Journal of Mammalogy, 96(1):185–193, 2015
DOI:10.1093/jmamma/gyu022

Dead mice can grow – variation of standard external mammal measurements from live and three postmortem body states

Ryan B. Stephens,* Krishun H. Karau, Christopher J. Yahnke, Sara R. Wendt, and Rebecca J. Rowe

Department of Biology and Museum of Natural History, University of Wisconsin Stevens Point, 167 Daniel O. Trainer Natural Resources Building, 800 Reserve Street, Stevens Point, WI 54481-3897, USA (RBS, KHK, CJY, SRW)
Department of Natural Resources and the Environment, University of New Hampshire, 114 James Hall, 56 College Road, Durham, NH 03824-3534, USA (RBS, RJR)

* *Correspondent: ryan.stephens@unh.edu*

The use of standard external measurements is pervasive in mammalogy, due in part to their applicability to both living and dead animals. Nevertheless, comparisons of measurements made between animals in different pre- and postmortem body states may be problematic. To investigate the impact of body state on standard measurements and their associated variances, we took 5 external measurements on 70 *Mus musculus* while living and in 3 postmortem body states (primary flaccidity, rigor mortis, and secondary flaccidity). Total length, tail length, and hind-foot length were significantly longer in states of primary and secondary flaccidity than when measured on live individuals or those in rigor mortis. Ear length increased from living to primary flaccidity, after which it decreased. Weight decreased between each postmortem body state, likely as an artifact of desiccation. Variance was always greater for measurements taken in the living state than during postmortem body states. Irrespective of body state, variance was particularly high for ear length and hind-foot length, which are prone to observer bias and should be used with caution. Additional tests using field-collected data from populations of *Peromyscus leucopus* and *Peromyscus maniculatus* confirmed our lab-based results from *M. musculus*. External measurements taken during a postmortem state should not be compared to measurements from live animals and could lead to incorrect species identification. Additionally, comparisons among individuals measured in different postmortem body states may confound efforts to assess changes in body size over space or time.

Key words: body size, body state, measurement error, museum specimens, *Mus musculus*, primary flaccidity, rigor mortis, secondary flaccidity, standard mammal measurements

Body measurements of vertebrates can be made on skeletal material or from external features (Sumner 1927). The recording of external metrics is particularly common in mammalogy where sets of standard external measurements have been developed for terrestrial nonvolant mammals and include: total length, tail length, hind-foot length, ear (pinna) length, and weight (Merriam 1889; Miller 1899; Hall 1962). These measurements are ubiquitous in the mammalian taxonomic literature (Schmidly et al. 1988; Rasolooarison et al. 2000; Dumbacher et al. 2014) and are used to answer diverse questions in physiology (McNab 1966), population ecology (Harestad and Bunnel 1979), community ecology (Bowers and Brown 1982; Rowe et al. 2011), and evolution (Mullen and Hoekstra 2008). At the time of measurement, the body state of the animal (i.e., living or postmortem) often is not recorded and may vary. Many field studies record measurements on live individuals whereas measurements associated with museum specimens may have been taken immediately upon death (primary flaccidity), during rigor mortis, or from specimens that were frozen prior to vouchering and are measured once thawed in the condition of secondary flaccidity. Thus, standard metrics are flexible in that they can be applied to a wide array of body states. However, bias or measurement error (variance) introduced in a given pre- or postmortem body state may reduce the practical application or comparability of metrics among body states. This can be problematic when results of a comparison are interpreted as biologically meaningful, but factors affecting bias or error are not accounted for. Differences among body states increase the risk of committing type I errors when different body states are pooled and type II errors when groups measured in different pre- or postmortem body states are compared.

The degree of bias associated with measurements from animals in different body states has been little explored, and the putative degree of measurement change among body states is

largely anecdotal (Sumner 1927; Palmer 1947). For example, contraction of muscles during the body state of rigor mortis is assumed to result in shorter measurements than at other postmortem body states (e.g., Palmer 1947). Additionally, live animals may contract muscles during handling and have shorter lengths than at a postmortem body state. Using measurements from field guides and mammalian species accounts to identify live cryptic species in the field may be problematic, since such references often use external measurements taken from museum specimen labels (e.g., Lackey et al. 1985; Kurta 1995; Reid 2006), and could lead to erroneous conclusions of species identity.

Intra- or interobserver error associated with external measurements may also increase the chances of committing type I and II errors when using inferential statistics. Observer measurement error may come from a number of sources including poorly or inconsistently defined metrics, lack of observer experience, plasticity in the object being measured, and the size of the object, such that smaller metrics have proportionately larger errors (Bailey and Byrnes 1990; Blackwell et al. 2006). Measurement variability may also be affected by body state of the animal (Sumner 1927; Blackwell et al. 2006). For instance, Blackwell et al. (2006) found that measurement error was greater for live animals than for specimens that were preserved in ethanol (which results in a stiffened condition similar to rigor mortis).

Considering the widespread use of standard measurements in mammalian studies, understanding the relationship of these measurements among body states is important to account for and reduce bias and variation, allowing for more precise comparisons. To address bias and variation among body states for 5 standard measurements, we measured both laboratory-reared and wild-caught mice in 4 body states (living, primary flaccidity, rigor mortis, and secondary flaccidity). Specifically, our objectives were to determine if measurement results are similar among the 4 body states, to assess error rates among the 4 body states, and to determine if differences in body state are detectable in field-collected data at the population level.

Methods

Lab measurements.—In order to investigate the impact of body state on change and variation of standard mammal measurements, we took advantage of an opportunity to use domestic house mice (*Mus musculus*) which were to be euthanized for reptile food. *M. musculus* is an excellent study taxon because it is a common laboratory animal and is proximate in size to many species of small mammals encountered during field studies. We acquired 70 adult mice from a local breeder and temporarily housed them at the University of Wisconsin Stevens Point (UWSP) animal care facility. Following completion of the study, mice were donated to the UWSP herpetology society. Our protocol was approved by the UWSP Animal Care and Use Committee (protocol #2011.11.13) and followed guidelines outlined by the American Society of Mammalogists (Sikes et al. 2011).

Four observers (CJY, KHK, RBS, and SRW) measured each of the 70 mice in 4 body states commonly encountered during field studies: living (conscious), primary flaccidity (immediately following euthanasia), rigor mortis (temporarily stored on ice), and secondary flaccidity (frozen and thawed). Standard measurements commonly used by researchers trained at North American institutions were taken following Hall (1962): total length (tip of nose to distal tip of the terminal caudal bone), tail length (sacrum to terminal caudal bones), right hind-foot length (calcaneus to distal tip of the longest claw), right pinna length (intertragic notch to tip of helix, excluding hairs), and weight. For total length, we measured postmortem mice in the supine position (along the dorsum); however, because live mice are handled by grasping the nape of the neck, we measured them in the prone position (along the ventrum). We address potential differences in total length measurements between live and postmortem animals in the "Discussion." All other measurements were taken in the same fashion for all body states. Length measurements were taken using a clear plastic ruler to the nearest 0.5 mm. To reflect differences in field and lab methods, respectively, we measured live animals and those in primary flaccidity using a digital pull scale (accuracy 0.1 g) and animals in rigor mortis and secondary flaccidity using a stationary bench scale (accuracy 0.001). We compared scales before and after each measuring session by weighing a number of mice on each scale to make certain that differences in the types of scale did not affect recorded weights. All weights were recorded to the nearest 0.1 g. Live mice were weighed in a clear plastic bag, with final weight tared for bag weight.

Before the initiation of this study, observers varied in the number of small mammals measured (70–2,500 individuals), representing the range of experience commonly encountered in a field or museum setting. During the study, each observer (n = 4) measured each mouse (n = 70) at every body state (n = 4). In order to better quantify inter- and intraobserver variation, each observer took 2 additional sets of measurements on 10 randomly selected mice during each body state. Thus, for each body state, each observer took 3 sets of measurements from 10 mice and measured the remaining 60 mice once. Randomization helped reduce handling stress on a given mouse while alive. Although the 10 mice were randomly selected for each person at each body state, there was considerable overlap of mice measured among body states and observers. This interdigitating structure of the data set allowed for estimates of both inter- and intraobserver variation. Mice were measured in batches of approximately 17 so that all observers could complete measurements within a 1-h timeframe. Following measurements, live animals were euthanized using a CO_2 chamber and remeasured within 30 min during primary flaccidity. Mice were then placed in individual plastic bags in a cooler of crushed ice for 4 h and subsequently remeasured during rigor mortis (within 30 min of removal from ice). Finally, mice were double bagged and frozen in an upright freezer for 14 days at −18°C, thawed for 3 h at approximately 20°C, and remeasured during secondary flaccidity.

Statistical analysis of lab measurements.—We used linear mixed effects models to determine whether standard measurements differed by body state. Our data were nested and only some mice were measured repeatedly within a body state by an observer. Linear mixed effects models are robust to such hierarchical data structure and missing observations (i.e., differences in the number of measurements on a given mouse by an observer within a body state) and allow for within-mouse and observer covariance (i.e., the same mice measured by the same observer at multiple body states—Zuur et al. 2009). Separate models were created for each measurement; however, we followed the same steps for model selection and validation. We modeled each measurement as a function of body state, treating body state as a fixed effect. Observer and mouse were treated as random effects to account for nonindependence among body states. This base model was then compared using a log-likelihood test to a model that allowed residual error to vary by body state. When significant, this multiple variance for body state was retained in the model. We validated model fit by plotting residuals versus fitted values and by evidence of homogeneity of variances and normality of both the residuals and random effects (Pinheiro and Bates 2000; Zuur et al. 2009). Pairwise differences between body states were tested using Tukey's contrasts.

The hierarchical models partitioned variance associated with the random effects (observer and mouse) into causal components of among mouse variation (a_k), among observer variation in the same mouse ($a_{j|k}$), and residual error (ε_{ijk}). For each measurement, the optimal model included a residual error component (ε_{ijk}) that differed by body state (see "Results"). We are aware of no formal statistical methods to compare different ε_{ijk} among body states. Thus, we took a conservative approach, using lack of overlap of the 95% *CI*s of *SD* of residual error between body states as an ad hoc approximate test for differences. *CI*s were calculated from *SE*. We calculated intraclass correlation coefficient (ICC) at the mouse level ($\text{ICC}_{\text{intraobserver}}$, equation 1), which is the correlation among measurements from the same observer measuring the same mouse multiple times at a given body state, and at the observer level ($\text{ICC}_{\text{interobserver}}$, equation 2), which is the correlation among observers measuring the same mouse at a given body state (Hayen et al. 2007).

$$\text{ICC}_{\text{intraobserver}} = \frac{\sigma^2_{\text{mouse}} + \sigma^2_{\text{observer}}}{\sigma^2_{\text{mouse}} + \sigma^2_{\text{observer}} + \sigma^2} \quad (1)$$

$$\text{ICC}_{\text{interobserver}} = \frac{\sigma^2_{\text{mouse}}}{\sigma^2_{\text{mouse}} + \sigma^2_{\text{observer}} + \sigma^2} \quad (2)$$

We used ICCs to qualitatively assess the repeatability of intra- and interobserver measurements. Finally, to determine how differences among observers might impact error, we assessed the proportion of each random variance component to the total variance.

Field measurements.—We included field-based data from 2 studies to investigate whether the findings from our lab study, using the same individuals at multiple body states, are representative of those encountered during small mammal surveys when different individuals from the same population are measured at different body states. Specifically, we compared measurements from a series of live white-footed mice (*Peromyscus leucopus*) to a second series in secondary flaccidity, and we compared a series of deer mice (*Peromyscus maniculatus*) in primary flaccidity to a series in rigor mortis. *P. leucopus* (n = 34) from Bartlett Experimental Forest, White Mountain National Forest, New Hampshire (44°3′7.2″N, −71°17′25.1″W) were captured using Sherman live traps (7.6 × 8.9 × 23.9 cm; H.B. Sherman Inc., Tallahassee, Florida) as part of an ongoing study investigating community ecology and were measured alive. Postmortem specimens (n = 18) from Bartlett Experimental Forest were obtained from the United States Forest Service as part of a long-term monitoring project and were collected using wet pitfall arrays. Vouchers were frozen for < 3 months prior to measurement and preparation. All specimens were collected within a 1.5-km radius between July and August 2013. Measurements were recorded by the first author using the same methods outlined for measuring laboratory mice, with the exception that field weights were taken using a Pesola spring scale (Pesola, Baar, Switzerland) to the nearest 0.5 g. We pooled males and females because *P. leucopus* does not show sexual dimorphism in external measurements (Stephens et al. 2014). Because pitfall traps tend to capture smaller *P. leucopus* (Stephens and Anderson 2014), we made the groups comparable by excluding pitfall-captured juveniles (mass ≤ 15 g—Wolff 1985) and Sherman live trap captures > 25 g (n = 6), a range of weights (> 24.5 g) not captured by pitfall traps.

Our field data on *P. maniculatus* come from a resurvey study in the Humboldt-Toiyabe National Forest, Nevada (38°53′27.2″N, −117°14′30.8″W) that used snap traps and Sherman live traps. All animals were collected within a 300-m radius 3–6 May 2011. During the time of trapping, the nights were cool (approximately 6–10°C by 0600 h), and snap-trapped animals were retrieved shortly after sunrise, stored on ice, and measured by RJR. Snap-trapped animals (n = 85) were in rigor mortis when measured, and animals with broken spines were excluded. Animals captured in Sherman traps (n = 87) were euthanized immediately before vouchering and were measured in primary flaccidity by D. Rogers (Brigham Young University, Salt Lake City, Utah). Measurement methods followed those outlined for laboratory mice, with the exception that field weights were taken using a Pesola spring scale to the nearest 1.0 g. To reduce bias resulting from smaller individuals being captured by Sherman traps, we excluded juveniles (≤ 15 g—Wolff 1985) from the dataset.

For both *P. leucopus* and *P. maniculatus*, we compared measurements by body state using a Welch 2-sample t-test with α = 0.05. Because we used weight to adjust for bias resulting from trap type, we did not use weight to infer differences between body states. All voucher specimens from New Hampshire were deposited in the Yale Peabody Museum of Natural History (New Haven, Connecticut) and National Museum of Natural History (Washington, D.C.), and vouchers

from Nevada were deposited in the Natural History Museum of Utah (Salt Lake City, Utah) and Monte L. Bean Life Science Museum (Provo, Utah). The New Hampshire live trapping protocol was approved by the University of New Hampshire Animal Care and Use Committee (#120708) and Nevada sampling by the University of Utah Animal Care and Use Committee (#09-02004). All trapping procedures followed guidelines outlined by the American Society of Mammalogists (Sikes et al. 2011).

All statistical analyses were conducted in R 3.0.2 (R Development Core Team 2013). Mixed effects models were performed with lme using the nlme package (Pinheiro et al. 2012), and Tukey's contrasts were calculated using glht in the multcomp package (Hothorn et al. 2008).

Results

Lab measurements.—For all measurements, the optimal mixed effects model included a multiple residual variance structure based on body state (total length: $\chi^2_{[1]} = 122.3$, $P < 0.001$; tail length: $\chi^2_{[1]} = 184.6$, $P < 0.001$; hind-foot length: $\chi^2_{[1]} = 126.1$, $P < 0.001$; ear length: $\chi^2_{[1]} = 205.1$, $P < 0.001$; and weight: $\chi^2_{[1]} = 747.2$, $P < 0.001$). Plots of fitted versus observed values for the optimal models indicated excellent overall model fits. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. For a given standard measurement, we recorded 360 observations (including remeasured mice) in each body state, totaling 1,440 observations for all body states. Although we recorded length measurements to the nearest 0.5 mm, our models are based on composite samples and thus derive estimates with greater precision than a single observation (Taylor 1997), and therefore permit the reporting of statistically significant differences at the 0.1 mm level. Our 5 measurements vary greatly in overall size (e.g., ears compared to total length), and to better capture the magnitude of a given difference, we report differences by body state both in values measured (mm) and as percent change in the measurement.

Trends and pairwise significance among the 5 standard metrics at each body state are shown in Fig. 1 and multiplicative factors used to convert measurements between body states are shown in Table 1. Although the overall magnitudes were different, total length, tail length, and hind-foot length all showed similar trends in measurements across body states—increasing from living to primary flaccidity, decreasing in rigor mortis, and finally increasing again during secondary flaccidity. Of these measurements, mean live total length ($\bar{X} = 174.3$ mm) showed the largest change, increasing by 7.3 mm (4.2%) to primary flaccidity, whereas mean live tail length (86.4 mm) and mean hind-foot length (18.7 mm) increased by 2.1 mm (2.4%) and 0.5 mm (2.5%), respectively. From primary flaccidity to rigor mortis, mean total length, mean tail length, and mean hind-foot length decreased by 6.4 mm (3.7%), 2.0 mm (2.3%), 0.3 mm (1.4%), respectively. From rigor mortis to secondary flaccidity, mean total length, mean tail length, and mean hind-foot length increased by 6.7 mm (3.8%), 1.1 mm (1.2%), and 0.2 mm (1.0%), respectively. Tail lengths between live and rigor mortis

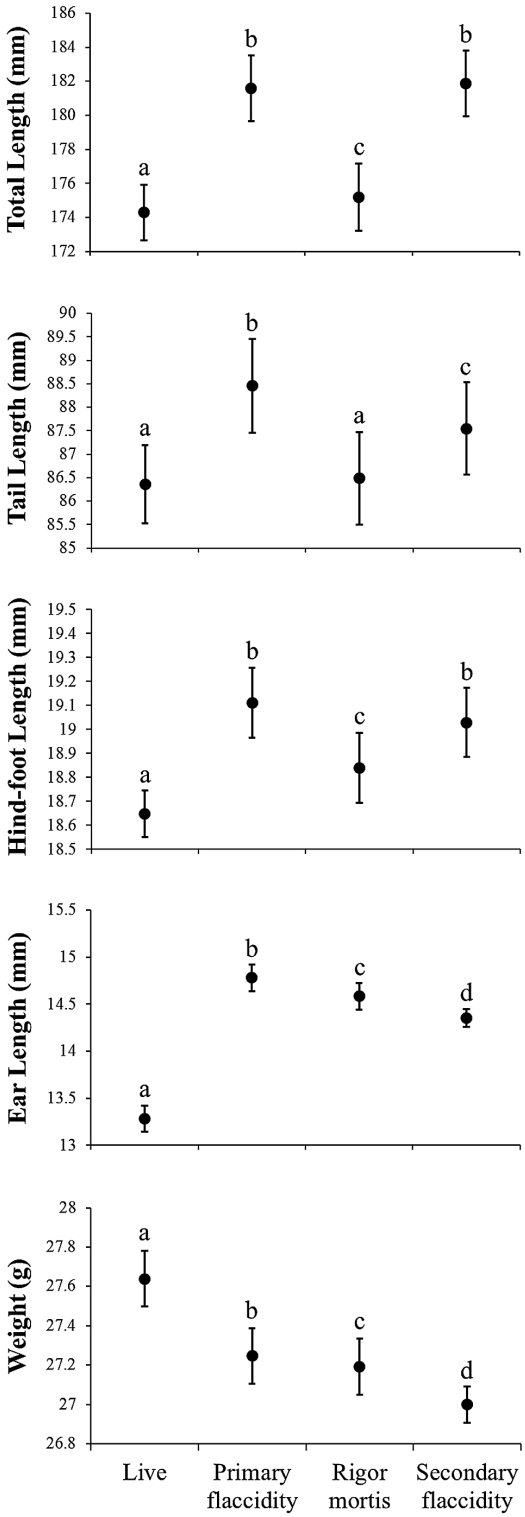

Fig. 1.—Mean values of 5 standard external measurements taken on 70 *Mus musculus* at 4 body states by 4 observers. Observers measured all 70 mice at each body state and took an additional 2 sets of measurements from 10 randomly selected mice at each body state. Data were fit using linear mixed effects models. Error bars represent *SE* around estimates after accounting for random effects of mouse and observer. Post hoc comparisons among body states were performed using Tukey's contrasts and values with the same letter are not significantly different, whereas different letters denote significant differences at $\alpha = 0.05$.

Table 1.—Multiplication factors for converting standard external measurements among living (L) and 3 postmortem body states (primary flaccidity [PF], rigor mortis [RM], and secondary flaccidity [SF]). Measurements used to create conversions were from 70 *Mus musculus* measured at each body state by 4 observers. For each standard external measurement, a pairwise matrix is provided with the appropriate multiplier for each body state comparison (starting with body state on left); self-similarity is shown along the diagonal. Note that weight conversions are not listed because weight is a function of desiccation over time and not body state per se (see "Discussion").

Body state	Total length				Tail length				Hind-foot length				Ear length			
	L	PF	RM	SF	L	PF	RM	SF	L	PF	RM	SF	L	PF	RM	SF
L	1.000	1.042	1.005	1.043	1.000	1.024	1.001	1.014	1.000	1.025	1.010	1.020	1.000	1.113	1.098	1.081
PF	0.960	1.000	0.965	1.001	0.976	1.000	0.978	0.990	0.976	1.000	0.986	0.996	0.899	1.000	0.987	0.971
RM	0.995	1.037	1.000	1.038	0.999	1.023	1.000	1.012	0.990	1.014	1.000	1.010	0.911	1.013	1.000	0.984
SF	0.958	0.999	0.963	1.000	0.986	1.010	0.988	1.000	0.980	1.004	0.990	1.000	0.925	1.030	1.016	1.000

body states were not significantly different ($P > 0.05$), but mean total length and mean hind-foot length were significantly larger in rigor mortis by 0.9 mm (0.5%) and 0.2 mm (1.0%), respectively. Mean total length and mean hind-foot length were not significantly different ($P > 0.05$) between primary and secondary flaccidity, but tail length was smaller during secondary flaccidity by 0.9 mm (1.0%).

Ear length showed the largest variation of any of the measurements among body states. Mean ear length at primary flaccidity was 1.5 mm (11.3%) larger than live ear measurements (13.3 mm) and decreased by 0.2 mm (1.3%) from primary flaccidity to rigor mortis. Unlike other length measurements that increased following rigor mortis, ear length continued to decrease from rigor mortis to secondary flaccidity by 0.2 mm (1.6%). Weight decreased by 0.4 g from live measurements to secondary flaccidity, likely a result of urination following euthanasia. Although we detected significant differences in weights between mice at primary flaccidity and rigor mortis, there was < 0.1 g (0.20%) difference. From rigor mortis (27.2 g) to secondary flaccidity, there was a 0.2 g (0.72%) decrease in body weight, indicating that after death, mice gradually decrease in weight.

Error effects.—Error estimates for the random effects are listed in Table 2; measurement error variance is the sum of the observer and residual variances. For each measurement, we found similar trends in unexplained residual error across body states. For all variables, residual error was greater for live measurements than for postmortem body states except for total length, which had similar residual errors between live and rigor mortis body states. There were no differences in residual error among any of the length measurements across the 3 postmortem states; however, weight had greater variability during secondary flaccidity. Total length and tail length had intraobserver and interobserver correlations close to 1.00, indicating they are highly reproducible, whereas hind-foot length and ear length had lower values, indicating that they are less reproducible both within and among observers. Weight had extremely high values for both (1.00), indicating that measurements were almost identical within and among observers. Based on proportion of residual error, the majority of random variance (88.99–99.98%) was explained by mouse (among mouse variation) for total length, tail length, and weight. Live measurements of hind-foot length and ear length were extremely variable, with 51.18% and 56.74% of the random variance unexplained, respectively. Variance of ear length and hind-foot length were more consistent during postmortem body states, with 60.31–73.85% of variance attributed to mouse. Variance due to observer, or the effect of an observer, composed a small portion of total variance (< 0.001–4.00%) for most measurements, with the exception of hind-foot length, of which observer error comprised 9.06–13.98% of total random variance.

Field measurements.—Results of measurements from wild *P. leucopus* closely mirrored those of our lab experiment with generally shorter measurements for live animals than specimens in secondary flaccidity. Mean measurements of total length, hind-foot length, and ear length were significantly smaller for live animals than in secondary flaccidity by 5.9 mm (3.7%), 0.4 mm (2.0%), and 1.1 mm (7.1%), respectively (Table 3). These differences in total length, hind-foot length, and ear length closely resemble those detected during our lab experiment of 4.3%, 2.0%, and 8.1%, respectfully. We found no significant differences for tail length or weight (Table 3).

Results of total length and tail length from wild *P. maniculatus* measured in primary flaccidity and rigor mortis were also similar to those of our lab experiment. Mean measurements of total length and tail length were smaller for animals in rigor mortis than for those in primary flaccidity by 3.7 mm (2.4%) and 3.7 mm (5.9%), respectively, despite being larger in weight (Table 3). Contrary to what we expected based on our experimental results, hind-foot length and ear length were larger in rigor mortis than in primary flaccidity by 1.1 mm (5.1%) and 2.2 mm (11.2%), respectively (Table 3). Yet, for the same observers, total length and tail length were consistent with our lab experimental results, indicating that bias caused by differences in observers can lead to statistically different results for these small measurements.

Discussion

This study is among the first to quantify the influence of body state on mean external measurements and their variances. We found statistical differences in all 5 standard measurements among 4 commonly encountered body states. Furthermore, the variance associated with measurement error was not consistent across body states and was particularly high for living compared to postmortem states. Moreover, we detect similar trends in field-collected data, underscoring the need to account for body state when using standard external measurements in research.

Table 2.—*SE* estimates due to mouse, observer, and residual for 5 standard external measurements taken on 70 *Mus musculus* at 4 body states by 4 observers. Observers measured all 70 mice at each body state and took an additional 2 sets of measurements from 10 randomly selected mice at each body state. Data were fit using linear mixed effects models with mouse and observer as random effects. Variance was partitioned into 3 components: among mouse variation (mouse), among observer variation in the same mouse (observer), and residual error (residual). Within a measurement, superscripts refer to overlap of 95% *CI*s, among body states, for *SD* of residual error. Values with the same superscript have overlapping 95% *CI*s, whereas different superscripts do not have overlapping *CI*s. $ICC_{interobserver}$ is the correlation among observers measuring the same mouse at a given body state and $ICC_{intraobserver}$ is the correlation among measurements from the same observer measuring the same mouse multiple times at a given body state. Measurement error variance is defined as the sum of residual and observer variances.

Measurement	Body state	*SE*			Random variance (%)			$ICC_{interobserver}$	$ICC_{intraobserver}$
		Mouse	Observer	Residual	Mouse	Observer	Residual		
Total length	Living	184.33	0.35	22.46[a]	88.99	0.17	10.84	0.89	0.89
	Primary flaccidity	184.33	0.35	8.70[b]	95.32	0.18	4.50	0.95	0.96
	Rigor mortis	184.33	0.35	17.23[a,b]	91.29	0.18	8.53	0.91	0.91
	Secondary flaccidity	184.33	0.35	7.39[b]	95.97	0.18	3.85	0.96	0.96
Tail length	Living	46.30	1.99	7.00[a]	83.74	3.61	12.65	0.84	0.87
	Primary flaccidity	46.30	1.99	2.60[b]	90.97	3.92	5.12	0.91	0.95
	Rigor mortis	46.30	1.99	2.04[b]	91.98	3.96	4.06	0.92	0.96
	Secondary flaccidity	46.30	1.99	1.47[b]	93.03	4.00	2.96	0.93	0.97
Hind-foot length	Living	0.49	0.11	0.63[a]	39.76	9.06	51.18	0.40	0.49
	Primary flaccidity	0.49	0.11	0.21[b]	60.31	13.74	25.95	0.60	0.74
	Rigor mortis	0.49	0.11	0.23[b]	58.92	13.43	27.65	0.59	0.72
	Secondary flaccidity	0.49	0.11	0.20[b]	61.31	13.98	24.71	0.61	0.75
Ear length	Living	0.47	0.02	0.65[a]	41.24	2.02	56.74	0.41	0.43
	Primary flaccidity	0.47	0.02	0.14[b]	73.85	3.61	22.53	0.74	0.77
	Rigor mortis	0.47	0.02	0.20[b]	67.62	3.31	29.07	0.68	0.71
	Secondary flaccidity	0.47	0.02	0.23[b]	64.86	3.17	31.97	0.65	0.68
Weight	Living	62.74	< 0.001	0.16[a]	99.74	0.00	0.25	1.00	1.00
	Primary flaccidity	62.74	< 0.001	0.01[b]	99.98	0.00	0.02	1.00	1.00
	Rigor mortis	62.74	< 0.001	0.01[b]	99.98	0.00	0.02	1.00	1.00
	Secondary flaccidity	62.74	< 0.001	0.04[c]	99.94	0.00	0.06	1.00	1.00

Table 3.—Mean, *SD*, range, and results of *t*-tests comparing 5 external measurements taken on 2 groups of individuals in 2 body states for *Peromyscus leucopus* (body states = living and secondary flaccidity) and *Peromyscus maniculatus* (body states = primary flaccidity and rigor mortis). *P. leucopus* were from Bartlett Experimental Forest, White Mountain National Forest, New Hampshire, 2013, and *P. maniculatus* were collected from Humboldt - Toiyabe National Forest, Nevada, 2011. All *P. leucopus* were measured by RBS. *P. maniculatus* in primary flaccidity were measured by D. Rogers and those in rigor mortis by RJR. An asterisk indicates significant differences at $\alpha = 0.05$.

Measurement	Body state														
	Living (n = 34)			Primary flaccidity (n = 87)			Rigor mortis (n = 85[a])			Secondary flaccidity (n = 18)					
	$\bar{X}$	*SD*	Range	$\bar{X}$	*SD*	Range	$\bar{X}$	*SD*	Range	$\bar{X}$	*SD*	Range	*t*	*d.f.*	*P*-value
P. leucopus															
Total length	158.0	11.3	132–179							163.9	7.9	147–174	−2.1983	45.945	0.0330*
Tail length	78.3	6.7	63.5–92							76.8	5.8	65–84.5	0.8609	39.416	0.3945
Hind-foot length	20.5	0.5	19–21							20.9	0.6	20–22	−2.1851	31.29	0.0365*
Ear length	15.4	0.8	14–18							16.5	0.7	15–18	−5.0196	36.659	< 0.0001*
Weight	19.7	2.5	15–24.5							19.1	2.6	15.1–24.8	0.7704	33.994	0.4464
P. maniculatus															
Total length				155.0	9.5	134–174	151.3	7.7	131–167				2.7672	163.262	0.0063*
Tail length				63.7	6.1	44–78	60.0	4.5	47–70				4.5484	157.165	< 0.0001*
Hind-foot length				19.9	0.8	18–21	20.9	0.7	19–22				−9.3424	169.799	< 0.0001*
Ear length				17.7	1.0	15–20	19.9	1.1	17–23				−14.249	167.338	< 0.0001*
Weight				18.4	2.5	15–25	20.0	3.3	15–29				−3.5457	156.878	0.0005*

[a] Total length and tail length for *P. maniculatus* in rigor mortis have samples sizes of 81 due to docked tails.

Standard measurements may be taken on mammals that are living or in 1 of 3 postmortem body states: primary flaccidity, rigor mortis, or secondary flaccidity. Living animals measured during capture-recapture studies have muscles that are rigid and flexed due to handling. Primary flaccidity occurs directly after death and is marked by muscle relaxation (Ota et al. 1973; Shkrum and Ramsay 2007) and is often encountered by field biologists who take measurements directly after euthanasia. Incidental mortalities or animals that are kill-trapped are often in rigor mortis. Rigor mortis occurs when the blood is depleted of oxygen, causing rigidity of muscles, and can last 5–6 h at 20°C and up to 15 h at 5°C (Ota et al. 1973; Kobayashi et al. 1996). Finally, secondary flaccidity is a relaxation of the muscles following rigor mortis but before putrefaction (Shkrum and Ramsay 2007) and is commonly observed in specimens that have been frozen and thawed for later preparation.

Measurement differences among body states.—Three of the 5 standard measurements (total length, tail length, and hind-foot length) followed similar trends across body states, increasing from living to primary flaccidity, decreasing to rigor mortis, and increasing again to secondary flaccidity. Measurement values during primary and secondary flaccidity were generally quite similar for both total length and hind-foot length. These measurements all involve skeletal muscle components and seem to be impacted by the resistance of muscles to stretching. Although total length was measured on live mice in a prone position and postmortem mice in a supine position, the overall changes were similar to those of other length measurements, indicating that the difference in position likely had little effect. Jewell and Fullagar (1966) also found no appreciable difference in body length when mice were measured in a prone and supine position. Ear length showed a similar increase from living to primary flaccidity; however, the difference from primary flaccidity to rigor mortis was minimal compared to that of the other length measurements. Ear muscles are small and only a few fibers thick (Murray et al. 2010), making them easy to manipulate, even during rigor mortis. Following freezing for 2 weeks, we found that ear length was actually shorter than at rigor mortis. The pinna has a large surface area and, along with the tail (tail decreased by 1% from primary to secondary flaccidity), likely started to desiccate during freezing, decreasing its overall length (Sumner 1927). Indeed, we noted a subtle but gradual decrease in body weight in postmortem states indicating that desiccation was occurring. Given these circumstances, both total length and hind-foot length are likely consistent between primary and secondary flaccidity, whereas it may not be advisable to use ear length, tail length, and weight recorded following prolonged freezing. Very small animals such as *Sorex* spp. may be especially vulnerable to desiccation because of their high surface area to volume ratio.

Measurement error.—For all 5 metrics, measurements taken on live animals were consistently more variable than in postmortem measurements. Live animals frequently move or try to bite while being measured, complicating measurement efforts. Total length was the only measurement to have similar variation between the live state and a postmortem body state. This may be a function of stretching the sinuous spinal column to varying degrees during rigor mortis (Jewell and Fullagar 1966). Hind-foot length and ear length had a higher proportion of residual errors and observer errors than total length and tail length. This is likely a function of the short length of hind-foot and ear measurements and greater difficulty in identifying the landmarks for these measures (Blackwell et al. 2006; Martin et al. 2013). Rounding errors to the nearest 0.5 mm also could have played a role in this variability, although Blackwell et al. (2006) used vernier calipers to the nearest 0.1 mm and found similar error trends.

Implications of external measurements in research.—Using field-collected data, we corroborated the patterns observed in our lab experiment. For *P. leucopus*, we found a significant decrease in total length, hind-foot length, and ear length measurements for live animals compared to those in secondary flaccidity. Similarly, *P. maniculatus* in primary flaccidity had greater total length and tail length measurements than those in rigor mortis. These findings suggest that when standard measurements from field studies or museum voucher labels are used, the body state under which they were measured should be considered. Many field guides list standard measurements for identification purposes, particularly for small mammals where they are intended to be used to distinguish superficially similar species (e.g., Kurta 1995; Reid 2006). The range of measurements given for these guides is generally taken from museum specimens that were prepared in a postmortem body state. Consequently, using measurements from field guides to distinguish live animals could lead to erroneous species identification. This may be particularly problematic for ear length, which showed a 1.5 mm (11.3%) increase from the live body state to secondary flaccidity (*M. musculus*; Fig. 1) and is often used for species differentiation (e.g., Reid et al. 2013; Stephens et al. 2014). In cases where standard measurements are the primary means of species identification in the field, it is advisable to take voucher specimens or tissue biopsies for confirmation. In cases when rare or sensitive species are involved, adjusting measurements by conversion factors in Table 1 may be satisfactory for making comparisons between field guides and live animal measurements.

External measurements are often used to calculate an index for animal size or health (e.g., weight or body size: total length minus tail length). If these measures are taken from different body states, they may confound comparisons among populations over space and time. This may be especially true in studies that use vouchered museum specimens generated by numerous collectors. Often animals collected by a researcher are captured and prepared in a similar manner. If a collector uses kill traps and specimens are measured and prepared in the field, animals may be disproportionately measured in rigor mortis compared to collectors who live trap and measure animals directly after euthanasia (primary flaccidity) or freeze specimens and prepare them later in the lab (secondary flaccidity). Unknowingly comparing measurements from a series of specimens prepared in rigor mortis to a series collected during secondary or primary flaccidity could lead to conclusions based on differences

due to body state and not actual morphological differences. Additionally, pooled data from different body states may give higher variance in measurements than is actually present. Accordingly, when vouchering specimens, we argue it is important to note if an animal is measured during rigor mortis. It is also important to record both the date of collection and the date of preparation, because measurements such as tail length, ear length, and weight may decrease from desiccation while frozen. These data will provide future researchers the flexibility to account for possible differences during statistical analysis or the prerogative of whether to include a specimen in a study. Although many collectors and institutions record such data, doing so is not standard practice.

We found that for all 5 measurements, the living body state had the highest amount of unexplained error compared with postmortem body states. Nevertheless, for total length, tail length, and weight, the proportion of observer and unexplained error was low and intra- and interobserver correlations were high, indicating that these metrics can be consistently measured among researchers. Hind-foot length and ear length are highly variable, especially on live mice. Blackwell et al. (2006) also found high error rates associated with these metrics. Our field measurements on *P. maniculatus* in rigor mortis and primary flaccidity confirm that differences between observers may cause significant differences between measurements. Caution should be exercised when using these measures to decipher small differences between populations or species. It is also important to avoid practices that may bias measurements, such as stretching of the pinna when taking an ear measurement or neglecting to measure to the end of the longest claw, which are often transparent and difficult to see in low light, when taking a hind-foot measurement. Finally, it is likely that our estimates of variances associated with live animal measurements are conservative compared to those encountered during field studies of live wild animals. The *M. musculus* used in our study were accustomed to humans and were easy to handle. More active species would have greater variance as they struggle during the measuring process. In particular, animals that are venomous (e.g., short-tailed shrews [*Blarina* spp.]) or those that are vigorous and difficult to handle, such as southern flying squirrels (*Glaucomys volans*), may have considerable variation associated with live measurements (Luiselli 2005).

External measurements have a number of qualities that make them well suited for addressing questions in mammalogy. Unlike many skeletal features, external measurements can be taken on live animals in the field, making them useful for distinguishing superficially similar species or describing phenotypic variation among populations without directly sacrificing animals (e.g., Stephens et al. 2014). Additionally, standard external measurements are taken during specimen preparation and are generally available for all museum voucher specimens. Moreover, these data are becoming more readily available via online databases such as VertNet (Constable et al. 2010). The increased accessibility of standard measurements and general coverage across both geographic and temporal scales makes them particularly useful for describing broad-scale morphologic patterns (e.g., Yom-Tov and Yom-Tov 2005; Mullen and Hoekstra 2008; Gardner et al. 2011). Nevertheless, our study found statistical differences in means and associated errors of external measurements depending on body state, and it is important to understand the limitations of external measurements for both live and postmortem mammals before using them in research. Specifically, the ability to use external measurements compiled from disparate sources (i.e., different researchers) to detect changes in body size over space and time may be limited due to differences in body state or observer. If statistical differences are detected for standard measurements from a compiled museum sample, but are within the range of variation found herein, they may be an artifact of bias and should be interpreted with caution.

Acknowledgments

We would like to thank E. Starke for tending to mice during the lab study. RBS would like to thank M. Mossman for early discussions that helped stimulate the impetus for this project. D. Rogers provided measurements for *Peromyscus maniculatus* in primary flaccidity. M. Yamaski and C. Costello provided postmortem specimens of *Peromyscus leucopus*. D. Hocking and M. Ducey helped with statistical analysis and interpretation of mixed effects models. This manuscript was greatly improved by comments from N. Woodman, L. Olson, J. Martin, and C. Dickman. The lab portion of this project was funded by the University of Wisconsin – Stevens Point Student Research Fund Program.

Literature Cited

Bailey, R. C., and J. Byrnes. 1990. A new, old method for assessing measurement error in both univariate and multivariate morphometric studies. Systematic Biology 39:124–130.

Blackwell, G. L., S. M. Bassett, and C. R. Dickman. 2006. Measurement error associated with external measurements commonly used in small-mammal studies. Journal of Mammalogy 87:216–223.

Bowers, M. A., and J. H. Brown. 1982. Body size and coexistence in desert rodents: chance or community structure? Ecology 63:391–400.

Constable, H., R. Guralnick, J. Wieczorek, C. Spencer, A. T. Peterson, and VertNet Steering Committee. 2010. VertNet: a new model for biodiversity data sharing. PLoS Biology 8:e1000309.

Dumbacher, J. P., G. B. Rathbun, T. O. Osborne, M. Griffin, and S. J. Eiseb. 2014. A new species of round-eared sengi (genus *Macroscelides*) from Namibia. Journal of Mammalogy 95:443–454.

Gardner, J. L., A. Peters, M. R. Kearney, L. Joseph, and R. Heinsohn. 2011. Declining body size: a third universal response to warming? Trends in Ecology and Evolution 26:285–291.

Hall, R. E. 1962. Collecting and preparing study specimens of vertebrates. Miscellaneous Publications - University of Kansas, Museum of Natural History 30:1–46.

Harestad, A. S., and F. Bunnel. 1979. Home range and body weight - a reevaluation. Ecology 60:389–402.

Hayen, A., R. J. Dennis, and C. F. Finch. 2007. Determining the intra-and inter-observer reliability of screening tools used in sports injury research. Journal of Science and Medicine in Sport 10:201–210.

Hothorn, T., F. Bretz, and P. Westfall. 2008. Simultaneous inference in general parametric models. Biometrical Journal 50:346–363.

Jewell, P., and P. Fullagar. 1966. Body measurements of small mammals: sources of error and anatomical changes. Journal of Zoology (London) 150:501–509.

Kobayashi, M., T. Takatori, K. Iwadate, and M. Nakajima. 1996. Reconsideration of the sequence of rigor mortis through postmortem changes in adenosine nucleotides and lactic acid in different rat muscles. Forensic Science International 82:243–253.

Kurta, A. 1995. Mammals of the Great Lakes region. University of Michigan Press, Ann Arbor.

Lackey, J. A., D. G. Huckaby, and B. G. Ormiston. 1985. *Peromyscus leucopus*. Mammalian Species 247:1–10.

Luiselli, L. 2005. Snakes don't shrink, but 'shrinkage' is an almost inevitable outcome of measurement error by the experimenters. Oikos 110:199–202.

Martin, J., M. Festa-Bianchet, S. Côté, and D. Blumstein. 2013. Detecting between-individual differences in hind-foot length in populations of wild mammals. Canadian Journal of Zoology (London) 91:118–123.

McNab, B. K. 1966. The metabolism of fossorial rodents: a study of convergence. Ecology 47:712–733.

Merriam, C. H. 1889. Brief directions for the measurement of small mammals and the preparation of museum skins. U.S. Department of Agriculture, Division of Economic Ornithology and Mammalogy, Circular 11:1–4.

Miller, G. S., Jr. 1899. Directions for preparing specimens of small mammals. 3rd ed. Bulletin of the United States National Museum 39:1–10.

Mullen, L. M., and H. E. Hoekstra. 2008. Natural selection along an environmental gradient: a classic cline in mouse pigmentation. Evolution 62:1555–1570.

Murray, L., T. Gillingwater, and S. Parson. 2010. Using mouse cranial muscles to investigate neuromuscular pathology in vivo. Neuromuscular Disorders 20:740–743.

Ota, S., Y. Furuya, and K. Shintaku. 1973. Studies on rigor mortis. Forensic Science 2:207–219.

Palmer, R. S. 1947. Notes on some Maine shrews. Journal of Mammalogy 28:13–16.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effects models in S and S-plus. Springer-Verlag, New York.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Development Core Team. 2012. nlme: linear and nonlinear mixed effects models. R package, version 3.1-105. http://cran.r-project.org/web/packages/nlme/index.html. Accessed 25 February 2013.

Rasolooarison, R. M., S. M. Goodman, and J. U. Ganzhorn. 2000. Taxonomic revision of mouse lemurs (*Microcebus*) in the western portions of Madagascar. International Journal of Primatology 21:963–1019.

R Development Core Team. 2013. R: a language and environment for statistical computing [version 2.15]. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org. Accessed 25 February 2013.

Reid, F. 2006. Peterson field guide to mammals of North America. 4th ed. Houghton Mifflin, New York.

Reid, R. E., E. N. Greenwald, Y. Wang, and C. C. Wilmers. 2013. Dietary niche partitioning by sympatric *Peromyscus boylii* and *P. californicus* in a mixed evergreen forest. Journal of Mammalogy 94:1248–1257.

Rowe, R. J., R. C. Terry, and E. A. Rickart. 2011. Environmental change and declining resource availability for small-mammal communities in the Great Basin. Ecology 92:1366–1375.

Schmidly, D. J., R. D. Bradley, and P. S. Cato. 1988. Morphometric differentiation and taxonomy of three chromosomally characterized groups of *Peromyscus boylii* from east-central Mexico. Journal of Mammalogy 69:462–480.

Shkrum, M. J., and D. A. Ramsay. 2007. Postmortem changes. Humana Press, Totowa, New Jersey.

Sikes, R. S., W. L. Gannon, and The Animal Care and Use Committee of the American Society of Mammalogists. 2011. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. Journal of Mammalogy 92:235–253.

Stephens, R. B., and E. M. Anderson. 2014. Effects of trap type on small mammal richness, diversity, and mortality. Wildlife Society Bulletin 38:619–627.

Stephens, R. B., E. M. Anderson, S. R. Wendt, and J. K. Meece. 2014. Field identification of sympatric *Peromyscus leucopus noveboracensis* and *P. maniculatus gracilis* in Wisconsin from external measurements. American Midland Naturalist 171:139–146.

Sumner, F. B. 1927. Linear and colorimetric measurements of small mammals. Journal of Mammalogy 8:177–206.

Taylor, J. R. 1997. An introduction to error analysis: the study of uncertainties in physical measurements. 2nd ed. University Science Books, Sausalito, California.

Wolff, J. O. 1985. The effects of density, food, and interspecific interference on home range size in *Peromyscus leucopus* and *Peromyscus maniculatus*. Canadian Journal of Zoology 63:2657–2662.

Yom-Tov, Y., and J. Yom-Tov. 2005. Global warming, Bergmann's rule and body size in the masked shrew *Sorex cinereus* Kerr in Alaska. Journal of Animal Ecology 74:803–808.

Zuur, A. F., E. N. Ieno, N. J. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology. R. Springer, New York.

Submitted 27 May 2014. Accepted 8 September 2014.

Associate Editor was Neal Woodman.