems result in an urge to slow,[63] whilst psychological drives to continue may include the desire to complete the marathon within a target time (e.g. under 3 hours). Allen et al.[64] used a very large dataset with circa 10 million marathon performances to demonstrate the frequency of marathon finishing time increased close to popular target times (Figure 2). Specifically, marathon runners appear to aim, or pace toward, a specific time (e.g. a ‘sub 3-h marathon’), and marathon finishing times immediately before such reference times (e.g. 2:58 or 2:59) are disproportionally more frequent than immediately after (e.g. 3:01

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

or 3:02). The data demonstrate the behavioural contributions to marathon performance, with goal attainment being an important driver of pacing regulation.

**** Figure 2 around here ****

2.2.3. Training strategies and marathon performance

Understanding marathon performance requires examining the characteristics of the training that athletes have undertaken in the weeks, months, or years, leading up to the event. However, systematically studying the causal relationship between training habits and performance is challenging. Elite athletes, for example, typically get involved in research studies for relatively short periods of time (e.g., for a few weeks or during a training camp),[12] and may be reluctant to participate in research studies that require a substantial change to their training schedule, fearing that such involvement could interfere with their preparation. As a result, much of the available evidence comes from studies reporting 'best-practice' or 'results-driven' approaches to training.[8]

Such 'best-practice' studies are informative and provide unique insights into the training characteristics of, often, world-class distance runners, but have some methodological caveats. For example, it is unclear whether these athletes could have achieved a better performance had another training approach been taken. The ongoing debate on which training intensity distribution paradigm is best for endurance performance can illustrate this point.[65,66] Training intensity distribution describes the allocation of training resources, typically quantified as the proportion of training time spent across different intensity zones. Two competing paradigms dominate this narrative: the polarised vs. pyramidal. The proponents of the polarised paradigm suggest athletes complete around 80% of their training at low intensity (i.e., in the moderate domain of exercise, typically below their lactate or ventilatory threshold), with the majority of the remaining ~20% allocated to train at high intensity (i.e., in the severe domain, above CS) and little training competed in between. The supporters of the pyramidal paradigm suggest

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

much of the training should be completed at low intensities alongside a higher proportion of training allocated to intensities in the heavy domain, between the lactate threshold and CS.

The effectiveness of different training intensity distribution approaches at increasing endurance performance has been studied either as a retrospective analysis of training patterns of elite athletes or in prospective studies in relatively short interventions [67–69]. More recently, an analysis of the intensity distribution in >100,000 marathon runners across performance levels reported that a pyramidal approach was the most popular approach for all runners, and faster runners were more likely to adopt a pyramidal approach [70]. Furthermore, the authors noted that markers of training volume, such as number of training sessions, total distance covered, or distance covered at moderate intensity, strongly predicted subsequent marathon performance.[70] These findings partially support evidence-based practices emphasising the importance of high training volume for endurance athletes. However, they also underscore limitations in current approaches and demonstrate the need for further research, particularly regarding the optimal type and quantity of high-intensity training required to achieve peak marathon performance.

3. Practical applications, limitations, and future research

3.1. Practical applications

The findings from the natural experiments discussed above, derived from large datasets of training and performance data, could be of interest to participants, decision-makers, and other stakeholders involved in endurance running. The ability to estimate CS from training data and predict marathon performance or inform pacing has obvious applications to athletes and coaches. Indeed, the association between CS and endurance performance extends beyond marathon running,[71] and training programmes based on CS may be more effective than other approaches.[72] Furthermore, in shorter running events such as 5,000-m and 10,000-m, where the intensity is likely to exceed CS at least for some sections of the race, modelling the finite amount of work (expressed in units of distance covered) that can be completed above CS, referred to as D′, has been shown to explain differences in performance between elite

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

athletes.[73] Similarly, continuous monitoring of the available W′, the cycling equivalent of D′, has been used to monitor performance during exercise, and inform pacing strategies.[74] The results from natural experiments discussed herein can help athletes optimise their chances of a world record or personal best (i.e. aim for races held close to sea level, in cold and dry weather, and at locations with low pollution), as well as help in the decision making of race-organisers assessing whether a race should go ahead. Incidentally, the two attempts to break the symbolic 2-h barrier for the marathon incorporate some of the points discussed above: events took place close to sea level, in cold dry weather, and runners were carefully paced by an electric car, which might minimise the detrimental effect of $PM_{2.5}$.[75]

3.2. Methodological Limitations

As participation in endurance and ultra-endurance events grows and more performance variables are routinely measured, natural experiments using large datasets of training and physiological data like those described in this review will become more common. While these approaches address limitations of laboratory-based studies, they also introduce methodological challenges that researchers should consider. A frequent criticism is that variables derived from training or performance data may not conform to gold-standard laboratory procedures. For example, estimating CS from training data assumes that an athlete's recorded best efforts across multiple distances truly reflect maximal capacity.[27] Similar CS values derived from habitual training and intentional maximal efforts have been reported,[30] but this assumption might require further scrutiny, particularly for relatively less experienced recreational athletes. Furthermore, large training datasets are often complex, noisy, or incomplete because devices that capture running data may not be able to record other modes of training that influence endurance performance, such as strength training. While these datasets can be large enough to allow meaningful patterns to emerge despite the noise, it does not eliminate issues around missing or incomplete information. Lastly, it is worth highlighting that natural experiments are inherently descriptive, and cannot determine causality in the observed relationships. Large datasets also carry a greater risk of producing

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

spurious associations, particularly if many variables are analysed simultaneously. Working with large datasets requires interdisciplinary expertise, and researchers must consider issues related to data privacy, consent, data ownership, and intellectual property.

3.3. Future Research

Future research should continue to assess whether metrics derived from natural experiments align with laboratory-based gold standards. Using CS as an example, further work could determine whether estimates obtained from training or performance data truly represent the physiological boundary between steady state and non-steady state exercise, extending the type of validation already demonstrated in cycling.[76] In addition, whether estimates of CS (or its equivalent in cycling, critical power) derived from training data in recreational runners (or cyclists) accurately capture the maximal metabolic steady state, as effectively as those derived from well-trained athletes, should be examined.

Research should also seek to further our understanding of the relationship between training inputs and performance outputs using large-scale datasets, including work incorporating machine learning or artificial intelligence.[77–80] Collaboration between researchers and commercial platforms with access to large athlete populations could support prospective studies in which participants volunteer to be randomly allocated to different training approaches. Users, for example, may opt in to take part in a research study and then be randomly allocated to different training programmes, with field- or laboratory-based testing used to assess the effectiveness of the intervention. Although this would not constitute a natural experiment, it would combine the advantages of large sample sizes with relatively controlled conditions, and thus allow for strong study design to assess competing hypotheses. While we think laboratory studies will remain essential for validating and contextualising findings from natural experiments, combining controlled trials with large-scale observational data may offer a more efficient use of resources, bringing together the precision of laboratory measurements and the scale of real-world datasets.

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

4. Conclusion

Experiments of nature and natural experiments are two types of passive experiments. The former investigates individuals with rare genetic or acquired conditions, whereas natural experiments focus on populations exposed to unusual environmental or behavioural stressors. In this paper, we argued that, despite its growing popularity, running a marathon remains an unusual physiological stressor. We presented natural experiments reporting on novel markers of endurance performance, and how environmental conditions (e.g. altitude, climate, or pollution) and behaviours (e.g. training before the marathon or pacing during the event) influence the observed outcome: marathon performance. These studies often rely on performance and physiological metrics that are analysed retrospectively and compiled into large datasets. Whilst the results discussed here often align with findings from previous studies conducted on small cohorts under controlled laboratory conditions, the large sample sizes used in these studies can provide insights that could otherwise fail to be detectable in laboratory-controlled studies with smaller samples. Nonetheless, we recognise the value of laboratory testing and data collection under controlled conditions, which complement findings from large datasets. Therefore, future studies may explore a hybrid approach in which researchers conducting well-controlled laboratory experiments work with running platforms to design integrated hybrid studies that combine large sample sizes with laboratory-controlled interventions.

5. Acknowledgements

The authors would like to thank Eric J. Allen and co-authors for providing the data used to redraw Figure 2.

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

References

1. Joyner MJ, Wiggins CC, Baker SE, Klassen SA, Senefeld JW. Exercise and Experiments of Nature. *Compr Physiol*. 2023;13(3):4879-4907. doi:10.1002/CPHY.C220027
2. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43(7):1334-1359. doi:10.1249/MSS.0B013E318213FEFB
3. Bramble DM, Lieberman DE. Endurance running and the evolution of Homo. *Nature*. 2004;432(7015):345-352. doi:10.1038/NATURE03052
4. Sayer A. How Many People Have Run A Marathon?: A Global Analysis. April 17, 2024. Accessed August 14, 2025. https://marathonhandbook.com/how-many-people-have-run-a-marathon/
5. Lieberman DE, Bramble DM. The Evolution of Marathon Running. *Sports Medicine 2007 37:4*. 2012;37(4):288-290. doi:10.2165/00007256-200737040-00004
6. Jones AM, Kirby BS, Clark IE, et al. Physiological demands of running at 2-hour marathon race pace. *J Appl Physiol (1985)*. 2021;130(2):369-379. doi:10.1152/JAPPLPHYSIOL.00647.2020
7. Joyner MJ, Hunter SK, Lucia A, Jones AM. Physiology and fast marathons. *J Appl Physiol (1985)*. 2020;128(4):1065-1068. doi:10.1152/JAPPLPHYSIOL.00793.2019
8. Haugen T, Sandbakk Ø, Seiler S, Tønnessen E. The Training Characteristics of World-Class Distance Runners: An Integration of Scientific Literature and Results-Proven Practice. *Sports Medicine - Open 2022 8:1*. 2022;8(1):1-18. doi:10.1186/S40798-022-00438-7
9. Hunter B, Muniz-Pumares D. Durability of Parameters Associated With Endurance Running in Marathoners. *Eur J Sport Sci*. 2025;25(11). doi:10.1002/EJSC.70073
10. Hunter B, Lena A, Muniz-Pumares D. Durability of physiological and biomechanical variables during a marathon. *J Sports Sci*. 2026;44(2):131-145. doi:10.1080/02640414.2025.2567780
11. McKay AKA, Stellingwerff T, Smith ES, et al. Defining Training and Performance Caliber: A Participant Classification Framework. *Int J Sports Physiol Perform*. 2022;17(2):317-331. doi:10.1123/IJSPP.2021-0451
12. Stellingwerff T, Burke LM, Caldwell HG, et al. Integrative Field-Based Health and Performance Research: A Narrative Review on Experimental Methods and Logistics to Conduct Competition and Training Camp Studies in Athletes. *Sports Med*. 2025;55(6):1377-1403. doi:10.1007/S40279-025-02227-0
13. Ferraz A, Duarte-Mendes P, Sarmento H, Valente-Dos-Santos J, Travassos B. Tracking devices and physical performance analysis in team sports: a comprehensive framework for research—trends and future directions. *Front Sports Act Living*. 2023;5:1284086. doi:10.3389/FSPOR.2023.1284086

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

14. Rogers B, Mourot L, Doucende G, Gronwald T. Fractal correlation properties of heart rate variability as a biomarker of endurance exercise fatigue in ultramarathon runners. *Physiol Rep*. 2021;9(14). doi:10.14814/PHY2.14956

15. Romano C, Lo Presti D, Silvestri S, Schena E, Massaroni C. Flexible Textile Sensors-Based Smart T-Shirt for Respiratory Monitoring: Design, Development, and Preliminary Validation. *Sensors (Basel)*. 2024;24(6):2018. doi:10.3390/S24062018

16. Ishihara K, Uchiyama N, Kizaki S, Mori E, Nonaka T, Oneda H. Application of Continuous Glucose Monitoring for Assessment of Individual Carbohydrate Requirement during Ultramarathon Race. *Nutrients*. 2020;12(4):1121. doi:10.3390/NU12041121

17. Okawara H, Sawada T, Nakashima D, et al. Lactate threshold evaluation in swimming using a sweat lactate sensor: A prospective study. *Eur J Sport Sci*. 2024;24(9):1302. doi:10.1002/EJSC.12179

18. Lee T, Cho Y, Cha KS, et al. Accuracy of 11 Wearable, Nearable, and Airable Consumer Sleep Trackers: Prospective Multicenter Validation Study. *JMIR Mhealth Uhealth*. 2023;11:e50983. doi:10.2196/50983

19. Strava Inc. Strava Releases Year In Sport Trend Report, Showing What Makes and Breaks Motivation Across Generations. January 3, 2024. Accessed August 14, 2025. https://press.strava.com/articles/strava-releases-year-in-sport-trend-report

20. Coates AM, Joyner MJ, Little JP, Jones AM, Gibala MJ. A Perspective on High-Intensity Interval Training for Performance and Health. *Sports Med*. 2023;53(Suppl 1):85-96. doi:10.1007/S40279-023-01938-6

21. Black MI, Jones AM, Blackwell JR, et al. Muscle metabolic and neuromuscular determinants of fatigue during cycling in different exercise intensity domains. *J Appl Physiol (1985)*. 2017;122(3):446-459. doi:10.1152/JAPPLPHYSIOL.00942.2016

22. Poole DC, Burnley M, Vanhatalo A, Rossiter HB, Jones AM. Critical power: An important fatigue threshold in exercise physiology. *Med Sci Sports Exerc*. 2016;48(11):2320-2334. doi:10.1249/MSS.0000000000000939

23. Dotan R. A critical review of critical power. *Eur J Appl Physiol*. 2022;122(7):1559-1588. doi:10.1007/S00421-022-04922-6

24. Gorostiaga EM, Sánchez-Medina L, Garcia-Tabar I. Over 55 years of critical power: Fact or artifact? *Scand J Med Sci Sports*. 2022;32(1):116-124. doi:10.1111/SMS.14074

25. Jones AM, Burnley M, Black MI, Poole DC, Vanhatalo A. The maximal metabolic steady state: redefining the “gold standard.” *Physiol Rep*. 2019;7(10). doi:10.14814/PHY2.14098

26. Nixon RJ, Kranen SH, Vanhatalo A, Jones AM. Steady-state $\dot{V}O_2$ above MLSS: evidence that critical speed better represents maximal metabolic steady state in well-trained runners. *Eur J Appl Physiol*. 2021;121(11):3133. doi:10.1007/S00421-021-04780-8

27. Caen K, Poole DC, Vanhatalo A, Jones AM. Critical Power and Maximal Lactate Steady State in Cycling: “Watts” the Difference? *Sports Med*. 2024;54(10):2497-2513. doi:10.1007/S40279-024-02075-4

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

28. Muniz-Pumares D, Karsten B, Triska C, Glaister M. Methodological approaches and related challenges associated with the determination of critical power and curvature constant. *J Strength Cond Res*. 2019;33(2):584-596. doi:10.1519/JSC.0000000000002977

29. Hunter B, Meyler S, Maunder E, Cox T, Muniz-Pumares D. The relationship between the moderate-heavy boundary and critical speed in running. In: *28th Annual Congress of the European College of Sport Science*. 2023.

30. Hunter B, Ledger A, Muniz-Pumares D. Remote Determination of Critical Speed and Critical Power in Recreational Runners. *Int J Sports Physiol Perform*. Published online 2023. doi:10.1123/ijspp.2023-0276

31. Jones AM, Vanhatalo A. The 'Critical Power' Concept: Applications to Sports Performance with a Focus on Intermittent High-Intensity Exercise. *Sports Medicine*. 2017;47:65-78. doi:10.1007/s40279-017-0688-0

32. Smyth B, Muniz-Pumares D. Calculation of Critical Speed from Raw Training Data in Recreational Marathon Runners. *Med Sci Sports Exerc*. 2020;52(12):2637-2645. doi:10.1249/MSS.0000000000002412

33. Clark IE, Vanhatalo A, Thompson C, et al. Dynamics of the power-duration relationship during prolonged endurance exercise and influence of carbohydrate ingestion. *J Appl Physiol*. 2019;127(3):726-736. doi:10.1152

34. Clark IE, Vanhatalo A, Thompson C, et al. Changes in the power-duration relationship following prolonged exercise: Estimation using conventional and all-out protocols and relationship with muscle glycogen. *Am J Physiol Regul Integr Comp Physiol*. 2019;317(1):R59-R67. doi:10.1152/AJPREGU.00031.2019/

35. Clark IE, Vanhatalo A, Bailey SJ, et al. Effects of Two Hours of Heavy-Intensity Exercise on the Power-Duration Relationship. *Med Sci Sports Exerc*. 2018;50(8):1658-1668. doi:10.1249/MSS.0000000000001601

36. Maunder E, Seiler S, Mildenhall MJ, Kilding AE, Plews DJ. The Importance of 'Durability' in the Physiological Profiling of Endurance Athletes. *Sports Medicine*. *Springer Science and Business Media Deutschland GmbH*. 2021;51(8):1619-1628. doi:10.1007/s40279-021-01459-0

37. Jones AM. The fourth dimension: physiological resilience as an independent determinant of endurance exercise performance. *J Physiol*. 2023;0. doi:10.1113/JP284205

38. Meixner B, Joyner MJ, Sperlich B. Durability, Fatigability, Repeatability and Resilience in Endurance Sports: Definitions, Distinctions, and Implications. *J Appl Physiol (1985)*. Published online July 1, 2025. doi:10.1152/JAPPLPHYSIOL.00343.2025

39. Hunter B, Maunder E, Jones AM, Gallo G, Muniz-Pumares D. Durability as an index of endurance exercise performance: Methodological considerations. *Exp Physiol*. Published online March 27, 2025. doi:10.1113/EP092120

40. Rothschild JA, Gallo G, Hamilton K, et al. Durability of the moderate-to-heavy intensity transition can be predicted using readily available markers of physiological decoupling. *Eur J Appl Physiol*. Published online May 22, 2025:1-10. doi:10.1007/S00421-025-05815-0/FIGURES/4

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

41. Smyth B, Maunder E, Meyler S, Hunter B, Muniz-Pumares D. Decoupling of Internal and External Workload During a Marathon: An Analysis of Durability in 82,303 Recreational Runners. *Sports Medicine*. 2022;52(9):2283-2295. doi:10.1007/S40279-022-01680-5,

42. Hamlin MJ, Hopkins WG, Hollings SC. Effects of Altitude on Performance of Elite Track-and-Field Athletes. *Int J Sports Physiol Perform*. 2015;10(7):881-887. doi:10.1123/IJSPP.2014-0261

43. Lara B, Salinero JJ, Del Coso J. Altitude is positively correlated to race time during the marathon. *High Alt Med Biol*. 2014;15(1):64-69. doi:10.1089/HAM.2013.1060

44. Galloway SDR, Maughan RJ. Effects of ambient temperature on the capacity to perform prolonged cycle exercise in man. *Med Sci Sports Exerc*. 1997;29(9):1240-1249. doi:10.1097/00005768-199709000-00018

45. Mantzios K, Ioannou LG, Panagiotaki ZOE, et al. Effects of Weather Parameters on Endurance Running Performance: Discipline-specific Analysis of 1258 Races. *Med Sci Sports Exerc*. 2021;54(1):153. doi:10.1249/MSS.0000000000002769

46. Ely MR, Cheuvront SN, Roberts WO, Montain SJ. Impact of weather on marathon-running performance. *Med Sci Sports Exerc*. 2007;39(3):487-493. doi:10.1249/mss.0b013e31802d3aba

47. Knechtle B, Valero D, Villiger E, et al. Trends in Weather Conditions and Performance by Age Groups Over the History of the Berlin Marathon. *Front Physiol*. 2021;12. doi:10.3389/FPHYS.2021.654544

48. Knechtle B, Gangi S Di, Rust CA, Villiger E, Rosemann T, Nikolaidis PT. The role of weather conditions on running performance in the Boston Marathon from 1972 to 2018. *PLoS One*. 2019;14(3). doi:10.1371/JOURNAL.PONE.0212797

49. Weiss K, Valero D, Villiger E, et al. Relationship between running performance and weather in elite marathoners competing in the New York City Marathon. *Sci Rep*. 2022;12(1). doi:10.1038/s41598-022-25901-z

50. Breslow RG, Collins JE, Troyanos C, et al. Exertional Heat Stroke at the Boston Marathon: Demographics and the Environment. *Med Sci Sports Exerc*. 2021;53(9):1818-1825. doi:10.1249/MSS.0000000000002652

51. Roberts WO. Determining a “do not start” temperature for a marathon on the basis of adverse outcomes. *Med Sci Sports Exerc*. 2010;42(2):226-232. doi:10.1249/MSS.0B013E3181B1CDCF

52. Mougin L, Bougault V, Racinais S, et al. Environmental challenges facing athletes, stakeholders and spectators at Paris 2024 Olympic and Paralympic Games: an evidence-based review of mitigation strategies and recommendations. *Br J Sports Med*. 2024;58(15):870-881. doi:10.1136/BJSPORTS-2024-108281

53. Xing YF, Xu YH, Shi MH, Lian YX. The impact of PM2.5 on the human respiratory system. *J Thorac Dis*. 2016;8(1):E69. doi:10.3978/J.ISSN.2072-1439.2016.01.19

54. Cavalcante De Sá M, Nakagawa NK, Saldiva De André CD, et al. Aerobic exercise in polluted urban environments: effects on airway defense mechanisms in young healthy amateur runners. *J Breath Res*. 2016;10(4):046018. doi:10.1088/1752-7163/10/4/046018

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

55. Marr LC, Ely MR. Effect of air pollution on marathon running performance. *Med Sci Sports Exerc*. 2010;42(3):585-591. doi:10.1249/MSS.0B013E3181B84A85

56. Fleury ES, Bittker GS, Just AC, Braun JM. Running on Fumes: An Analysis of Fine Particulate Matter's Impact on Finish Times in Nine Major US Marathons, 2003–2019. *Sports Medicine 2024 55:6*. 2024;55(6):1505-1514. doi:10.1007/S40279-024-02160-8

57. Abbiss CR, Laursen PB. Describing and understanding pacing strategies during athletic competition. *Sports Medicine*. 2008;38(3):239-252. doi:10.2165/00007256-200838030-00004

58. Berndsen J, Lawlor A, Smyth B. Exploring the wall in marathon running. *Journal of Sports Analytics*. 2020;6(3):173-186. doi:10.3233/JSA-200354

59. Smyth B. Fast starters and slow finishers: A large-scale data analysis of pacing at the beginning and end of the marathon for recreational runners. *Journal of Sports Analytics*. 2018;4(3):229-242. doi:10.3233/JSA-170205

60. Hubble C, Zhao J. Gender differences in marathon pacing and performance prediction. *Journal of Sports Analytics*. 2015;2(1):19-36. doi:10.3233/JSA-150008

61. Noakes TD, St. Clair Gibson A, Lambert E V. From catastrophe to complexity: a novel model of integrative central neural regulation of effort and fatigue during exercise in humans: summary and conclusions. *Br J Sports Med*. 2005;39(2):120-124. doi:10.1136/BJSM.2003.010330

62. St Clair Gibson A, Swart J, Tucker R. The interaction of psychological and physiological homeostatic drives and role of general control principles in the regulation of physiological systems, exercise and the fatigue process–The Integrative Governor theory. *Eur J Sport Sci*. 2018;18(1):25-36. doi:10.1080/17461391.2017.1321688

63. Hureau TJ, Romer LM, Amann M. The "sensory tolerance limit": A hypothetical construct determining exercise performance? *Eur J Sport Sci*. 2018;18(1):13-24. doi:10.1080/17461391.2016.1252428

64. Allen EJ, Dechow PM, Pope DG, Wu G. Reference-dependent preferences: Evidence from marathon runners. *Manage Sci*. 2017;63(6):1657-1672. doi:10.1287/MNSC.2015.2417;SUBPAGE:STRING:ACCESS

65. Burnley M, Bearden SE, Jones AM. Polarized Training Is Not Optimal for Endurance Athletes: Response to Foster and Colleagues. *Med Sci Sports Exerc*. 2022;54(6):1038-1040. doi:10.1249/MSS.0000000000002924

66. Foster C, Casado A, Esteve-Lanao J, Haugen T, Seiler S. Polarized Training Is Optimal for Endurance Athletes. *Med Sci Sports Exerc*. 2022;54(6):1028-1031. doi:10.1249/MSS.0000000000002871

67. Casado A, González-Mohíno F, González-Ravé JM, Foster C. Training Periodization, Methods, Intensity Distribution, and Volume in Highly Trained and Elite Distance Runners: A Systematic Review. *Int J Sports Physiol Perform*. 2022;17(6):820-833. doi:10.1123/IJSPP.2021-0435

68. Muñoz I, Seiler S, Bautista J, España J, Larumbe E, Esteve-Lanao J. Does Polarized Training Improve Performance in Recreational Runners? *Int J Sports Physiol Perform*. 2014;9(2):265-272. doi:10.1123/IJSPP.2012-0350

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

69. Festa L, Tarperi C, Skroce K, La Torre A, Schena F. Effects of Different Training Intensity Distribution in Recreational Runners. *Front Sports Act Living*. 2019;1:495162. doi:10.3389/FSPOR.2019.00070/BIBTEX

70. Muniz-Pumares D, Hunter B, Meyler S, Maunder E, Smyth B. The Training Intensity Distribution of Marathon Runners Across Performance Levels. *Sports Medicine*. 2025;55(4):1023-1035. doi:10.1007/S40279-024-02137-7/FIGURES/5

71. Goulding RP, Marwood S. Interaction of Factors Determining Critical Power. *Sports Medicine 2023 53:3*. 2023;53(3):595-613. doi:10.1007/s40279-022-01805-w

72. Meyler SJR, Swinton PA, Bottoms L, et al. Changes in Cardiorespiratory Fitness Following Exercise Training Prescribed Relative to Traditional Intensity Anchors and Physiological Thresholds: A Systematic Review with Meta-analysis of Individual Participant Data. *Sports Medicine*. 2024;55(2):301-323. doi:10.1007/S40279-024-02125-X/TABLES/2

73. Kirby BS, Winn BJ, Wilkins BW, Jones AM. Interaction of exercise bioenergetics with pacing behavior predicts track distance running performance. *J Appl Physiol (1985)*. 2021;131(5):1532-1542. doi:10.1152/JAPPLPHYSIOL.00223.2021

74. Skiba PF, Clarke D, Vanhatalo A, Jones AM. Validation of a novel intermittent w' model for cycling using field data. *Int J Sports Physiol Perform*. 2014;9(6):900-904. doi:10.1123/IJSPP.2013-0471

75. Hutchinson A. Research Shows Air Pollution Slows Down Marathoners. January 9, 2025. Accessed August 14, 2025. https://www.outsideonline.com/health/training-performance/air-pollution-and-exercise

76. Maunder E, Rothschild JA, Ramonas A, Delcourt M, Kilding AE. A three-minute all-out test performed in a remote setting does not provide a valid estimate of the maximum metabolic steady state. *Eur J Appl Physiol*. 2022;122(11):2385-2392. doi:10.1007/S00421-022-05020-3/FIGURES/4

77. Rothschild JA, Stewart T, Kilding AE, Plews DJ. Predicting daily recovery during long-term endurance training using machine learning analysis. *Eur J Appl Physiol*. 2024;124(11):3279-3290. doi:10.1007/S00421-024-05530-2/FIGURES/4

78. Jianjun Q, Isleem HF, Almoghayer WJK, Khishe M. Predictive athlete performance modeling with machine learning and biometric data integration. *Scientific Reports 2025 15:1*. 2025;15(1):1-22. doi:10.1038/s41598-025-01438-9

79. Boudry F, Durand F, Meric H, Mouakher A. The role of machine learning methods in physiological explorations of endurance trained athletes: a mini-review. *Front Sports Act Living*. 2024;6:1440652. doi:10.3389/FSPOR.2024.1440652/BIBTEX

80. Vos L, Vergeer R, Goulding R, et al. Predicting physical performance after training: insights from machine learning using small samples. Published online August 6, 2024. doi:10.21203/RS.3.RS-4707433/V1

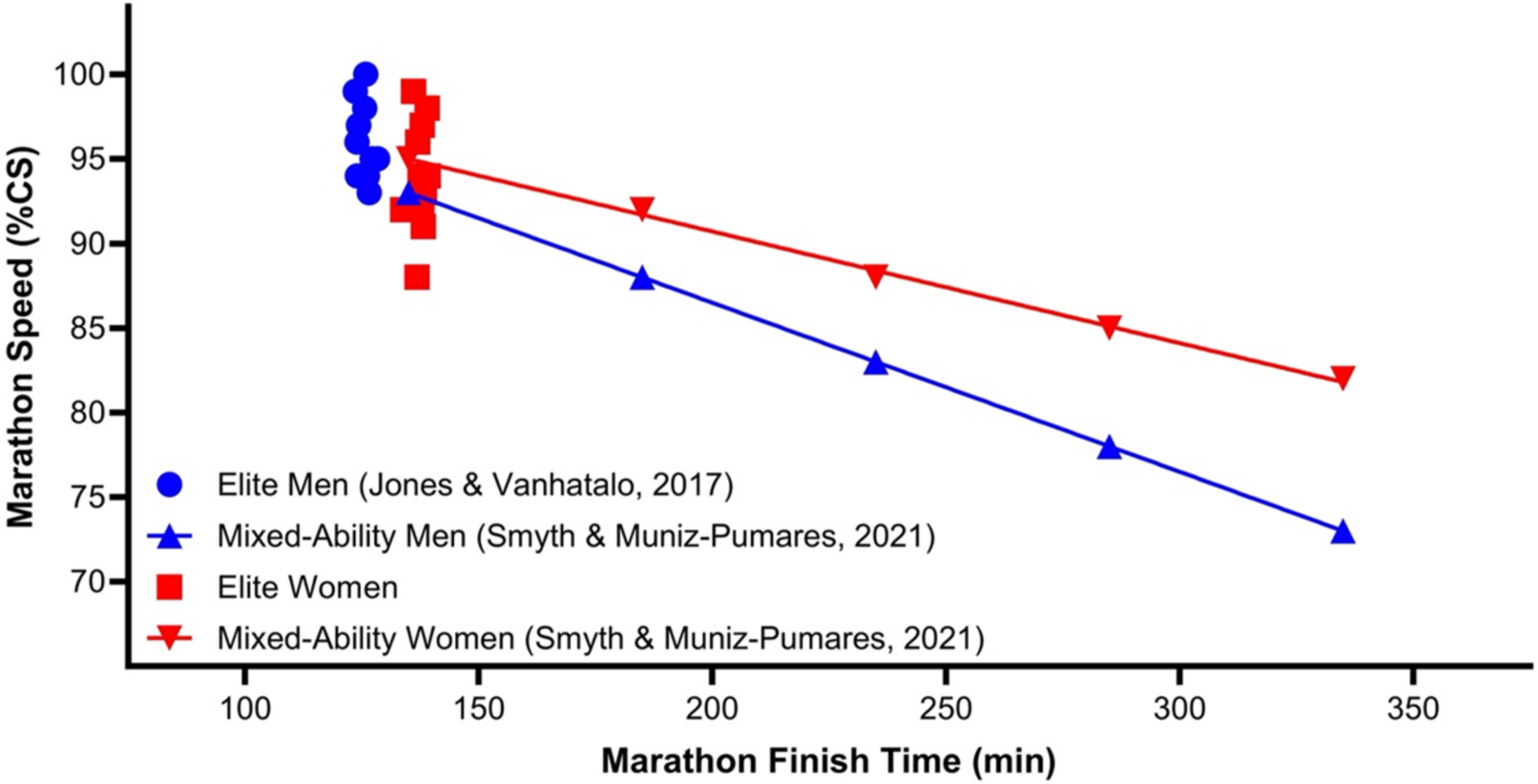
Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

Figure Legends

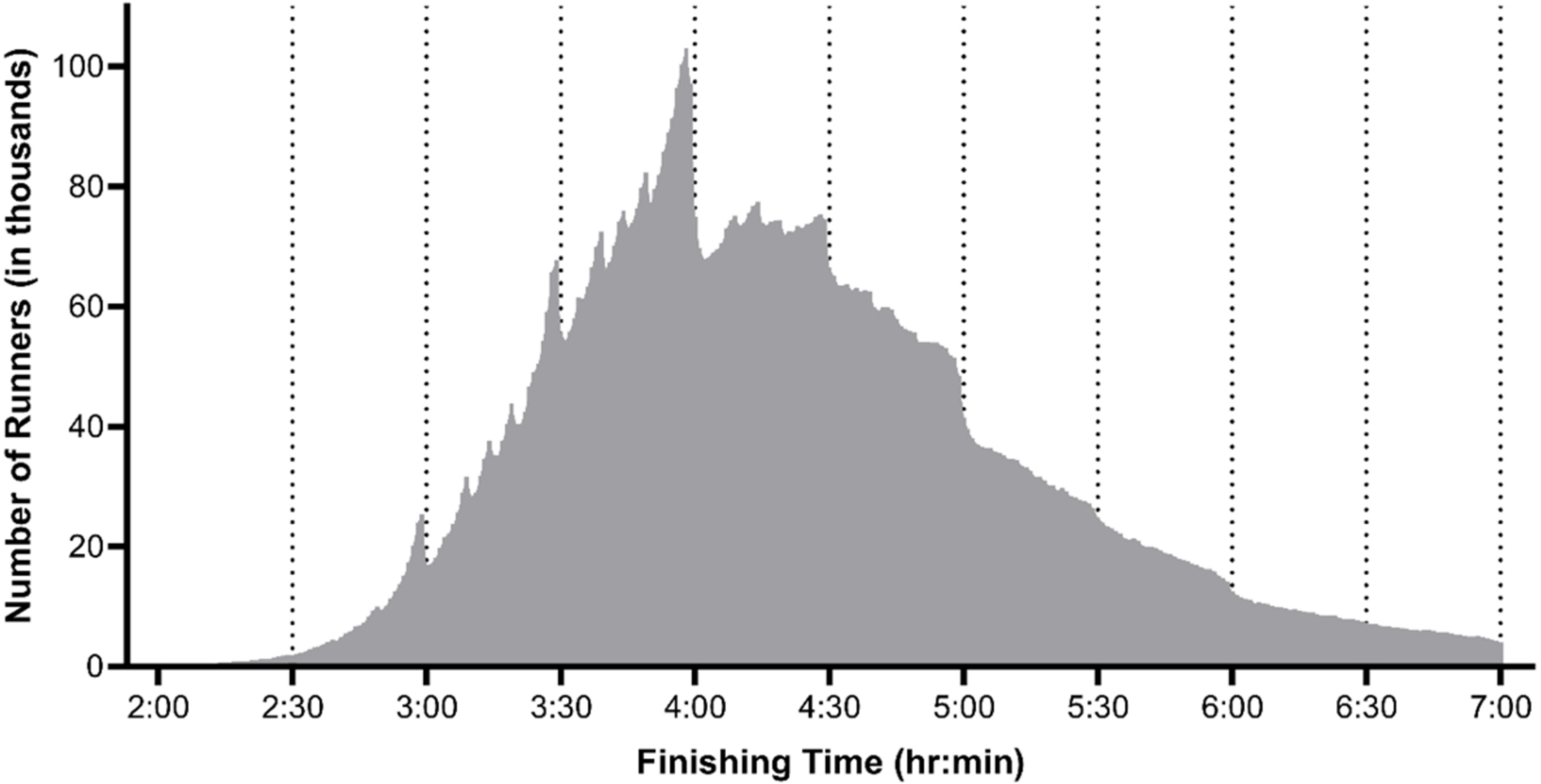
Figure 1. Marathon speed, expressed as percentage of the athlete's estimated critical speed (%CS). Triangles show data obtained from ~18000 male and ~6000 female recreational runners, and data in circles represent 12 elite male and female elite athletes. CS was determined from personal bests in elite athletes,[31] and from the best performance recorded during training in recreational runners.[32] The ability of athletes with faster finishing times to complete the marathon closer to their critical speed could be explained by their superior durability. This hypothesis was subsequently tested in another experiment of nature.

Figure 2. Natural experiment investigating the effect of reference times on marathon finishing. Results from ~10 million races demonstrated that athletes overcome the physiological drive to stop in order to achieve the goal of completing the race within a specific reference time. This is evidenced by the disproportionately high number of marathon finishing times recorded before common target times, e.g. 3:30 and 4:00. Redrawn with permission from Allen et al.[64]

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.



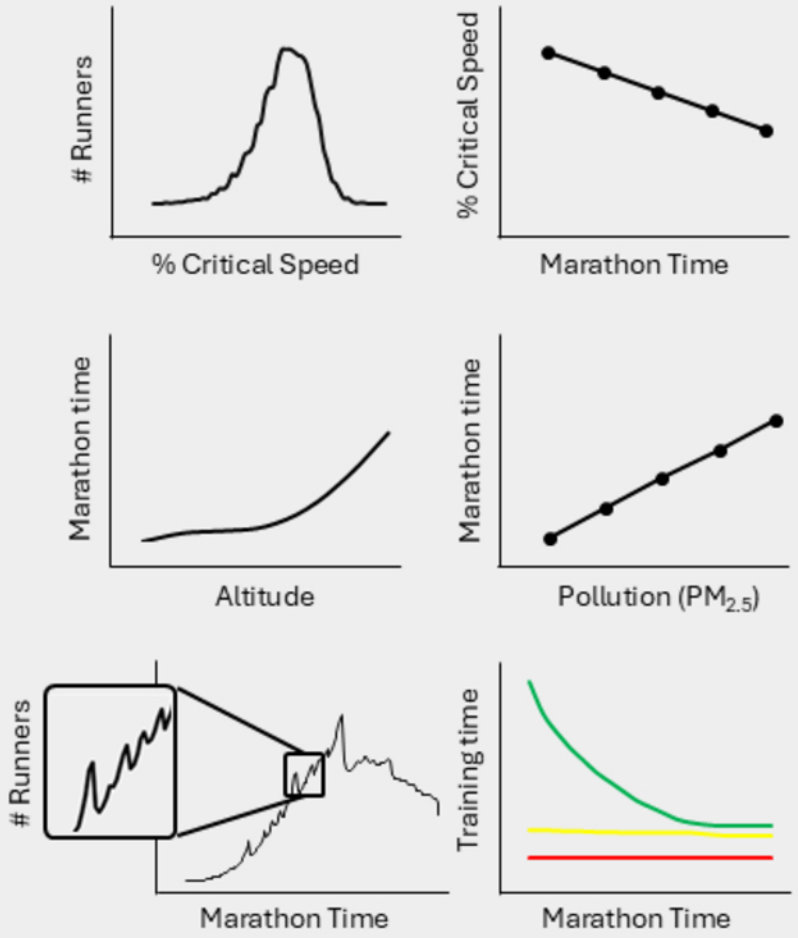
Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.



Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.

Natural Experiments and Marathon Running

AIM:
Review natural experiments in marathon running.



RESULTS:

Markers of performance

- Marathons are run at about 86% of critical speed.
- Faster runners run closer to their critical speed due to better durability.

Environmental factors

- Altitude reduces marathon performance, even at relatively low elevations.
- Air pollution, even within healthy limits, slows marathon times.

Behavior and marathon

- Finish times cluster around reference times, such as 3 hours.
- Better marathon runners train more overall, especially in zone 1.

CONCLUSION Natural experiments are useful for generating new and testing existing hypotheses, providing insights into marathon performance that may otherwise not be captured by smaller laboratory-based studies.

Downloaded from journals.physiology.org/journal/jappl (222.155.011.200) on August 17, 2026.