Editorial:

Drowning in a river with an average depth of three feet: Interpreting athletic performance gains

Running Title:

Interpreting athletic performance gains

Author:

Connie C.W. Hsia, M.D.
Department of Internal Medicine
University of Texas Southwestern Medical Center
5323 Harry Hines Blvd.,
Dallas, TX 75390-9034
Email: Connie.Hsia@utsouthwestern.edu
Performance optimization in competitive sports is about out-racing fellow competitors, setting the next official record, and striving for excellence under any prevailing conditions. Athletic events held at a moderately high altitude present extra challenges for performance enhancement, as a multitude of hypoxia-induced physiological perturbations – fluid shifts, acid-base dysequilibrium, sympathetic stimulation, ventilatory acclimatization, gas exchange limitation, and cerebro-cardiovascular and metabolic adjustments – arise concurrently. For those athletes who are unable to spend weeks at altitude to fully acclimatize before a race, is there any competitive advantage to be gained by spending one night at altitude before competition to allow some degree of acclimatization and performance advantage, compared to arriving on race day just before competition and minimizing the negative effects of overnight hypoxia exposure? Optimizing the timing of altitude arrival strategy holds specialized significance for athletes in a variety of sporting events.

In this issue, Foss et al (1) compared cycling performance after two short-term arrival strategies at simulated altitude. Using a paired design, ten cyclists were exposed to normobaric hypoxia (16% O$_2$, equivalent to 2,500 m altitude) for 14 h overnight or 2 h before undergoing a 20-km cycle ergometry time trial in normoxia or hypoxia. Subjects were blinded to the inspired gas and the trial order was randomized. The goal was to determine which arrival strategy yield the superior competitive advantage. Results showed no significant difference in average trial completion time between 2 h and 14 h of pre-race hypoxia exposure (mean ~36.7 min, range ~33 to 41 min), leading to the conclusion that the two arrival strategies are equivalent. A closer look at individual data points reveals that 7 of 10 athletes showed little difference in trial times with either arrival strategy, thus supporting the conclusion. However, 3 athletes exhibited 1-2 min of absolute difference in trial times (two faster with 14 h and one faster with 2 h pre-race hypoxia exposure). Changes in plasma volume, ventilatory parameters, and overnight sleep
quality assessed by a questionnaire, activity monitor and pulse oximetry, could not explain the divergent responses among subjects.

This study was carefully conducted and the results clearly reported. Nonetheless, there may be more than one way to interpret the data. A finding of non-difference in a small cohort raises the question of statistical power. Was the study adequately powered to detect an appropriate effect size that addresses the outcome goal? The “appropriate” effect size must be quite small as even split-second difference could separate victory from defeat. Deriving from the literature a 1.2% coefficient of variation in repetitive 20-km trial times conducted in normoxia, the investigators estimated a need of 12 subjects to detect 1% average time trial difference (about 22 s) between the two arrival strategies in hypoxia. The report of 10 subjects fell short of this target. The investigators expressed confidence that given a high p value (0.998) for the average trial times between the two arrival strategies, the lack of mean performance difference cannot be due to a lack of sample size. The sample size estimate assumed similar performance reproducibility in hypoxia as in normoxia; which has not been established. Upon acute hypoxic exposure, every step of oxygen transport is thrown into non-steady state; subjects exhibit differential sensitivity to these perturbations and acclimatize at dissimilar rates. Multiple minute hypoxia-induced perturbations could cumulatively undermine the predictability of response to intervention regardless of group means. More importantly, reliance on statistical characterization can potentially obscure the essential feature of athletic competition, i.e., recognition of individual deviation from the mean. The 1-2 min performance gain with one or the other arrival strategy in 3 out of 10 “outlier subjects”, as termed by the investigators, may signify the winning edge sought by this study.

In terms of practical application, which arrival strategy should be recommended for endurance race at moderate altitude? The study is informative in indicating that there is no superiority in either method, which holds for 7 out of 10 athletes. The other 3 “outlier” athletes would disagree
with the conclusion based on group means, but how could they be identified unless the two
strategies are compared in all athletes who are candidates for this enhancement? If 30% of
athletes can selectively gain a meaningful advantage, perhaps the optimal arrival strategy
should be individually determined for all athletes. Coaches may find it desirable to offer more
than one arrival strategy in a large team. The investigators did acknowledge that extreme
individual responses may ultimately influence choice of the arrival strategy. The study results
are not entirely conclusive because the emphasis on aggregate statistics did not match the
personalized outcome goal. Future investigation could exert greater impact by recruiting a
larger cohort, validation of performance reproducibility in acute hypoxia, determining individual
variance in time trial performance at both levels of inspired oxygen, and applying more sensitive
measures of sleep architecture, disordered breathing, and hypoxic ventilatory drive to discern
the sources of intra- and inter-subject variability.

The above discussion underscores two fundamental caveats in general data interpretation:
First, group means are routinely used to draw conclusions and infer the course of action for
individuals. "Does the good of the many outweigh the good of the one?" (Spock’s mother, Star
Trek IV). In this case at least, the answer is “no”; supra-average response is pre-requisite in
pursuit of the gold medal. Second, in an era of information overload and limited attention span
where published results are increasingly disseminated and consumed as snazzy headlines and
condensed summaries, it remains important to scrutinize discrete absolute measurements,
especially the “outliers”, as they offer clues and insight not available from averages, percent
changes and arbitrary p values. One could indeed drown while wading across an averagely
shallow river, getting in trouble with variance at the second moment.