Pediatric Exercise Science, 2019, 31, 382-385
https://doi.org/10.1123/pes.2019-0119

Human Kinetics
LETTER TO THE EDITOR

Interpreting Youth Aerobic Fitness: Promoting Evidence-Based Discussion–A Response to Dotan (2019)

Neil Armstrong and Jo Welsman
University of Exeter

We welcome Raffy Dotan's Letter to the Editor (14) as it gives us another opportunity to promote evidence-based discussion of the development of youth aerobic fitness. Readers of our contributions to the 2019 Special Issue of *Pediatric Exercise Science* (6,27,28) will recall that we concluded with, "The authors encourage all pediatric exercise scientists to engage with this discussion, to share ideas and methods, and be willing to explore alternatives. There are many issues to resolve and constructive, collaborative debate will speed our collective aim toward a better understanding of pediatric aerobic fitness in health and disease" (27, p. 256). Not the words of authors preaching a "gospel" with "evangelistic persistence" as Dotan (14) suggests, but of scientists genuinely seeking to stimulate evidence-based discussion of the development of youth aerobic fitness and its relationship with health and well-being.

Dotan raises a number of issues which we will address in the order he presents them. We will illustrate discrepancies between his Letter and papers in the Special Issue through direct quotations and clarify points arising from his exploration of scaling peak $\dot{V}O_2$ for body size.

Dotan states that we "ought to be credited with taking up the allometric challenge that had been raised, with little consequence, by Tanner" (14). Not so, we acknowledged that "The issues we raise are not new, having been eloquently and completely described by visionary scientists since 1949" (28, p. 189). Our early work in this area was enriched by collaboration with several visionary scientists including Drs Nevill and Winter (eg, 11,34,35), who have themselves provided numerous innovative contributions to the debate (eg, 19,20,37).

Dotan suggests that ratio scaling is "still the de facto scaling method of choice in 2019" because of its "hard-to-resist conceptual and practical simplicity" (14). As we stated in our Commentary, "We know of no other scientific discipline where an assumed relationship albeit 'convenient and traditional' has become the acceptable alternative to rigorous scientific justification" (28, p. 189). Ultimately, however, the principal reason that ratio scaling is still used to control for body mass in the vast majority of publications is because journal reviewers and editors seldom (if ever) require contributors to provide a scientific rationale and statistical justification for its use in the context of the submitted paper. By contrast, in our experience, journals always (quite rightly) require a scientific rationale and statistical justification for the use of other scaling methods such as multilevel allometric modeling. Why do editors and peer reviewers recognize peak $\dot{V}O_2$ ratio scaled with body mass as an acceptable alternative to rigorous statistical analysis despite, as noted in the Special Issue by Blais et al, "overwhelming scientific evidence of its many drawbacks" (12, p. 254)? We challenge editors and peer reviewers to be consistent, to treat all scaling methods with the same scrutiny, and to require authors to justify all statistical methods of size-adjusting peak $\dot{V}O_2$ (and other physiological variables) before publication.

Dotan asserts that potential users shy away from allometric scaling as, "Currently, there simply is no credible method of doing so, and Armstrong and Welsman do not bring us any closer to having one" (14). But inexplicably in his own paper in the very same Special Issue Dotan asserts, "A recent commentary by Welsman and Armstrong provides the background and methodological justification for the use of allometric scaling in children. The authors provide the scientific and statistical rationale and demonstrate that with the use of appropriate body mass exponents, the effect of body mass can effectively be removed" (15, p. 149). Which assertion is correct? We and others demonstrated the credibility and simplicity of allometric scaling of cross-sectional data with multiple covariates based in log-linear regression in this journal decades ago (eg, 18,33,36) and followed these contributions with regular tutorial updates (eg, 19,32,37). In an undergraduate textbook, Winter and Nevill (38) provide a scaling tutorial followed by step-by-step instructions of how to allometrically scale cross-sectional data using SPSS (IBM, Portsmouth, United Kingdom).

Dotan refers to our use of multilevel allometric modeling as "rather curious". But, we have been publishing multilevel allometric modeling analyses of longitudinal aerobic, anaerobic, physical activity, and muscle strength data in this journal and others since the turn of the century (eg, 9,10,13,23). It is a well-established and rigorous method of analysis which enables the effects of variables such as age, body mass, fat-free mass (FFM), and maturity status to be partitioned concurrently within an allometric framework to provide a flexible and sensitive interpretation of, in the present context, peak $\dot{V}O_2$ (eg, 2,11,20). It is also easily performed using a freely available computer program, and the regularly revised and updated User's Guide to MLwiN (21).

Dotan queries (and picks up the same thread later) why factors other than body mass were not addressed in our Commentary. The reason could not be clearer as we stated that, "Our Commentary specifically aimed to highlight the pitfalls of ratio scaling for a single anthropometric variable—body mass—in cross-sectional data, as this is the method that underpins the vast majority of published studies describing pediatric aerobic fitness. Our concern is that these studies are being used in the reporting of spurious relationships with health-related variables and to generate erroneous norms and 'clinical red flags'" (27, p. 256). Moreover, we stressed that "It is not the aim of this paper to provide a detailed statistical tutorial for scaling methods, as these are available

The authors are with the Children's Health and Exercise Research Centre, University of Exeter, Exeter, United Kingdom. Armstrong (N.Armstrong@exeter.ac.uk) is corresponding author.

elsewhere, or to extensively review the literature" (28, p. 185), and we provided for interested readers references to 3 tutorial papers where we have discussed a range of methods of scaling both cross-sectional and longitudinal data from first principles (30–32).

Dotan's comment on our factually correct statement, "If the simple ratio, peak $\dot{V}O_2$ divided by body mass, effectively removes the effect of body mass then the ratio standard, $mL \cdot kg^{-1} \cdot min^{-1}$, should not remain significantly correlated with body mass" (28, p. 187), that "the statement is unnecessary or even misleading" (14) is baffling. Quite simply, some researchers incorrectly assume that division by body mass removes the effect of body mass and exploring this correlation is a simple first check to see if the influence of body mass has indeed been removed or controlled for. The same principle applies regardless of whether we are interested in ratio-scaling peak $\dot{V}O_2$ (or other physiological variables) with body mass, fat mass, or FFM. Moreover, this and Dotan's subsequent comment, "It would have been more to the point if the authors stated that ratio scaling does not remove the influence of fat mass, rather than body mass" (14), do not resonate with comments in his own contribution to the very same Special Issue where he wrote that, when reporting $\dot{V}O_2max/peak$, "Due to the large changes it undergoes during maturation, properly accounting for body mass is of pronounced importance in youth" (15).

Dotan wonders why we did not factor in other covariates and indicates that this is something we have not addressed anywhere. As is absolutely clear from reading our Special Issue contributions, we do not recommend scaling youth aerobic fitness with one covariate (eg, body mass). We not only explicitly make this point (eg, "there is a need to concurrently control for age and maturational effects" [27, p. 256]), but also provide a specific example in our Commentary to show how adding age into a log-linear regression equation as an additional covariate describing 11- to 16-year-olds' peak $\dot{V}O_2$ yielded a significant term and the value of the body mass exponent reduced from 0.94 to 0.74 (29, p. 188). The addition of covariates other than body mass into allometric analyses has been in common use in pediatric exercise science by ourselves and others for over 20 years (eg, 16,17,34). Notably, in his Special Issue paper (15), Dotan cited one of our papers from 1996 in which we factor in stature as an additional covariate to body mass. In that paper, we explained the conditions necessary for intergroup comparisons (see later for discussion of comparisons of boys and girls) and demonstrated that, in direct conflict with ratio scaling, body size appropriately controlled for peak $\dot{V}O_2$ in boys increased with age through the teen years to age 21 years and peak $\dot{V}O_2$ in girls increased from 11 to 13 years before leveling-off to age 22 years (34).

To progress the debate and to highlight the limitations of interpreting youth aerobic fitness in ratio with a single covariate, we emphasized that youth aerobic fitness develops in accord with sex-specific, age-, and maturity-status-driven concurrent changes in a range of morphological and physiological covariates whose timing and tempo of changes are specific to individuals. We stated clearly that, "In our opinion, to fully understand pediatric aerobic fitness, longitudinal data are required" (27, p. 256), and we addressed this empirically through the use of multilevel allometric modeling of large longitudinal data sets where multiple, individual growth trajectories were examined in relation to the development of aerobic fitness (2,6). In our multilevel allometric models, we analyzed concurrently age, maturity status, body mass, and skinfold thicknesses. The outcomes were that age- and maturity-status-driven changes in body mass and skinfold thicknesses (as a surrogate of FFM) or estimated FFM (25) were the most powerful morphological influences on both aerobic fitness and short-term power output, independent of whether they were determined using treadmill running or cycle ergometry (2,5,6,8).

Dotan reports previously published body mass exponents of the resting metabolic rate (RMR) of birds and animals, which he then compares to RMR and $\dot{V}O_2max$ in human adults and children. We recommend Schmidt-Neilsen's (24) excellent book to those interested in the importance of animal size but fail to see the direct relevance of "the mean allometric body mass exponent, for 23 species of wild and domestic animals, to be 0.845, which is quite different than the 0.67 to 0.75 typically observed range of the corresponding RMR exponents" (14) to our papers. We have not empirically examined body mass exponents relating to RMR and peak $\dot{V}O_2$ in wild and domestic animals but, according to Rowland (22), there is sufficient ambiguity in the interpretation of changes in metabolic scope in human children to preclude its use as a size-independent marker of fitness.

Dotan (14) asks why we did not compare "derived exponents between highly fit or trained individuals and their untrained or low-fitness peers". The answer again is quite simply that it was not an objective of the Commentary, which was clearly stated—see earlier quotation (27, p. 256). In addition, we described the recruitment of participants in our studies, and none of the reported data sets consisted of highly fit or trained children. However, in their seminal paper on multilevel allometric modeling of elite youth athletes in training, Nevill et al (20) reported on the effects of sport training status on aerobic fitness, and body mass exponents were identified of 0.67 for boys and 0.70 for girls. Any effect of fitness between the differing sports was observed in either the magnitude of the intercept term or the interaction of sport × age.

Dotan claims that we convey the message that allometry is "*the ultimate and the only appropriate scaling methodology*" (14). Absolutely not. We encourage engagement with evidence-based discussion, sharing of ideas, and willingness to explore alternatives (see earlier quotation, ref 27, p. 256), and we welcome with open arms the introduction and development of rigorous alternative methodology. This is illustrated in our response to Blais et al (12), "The normalization method suggested by Blanchard et al warrants attention as it clearly addresses many of the statistical pitfalls of ratio scaling and enables estimation of submaximal as well as maximal indices of fitness derived from cardiopulmonary exercise testing" (27, p. 256).

Dotan then moves into another area outside the scope of our papers and he comments on "performance capacity in activities in which the carried weight of the body is a major determining factor" (14). We are, however, happy to respond and agree that body mass per se is informative in relation to the performance of youth track athletes as the greater the body mass the more work done transporting it over a set distance. The 20-m shuttle run test provides us with a prime example of the difference between performance and aerobic fitness. Dotan points out in his Special Issue paper that only muscle mass "directly contributes to oxygen uptake, whereas fat mass constitutes functional 'dead weight'" (15, p. 149). So, in a 20-m shuttle run test, carrying increased fat mass (dead weight) increases the work done per shuttle and adversely affects shuttle running performance but does not affect aerobic fitness. This, of course, confounds any alleged relationship between peak $\dot{V}O_2$ predicted from 20-m shuttle run test performance and true peak $\dot{V}O_2$ (4,7,26).

Dotan lingers with "performance capacity" and distinguishes between exercise in which body mass is carried and exercise in which body mass is supported. He asserts that, "Although body mass may not be the only relevant factor in activities of this nature, simple ratio scaling is still the best scaling methodology to reflect functional