# Variability of sub2:30 Athens Classic marathon race performances

**Konstantinos Patras,** [1 2] **Jim-Dimitris Georgoulis,** [3] **Koulla Parpa,** [4] **Marcos Michalides** [4]

[1]Laboratory of Microbiology, Department of Laboratory Medicine, University Hospital of Ioannina, Ioannina 45110, Greece
[2]s11 Ltd, Nicosia 1096, Cyprus
[3]First Department of Orthopaedic Surgery, National and Kapodistrian University of Athens, Athens, Greece
[4]Sport and Exercise Science, UCLan University of Cyprus, Pyla 7080, Cyprus

**Bayesian modeling | Race to race variability | Performance enhancement**

## Headline

Analysis of the smallest worthwhile performance enhancement provides the opportunity to evaluate the magnitude and meaningfulness of different types of interventions (training, tapering, ergogenic aids). The smallest worthwhile enhancement is considered to be 0.3 x within athlete race-to-race variability (1). However, unlike standardized tracks running events, virtually all marathon courses, inevitably display differences (flat vs. undulating vs. hilly courses and courses with turns vs. unidirectional courses). These course-specific differences may necessitate race-specific CV's and subsequently, SWC's (2, 3).

## Aim

To estimate and evaluate the variability in overall race time for the Athens Classic marathon.

## Design

Analysis of official split times for ≤2:30 (hh:min) race times for male runners in the Classic Athens Marathon route during 2002-2022.

## Methods

Race times were downloaded from the official site of the Classic Athens Marathon (`https://www.athensauthenticmarathon.gr`) between December 2022 and January 2023. The final data set consisted of 259 sub2:30 race-performances from 185 runners during 20 Athens Classic Marathon races held during 2002-2022; the 2020 Athens Classic Marathon was canceled due to Covid-19. The current datasets were freely available on public domain; therefore we did not seek to obtain ethical approval. Race times analyses were performed in R (4) using tidyverse for data wrangling and visualization (5) and brms for hierarchical Bayesian modeling (6). Hierarchical modeling accounts for different sources of variation (7), and Bayesian modeling incorporates background information in the form of prior distributions to inform model estimation (7). These approaches were well-suited to the present study because they enabled estimation and evaluation within a multilevel framework. The population- (aka fixed-) effects included a linear trend for calendar year, accounting for a general improvement of performance as result of better training and technology, an effect for nationality (domestic vs. foreign runners) and the interaction between the linear trend for calendar year and nationality. The group- (aka random-) effects in the model were SD's and included athlete ID, to estimate pure differences between athletes' mean ability, year ID to account for the mean effect of environmental factors on performance times as well as differences between competitions mean times not accounted for by the fixed effects and the model residual (within-athlete race-to-race variability). Performance times were log-transformed to yield the effects and SD's in percent change of the mean. Within-athlete race-to-race, between-athletes and between-races CV's (variability) were calculated as the square root of the model residual, the square root of the athlete ID variance, and the square root of the year ID variance, respectively. We modified the ROPE approach (8) to derive exact probabilities as per the mass of the posterior distribution of the population-effects contained within thresholds of 0.3, 0.9, 1.6, 2.5, and 4.0, respectively, of the within athlete race-to-race CV (residual); when 95% of the posterior distribution was outside the above thresholds, the effects were deemed being small, moderate, large, very large, and extremely large (1). We sampled the posterior distribution using Hamiltonian Monte Carlo with four chains and 1500 postwarm-up samples per chain. The model passed all diagnostic statistics (all $\hat{R}$ values $\leq 1.01$, all effective sample sizes $> 400$, zero divergent iterations) (7).

## Results

The observed mean sub2:30 race performances were 145.75 ± 2.73min for domestic and 139.51 ± 4.77min for foreign runners (Figure 1A), whilst mean sub2:30 race performance across years ranged from 138.94±5.56min to 146.12±1.50min (Figure 1B). Group-level effects as CV's are presented in Table 1. The variability in total race to race time provided a smallest worthwhile enhancement of 0.75% (0.3 x 2.5%). The population mean expected race time at baseline was 144.88min for domestic runners and -2.2% (-3.7%; -0.7%, small effect) lower for foreign runners (Table 2). The population mean expected linear trend was 0.9%(-1.3%; 3.3%, trivial effect) for domestic runners and -4.5%(-6.7%; -2.4%, moderate effect) for foreign runners (Table 2). The mean expected difference in population linear trends between domestic and foreign runners was -5.6%(-8.5%; -2.8%, moderate effect) (Figure 2). The expected mean performance across years for the sub-population of domestic and the sub-population of foreign runners is presented in Figure 3.

## Discussion

There has been an emerging interest in estimating the variability of elite race performance (9). The substantial between-sports variation (0.3–7.0%) warrants estimation of sport-specific variability in order to derive the smallest worthwhile enhancement (9); however there may be also a need for course-specific SWC in marathon (3). The within-athlete race-to-race variability in total race time for the Athens Classic marathon is in the upper range to what has been reported for most endurance sports (0.3–2.4%) (9). It has been shown that proficiency influences within-athlete variability, with the best athletes showing lower variability than lower ranked athletes but here we did not tested for a differential within-athlete variability. Our simple model provided a clear indication for the relative performance progression (or lack thereof) in a nation's most prestigious marathon race (10). Despite the growing interest and promotion of the historical marathon route as well as advancements in sports science support and technology, the last 20 years have been characterized by no apparent improvement in domestic mean race performance. In fact our analysis revealed a ~56% probability for a small deterioration. On the other hand mean race performance for the foreign runners has produced a moderate improvement (Table 2, Figure 2, and Figure 3); potentially the organizers have been able to attract better foreign runners over the years.

**Table 1. Within- and Between-Athletes Variability in Performance Times Expressed as CV (%).**

| Within-athlete SD (%) | | |
|---|---|---|
| Race to race (mean±95%CI) | Between-athletes SD (%) (mean±95%CI) | Between-races SD (%) (mean±95%CI) |
| 2.5(2.0;2.9) | 1.3(0.2;2.2) | 0.7(0.2;1.2) |

**Table 2. Effects (back-transformed) provided by the Bayesian linear mixed model.**

| Effect | | ROPE small | ROPE moderate | ROPE large | ROPE Very large | ROPE Extremely large |
|---|---|---|---|---|---|---|
| Intercept | 144.88 | n/a | n/a | n/a | n/a | n/a |
| Nationality.foreign | -2.2% (-3.8%; -0.7%) | 0.6% | 51.3% | 100% | 100% | 100% |
| Linear year trend | 0.9% (-1.3%; 3.3%) | 38.8% | 88.2% | 100% | 100% | 100% |
| Nationality*linear year trend | -5.6% (-8.5%; -2.8%) | 0.0 % | 0.0% | 11.5% | 72.6% | 100% |

**Intercept**: mean expected performance at baseline race (2002) by the domestic runners; Nationality.foreign: the difference in mean expected performance at baseline race (2002) between domestic and foreign runners; Linear year trend: the (linear) change in mean expected performance between baseline race (2002) and last race (2022) for domestic runners; nationality*linear year trend: the difference in the (linear) change in mean expected performance between baseline race (2002) and last race (2022) between domestic and foreign runners. Percentages in ROPE represent the % of posterior distribution of each effect that lies within each limit. The magnitude of the effect is the one where coverage inside ROPE is <5%.

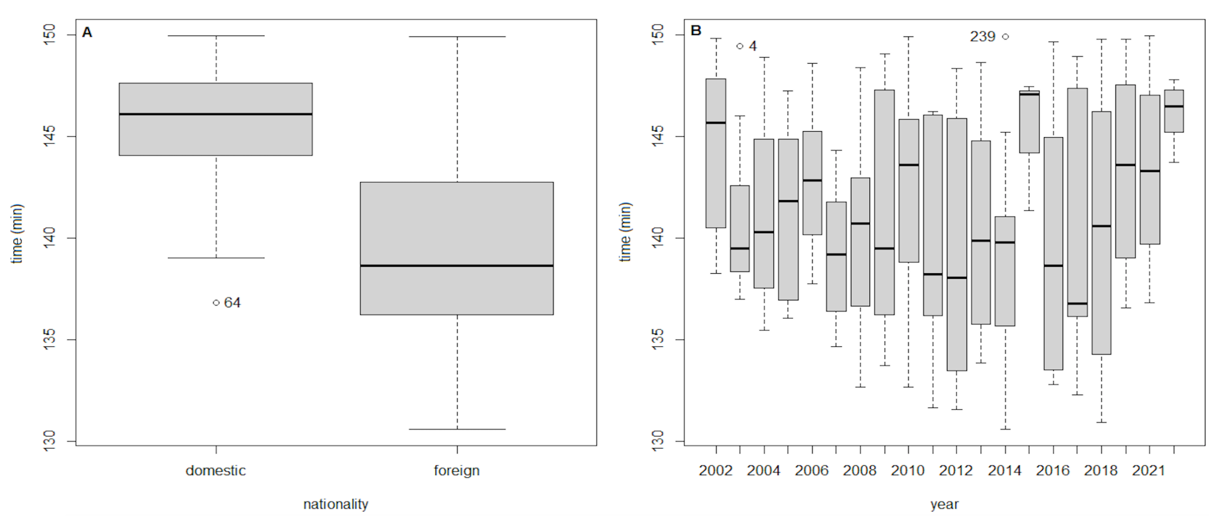


**Fig. 1. Central tendency and dispersion of observed mean performance for domestic and foreign runners.**

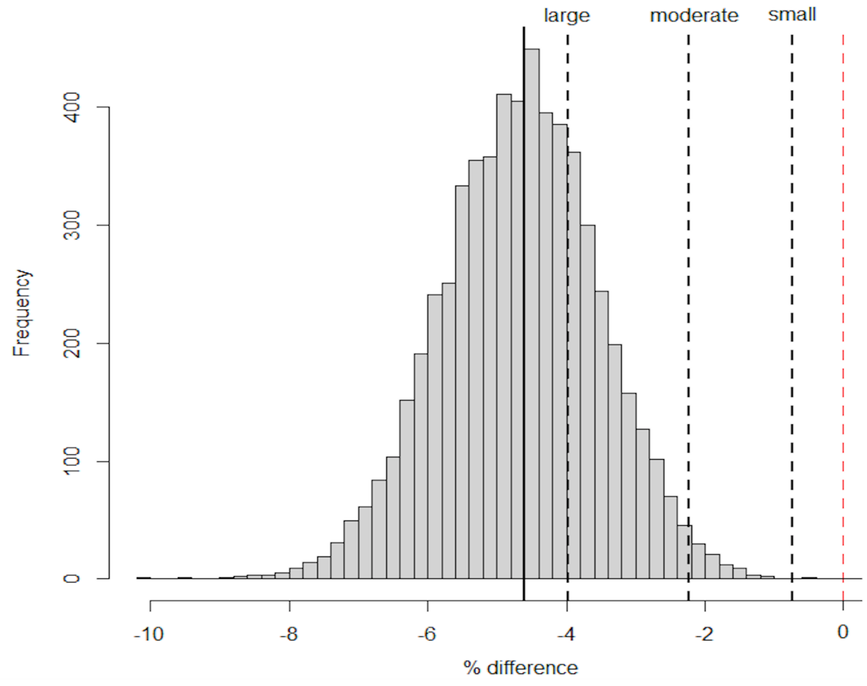


**Fig. 2. Posterior distribution of plausible values for the difference between the linear trend for domestic and foreign runners. The solid black line is mean of the posterior distribution of the effect, the red dashed line denotes zero and the black dashed lines denote small, moderate and large thresholds.**

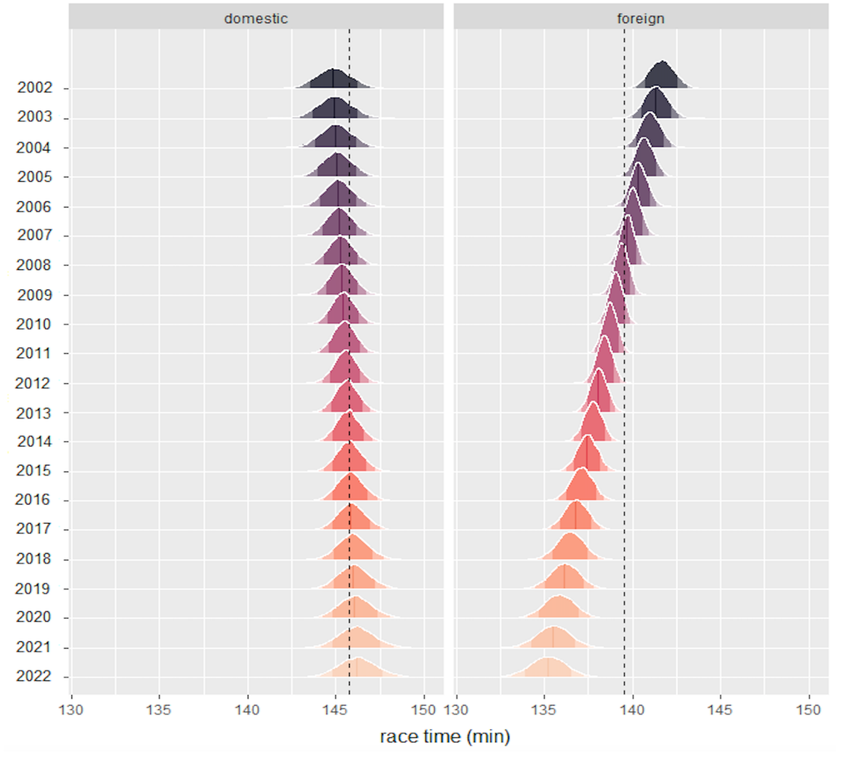


**Fig. 3. Posterior distribution of the expected year-to-year mean 2:30 performance for the domestic and foreign groups. The dashed lines denote the observed mean for each group respectively.**

## Practical applications

- The typical variation in sub2:30 Athens Classic marathon performances is $\sim$2.5%; physiological tests suitable for tracking SWC in performance need typical errors of measurement $\leq$2.5%.
- Sub2:30 Athens Classic male marathoners should focus on training interventions, ergogenic aids equipment that that enhance performance by at least 0.75% for a substantial effect on race time.
- Greek athletic governing bodies could scrutinize physiological characteristics of contemporary and past national level runners as well as current and past selection processes and training and support practices in an attempt to increase the pool of 2:30 marathon runners (10).

## Limitations

- The limited number of races for many foreign athletes may have impacted the results.
- The effect of age on athlete performance could not be recorded properly, thus it was not included in the model (3).
- The reported within-athlete variability accounts specifically for the race to race performance for consecutive AMA's and cannot be generalized to other races or courses.

**Twitter:** Konstantinos Patras (@KostasPatras_MD)

## Conflicts of interest

The authors declare no financial, institutional, and/or personal conflicts of interest that might inappropriately influence actions or statements of the present study.

## Accompanying dataset

Data and code will be made available to all interested researchers upon request from the corresponding author (KP).

## References

**1.** Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc, 2009; 41(1):3-13.

**2.** Hopkins WG, Hewson DJ. Variability of competitive performance of distance runners. Med Sci Sports Exerc, 2001; 33(9):1588-1592

**3.** Maffetone PB, Malcata R, Rivera I, Laursen PB. The Boston Marathon versus the World Marathon Majors. PLoS One, 2017; 12(9):e0184024

**4.** R Core Team. R: A language and environment for statistical computing, (2021).

**5.** Wickham, H. et al. Welcome to the Tidyverse. J. Open Source Softw, 2019; 4: 1686.

**6.** Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. J Stat Softw, 2017; 80(1):1-28.

**7.** Johnson AA, Ott MQ, Dogucu M. Bayes Rules! An introduction to applied Bayesian modeling. Chapman and Hall/CRC, 2022.

**8.** Kruschke JK. Doing Bayesian data analysis: A tutorial with R and BUGS. (2nd edition). Elsevier Academic Press, 2014.

**9.** Malcata RM, Hopkins WG. Variability of competitive performance of elite athletes: a systematic review. Sports Med, 2014; 44(12):1763-1774

**10.** Allen SV, Vandenbogaerde TJ, Hopkins WG. The performance effect of centralizing a nation's elite swim program. Int J Sports Physiol Perform, 2015; 10(2):198-203.