Cautionary comments on the measurement of maximum locomotor capabilities

Jonathan B. Losos*, Douglas A. Creer and James A. Schulte II**

Department of Biology, Campus Box 1137, Washington University, St Louis, MO 63130-4899, U.S.A.

Abstract

In recent years, laboratory measurements of locomotor performance of small organisms have been used increasingly to study questions in behavioural, physiological, and evolutionary ecology. Although a valuable tool that can provide insight into organismal function, behaviour, and adaptation, this approach is potentially limited by the quality of the data collected. Using the study of maximum sprint speed in lizards as an example, common problems in methodology that may confound interpretation of performance studies are discussed.

Key words: lizard, locomotion, performance, sprint speed, methodology

INTRODUCTION

An important development in the field of functional morphology has been emphasis on the measurement of ecologically relevant aspects of whole organism performance, such as maximum sprinting speed and jumping ability, particularly in small vertebrates (Huey & Stevenson, 1979; Bennett, 1980; Huey & Bennett, 1986). A central theme in this work has been that the relationship between phenotype and environment can best be discerned by understanding the effect that variation in phenotype has on whole organism functional capabilities and, in turn, the effect that variation in these capabilities has on how organisms interact with the environment (Huey & Stevenson, 1979; Arnold, 1983; Wainwright, 1988). This approach has been applied particularly extensively in the study of the ecological and evolutionary context of locomotion in a wide variety of organisms (much of this literature is reviewed in Wainwright & Reilly, 1994). These studies have focused on maximal performance capability because one can predict how variation in phenotype should translate into variation in maximal capabilities, which, in turn, should translate into differences in how the organism interacts with the environment; by extension, over evolutionary time, one would expect that selection thus would operate on maximal performance capabilities. The success of this approach is evident in the many studies that have established mechanistic relationships linking phenotype to environment with performance as the intermediate link (Wainwright & Reilly, 1994).

Nevertheless, estimating the maximal capabilities of an organism is not a straightforward task. We believe that serious problems in estimating such capabilities frequently occur in the literature and potentially call into question the results reported therein. Here these problems are discussed, focusing on the measurement of maximal sprint capabilities in lizards. Despite the many studies of lizard sprint speed (reviewed in Bennett, 1987; Bennett & Huey, 1990; Garland & Losos, 1994), the problems discussed in this paper are rarely mentioned in the literature and have never been discussed extensively. Although our comments pertain particularly to these studies, the points are general and should pertain to most measurements of ecologically relevant performance capabilities in many organisms.

METHODS FOR STUDYING MAXIMAL SPRINT ABILITY IN LIZARDS

Researchers usually measure maximal sprint ability by inducing a lizard to run along a narrow track for a distance of 1–4 m. Sprint speed is measured either by frame-by-frame analysis of video images (e.g. Fieler & Jayne, 1998; Irschick & Jayne, 1998; Jayne & Ellis, 1998) or by computers calculating the time elapsed as the lizard breaks light beams arranged at regular intervals along the track (Huey *et al.*, 1981; Miles & Smith, 1987). Less frequently, speed is measured by placing a lizard on a treadmill and increasing the speed until the lizard cannot maintain position (e.g. John-Alder, Garland & Bennett, 1986; Dohm *et al.*, 1998). One to six

*All correspondence to: J. B. Losos.
E-mail: losos@biology.wustl.edu
** Present address: Division of Amphibians & Reptiles, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, U.S.A.

trials are conducted per day, sometimes with additional days of similar testing. Occasionally, lizards are trained for 1 to several days before beginning data collection. The fastest recorded speed in any trial is usually used as the estimate of maximal sprinting ability for each individual.

PROBLEMS WITH THE ESTIMATION OF MAXIMAL SPRINTING ABILITY

The ease and simplicity of this approach surely has been one of the reasons it has been so widely used. None the less, two simple flaws can be seen in many published papers: (1) data are included from individuals obviously performing sub-maximally; (2) not enough trials are conducted to provide an accurate estimate of maximal capabilities.

INCLUSION OF INDIVIDUALS PERFORMING SUB-MAXIMALLY

The major problem bedevilling studies of performance is motivation – for whatever reason, some individuals may not use their maximal capabilities in laboratory trials (Bennett & Huey, 1990; Garland & Losos, 1994; see also Pough, 1989). Anyone who has conducted sprint trials with lizards is aware that some lizards in at least some trials will run at clearly sub-maximal speeds. This conclusion is obvious from observations of the gait and dynamics of limb movement, just as it is obvious that a human jogger is not sprinting at top speed. Moreover, some individuals are consistently unmotivated and perform sub-maximally trial after trial.

Our primary point here is simple: if the goal of performance studies is to estimate maximal capabilities, data from individuals obviously performing at less than their maximal capabilities must be excluded. None the less, a review of the literature indicates that many published studies do not follow this prescription. In a review of 65 studies on lizard sprinting published during 1979–99, some studies clearly indicate that individuals performing sub-maximally are excluded from the analysis; in others, the reader can infer that this has occurred. However, in nearly half (49%) of the studies, no mention is made of how individuals performing sub-maximally were treated and some papers make clear that data from all individuals are included. Of course, it is possible that some species never produce sub-maximal trials. In our experience with a wide variety of species, including some of those included in these studies, such an occurrence is rare.

The inclusion of data from individuals performing sub-maximally can invalidate studies in several ways. First, such data may obscure real relationships between phenotype and maximal capability. In many studies that do not exclude sub-maximal trials, hypothesized relationships are not found. Such findings may be correct, but they also could result from the inclusion of poor data. For example, as part of a larger study (Losos, 1990), the sprint speeds of three species of lizards, *Anolis cuvieri*, *A. evermanni*, and *A. krugi*, were measured (we use data from our own studies rather than criticizing particular studies for what seems to be a widespread practice). Among species of anoles, sprint speed and body size are strongly related, and *A. cuvieri* is eight times the mass of *A. evermanni* and 20 times that of *A. krugi*. None the less, if sprint measurements from all individuals are included, no interspecific difference is detected $(F_{2,51} = 0.05, P = 0.95)$. By contrast, when individuals that produced only sub-maximal trials are excluded, *A. cuvieri* is found to be significantly faster than the other two species $(F_{2,46} = 3.42, P = 0.04)$. In this case, only a few individuals consistently performed sub-maximally, but in some species, many, even a majority, do so.

Second, to the extent that some classes of individuals (e.g. sexes, species, populations) differ in motivation in a laboratory setting, then artefactual relationships may be created. For example, if two species actually do not differ in maximal sprint capabilities, but if the experimental set-up motivates only one of them to sprint at maximal capacity, then one might mistakenly conclude that one species was faster than the other (see also Sorci *et al.*, 1995).

This is not to say that data from individuals performing sub-maximally has no value. Quite the contrary, differences in laboratory performance sometimes may reflect important behavioural differences relevant to the conduct of organisms in nature. For example, some arboreal lizards hop, rather than run, on a horizontal racetrack (van Berkum, 1986). Similarly, cold lizards of several species will stand their ground and attempt to bite, whereas warm lizards are much more likely to run quickly down a laboratory racetrack (Hertz, Huey & Nevo, 1982; Crowley & Pietruszka, 1983). If differences in performance in the laboratory reflect behavioural differences in the field, such findings are informative, even if they do not represent estimates of maximal capabilities.

The obvious conclusion is that one must be clear about what one is trying to estimate in laboratory studies of performance. If estimating maximal capabilities is the goal, then sub-maximal trials need to be excluded. Some individuals may produce nothing but sub-maximal trials; those individuals must be removed from the study. This conclusion leads to the obvious question: can researchers objectively determine when a lizard is performing sub-maximally? We believe that the answer is 'yes'. Often, sub-maximal trials will be obvious, just as anyone can distinguish horses that are trotting from those that are galloping. Thus, the adoption of a set of criteria to distinguish sub-maximal trials from good trials is recommended. These criteria, which must be based on the manner in which the lizard runs and not the speed it attains, will vary from species to species. Possible criteria might include factors such as the placement of the limbs and the extent to which the body is lifted off the substrate (many lizards when

running rapidly will use a moderately erect posture with the body well off the substrate, whereas conspecifics running at lower speeds will use a more sprawling posture), the consistency of the run (as opposed to trials in which the lizard moves in a jerky, start-and-stop fashion), and whether the lizard ran into the wall of the track or stumbled. Video technology has improved greatly in recent years (e.g. Irschick & Jayne, 1999) and could be used to indicate whether, for example, lizards were using digitigrade posture and maximizing effective limb length during a run (although this would make sprint performance trials considerably more labour intensive).

This approach will not be perfect, of course. Identifying a lizard running at 50% of its maximal capabilities may be easy, but distinguishing individuals running at 90% will be more difficult. Inevitably, some lizards running close to, but not at, top speed will be included. None the less, even if eliminating all sub-maximal trials is not possible, getting rid of some bad trials is better than none at all, particularly given that the worst trials (relative to maximal abilities) will be the ones most likely to be excluded. Thus, as long as reasonable criteria can be developed and consistently applied, this approach is preferable to simply using all data when some are clearly faulty.

To some, the idea that sub-maximal trials should be identified and excluded may seem fraught with subjectivity. One researcher in the field commented, 'What the authors are in fact suggesting is tailoring the data, using subjective criteria to accept or reject an individual observation. I do not think that you can or should reject an animal's performance because it does not meet some internal standard of grace and fluidity.' This is not an unreasonable position. None the less, the same researcher acknowledged that some lizards in at least some trials produce runs that clearly are well below their maximal capabilities. In this light, we question how desirable is a research protocol that requires that data that are obviously deficient and inaccurate be retained and analysed. We feel that researchers can be unbiased in deciding when to exclude data and that such efforts are preferable to including bad data.

The researcher cited above disagrees with this prescription, fearing that it is an invitation for workers, either consciously or unconsciously, to discard data that do not support their ideas. This is a valid concern, but the establishment of specific criteria should be sufficient to prevent at least unconscious bias. Moreover, many experiments can be designed such that the decision to include or exclude a trial is made by someone blind to whether exclusion of the trial would favour or work against the hypothesis being tested.

Although one can never know that measurements do represent maximal capabilities, retrospective confirmation may exist if relationships between phenotype and performance, which are predicted to exist on mechanistic grounds, are actually discovered (Bennett & Huey, 1990; Garland & Losos, 1994). For example, as one would expect, sprinting ability is correlated with amount of muscle fibre in desert iguanas (Gleeson & Harrison, 1988) and with relative limb length in several iguanid lizards (Miles, 1994). Unless motivation is correlated with characteristics such as muscle fibre or relative limb length (which is at least conceivable, but perhaps not likely), then a statistically robust relationship between morphology and performance would strongly suggest that maximal sprint speed has been measured. Of course, we do not mean to imply the converse: failure to find such a correlation should not be taken as retrospective evidence that lizards were performing sub-maximally.

One might think that measurements of repeatability would get around the problem of recognizing sub-maximal performance. However, this may not be a solution, because motivation may well be highly repeatable, as van Berkum & Tsuji (1987) demonstrated for hatchling *Sceloporus occidentalis*. Hence, a high repeatability may simply indicate that some lizards repeatedly run sub-maximally. For example, in *A. krugi*, 56% of all trials were scored as 'good' (the highest rating). Given that five trials were conducted per individual, we would expect, based on a binomial distribution, that if the likelihood of performing maximally did not differ among individuals, most lizards would produce two to three 'good' trials. In fact, a majority of lizards produced either five or fewer than two 'good' runs. Clearly, motivation varies among individuals; some lizards always run well and others rarely do so. Studies that intend to investigate the relationship between phenotype and maximal abilities thus must exclude the poorly motivated lizards because their slow speeds result not from phenotypic deficiencies (e.g. shorter legs, smaller muscle mass), but from behavioural tendencies.

INSUFFICIENT NUMBER OF TRIALS

As any fan of track and field knows, even individuals attempting to perform maximally will produce variation in realized performance from trial to trial. Consequently, if one wants to produce an estimate of an individual's maximal capabilities, then multiple trials should be conducted. Ideally, one would repeatedly measure performance until an asymptote is reached (assuming that performance does not decline with time, e.g. Hertz *et al.*, 1983). In reality, logistic constraints limit the number of trials that are feasible, but many studies measure performance in five or more trials. By contrast, more than one-fifth of the studies we surveyed used only one or two trials. It seems probable that in studies with few trials, maximal performance for some individuals may be severely underestimated. For example, in *A. krugi* the correlation between maximum speed estimated from the first two trials *vs* that estimated from all five trials is only 0.79 (using only 'good' trials). Although significant, this correlation indicates that more than one-third of the variation in maximum sprint speed estimated from five trials cannot be predicted by the estimates based on two trials.

The consequences of such underestimates of maximal capability are fairly obvious. First, within species, attempts to correlate maximum performance capability with phenotypic characteristics will be thwarted if maximal abilities are underestimated, particularly given that by chance in a limited number of trials, some individuals will perform near their absolute maximum and others may not. Second, comparisons among groups may be affected. On the one hand, real differences may be obscured. For example, when maximum performance is estimated for the three anole species discussed above using only the first two trials (only including those individuals that produced 'good' trials), then interspecific differences disappear ($F_{2,40} = 1.18$, $P = 0.31$; sample sizes smaller than the previous analysis because some individuals did not produce 'good' runs in the first two trials). On the other hand, as discussed previously, artefactual differences may be created if some species are more consistent than others in performing at their peak such that, in relatively few trials, they are more likely to produce speeds close to their maximal capabilities.

We do not believe that any magic number of trials exists that will guarantee acceptable measurements of maximal sprint speed. Rather, it is recommended that researchers conduct as many trials as logistically possible (with the welfare of the animal being one important consideration). None the less, four to six trials, the number of trials used in early lizard studies, certainly must be superior to conducting only one or two trials.

CONCLUSIONS

One might argue that these problems are so obvious that they must no longer occur. Our review of the literature does not support this belief. Of 11 studies of sprint performance in 1997–99, only four explicitly excluded trials in which lizards were running sub-maximally and six used only two trials per lizard. Thus, the issues raised are very much evident in the work currently being published.

Alternative interpretations of the data obtained are available for the major conclusions of many of these recent papers, although none of the papers acknowledged the existence of possible alternative explanations. For example, the failure to find differences in maximal capabilities between species or between lizards with and without tails could result, not because no real differences in maximal capabilities exist, but because such differences were obscured by data from individuals not exhibiting maximal capabilities or because the capabilities of some individuals were poorly estimated from too few trials.

Similarly, some studies found that two morphologically very similar species displayed great differences in sprint speed or that the sprint speed of females of two very similar populations responded to gravity in fundamentally different ways (gravity increases sprint speed in one population and decreases it in the other). These differences might result from phenotypic differences undetected by the investigator. But, an equally plausible alternative explanation is that the taxa are behaviourally distinct such that they respond to laboratory conditions differently, producing apparent differences in maximal sprint capabilities. Although the genesis of such behavioural differences could be an interesting topic in its own right, this possibility is rarely mentioned, much less investigated.

One of the attractions of studying whole-organism measures of performance is that it is a relatively easy means by which field biologists can attempt to understand phenotype–environment correlations (Huey *et al.*, 1981). Many applications of this methodology have been successful in providing insights into ecomorphological correlations and the workings of natural selection (e.g. Wainwright & Reilly, 1994). None the less, workers should endeavour, when possible, to address these potential pitfalls more carefully; estimates of maximal capabilities are only as reliable as the data upon which they are based.

REFERENCES

Arnold, S. J. (1983). Morphology, performance, and fitness. *Am. Zool.* **23**: 347–361.

Bennett, A. F. (1980). The thermal dependence of lizard behaviour. *Anim. Behav.* **28**: 752–762.

Bennett, A. F. (1987). The accomplishments of ecological physiology. In *New directions in ecological physiology*: 1–8. Feder, M. E., Bennett, A. F., Huey, R. B. & Burggren, W. (Eds). New York: Cambridge University Press.

Bennett, A. F. & Huey, R. B. (1990). Studying the evolution of physiological performance. In *Oxford surveys in evolutionary biology*: 251–284. Futuyma, D. & Antonovics, J. (Eds). Oxford: Oxford University Press.

Crowley, S. R. & Pietruszka, R. D. (1983). Aggressiveness and vocalization in the leopard lizard (*Gambelia wislizenii*): the influence of temperature. *Anim. Behav.* **31**: 1055–1060.

Dohm, M. R., Garland, T. Jr, Cole, C. J. & Townsend, C. R. (1998). Physiological variation and allometry in western whiptail lizards (*Cnemidophorus tigris*) from a transect across a persistent hybrid zone. *Copeia* **1998**: 1–13.

Fieler, C. L. & Jayne, B. C. (1998). Effects of speed on the hindlimb kinematics of the lizard *Dipsosaurus dorsalis*. *J. exp. Biol.* **201**: 609–622.

Garland, T. Jr & Losos, J. B. (1994). Ecological morphology of locomotor performance in squamate reptiles. In *Ecological morphology: integrative organismal biology*: 240–302. Wainwright, P. C. & Reilly, S. M. (Eds). Chicago: University of Chicago Press.

Gleeson, T. T. & Harrison, J. M. (1988). Muscle composition and its relation to sprint running in the lizard *Dipsosaurus dorsalis*. *Am. J. Physiol.* **255**: R470–R477.

Hertz, P. E., Huey, R. B. & Nevo, E. (1982). Fight *vs* flight: body temperature influences defensive responses in lizards. *Anim. Behav.* **30**: 676–679.

Hertz, P. E., Huey, R. B. & Nevo, E. (1983). Homage to Santa Anita: thermal sensitivity of sprint speed in agamid lizards. *Evolution* **37**: 1075–1084.

Huey, R. B. & Bennett, A. F. (1986). A comparative approach to field and laboratory studies in evolutionary biology. In *Predator–prey relationships: perspectives and approaches from the study of lower vertebrates*: 82–99. Feder, M. E. & Lauder, G. V. (Eds). Chicago: University of Chicago Press.

Huey, R. B., Schneider, W., Erie, G. L. & Stevenson, R. D. (1981). A field-portable racetrack and timer for measuring acceleration and speed of small cursorial animals. *Experientia* **37**: 1357.

Huey, R. B. & Stevenson, R. D. (1979). Integrating thermal physiology and ecology of ectotherms: a discussion of approaches. *Am. Zool.* **19**: 357–366.

Irschick, D. J. & Jayne, B. C. (1998). Effects of incline on speed, acceleration, body posture and hindlimb kinematics in two species of lizard *Callisaurus draconoides* and *Uma scoparia*. *J. exp. Biol.* **201**: 273–287.

Irschick, D. J. & Jayne, B. C. (1999). Comparative three-dimensional kinematics of the hindlimb for high-speed bipedal and quadrupedal locomotion of lizards. *J. exp. Biol.* **202**: 1047–1065.

Jayne, B. C. & Ellis, R. V. (1998). How inclines affect the escape behaviour of a dune-dwelling lizard, *Uma scoparia*. *Anim. Behav.* **55**: 1115–1130.

John-Alder, H. B., Garland, T. Jr & Bennett, A. F. (1986). Locomotory capacities, oxygen consumption, and the cost of locomotion of the shingle-back lizard (*Trachydosaurus rugosus*). *Physiol. Zool.* **59**: 523–531.

Losos, J. B. (1990). Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. *Ecol. Monogr.* **60**: 369–388.

Miles, D. B. (1994). Population differentiation in locomotor performance and the potential response of a terrestrial organism to global environmental change. *Am. Zool.* **34**: 422–436.

Miles, D. B. & Smith, R. G. (1987). A microcomputer-based timer and data acquisition device for measuring sprint speed and acceleration in cursorial animals. *Funct. Ecol.* **1**: 281–286.

Pough, F. H. (1989). Organismal peformance and Darwinian fitness: approaches and interpretations. *Physiol. Zool.* **62**: 199–236.

Sorci, G., Swallow, J. G., Garland, T. Jr & Clobert, J. (1995). Quantitative genetics of locomotor speed and endurance in the lizard *Lacerta vivipara*. *Physiol. Zool.* **68**: 698–720.

van Berkum, F. H. (1986). Evolutionary patterns of the thermal sensitivity of sprint speed in *Anolis* lizards. *Evolution* **40**: 495–604.

van Berkum, F. H. & Tsuji, J. S. (1987). Inter-familial diffrences in sprint speed of hatchling *Sceloporus occidentalis* (Reptilia: Iguanidae). *J. Zool. (Lond.)* **212**: 511–519.

Wainwright, P. C. (1988). Morphology and ecology: functional basis of feeding constraints in Caribbean labrid fishes. *Ecology* **69**: 635–645.

Wainwright, P. C. & Reilly, S. M. (Eds) (1994). *Ecological morphology: integrative organismal biology*. Chicago: University of Chicago Press.