Testing Reliability of Body Size Measurements Using Hind Foot Length in Roe Deer

Mathieu Garel, Jean-Michel Gaillard, Thierry Chevrier, Jacques Michallet, Daniel Delorme, Guy Van Laere

Testing Reliability of Body Size Measurements Using Hind Foot Length in Roe Deer

MATHIEU GAREL,[1] *Office National de la Chasse et de la Faune Sauvage, Centre National d'Étude et de Recherche Appliquée Faune de Montagne, 5 allée de Bethléem, Z.I. Mayencin, F-38610, Gières, France*

JEAN-MICHEL GAILLARD, *Université de Lyon, Université Lyon 1, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5558, Laboratoire de Biométrie et Biologie Evolutive, Bâtiment G. Mendel, 43 boulevard du 11 novembre 1918, F-69622, Villeurbanne, France*

THIERRY CHEVRIER, *Office National de la Chasse et de la Faune Sauvage, Centre National d'Étude et de Recherche Appliquée Cervidés-Sanglier, 1 Place Exelmans, F-55000, Bar-le-Duc, France*

JACQUES MICHALLET, *Office National de la Chasse et de la Faune Sauvage, Centre National d'Étude et de Recherche Appliquée Cervidés-Sanglier, 1 Place Exelmans, F-55000, Bar-le-Duc, France*

DANIEL DELORME, *Office National de la Chasse et de la Faune Sauvage, Centre National d'Étude et de Recherche Appliquée Cervidés-Sanglier, 1 Place Exelmans, F-55000, Bar-le-Duc, France*

GUY VAN LAERE, *Office National de la Chasse et de la Faune Sauvage, Centre National d'Étude et de Recherche Appliquée Cervidés-Sanglier, 1 Place Exelmans, F-55000, Bar-le-Duc, France*

ABSTRACT We quantified the repeatability of >900 individual measures of hind foot length from 2 French populations of roe deer (*Capreolus capreolus*) monitored by capture–recapture. We found a high repeatability (i.e., high intra-class correlation, 0.76, 95% CI = 0.72–0.83 and 0.92, 95% CI = 0.91–0.95) in both populations. We also found that inexperienced observers reached a high level of intra- (1.00, 95% CI = 0.96–1.00) and inter-observer repeatability (0.99, 95% CI = 0.98–1.00) when measuring hind foot length of harvested animals with a tool specifically designed for this task. Managers should pay particular attention to limit measurement errors because unreliable measurements require an increased sample size to assess individual variation and can mask biological patterns.

KEY WORDS *Capreolus capreolus*, hind foot length, intra-class correlation, observer reliability, repeatability, reproducibility.

The study of individual variation in life history traits (LHT) occupies a central place in evolutionary ecology (Hayes and Jenkins 1997). However, the ability of ecological studies to identify individual variation and factors shaping them depends on the accuracy with which individual measurements are made. For instance, LHT such as fluctuating asymmetry (i.e., differences that occur between the right and left sides in bilateral characters) is known to be prone to measurement errors (Palmer and Strobeck 1986, Merilä and Björklund 1995). Given the usually low level of fluctuating asymmetry, reliability of published estimates has been hotly debated and several authors have stressed the necessity of considering measurement errors when analyzing fluctuating asymmetry (e.g., Palmer and Strobeck 1986).

Although numerous studies in large herbivores have shown the importance of studying body size variation to obtain insight into mechanisms causing fitness variation among individuals (e.g., Bonenfant et al. 2009; for a review see Gaillard et al. 2000), there are few studies addressing reliability of body size measurements (but see McLaren and Curran 2001 in moose [*Alces alces*]). Beyond biological aspects, accuracy of measurements has implications in terms of monitoring because imprecise measurements should require an increased sample size to detect biological patterns.

Hind foot length has recently been shown to be a relevant indicator of phenotypic quality in roe deer (*Capreolus capreolus*; Toïgo et al. 2006, Zannèse et al. 2006). In addition, hind foot length can be easily measured and collected over large areas and is not subject to temporal variation caused by varying degrees of fullness of the digestive tract or loss of body fluids as is body mass (Klein 1964, Zannèse et al. 2006). However, reliability of hind foot length measures has not yet been thoroughly investigated. We sought to fill the gap from longitudinal studies of 2 populations intensively monitored by capture–mark–recapture (see Gaillard et al. 2003). More specifically, we quantified repeatability of individual measures of hind foot length. A few experienced observers performed measurements of body size in one population, whereas several people including professionals and volunteers were involved in measurements in the other population. We therefore assessed the effect of observer qualification when measuring hind foot length by comparing both populations. Because hind foot does not experience negative growth during periods of physiological stress, has a high priority during early growth, and stops growing early, we can use repeated measurements of adults to assess measurement errors independently of density-dependent and -independent factors experienced by animals (Klein 1964, Navarre 1993).

We also performed an experiment involving inexperienced observers to assess both repeatability (i.e., intra-observer reliability) and reproducibility (i.e., inter-observer reliability; see Hayen et al. 2007) when measuring hind foot length of harvested roe deer with a tool specifically designed for this task.

STUDY AREA

We studied 2 French populations of roe deer intensively monitored for >30 years and managed by the Office

[1] *E-mail: mathieu.garel@oncfs.gouv.fr*

National de la Chasse et de la Faune Sauvage. The forests, both managed by the Office National des Forêts, were highly contrasting in habitat quality and, thereby, roe deer survival and reproduction markedly differed between populations (e.g., Gaillard et al. 2003). Roe deer at Trois Fontaines (TF), a 1,360-ha Territoire d'Étude et d'Expérimentation (48°42′N, 4°55′E) covered by a productive oak–beech (*Quercus* spp.–*Fagus sylvatica*) forest under continental climatic influences, had high survival and reproduction during most of the study period except for the last 5 years due to spring–summer droughts and experimental increase of population density (Pettorelli et al. 2006, Nilsen et al. 2009). On the other hand, roe deer at Chizé, a 2,614-ha Réserve Nationale de Chasse (CH; 46°07′N, 0°25′W) covered by a less productive oak–beech forest under oceanic and Mediterranean influences, had low and variable survival and reproduction during the study period because of frequent spring–summer droughts and episodes of density-dependence (Pettorelli et al. 2006, Nilsen et al. 2009).

METHODS

We intensively monitored both study populations using capture–mark–recapture methods starting in 1976 (TF) and 1978 (CH). We caught roe deer annually in January–February using drive-netting (i.e., about 5 km of vertical nets/capture day, 10–12 capture days/yr), a method approved by the French Environment Ministry (articles L.424-1, R.411-14, and R.422-87 of the French code of environment). Such captures allowed managers to control the size of these enclosed populations by removals. In addition, we ear-tagged newborns during the fawning period (May–Jun; Gaillard et al. 1993). We marked known-aged roe deer (either previously marked as newborn or caught first when <1 yr of age) we caught in winter with both ear tags and numbered collars, allowing long-term individual surveys by identification at a distance. We measured hind foot length of captured and recaptured animals each year between 1986 and 2009 (no data available in 1987 and 1992) and between 1985 and 2002 (no data available in 1986) at TF and CH, respectively. We measured the outstretched hind foot from the heel (top of the calcaneum) to the tip of the hoof (±1 mm at TF and ±0.5 mm at CH). We took measurements with a large caliper (Fig. 1a). Only a few experienced observers performed the measurements at CH, whereas >10 observers including professionals and volunteers were involved at TF.

We experimentally assessed the intra- and inter-observer reliability when measuring hind foot length using 10 observers and 4 hind feet collected on harvested roe deer (>1.5 yr old; massif des Bauges, 45°39′N, 6°5′E, 2008–2009). We took measurements using a tool specifically designed to improve measurement reliability (Fig. 1b). Each observer took 2 measures of every hind foot. Only one observer was familiar with the tool and all observers were inexperienced, because none of them routinely measured body size of roe deer as did professionals working regularly in the 2 study sites.

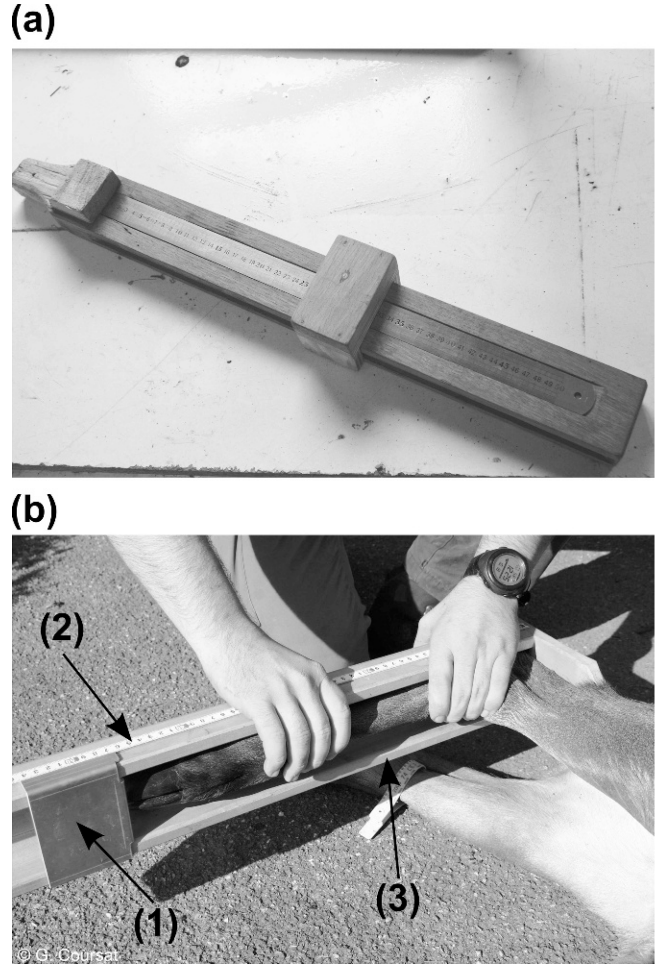

Figure 1. (a) The tool we used to measure hind foot length of roe deer caught at Chizé (1985–2002) and Trois Fontaines (1986–2009), France. Note that we put captured roe deer on a table, where they were held by 3 people during measurement. (b) The tool developed to make the measurement of the hind foot length easier and more accurate when measuring harvested ungulates. (1) Cursor in aluminum. (2) Flexible steel tape fixed on the wood. (3) Piece of lime tree of 670 mm × 100 mm × 70 mm; total weight of 400 g.

We did not record observer identity at either site so we could only assess repeatability of hind-foot length measurements of a given animal. We therefore used a one-way random-effects model, with animal identity as a random factor, to compute intra-class correlation coefficient at both sites (noted ICC(1) sensu Shrout and Fleiss 1979, McGraw and Wong 1996). We used statistical developments for unbalanced design to compute ICC(1) values and confidence intervals because number of measurements differed among animals (Burdick et al. 2006). The intra-class correlation coefficient represents the ratio between inter-individual variance and the sum of intra (i.e., measurement error) and inter-individual variances (McGraw and Wong 1996). The intra-class correlation coefficient is, therefore, close to 1 when there is high repeatability among hind foot measurements performed on the same animal and can, thus, be interpreted as a reliability index. We expected ICC(1) values to be greater at CH compared to TF due to differences of qualification among observers performing

Table 1. Number of roe deer ≥32 months old at Chizé (CH, 1985–2002) and Trois Fontaines (TF, 1986–2009), France, according to the number of repeated measurements of their hind foot length.

	Repeated measurements							
Study site	2	3	4	5	6	7	8	9
CH	46	26	14	13	7	2	2	2
TF	106	50	20	21	6	0	1	0

measurements of hind foot length. We restricted the analysis to roe deer for which >1 hind foot measurements were available. We considered animals ≥32 months old to avoid including any animal for which hind foot could still be growing (e.g., late-born fawn; Navarre 1993). Note, however, that performing the analyses on animals ≥20 months or ≥44 months old did not change estimates.

We analyzed data obtained from our experimental design using the approach of Hayen et al. (2007), assuming that observers were drawn randomly from a larger population. Indeed, we aimed to expand our results to a population of biologists larger than the 10 observers involved in the experiment. The model by Hayen et al. (2007) allows simultaneous consideration of intra- and inter-observer reliability by computing specific intra-class correlation coefficients, noted ICC_{intra} and ICC_{inter}, respectively. As for ICC(1), values close to 1 indicated high reliability. Intra-observer reliability measures the degree to which measurements taken by the same observer on a given hind foot are consistent (i.e., repeatability), whereas inter-observer reliability measures the degree to which measurements taken by different observers on a given hind foot are similar (i.e., reproducibility). We considered ICC values different ($P <$ 0.05) if 95% confidence intervals did not overlap.

We also computed descriptive statistics (e.g., maximal difference between repeated measurements, SD) at the individual level. In addition to standard deviation, we used the coefficient of variation of the hind foot length measurements. Coefficient of variation is a dimensionless number that allows for comparison of data sets with different means, due for instance to between-population differences in environmental conditions or sex structure.

We performed all analyses using R 2.6.0 (R Development Core Team 2007) and implemented the approaches of Burdick et al. (2006) and Hayen et al. (2007). The R codes are available upon request from the senior author.

RESULTS

We had 112 and 204 roe deer at CH and TF, respectively, with 2–9 measurements (CH: $\bar{x} = 3.4$, TF: $\bar{x} = 2.9$; Table 1). Average hind foot length was lower at CH (343.2 mm, CV = 2.8%) than at TF (356.7 mm, CV = 3.2%; $P < 0.001$ from a mixed model with animal identity as random factor). For CH, average maximal difference was 3.9 mm (maximal difference observed for a given hind foot of 31.5 mm) and the average standard deviation was 2.0 mm. For TF, average maximal difference was 7.3 mm (maximal difference of 45 mm) and average standard deviation was 4.0 mm. Average coefficient of variation was 2 times lower at CH (0.58%)

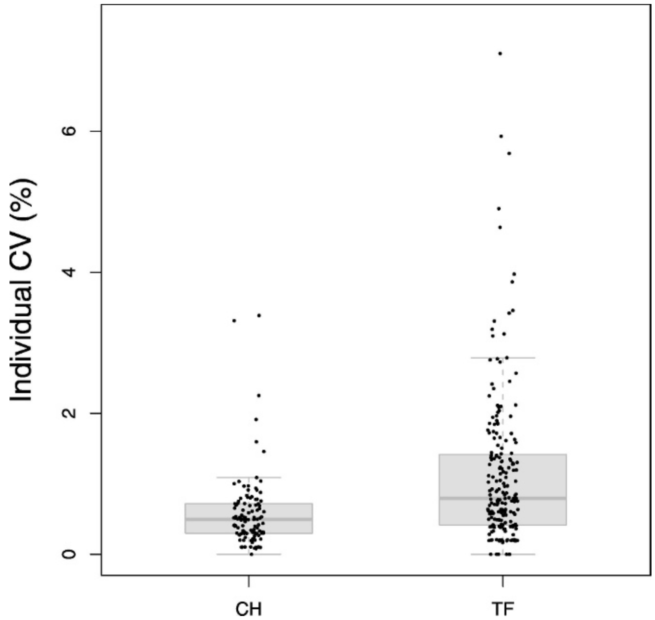

Figure 2. Coefficient of variation (filled circles) computed for each roe deer using repeated measurements of its hind foot length in the populations of Chizé (CH, 1985–2002) and Trois Fontaines (TF, 1986–2009; see Table 1), France. Boxes indicate, from bottom to top, the first, median, and third quartiles; vertical lines indicate the most extreme data points, which are <1.5 times within the interquartile range from the box.

compared to TF (1.13%; Fig. 2). Accordingly, ICC(1) was greater at CH (0.92, 95% CI = 0.91–0.95) than at TF (0.76, 95% CI = 0.72–0.83).

For our experiment, average hind foot length was 344.8 mm (CV = 2.3%). Maximal difference between observers for a given hind foot length was 4 mm, whereas maximal difference between 2 measurements for a given observer was 2 mm (Fig. 3). Average standard deviation (0.74 mm) and CV (0.22%) were lower than the best values reported for CH and TF. We found a higher ICC(1) (i.e., ignoring the observer effect; 0.99, 95% CI = 0.98–1.00) than observed at CH and TF. Repeatability within observers (ICC_{intra} = 1.00, 95% CI = 0.96–1.00) and reproducibility among observers (ICC_{inter} = 0.99, 95% CI = 0.98–1.00) were especially high, leading both intra- (0.59 mm, CV = 0.17%) and inter- (0.80 mm, CV = 0.23%) observer measurement errors to be low.

DISCUSSION

We found high repeatability of hind foot length measurements of roe deer in both populations we studied (e.g., Fig. 2). Our results also showed that differences in hind foot length <3–5 mm could be associated with measurement errors. For instance, at CH, a 2-fold increase in density led to a decrease of 16.6 mm in hind foot length of fawns, which therefore can be safely attributed to a density-dependent response of individuals (Toïgo et al. 2006).

Both the difference in reliability of measurements between CH (few experienced people) and TF (>10 people including professionals and volunteers) and the high repeatability in absolute terms we obtained at CH over >20 years emphasizes the importance of observer qualifi-

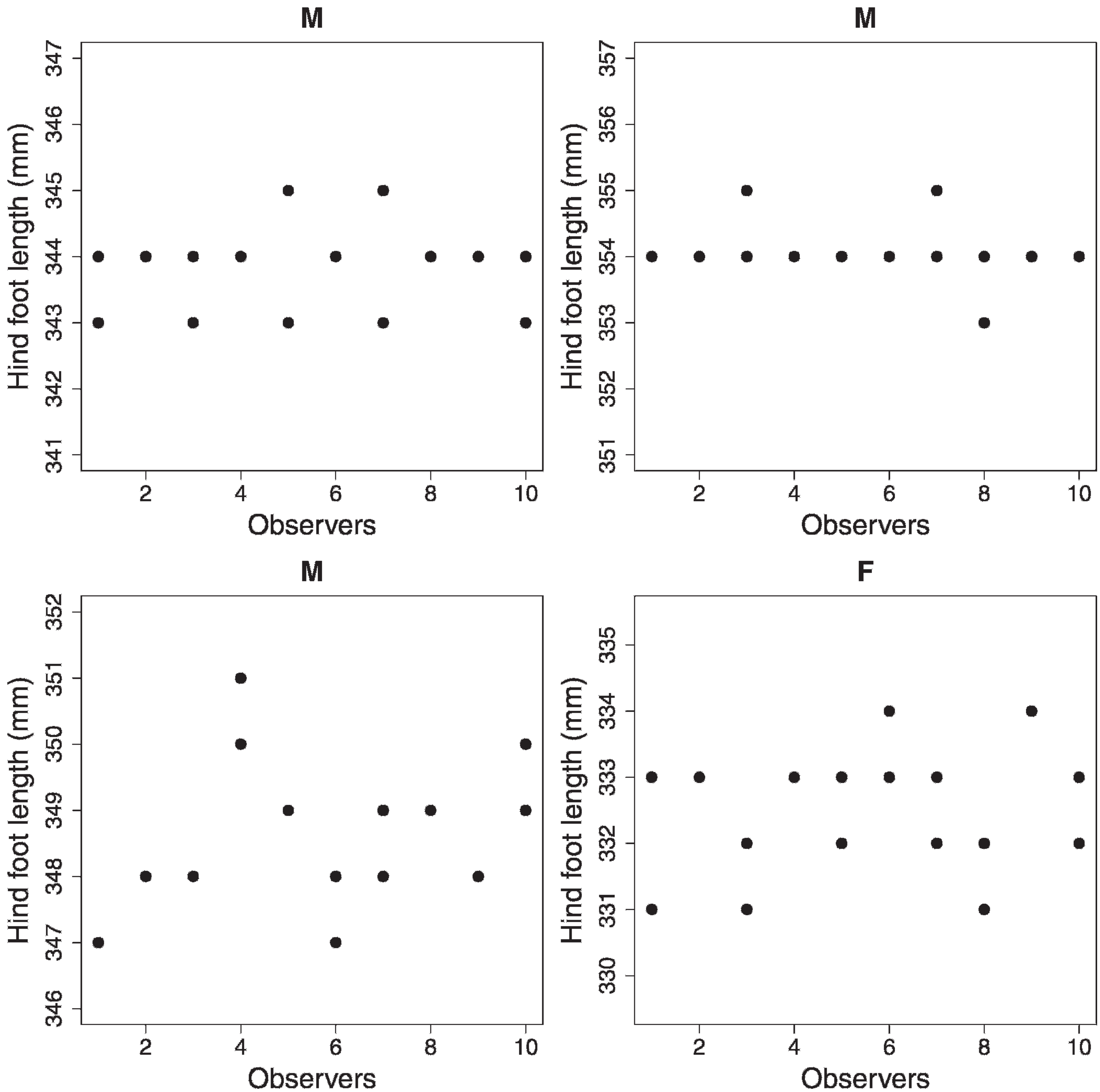

Figure 3. Hind foot length measurements taken by 10 observers on 4 hind feet collected on harvested roe deer, massif des Bauges, France, 2008–2009. Each observer randomly performed 2 measurements on every foot. The observer took 2 identical measures when only one length occurs on the plot.

cation. Part of this difference could also be explained by specific measurement procedure. We measured both hind feet of a given animal at CH so observers were able to detect a measurement error when comparing left and right foot measurements.

Management Implications

Increasing accuracy of measurements should help managers to detect variation in LHT that would have been overlooked by using unreliable measurements. Further, by increasing accuracy of measurement managers will need a lower sample size to detect a given biological process, thus reducing monitoring costs. Although our results highlight the effect of observer qualification on performance during body size measurements performed on the field, the practice of wildlife management depends upon long-term databases and collecting such data often requires employing numerous observers (Whitaker 2003). Therefore, many monitoring programs are undertaken with the assistance of volunteers interested in wildlife studies. Our field data suggest that measuring both hind feet might help to improve accuracy of measurements. When working on material collected from harvested animals, our experimental test showed that inexperienced observers reach a high level of repeatability and reproducibility. This high performance was probably largely explained by relying on a specific measurement tool designed to analyze data collected during hunting. As compared to other tools (e.g., Fig. 1a), the improvement was to add a gutter (Fig. 1b) that helps to lock the hind foot when taking measurements. Development of such tools may, therefore, be encouraged to standardize measurements over large spatial and temporal scales.

Acknowledgments

We thank V. Boulanger, G. Bourgoin, F. Couilloud, F. Klein, S. Marin, M. Montadert, E. Pfaff, and C. Redjaj for their participation in the experimental approach. We thank N. Guillon for helping to measure roe deer captured at Chizé. We are grateful to hundreds of volunteers that contributed to winter captures of roe deer in both sites. We gratefully acknowledge G. Domenge-Chenal for making the measurement tool of hind foot length. We also thank 2 anonymous referees for constructive comments on a previous draft of that paper.

LITERATURE CITED

Bonenfant, C., F. Pelletier, M. Garel, and P. Bergeron. 2009. Age-dependent relationship between horn growth and survival in wild sheep. Journal of Animal Ecology 78:161–171.

Burdick, R. K., J. Quiroz, and H. K. Iyer. 2006. The present status of confidence interval estimation for one-factor random models. Journal of Statistical Planning and Inference 136:4307–4325.

Gaillard, J.-M., D. Delorme, J.-M. Jullien, and D. Tatin. 1993. Timing and synchrony of births in roe deer. Journal of Mammalogy 74:738–744.

Gaillard, J.-M., P. Duncan, D. Delorme, G. Van Laere, N. Pettorelli, D. Maillard, and G. Renaud. 2003. Effects of hurricane Lothar on the population dynamics of European roe deer. Journal of Wildlife Management 67:767–773.

Gaillard, J.-M., M. Festa-Bianchet, N. G. Yoccoz, A. Loison, and C. Toïgo. 2000. Temporal variation in fitness components and population dynamics of large herbivores. Annual Review of Ecology and Systematics 31:367–393.

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.

Hayes, J. P., and S. H. Jenkins. 1997. Individual variation in mammals. Journal of Mammalogy 78:274–293.

Klein, D. R. 1964. Range related differences in growth of deer reflected in skeletal ratios. Journal of Mammalogy 45:601–622.

McGraw, K. O., and S. P. Wong. 1996. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1:30–46.

McLaren, B., and R. Curran. 2001. Accuracy in moose mandible size versus tooth wear assessments. Alces 37:13–17.

Merilä, J., and M. Björklund. 1995. Fluctuating asymmetry and measurement error. Systematic Biology 44:97–101.

Navarre, P. 1993. Distinction des chevrillards et des chevreuils (*Capreolus capreolus*) de plus d'un an par examen du métacarpe ou du métatarse. Gibier Faune Sauvage 10:135–142. [In French.]

Nilsen, E. B., J.-M. Gaillard, R. Andersen, J. Odden, D. Delorme, G. Van Laere, and J. D. C. Linnell. 2009. A slow life in hell or a fast life in heaven: demographic analyses of contrasting roe deer populations. Journal of Animal Ecology 78:585–594.

Palmer, R. A., and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis, patterns. Annual Review of Ecology and Systematics 17:391–421.

Pettorelli, N., J.-M. Gaillard, A. Mysterud, P. Duncan, N. C. Stenseth, D. Delorme, G. Van Laere, C. Toïgo, and F. Klein. 2006. Using a proxy of plant productivity (NDVI) to find key periods for animal performance: the case of roe deer. Oikos 112:565–572.

R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Shrout, P. E., and J. L. Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 86:420–428.

Toïgo, C., J.-M. Gaillard, G. Van Laere, A. J. M. Hewison, and N. Morellet. 2006. How does environmental variation influence body mass, body size, and body condition? Roe deer as a case study. Ecography 29:301–308.

Whitaker, D. M. 2003. The use of full-time volunteers and interns by natural-resource professionals. Conservation Biology 17:330–333.

Zannèse, A., A. Baïsse, J.-M. Gaillard, A. J. M. Hewison, K. Saint-Hilaire, C. Toïgo, G. Van Laere, and N. Morellet. 2006. Hind foot length: a new biological indicator for monitoring roe deer populations at a landscape scale. Wildlife Society Bulletin 34:351–358.