

Inter-observer and intra-observer differences in measuring body length: a test in the common lizard, *Zootoca vivipara*

Evgeny S. Roitberg^{*1}, Valentina F. Orlova², Valentina N. Kuranova³, Nina A. Bulakhova³, Oleksandr I. Zinenko⁴, Katarina Ljubisavljevic⁵, Regina R. Shamgunova⁶, Miguel A. Carretero⁷, Astrid Clasen⁸, Michael Fokt⁹, Wolfgang Böhme⁸

Abstract. The snout-vent length (SVL), a conventional measure of overall body size in lizards and snakes, is used in a wide variety of ecological, evolutionary, and taxonomical studies. Trends in SVL variation are often analysed using data from several researchers (observers), but possible confounding effects due to inter-observer differences in measurement protocols have never been appropriately examined. This study reports inter-observer biases between eleven herpetologists who measured the same specimens of the Eurasian common lizards (21 adult specimens were examined by eight observers and additional 192 specimens by two observers). Intra-observer bias over time (1.5-15 months between measuring sessions) was also estimated. In the vast majority of comparisons, mean difference between the first author and another observer varied from -1.0 to $+0.8$ mm, or from -1.9 to $+1.6\%$ if expressed as a percent of the specimen's SVL value. Some non-regular effects of sex and study sample on the studied bias were revealed, and their possible reasons are discussed. We are advising the researchers who intensively collect SVL and other morphometric data to consider testing intra-observer and inter-observer biases and to establish etalon samples available for re-examinations.

Keywords: body size, inter-observer bias, intra-observer bias, lizards, morphometrics, snout-vent length, *Zootoca vivipara*.

Introduction

Body size is a fundamental phenotypic characteristic of the organism, important in various biological contexts (e.g., Brown et al., 1999;

Blanckenhorn, 2000; Meiri, 2008). Squamate reptiles, snakes and particularly lizards, belong to favourite models of evolutionary and ecological studies which address body size variation. An overwhelming majority of studies on snakes and lizards use the snout-vent length (SVL) as a measure of overall size (Shine, 1994; Cox, Skelly and John-Alder, 2003; Cox, Butler and John-Alder, 2007). Indeed, for elongated animals, SVL is a reasonable body size estimator. Although body mass could be the most relevant indicator of overall size, it typically varies with reproductive status, fat storage, digestive state, and state of the tail. Therefore, body mass data are collected much less frequently than those on SVL (e.g., Meiri, 2010). Other traits of lizard morphometrics (abdomen length, leg length, and head dimensions) are also recorded much less often as SVL, especially in mark-recapture studies. SVL is thus the principal body size estimator which is used in extensive comparative analyses of multiple populations and/or species – e.g., when treating Bergmann's rule. For a better covering the studied variation, such analy-

1 - Institute of Integrated Sciences, Department of Biology, University of Koblenz-Landau, Universitätsstr. 1, D-56070 Koblenz, Germany

2 - Zoological Research Museum, Moscow M.V. Lomonosov State University, Bolshaya Nikitskaya 6, 125009 Moscow, Russia

3 - Tomsk State University, prosp. Lenina 36, 634050 Tomsk, Russia

4 - Museum of Nature, Karazin Kharkiv National University, vul. Trinklera 8, 61022 Kharkiv, Ukraine

5 - Department of Evolutionary Biology, Institute for Biological Research, University of Belgrade, Bulevar Despota Stefana 142, 11060 Belgrade, Serbia

6 - Surgut State University, ul. Energetikov 14, 628400 Surgut, Russia

7 - CIBIO-UP, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

8 - Zoologisches Forschungsmuseum A. Koenig, Adenauerallee 160, D-53113 Bonn, Germany

9 - Department of Zoology, Charles University, Vinicná 7, CZ-12844 Praha 2, Czech Republic

*Corresponding author; e-mail: eroit@web.de

ses often incorporate data from several independent studies (e.g., Niewiarowski, 1994; Tracy, 1999; Ashton and Feldman, 2003; Angilletta et al., 2004; Roitberg, 2007; Meiri, 2007).

However, individual observers may differ in the measurement protocol (e.g., one investigator can stretch a lizard's body more than another) resulting in inter-observer variability *sensu* J.C. Lee (1990). It is tacitly assumed that effects of this variability are minor as compared to the effects under study. However, data on inter-observer biases in lizard morphometrics are very scarce and, to our knowledge, restricted to one study (Pérez-Mellado and Gosá, 1988). Even for external morphology of reptiles, amphibians, and fish in general, relevant studies are rare and involve only two observers (Lee, 1990; Hayek, Heyer and Gascon, 2001; Mina, Levin and Mironovsky, 2005; Mironovskii, 2006).

In this paper we examine inter-observer biases among eleven herpetologists who recorded SVL on the same museum specimens of the European common lizard, *Zootoca vivipara*. The problem of inter-observer bias is particularly relevant for this species. Its huge geographic range – much of the northern Palearctic from Spain to Japan – can hardly be appropriately covered within a single investigation. At the same time, quite many students throughout Eurasia extensively collected SVL data from local or regional samples of *Z. vivipara* for various

research purposes. Careful summarizing these data could provide insights into phenotypic diversity of this widespread species and also contribute to several important topics of evolutionary ecology like eco-geographic clines and allometry of sexual size dimorphism.

Material and methods

Our data include two subsets. Twenty one adult lizards (Sample 1) were measured by seven of us and Antigoni Kaliontzopoulou (AK) within a relatively short time (Nov. 2009–September 2010). Additionally, ESR measured SVL in 192 museum specimens previously examined by one of the other observers (table 1); in this subset the time elapsed between the two examinations constituted a few days for OIZ, six years for MF, ten years for AC, and some 20 to 30 years for VFO. Furthermore, to estimate intra-observer bias over time, 47 of the 213 study specimens were repeatedly measured by the same observer, OIZ or/and ESR (table 2, fig. 1), 1.5–15 months after the first examination. Before being examined first time the study specimens were kept in preservative during at least two years, so that preparation-induced shrinkage of SVL (Verwust, Van Dongen and Van Damme, 2009) apparently ceased.

Sample 1 was collected in the middle stream of the Ob' River (West Siberia) during July–September 2007. The freshly euthanized animals were placed into 4% formalin for 60 days and then were kept in 70° alcohol. Some other study specimens were also fixed and held some time in formalin before being placed to alcohol, while the others have been prepared with alcohol only. The overall preparation quality of this material varied from very good to moderate. Further details are given in table 1 and the Appendix. No animals have been sacrificed for the present study.

Except two of us (NAB was supervised by VNK, and AK by MAC as graduate and then a PhD students) the observers listed in table 1 can well be regarded as independent

Table 1. Variation in SVL (measurements of ESR) in our study samples of the common lizard.

Study sample	Sex	n	Min	Max	Mean	SE	SD	Median
Sample 1 (West Siberia)	M	10	50	56	52.1	0.61	1.92	51.5
	F	11	52	67	60.8	1.33	4.41	62.0
Sample 2 (eastern Germany)	M	19	46	65	53.9	1.21	5.27	52.3
	F	7	52	64	56.5	1.57	4.16	57.3
Sample 3 (central Germany)	M	26	42	56	49.4	0.59	3.03	48.8
	F	20	43	60	52.6	1.06	4.73	52.0
Sample 4 (Middle Siberia)	M	10	42	55	48.1	1.55	4.89	46.8
	F	7	50	64	58.5	1.66	4.39	59.0
Sample 5 (north-eastern Kazakhstan)	M	12	43	53	48.1	0.85	2.96	48.5
	F	17	46	63	54.1	1.49	6.13	53.0
Sample 6 (Czech Republic)	M	11	42	54	49.4	0.93	3.09	49.8
	F	8	46	61	53.7	1.90	5.36	54.0
Sample 7 (western Germany)	M	27	45	56	50.1	0.55	2.84	50.0
	F	10	48	67	57.3	1.90	6.01	58.9

Table 2. Summary statistics for intra-observer and inter-observer biases in measuring snout-vent length (SVL) of the common lizard. In each row, positive bias indicates larger measurement value in the first observer vs the second observer, and negative bias indicates lower measurement value in the first observer vs the second observer. *P* is the probability of null-hypothesis that the mean bias does not differ from zero, and *P*(*m* + *f*) is the corresponding probability for the pooled sample of males and females. Bold values are those which remained significant (*P* < 0.05) after the sequential Bonferroni adjustment (Rice, 1989) for multiple tests.

Operators compared	Sample	Sex	n	Min	Max	Median	Mean	SE	SD	<i>P</i>	<i>P</i> (<i>m</i> + <i>f</i>)
ESR vs ESR2	Sample 1	M	10	-0.5	1.0	0.15	0.27	0.21	0.67	0.232	0.432
		F	11	-1.0	1.0	0.20	-0.01	0.23	0.75	0.969	
	Sample 2	M	13	-0.5	1.5	0.10	0.27	0.16	0.59	0.125	0.583
		F	13	-1.0	1.0	0.00	-0.14	0.15	0.56	0.388	
ESR vs ESR3	Sample 2	M	13	-1.0	2.1	0.30	0.54	0.21	0.76	0.026	0.010
		F	13	-0.5	1.0	0.00	0.16	0.12	0.44	0.207	
OIZ vs OIZ2	Sample 2	M	13	-1.5	0.5	-0.50	-0.46	0.18	0.66	0.027	0.350
		F	13	-2.0	2.0	0.00	0.12	0.30	1.08	0.708	
ESR vs RRS	Sample 1	M	10	-1.0	4.9	0.33	0.78	0.54	1.71	0.183	0.019
		F	11	-1.0	3.2	0.75	0.83	0.38	1.25	0.052	
ESR vs NAB	Sample 1	M	10	-0.4	2.0	0.50	0.64	0.28	0.88	0.046	0.000
		F	11	-0.5	1.5	1.10	0.82	0.20	0.65	0.002	
ESR vs VNK	Sample 1	M	10	-0.6	0.3	-0.19	-0.14	0.10	0.31	0.179	0.947
		F	11	-1.0	1.4	0.20	0.14	0.22	0.74	0.533	
ESR vs KL	Sample 1	M	10	-0.3	1.7	0.21	0.43	0.21	0.65	0.066	0.002
		F	11	-0.4	1.8	0.67	0.64	0.21	0.71	0.014	
ESR vs MAC	Sample 1	M	10	-0.5	3.2	1.05	1.20	0.35	1.10	0.007	0.000
		F	11	-0.7	3.3	2.39	2.02	0.34	1.13	0.000	
ESR vs AK	Sample 1	M	10	-1.2	0.3	-0.22	-0.40	0.19	0.62	0.071	0.059
		F	11	-1.3	1.2	-0.22	-0.18	0.21	0.68	0.413	
ESR vs OIZ	Sample 1	M	10	-1.0	1.0	-0.10	-0.17	0.18	0.57	0.372	0.003
		F	11	-1.3	0.0	-0.50	-0.62	0.14	0.46	0.001	
	Sample 2	M	13	-1.5	0.5	-0.20	-0.47	0.18	0.64	0.021	0.044
		F	13	-1.7	3.0	-0.50	-0.38	0.37	1.33	0.319	
ESR vs VFO	Sample 3	M	26	-2.0	0.5	-0.50	-0.63	0.13	0.65	0.000	0.000
		F	19	-2.5	0.5	-1.00	-1.01	0.17	0.74	0.000	
	Sample 4	M	10	-2.0	0.5	-1.00	-0.76	0.22	0.68	0.006	0.000
		F	7	-2.0	0.0	-1.00	-1.03	0.22	0.58	0.003	
	Sample 5	M	12	-1.0	3.0	0.00	0.21	0.33	1.15	0.543	0.504
		F	17	-3.5	1.8	-0.50	-0.41	0.31	1.26	0.197	
ESR vs MF	Sample 6	M	11	-0.2	3.6	0.35	0.75	0.36	1.20	0.065	0.023
		F	8	-1.2	1.8	0.53	0.42	0.32	0.91	0.232	
ESR vs AC	Sample 7	M	27	-1.2	1.1	-0.20	-0.25	0.11	0.59	0.038	0.001
		F	10	-2.5	-0.2	-0.48	-0.69	0.21	0.67	0.010	

researchers, and we assume them to form a representative sample from a community working on lacertid lizard morphometrics.

All operators were asked to record test measurements in their usual way (e.g., some operators use a rule, the others use dial callipers). As a great amount of data had already been collected before we started to collaborate, our intent was merely to estimate the existing biases rather than to standardize individual techniques of measuring the trait and reading the values from the device.

The main target variable was the bias (difference) between ESR (the observer who measured all study samples) and another observer. Similarly, the difference between the first and subsequent measurements of the same observer was used to quantify the intra-observer bias. For each test series – a unique combination of the operator pair, study sample,

and sex – we checked whether the mean bias differed significantly from zero (one-sample t-test, adjusted for multiple comparisons, table 2). Using several ANOVA models, we checked whether the above bias varied (1) between sexes and among study samples within a pair of observers, and (2) among individual observers and between the groups of observers which used different measuring device. For some of these analyses, a transformation $Y = \ln(x + 10)$ was used to meet an assumption of homogeneity of variances.

Additionally, for the SVL values taken on sample 1, we performed a repeated-measures ANOVA (rm-ANOVA), Observer × Sex. Also, for the first and subsequent measurements of the same observers (ESR, ESR2, ESR3 for Sample 1; ESR, ESR2 for Sample 2; OIZ, OIZ2 for Sample 2; cf. table 2) three rm-ANOVAs, Measuring session × Sex were run to analyse intra-observer variation. Because of a highly

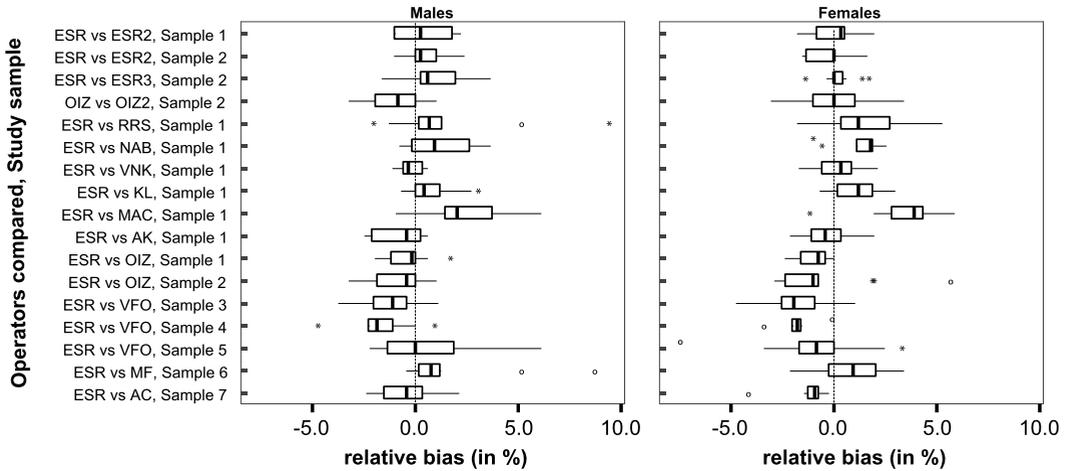


Figure 1. Intra- and inter-observer relative biases (in % of specimen’s SVL averaged for two observers) for different combinations of the Operator pair, Study sample and Sex. The individual data are summarized as boxplots including the median (bold horizontal line), the 25th and 75th percentiles (box), range without outliers (vertical line), and outliers (observations located more than two interquartile range above or below the median) indicated as open circles or asterisks.

significant deviation from sphericity in some of these analyses (Mouchly’s $W = 0.014$, $P < 0.001$), Greenhouse-Geisser test of within-subject effects was used. A transformation $Y = 1/SVL^2$ was used to reach homogeneity of variances.

Results

Table 1 describes the variation of SVL in the study samples. Table 2 provides sample statistics for raw biases within and between observers, and fig. 1 summarizes variation for relative biases, expressed as a percentage of the specimen’s SVL, averaged for a corresponding pair of observers (or measuring sessions of the same observer).

For most observers (Observer × Sample × Sex combinations) the mean raw bias varied from −1.0 to +0.8 mm, being somewhat higher (2.2 mm for females) for MAC measurements (table 2). Raw biases for individual specimens were mostly within −2.5 and +3.5 mm, with two deviating values of −3.5 and +4.9 mm (table 2). The relative bias varied mostly from −1.9 to 1.6% for mean values and from −7.3 to 9.5% (mostly from −4.7 to 6.1%) for individual specimens (fig. 1).

Intra-observer biases were clearly smaller than the largest third of the inter-observer biases: extreme values ranged from −1.0 to +1.5 mm for ESR and from −2.0 to +2.0 mm for OIZ; means and medians ranged from −0.14 to +0.54 mm for ESR and from −0.50 to 0.0 mm for OIZ (table 2; see also fig. 1 for relative values). The smallest third of the inter-observer biases were of the same magnitude as those within observers.

As no consistent correlation between the raw bias and the SVL value of the corresponding specimen was found within study series (in only two of the 26 inter-observer test series and one of the 8 intra-observer test series, the product-moment correlation coefficient reached $0.01 < P < 0.05$, and these become insignificant if adjusted for multiple comparisons; 17 coefficients were positive and the other 17 were negative), we restrict our analyses to raw biases.

For sample 1 which was examined by eight observers, a two-way rm-ANOVA revealed a significant effect of Observer ($F = 15.21$, $df = 2.98$, $P < 0.001$) but not the Observer × Sex interaction ($F = 0.49$, $df = 2.98$, $P = 0.693$). Significance of inter-observer differences can also be seen from table 2: in many test series (Observer × Sample × Sex and Observer ×

Sample combinations) the mean bias differed significantly from zero.

The rm-ANOVAs made for the three Observer \times Sample combinations for which repeated measurements had been made (see the Methods section) revealed a significant effect of Measuring session ($F = 5.06$, $df = 1.717$, $P = 0.014$) for ESR measurements of Sample 2. This P value remained significant after Bonferroni correction (multiplying by 3).

To estimate whether a bias between particular observers can differ for males and females and/or among study samples, ANOVAs with the raw bias as the dependent variable, and Sex and Sample as categorical effects were performed for those observer pairs who examined two or more samples or their data were not covered by the above analysis. A significant (or nearly so) effect of Sex was revealed for ESR-VFO and ESR-AC biases, and a strong effect of Sample was revealed for ESR-VFO biases (table 3).

A nested design ANOVA, Sex \times Measuring device, Observer(Measuring device) was run to check if the bias to ESR differs between observers which used ruler (ESR and OIZ), mechanical callipers (NAB, VFO, MF, and AC), and digital callipers (RSS, VNK, KL, MAC, and AK). To incorporate the reference observer in this analysis, biases of his repeated measurements vs first measurements were added. The effect of Observer (nested in Measuring device) was highly significant ($F_8 = 17.15$, $P < 0.001$), whereas that of Sex ($F_1 = 2.52$,

$P = 0.113$) and Measuring device ($F_2 = 0.798$, $P < 0.484$) was not.

Discussion

Our study revealed highly significant inter-observer biases for measuring SVL in the Eurasian common lizard. However, for the vast majority of the tested researchers, mean bias located within 2 mm in raw values and within 3.5% relative to the specimen's SVL. Mean intra-observer biases, as tested for two observers, ranged within 0.7 mm and hardly can solely explain the above inter-observer pattern. For several traits of external morphology in birds, inter-observer variability was also higher than that within observers (Goodenough et al., 2010).

Another factor which could contribute to the inter-observer variability is measuring device (e.g., Jordaens et al., 2002). Although the structure of our data does not allow a rigorous testing this effect, our analysis clearly argued for a minor impact of this factor to the overall inter-observer variability pattern (see also fig. 2).

The considered measurement biases are not negligible but they are much less than the known magnitude of the geographic variation in *Zootoca vivipara*: mean SVL of gravid females in the Asian part of its range varied from 54.3–73.2 mm (Orlova, Kuranova and Bulakhova, 2005). Pérez-Mellado and Gosá (1988) provided means and standard deviations for two test samples of adult *Podarcis muralis* ($n = 33$) and *P. liolepis* [formerly *P. hispanica*] ($n = 30$) measured by two observers. Mean inter-observer bias for SVL was 0.47 mm for the first and 0.35 mm for the second sample; these values constitute respectively 0.8% and 0.6% of mean SVL averaged for two observers. These values are close to the lower values from the range of biases among eleven observers in our study (table 2; fig. 1).

For another confounding factor due to data heterogeneity, the shrinkage due to fixation and preservation, a SVL bias of 3.7% was reported

Table 3. Two-way ANOVAs with inter-observer bias as the dependent variable, and Sex or Sex and Sample as the categorical effects. As no significant Sex \times Sample interaction was revealed, the model with main effects only is presented.

Inter-observer bias	Source	df	F	P
ESR-OIZ	Sex	1	0.377	0.542
	Sample	1	0.305	0.583
ESR-VFO	Sex	1	5.410	0.022
	Sample	2	6.764	0.002
ESR-MF	Sex	1	0.429	0.521
ESR-AC	Sex	1	3.802	0.059

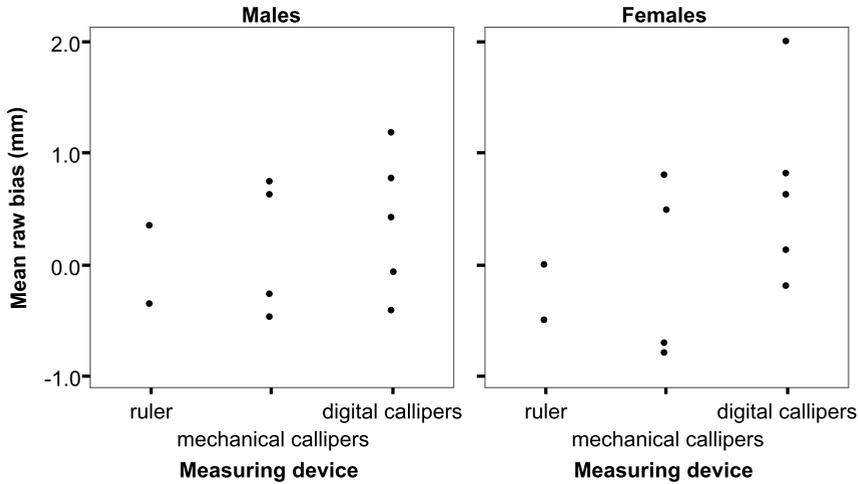


Figure 2. Distribution of mean inter-observer biases by measuring device. Unlike in table 2, data for different study samples are pooled to provide one data point for each observer. For ESR, bias of repeated measurements vs first measurements was used to incorporate the reference observer in this analysis. Ruler was used by ESR and OIZ; mechanical callipers by NAB, VFO, MF, and AC; digital callipers by RSS, VNK, KL, MAC, and AK.

for young iguanas (Vervust, Van Dongen and Van Damme, 2009).

Although moderate and without consistent fashion, inter-observer bias could differ between sexes (table 3). As the bias did not show correlation with specimen's SVL within sexes these differences can hardly be determined by sexual size dimorphism *per se*. We hypothesize they can be due to sex differences in overall elasticity of the pre-caudal part of lizard's body. Female body is likely to be more elastic because its hard part (head) is shorter while the softer part (abdomen) is longer than in males (Braña, 1996; Kratochvíl et al., 2003). Then, if one observer tends to stretch (straighten out) lizard's body stronger than the other, the resulting increment in the recorded SVL should be generally greater for female specimens. The bias is not expected to be sex-specific however, if two observers stretch the specimen at similar rates but differ in other aspects of their realized conception of the character. These aspects may concern actual landmarks for the tip of snout and for the vent, and reading the values from the device. Biases of the latter two kinds can be termed *terminal effects* as they relate to the endpoints of the body part to be measured. Such biases should be

largely independent of the specimen's SVL, and a lack of consistent relationship between bias and SVL within our test series (see Results) suggests their role to be considerable. In line with the latter point, in hard skeletal structures, in which the terminal effects are the only source of intra- and inter-observer biases, relative measurement error tends to be negatively related to the character mean (Yezerinak, Lougheed and Handford, 1992; Palmerim, 1998).

The bias between ESR and VFO was found to vary among study samples (tables 2, 3; fig. 1). Whereas Samples 3 and 4 showed a consistent negative bias of -0.63 to -1.04 mm, there is no consistent bias for Sample 5. This pattern is likely to explain with a (unconscious) temporal shift of observer's conception of character, i.e., *intra-observer bias* (Lee, 1982; this study): whereas ESR measured all these samples within a few days in October 2009, VFO examined Sample 5 some 20 years later than Samples 3 and 4. As the three samples did not differ considerably in preparation quality this potentially important factor is not a likely alternative reason for the above pattern.

As for the biases within pairs of a supervisor & her/his former student (VNK & NAB and

MAC & AK) they did not tend to be more similar than those of independent observers (table 2; fig. 1). Therefore, all observers in this study can well be treated as independent researchers in terms of their recording SVL.

Even though a causal analysis of inter- and intra-observer biases may be too complicated (in relation to the efforts we are ready to invest in it) and not strictly necessary, a merely descriptive investigation of this phenomenon should help to more appropriately consider external morphometric data. If, for instance, larger mean values come from an observer with a positive bias the magnitude of differences we can trust should be higher than in cases where the among-sample differences and the inter-observer bias show opposite directions. We are advising the researchers who intensively collect SVL and other morphometric data to consider testing intra- and inter-observer biases.

If a researcher is recording morphometric data for the same species over considerable time, particularly in a discontinuous fashion, a repeated measuring at least ten specimens of each sex is recommended to control for a temporal shift of his/her conception of character. A reasonable practice of monitoring inter-observer biases, particularly for widespread species or other species which are intensively studied by multiple observers, could involve establishing of test samples available for re-examination and subsequent publishing of the results on a website or as an appendix to a journal paper.

Acknowledgements. Antigoni Kaliontzopoulou participated in our test measurements of Sample 1. Uwe Fritz and Jiri Moravec provided access to museum series in Dresden and Prague. Critical comments of Daniel Berkowic, Shai Meiri, and an anonymous reviewer stimulated valuable improvements of the earlier version of the manuscript. This study is part of a larger project supported by the Deutsche Forschungsgemeinschaft (grant RO 4168/1-1 to ESR).

References

Angilletta, M.J., Niewiarowski, P.H., Dunham, A.E., Leaché, A., Porter, W.P. (2004): Bergmann's clines in

- ectotherms: Illustrating a life-historical perspective with sceloporine lizards. *Amer. Natur.* **164**: E168-E183.
- Ashton, K.G., Feldman, C.R. (2003): Bergmann's rule in nonavian reptiles: turtles follow it, lizards and snakes reverse it. *Evolution* **57**: 1151-1163.
- Blanckenhorn, W.U. (2000): The evolution of body size: what keeps organisms small? *Quarterly Review of Biology* **75**: 385-407.
- Braña, F. (1996): Sexual dimorphism in lacertid lizards: male head increase vs female abdomen increase? *Oikos* **75**: 511-523.
- Brown, R.P., Znari, M., El Mouden, E.N., Harris, P.E. (1999): Estimating asymptotic body size and testing geographic variation in *Agama impalearis*. *Ecography* **22**: 277-284.
- Cox, R.M., Skelly, S.L., John-Alder, H.B. (2003): A comparative test of adaptive hypotheses for sexual size dimorphism in lizards. *Evolution* **57**: 1653-1669.
- Cox, R.M., Butler, M.A., John-Alder, H.B. (2007): The evolution of sexual size dimorphism in reptiles. In: *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*, p. 38-49. Fairbairn, D.J., Blanckenhorn, W.U., Székely, T., Eds, New York, Oxford University Press.
- Goodenough, A.E., Stafford, R., Catlin-Groves, C.L., Smith, A.L., Hart, A.G. (2010): Within- and among-observer variation in measurements of animal biometrics and their influence on accurate quantification of common biometric-based condition indices. *Ann. Zool. Fennici* **47**: 323-334.
- Hayek, L.C., Heyer, W.R., Gascon, C. (2001): Frog morphometrics: a cautionary tale. *Alytes* **18**: 153-177.
- Jordaens, K., Van Dongen, S., Van Riel, P., Geenen, S., Verhagen, R., Backeljau, T. (2010): Multivariate morphometric soft body parts in terrestrial slugs: comparison between two datasets, error assessment and taxonomic implications. *Biol. J. Linn. Soc.* **75**: 533-542.
- Kratochvíl, L., Fokt, M., Rehak, I., Frynta, D. (2003): Misinterpretation of character scaling: a tale of sexual dimorphism in body shape of common lizards. *Canad. J. Zool.* **81**: 1112-1117.
- Lee, J.C. (1982): Accuracy and precision in anuran morphometrics: artifacts of preservation. *Syst. Zool.* **31**: 266-281.
- Lee, J.C. (1990): Sources of extraneous variation in the study of meristic characters: the effect of size and of inter-observer variability. *Syst. Zool.* **39**: 31-39.
- Meiri, S. (2008): Evolution and ecology of lizard body sizes. *Global Ecology and Biogeography* **17**: 724-734.
- Meiri, S. (2010): Length-weight allometries in lizards. *J. Zool.* **270**: 1-9.
- Mina, M.V., Levin, B.A., Mironovsky, A.N. (2005): On the possibility of using character estimates obtained by different operators in morphometric studies of fish. *Journal of Ichthyology* **45**: 284-294.
- Mironovskii, A.N. (2006): Factors determining the comparability of data obtained by estimation of morphometric characters in fish. *Journal of Ichthyology* **46**: 240-251.

- Niewiarowski, P. (1994): Understanding geographic life-history variation in lizards. In: *Lizard Ecology: Historical and Experimental Perspectives*, p. 31-49. Vitt, L.J., Pianka, E.R., Eds, Princeton, Princeton University Press.
- Orlova, V.F., Kuranova, V.N., Bulakhova, N.A. (2005): Some aspects of reproductive biology of *Zootoca vivipara* (Jacquin, 1787) in the Asian part of its area. In: *Herpetologia Petropolitana*, p. 255-258. Ananjeva, N., Tsinenko, O., Eds, St. Petersburg, Societas Europaea Herpetologica.
- Palmeirim, J.M. (1998): Analysis of skull measurements and measurers: can we use data obtained by different observers? *Journal of Mammalogy* **79**: 1021-1028.
- Pérez-Mellado, V., Gosá, A. (1988): Biometria y folidosis en Lacertidae (Sauria, Reptilia). Algunos aspectos metodológicos – *Revista Espanola Herpetologia* **3** (2): 105-119.
- Rice, W.R. (1989): Analyzing tables of statistical tests. *Evolution* **43**: 223-225.
- Roitberg, E.S. (2007): Variation in sexual size dimorphism within a widespread lizard species. In: *Sex, Size, and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*, p. 143-153, 216-217. Fairbairn, D.G., Blanckenhorn, W.U., Székely, T., Eds, New York, Oxford University Press.
- Shine, R. (1994): Sexual size dimorphism in snakes revisited. *Copeia* **1994**: 326-346.
- Tracy, C.R. (1999): Differences in body size among chuck-walla (*Sauromalus obesus*) populations. *Ecology* **80**: 259-271.
- Vervust, B., Van Dongen, S., Van Damme, R. (2009): The effect of preservation on lizard morphometrics – an experimental study. *Amphibia-Reptilia* **30**: 321-329.
- Yezerinac, S.M., Lougheed, S.C., Handford, P. (1992): Measurement error and morphometric studies – statistical power and observer experience. *Systematic Biology* **41**: 471-482.

Received: April 13, 2011. Accepted: September 3, 2011.

Appendix

Specimens examined

Sample 1: See Methods for details.

Sample 2: Zoological State Collections Dresden, MTD 4781, 5320, 5676, 8394, 8399, 9581, 9584-9586, 9931, 9932, 10001, 10015, 11612, 13490, 13673, 13741, 15153, 15154, 15155, 15156, 15157, 15159, 16923, 19512, 22464.

Sample 3: Zoological Research Museum of Moscow State University (ZM MGU): R3278(1-46).

Sample 4: ZM MGU: R3103(1-17);

Sample 5: ZM MGU: R8424(1-8), R8429(1-29).

Sample 6: Natural History Museum of Prague: 33742(1-11, 13-14), 35289(1-6).

Sample 7: Zoological Research Museum A. Koenig, Bonn: ZFMK 7747, 30651, 30652, 30653, 45796, 45799, 45801, 45802, 45806-45809, 45813-45816, 45819, 45824, 45827, 55989, 6384, 6389, 6390-6392, 6394, 6395, 6398-6401, 6404, 6409, 6411, 72951, 73685, 73686.