Evolutionary Anthropology 25:232–238 (2016)

ISSUES

# Transparency, Usability, and Reproducibility: Guiding Principles for Improving Comparative Databases Using Primates as Examples

CAROLA BORRIES, AARON A. SANDEL, ANDREAS KOENIG, EDUARDO FERNANDEZ-DUQUE, JASON M. KAMILAR, CAROLINE R. AMOROSO, ROBERT A. BARTON, JOEL BRAY, ANTHONY DI FIORE, IAN C. GILBY, ADAM D. GORDON, ROGER MUNDRY, MARKUS PORT, LAUREN E. POWELL, ANNE E. PUSEY, AMANDA SPRIGGS, AND CHARLES L. NUNN

Recent decades have seen rapid development of new analytical methods to investigate patterns of interspecific variation. Yet these cutting-edge statistical analyses often rely on data of questionable origin, varying accuracy, and weak comparability, which seem to have reduced the reproducibility of studies. It is time to improve the transparency of comparative data while also making these improved data more widely available. We, the authors, met to discuss how transparency, usability, and reproducibility of comparative data can best be achieved. We propose four guiding principles: 1) data identification with explicit operational definitions and complete descriptions of methods; 2) inclusion of metadata that capture key characteristics of the data, such as sample size, geographic coordinates, and nutrient availability (for example, captive versus wild animals); 3) documentation of the original reference for each datum; and 4) facilitation of effective interactions with the data via user friendly and transparent

Carola Borries, Department of Anthropology and Interdepartmental Doctoral Program in Anthropological Sciences, Stony Brook University, SUNY, Stony Brook, NY
Aaron A. Sandel, Department of Anthropology, University of Michigan, Ann Arbor, MI
Andreas Koenig, Department of Anthropology and Interdepartmental Doctoral Program in Anthropological Sciences, Stony Brook University, SUNY, Stony Brook, NY
Eduardo Fernandez-Duque, Department of Anthropology, Yale University, New Haven, CT and Facultad de Recursos Naturales, Universidad Nacional de Formosa, Formosa, Argentina
Jason M. Kamilar, Department of Anthropology and Graduate Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA
Caroline R. Amoroso, Department of Evolutionary Anthropology, Duke University, Durham, NC
Robert A. Barton, Evolutionary Anthropology Research Group, Durham University, Durham, UK
Joel Bray, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ
Anthony Di Fiore, Department of Anthropology, University of Texas, Austin, TX
Ian C. Gilby, School of Human Evolution and Social Change and Institute of Human Origins, Arizona State University, Tempe, AZ
Adam D. Gordon, Department of Anthropology, University at Albany, SUNY, Albany, NY
Roger Mundry, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Markus Port, Department of Behavioral Ecology, University of Goettingen, Goettingen Germany
Lauren E. Powell, Evolutionary Anthropology Research Group, Durham University, Durham, UK
Anne E. Pusey, Department of Evolutionary Anthropology, Duke University, Durham, NC
Amanda Spriggs, Department of Anthropology, University at Albany, SUNY, Albany, NY
Charles L. Nunn, Department of Evolutionary Anthropology and Duke Global Health Institute, Duke University, Durham, NC

Key words: data provenance; metadata; primary sources; procedure documentation; user interaction


DOI: 10.1002/evan.21502
Published online in Wiley Online Library (wileyonlinelibrary.com).

interfaces. We urge reviewers, editors, publishers, database developers and users, funding agencies, researchers publishing their primary data, and those performing comparative analyses to embrace these standards to increase the transparency, usability, and reproducibility of comparative studies.

From the beginning of evolutionary biology, the comparative method has been a major analytical tool,[1–3] allowing the examination of patterns and processes of evolutionary change.[4] Some of the main obstacles to overcome in comparative analyses have been statistical in nature: How should we control for confounding variables? What criteria should we use to assess whether patterns are statistically significant and biologically meaningful? How should we control for the nonindependence of comparative data that stems from phylogenetic relatedness? Much progress has been made with respect to these issues, especially in the development and use of phylogenetic comparative methods.[3,5,6] For example, building on initial descriptions of phylogenetically independent contrasts,[7] methods can now incorporate phylogenetic uncertainty,[8,9] intraspecific variation,[10–12] and different models of phenotypic evolution.[13,14]

Although phylogenetic and statistical methods are rapidly advancing, an increasing number of researchers argue that the data to which these methods are applied are "stuck in the Dark Ages."[15–17] It is imperative that before the specific methods used in a comparative study are considered, the suitability of the data be thoroughly evaluated. The time has come to bring our comparative databases into the modern age and to represent uncertainty in our data in the same way we might represent uncertainty in a statistical model or a phylogeny.[18] It is also important that we be able to evaluate which sources of uncertainty — in the data, the phylogeny, and the statistical methods — have the greatest influence on comparative results.

To approach these issues, the authors met on May 28, 2014 at the National Evolutionary Synthesis Center (NESCent, Durham, NC) and identified four guiding principles for improving comparative databases. We focused on primates, a relatively well-studied mammalian order that is the subject of many comparative studies. However, these concerns and suggestions are relevant to all taxonomic groups and disciplines.[19–21] Here, we begin by identifying the problems shared by investigators of a wide range of comparative questions involving morphology, life history, behavior, and ecology.

**It is imperative that before the specific methods used in a comparative study are considered, the suitability of the data be thoroughly evaluated. The time has come to bring our comparative databases into the modern age and to represent uncertainty in our data in the same way we might represent uncertainty in a statistical model or a phylogeny.**

Because data points used in comparative analyses cannot always be traced to actual measurements, one problem is related to data provenance. Gestation length in proboscis monkeys (*Nasalis larvatus*) appears to be such a case. As far as we know, the gestation length of this species has yet to be determined. Nevertheless, in almost all primate life-history compilations, gestation length for proboscis monkeys is reported as 166 days.[22–24] Moreover, this value has been used in many comparative studies.[25,26] The value of 166 days appears to originate from Schultz[27:281] who, 74 years ago, stated that "Nothing is known in regard to the duration of the various periods of growth in the proboscis monkey, but it may be assumed that these do not differ radically from the conditions in macaques. In the latter pregnancy is known to last 166 days . . . ."

Such a statement was acceptable at the time, when the strong allometric relationship between body mass and life history traits was less widely appreciated. Today, however, it is unreasonable to assume a similar gestation length in two species having such different adult female body masses (10.5 kg in the proboscis monkey versus 4.9 kg in rhesus[28]). Such erroneous claims may be perpetuated in any study, but comparative studies are particularly vulnerable, as the authors are unlikely to have in-depth knowledge of every taxon included in the analysis (see also Simoes and colleagues[29]). Unfortunately, such inappropriate attribution of data to particular sources is a recurring problem in comparative databases of primate life history (Borries and coworkers, unpublished compilation).

In other cases, data from the primary literature that percolate into comparative studies may reflect results that are of questionable value because of small sample sizes, short study periods, or a specific research design. For example, a group of wild long-tailed macaques (*Macaca fascicularis*) was characterized as 100% fruit-eating during five months of the year based on observations of 2, 3, 7, 7, and 20 instances of feeding per month. In another month, the same group was considered 100% grass-eating based on a single feeding observation.[30] Clearly, one cannot be confident in any characterization of the diet of a population when assessed by only a handful of isolated feeding observations, yet these values have subsequently been used in comparative tests (for example, by Tsuji, Hanya, and Grueter[31] and Jablonski and coworkers[32]).

Similarly, certain methods for recording data may lead to results of limited general use. This is so, for example, when sizes, compositions, or densities are estimated for unhabituated groups, because many members may have fled at the observer's approach, or during brief encounters on transects when many individuals are out of sight for similar reasons (for example, Patterson and coworkers[33]). Such approaches may result in group sizes being underestimated and group densities being overestimated by a factor of 2 or more.[34]

All these issues relate to the accuracy of individual data points; that is, how closely data match the "true" value that is free from systematic errors.[35] While it may be difficult to demonstrate the extent of bias introduced by a single inaccurate data point, the basic problem runs deeper. As scientists, we are obliged to provide and use accurate data. Inaccurate data have the potential to reduce reproducibility and lead to poor use of time and resources.[36] Importantly, in the few cases currently published in our field, existing databases were found to contain multiple data points with problems like those described.[15,33] Using such inaccurate data in comparative analyses may bias the results or lead to failure to detect existing patterns.

To evaluate the effect of accuracy, Borries, Gordon, and Koenig[15] compared gestation lengths in two primate taxa, Asian colobines and Asian macaques, with data drawn from four published life-history compilations. Gestation length is expected to be similar among closely related species and to vary with body mass.[37] However, the authors, using a phylogenetic generalized least squares model, found no statistically significant relationship between gestation length and body mass or taxon in any of the four datasets. In contrast, the model based on a fifth set of data containing only entries checked for accuracy produced the expected relationship ($R^2_{adj} = 0.91$, $P < 0.001$), with statistically significant effects for body mass ($P < 0.007$) and taxon ($P < 0.001$).[15]

Comparative studies may also be compromised when data collected under different criteria or by incompatible methods are pooled under the same trait. Even a trait as seemingly straightforward as body mass may generate substantial errors (beyond resolution and precision of the scales used[38]) when data using different definitions of "adult" are lumped together.[39] In a study of ape morphology, 28% of chimpanzee skeletons and 38% of gorilla skeletons with adult dentition had bones that were still growing and likely had not yet reached adult mass.[40] Thus, despite not being fully grown, these individuals would have been classified as adults if dentition were used as the defining characteristic for adulthood. Another trait for which a range of definitions has been applied is "weaning age," which in primates can include age at first intake of solid food (within the first weeks or months of life); observations of conflict over access to the nipple; the ability to survive as an orphan; or the age at cessation of nipple contact,[41,42] which in extreme cases can average 6.5 years (Bornean orangutan, *Pongo pygmaeus wurmbii*[43]).

Subsuming data based on such vastly different definitions into the single trait weaning age will be unlikely, in most cases, to produce a false-positive result (Type I error). Compared to the earlier example on accuracy in gestation length, in this case each data point may be accurate, but the researcher is implicitly using different definitions, thus compromising the compatibility of the data. Combining incompatible data, as in the case of weaning age, may prevent us from detecting real patterns[44] or from determining the relative strength of different effects.[45]

Another example, from the study on life history in Asian colobines and Asian macaques,[15] illustrates just this point. Data for age at first reproduction and reproductive rates were drawn from only primary sources and checked for accuracy, trait definitions, and data collection methods. Still, the comparison revealed no statistically significant relationship between body mass and either of these life-history variables. This result can be explained by the fact that data from different nutritional conditions, captive and wild, were combined in the dataset, since nutritional intake is known to have a great impact on maturation and reproductive output.[46] Including additional information on nutrient availability as a simple binary variable resulted in greatly improved models with an $R^2_{adj.}$ value of 0.42 for age at first reproduction ($P$ value for body mass = 0.01) and an $R^2_{adj.}$ value of 0.65 for reproductive rates ($P$ value for body mass $< 0.01$).

Together, these issues related to data accuracy and comparability may lead to conflicting results that cause more confusion than clarity and thus slow our progress toward finding general patterns. Consequently, reproducibility may become impossible and results from different studies may differ, producing less rather than more certainty in our conclusions. While it is often unclear why results vary among comparative studies, we suspect that many differences emerge because researchers not only use different data-collection protocols and divergent definitions for traits of interest, but also rely on inaccurate, incomplete, or imprecise compilations as datasets. All of these circumstances can and should be improved.

## FOUR GUIDING PRINCIPLES TO IMPROVE COMPARATIVE DATA

To tackle these issues, the authors discussed the status of comparative databases for the mammalian order Primates. Here, we outline four guiding principles that we believe have the potential to improve future comparative studies. We were guided by our experience with various kinds of datasets, analyses, and questions, and by Whitlock's[47:62] advice: "The central goal to have in mind ... is to ensure that a new user, perhaps someone unknown to you working with the data 20 years later, can correctly interpret the results and derive correct conclusions from the data." Thus, in addition to ensuring the transparency of a database and its usability in the present, we were also concerned about future reproducibility of the result. This "call to arms" agrees with several of the standards of the Transparency and Openness Promotion Guidelines,[48,49] which, among others, call for a standardization of research procedures and a clear description of all aspects of data collection and definition.

### Principle 1: Defining and Describing Data Precisely

The path to improvement begins with unequivocal, complete descriptions and definitions of all variables included in a comparative database, with details about how data were measured or determined.[50] Before naming a variable, it is recommended that researchers check the literature to identify and use previously published definitions. Using precise operational definitions also provides explicit criteria for including and excluding data from a comparative database. Importantly, once a definition has been chosen, only data matching it should be included in a database. We suggest reporting in comparative databases the means, standard deviations, medians, and ranges when available, or calculations of these measures when an original source provides raw data.

To guarantee transparency and reproducibility, data provenance is of crucial importance. It must be possible to trace each data point back to its original source. To this end, every alteration, even a mere conversion of dimensions (for example, from days to weeks or centimeters to millimeters), should be identified in the database by using, for example, a Boolean data type to indicate whether a certain action was performed or not.

When compiling data, the rate of transcription errors can be much reduced by using a relational database[3] or a "not only SQL" approach.[51] An additional essential aspect of generating high-quality databases is the proofreading and double-checking process.[52] Ideally, someone other than the person who entered the data would do the double-checking. As a final step, we recommend having an external expert examine the selected data. This could be an author of the underlying primary sources or somebody familiar with the taxon and its relevant literature.

### Principle 2: Tagging Data with Metadata

It is essential to include additional information (metadata) to further characterize and contextualize the primary data used in a comparative analysis. Beyond those categories summarized in the Dublin Core Metadata Initiative (http://dublincore.org/) and Darwin Core (http://rs.tdwg.org/dwc/),[53] we strongly recommend including location of sampling (geographical coordinates and their precision), time period, study duration, number of groups, number of individuals, and other measures of sample size. Some of these types of metadata are being included in proposed extensions to Dublin and Darwin Core, including PaleoCore

**To guarantee transparency and reproducibility, data provenance is of crucial importance. It must be possible to trace each data point back to its original source. To this end, every alteration, even a mere conversion of dimensions (for example, from days to weeks or centimeters to millimeters), should be identified in the database**

(http://paleocore.org/) for paleobiology datasets and EthoCore (http://ethoinformatics.org/) for behavioral and ecological datasets. Metadata are essential components of comparative databases to capture trait variation within and among species (see Strier and colleagues[50]). They furthermore allow for a gross quality assessment (for example, sample size, number of individuals) and enable users of the data to select particular types, such as only those studied for a specified minimum time period, for specific analyses.

Some metadata also help to categorize the core methods used for data collection (for example, gestation length based on conceptions estimated by hormonal concentrations as opposed to mating patterns) or data analysis (such as home-range sizes calculated using minimum convex polygons rather than local convex hull or kernel density methods). Information on ecological context and nutrient availability is also crucial, given that captive and wild animals have different nutritional regimes, which may affect key variables such as body mass or speed of growth and reproduction.[54–56] Although distinctions among captivity, provisioning, and food-enhanced conditions (crop raiding) are often possible,[39] it may suffice to indicate whether or not the study animals consumed any kind of human-made food.[15] This enables the compiler of comparative data to include, for example, data from captivity and the wild into a single database and then control for nutritional condition in the analysis.

### Principle 3: Documenting Procedures Thoroughly

To maintain reproducibility, a comparative database requires a written protocol that describes the specific search strategy used to locate data that subsequently were selected and included. Such documentation will also include the list of terms used in online search engines, how primary sources were located, and what other search variables or methods, such as searches within a given species, by study site, or by variable, were used. These protocols should be clearly written and linked to the database and/or provided as publications that describe or use the database.

In comparative databases, every datum is ideally documented by providing its source (the full reference for its first publication) together with the page number and/or Table or Figure number, as appropriate. This ensures that the primary source indeed exists (unlike the proboscis monkey example discussed earlier) and allows for speedy location of data even within extensive sources such as books or theses. Further, it enables users to reconcile discrepancies in existing datasets.

Trait definitions, sampling methods, and actual data values can be directly extracted only from the primary source. On occasion, relevant metadata may have to be retrieved from other primary sources, and it is important that the trail to those sources also be provided. Past compilations can be helpful in locating primary sources, but they themselves cannot serve as primary data sources. The only exceptions are databases assembled in accordance with the principles outlined here. When an analysis is published based on a comparative database, the version number of the database should be explicitly identified.[52] Before using such a database, however, we recommend conducting multiple, random spot checks against the primary sources provided for a taxon that the author is very familiar with, and to only use databases with very low error rates.

### Principle 4: Facilitating Effective Interactions with the Data

Ideally, comparative databases are made publicly available, leaving it to the user to decide if and which of its components are considered important. To facilitate access to a database by users who are unfamiliar with its basic contents and structure, all key components of the comparative database, including the schema that illustrates the relationships between the different components, can be summarized in a concise yet complete "read me" file. In addition, an index and table of contents will facilitate orientation and provide a first overview.

All elements of a database, including metadata, are best made available for download in a widely available format, such as text files, thus allowing a wide range of future uses via different programs. We recommend providing a clear, largely self-explanatory output design with an easy-to-understand web-browser-based Graphical User Interface (GUI) that allows for a limited set of query options. The interface will also help prevent accidental misuse of the database, so that fields can be combined only in the ways intended. Users will be unable to gain direct access to the underlying database (where errors could be introduced). A GUI can best be improved by running extensive test queries prior to release. Such queries should simulate data extractions required for already published analyses. These processes usually take more time than anticipated, but they are extremely important because their outcome may decide the success or failure of a database.

## ADDITIONAL MEASURES TO IMPROVE COMPARABILITY

Implementing the principles we have outlined requires the support of authors and reviewers of the primary literature, as well as editors and publishers, developers and users of comparative databases, and funding agencies. We now consider each of these in turn.

Authors publishing primary data that might later be collated into a comparative database play a key role in setting standards for the available data. In particular, principles 1 and 2 demonstrate the importance of clear variable definitions and associated contextual information that can be used later as metadata. Authors are encouraged to publish this information, even if it is unnecessary for their current manuscript. The extra information can be presented as supplemental material and by referencing published work containing these data. In the near future, we hope that we, as a discipline, will agree on explicit guidelines to standardize data even before they are collected.

Reviewers are also essential to improving the primary and comparative data reported. As manuscript reviewers, we can all contribute toward implementing new rigor by requesting additional explanatory information from our peers to meet the standards stated here. Compliance may be best achieved when reviewers articulate to authors why following standards will increase the impact of their research rather than simply setting a bar for authors to reach to achieve publication.

Editors and publishers can support the process by allowing the inclusion of additional information. They also can attach contingencies to acceptance of primary and comparative research papers, such as data upload in respective databases, before a publication can go online. A good example is in place for DNA sequences, which are submitted to and made available through NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/). This is also an elegant solution to keep existing databases updated in the long run. Recording and publishing data so that they can be used in comparative databases is an important core contribution and should be rewarded with recognition in the form of citations. This would require some changes to the data reporting process in comparative publications.[57] We support referencing the authors of the data compilations used, as well as all primary sources considered in the respective analysis. Change is already under way in several scientific fields. Multiple journals have adopted a catalogue of increasingly stringent standards developed by the Transparency and Openness Promotion Committee.[48]

Database developers can also help achieve these new standards by working closely with scientific experts from the targeted research areas. Standardization can be facilitated by providing access to the underlying metadata of a database; in this way, authors of primary data become aware of what to report and which metadata to include in their work. In addition, developers can play a major role in making databases more comprehensive and, at the same time, easy to use.

Finally, database users can play a major part by providing feedback on individual entries in the databases they use. Their input can be facilitated by web-based portals allowing flexible comments and information transfer, with the aim of completing or correcting specific database content.

Improving transparency and comparability is a slow process requiring recognition by research sponsors. We suggest that funding agencies approach this in several ways. Standardization of data reporting and sharing could be implemented in data management plans. In addition, by including these standards in calls for proposals, as well as specifically calling for development of comparative databases, incentives will exist to invest in the principles we have outlined. Funding agencies can also provide options for long-term maintenance and continued improvement of existing databases. Finally, any effort to improve transparency, usability, and reproducibility of data should be honored in decisions made by funding agencies, as well as tenure and promotion committees.

## MOVING FORWARD

It is the collective responsibility of all researchers building and using comparative datasets to assess the effect of data accuracy and compatibility on the results of comparative studies. We already have evidence that data for wild animals differ from those for captive or provisioned ones,[56,58–60] that body mass data are prone to large errors,[39,40] and that estimates of group size are sensitive to sampling methods.[33] We are also gaining a better understanding of the consequences of intraspecific variation on some key associations, such as those between neocortex size and group size.[61] We need more such studies. Unfortunately, achieving similar results based on different datasets[25] is no guarantee of accuracy, since existing compilations are often strongly interdependent and may contain similar or even identical flaws.[15]

We are aware that the suggested guiding principles toward transparency, usability, and reproducibility come at a price. The process requires a major time investment that will slow down comparative research until databases become available that are in compliance with the standards proposed here. We are also aware that more transparency of research methods and materials in comparative databases is just a first step. Different studies often use different methods for data collection and analysis, the results of which may be difficult or even impossible to compare. Reaching standardization at the level of data collection, as well as analyses, is an additional and important goal for the future. We are reminded of Felsenstein's[7; 14] original call to arms to use phylogenies in comparative studies. He noted that "Some reviewers of this paper felt that the message was 'rather nihilistic ....'" Yet in the past 30 years, a huge diversity of new methods and phylogenies has emerged to fill the gap he identified. We are now at a similar point with regard to improving comparative databases, so that these methods can be applied to their best effect and the findings are more certain. We urge authors, reviewers, editors, publishers, database developers, and users, as well as funding agencies and compilers of data, to embrace these guiding principles, to honor the accompanying efforts, and to help us generate new knowledge.

## ACKNOWLEDGMENTS

We gratefully acknowledge the National Evolutionary Synthesis Center (NESCent, Durham, NC), which is supported by the NSF (EF-0905606), for hosting and supporting our meeting on May 28, 2014. Additional support was provided by NSF Biological Anthropology (BCS-1355902) and Duke University. We thank Robyn Overstreet for fruitful discussions and contributions. We are also grateful to Charles Janson, an anonymous reviewer, and John Fleagle, for helpful comments and suggestions on the manuscript.

## REFERENCES

**1** Tinbergen N. 1963. On aims and methods of ethology. Z Tierpsychol 20:410–433.

**2** Darwin C. 1859. On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life. London: Murray.

**3** Nunn CL. 2011. The comparative approach in evolutionary anthropolgy and biology. Chicago: University of Chicago Press.

**4** Harvey PH, Pagel MD. 1991. The comparative method in evolutionary biology. New York: Oxford University Press.

**5** Garamszegi LZ, editor. 2014. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Berlin: Springer.

**6** Garland TJ, Bennett AF, Rezende EL. 2005. Phylogenetic approaches in comparative physiology. J Exp Biol 208:3015–3035.

**7** Felsenstein J. 1985. Phylogenies and the comparative method. Am Nat 125:1–15.

**8** Pagel MD, Lutzoni F. 2002. Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. In: Laessig M, Valleriani A, editors. Biological evolution and statistical physics. Berlin: Springer. p 148–161.

**9** Pagel MD, Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. Am Nat 167:808–825.

**10** Ives AR, Midford PE, Garland TJ. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Syst Biol 56:252–270.

**11** Hansen TF, Bartoszek K. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. Syst Biol 61:413–425.

**12** Garamszegi LZ. 2014. Uncertainties due to within-species variation in comparative studies: measurement errors and statistical weights. In: Garamszegi LZ, editor. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Berlin: Springer. p 157–199.

**13** Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

**14** Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164:683–695.

**15** Borries C, Gordon AD, Koenig A. 2013. Beware of primate life history data: a plea for data standards and a repository. PLoS One 8: e67200.

**16** Ioannidis JPA. 2005. Why most published research findings are false. PLoS Med 2:e124.

**17** Freedman LP, Cockburn IM, Simcoe TS. 2015. The economics of reproducibility in preclinical research. PLoS Bio 13:e1002165.

**18** Arnold C, Matthews LJ, Nunn CL. 2010. The 10kTrees website: a new online resource for primate phylogeny. Evol Anthropol 19:114–118.

**19** Drummond CG. 2009. Replicability is not reproducibility: nor is it good science. Proc Eval Methods Mach Learn Workshop 26th ICML, Montreal, Quebec, Canada. http://www.csi.uottawa.ca/~cdrummon/pubs/ICMLws09.pdf.

**20** Begley CG, Ellis LM. 2012. Raise standards for preclinical cancer research. Nature 483: 531–533.

**21** Open Science Collaboration. 2012. An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. Perspect Psychol Sci 7:652–655.

**22** Jones KE, Bielby J, Cardillo M, et al. 2009. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecology 90:2648.

**23** Harvey PH, Martin RD, Clutton-Brock TH. 1987. Life histories in comparative perspective. In: Smuts BB, Cheney DL, Seyfarth RM, et al., editors. Primate societies. Chicago: University of Chicago Press. p 181–196.

**24** Kappeler PM, Pereira ME, editors. 2003. Primate life histories and socioecology. Chicago: University of Chicago Press.

**25** Jackson G, Mooers AO, Dubman E, et al. 2014. Basal metabolic rate and maternal energetic investment durations in mammals. BMC Evol Biol 14:194.

**26** Bielby J, Mace GM, Bininda-Emonds ORP, et al. 2007. The fast-slow continuum in mammalian life history: an empirical reevaluation. Am Nat 169:748–757.

**27** Schultz AH. 1942. Growth and development of the proboscis monkey. Bull Mus Comp Zool 89:279–314.

**28** Gordon AD. 2006. Scaling of size and dimorphism in primates II: Macroevolution. Int J Primatol 27:63–105.

**29** Simoes TR, Caldwell MW, Palci A, et al. Early view. Giant taxon-character matrices:

quality of character constructions remains critical regardless of size. Cladistics. doi:10.1111/cla.12163

**30** Wheatley BP. 1980. Feeding and ranging of East Bornean *Macaca fascicularis*. In: Lindburg DG, editor. The macaques: studies in ecology, behavior and evolution. New York: van Nostrand Reinhold. p 215–246.

**31** Tsuji Y, Hanya G, Grueter CC. 2013. Feeding strategies of primates in temperate and alpine forests: comparison of Asian macaques and colobines. Primates 54:201–215.

**32** Jablonski NG, Whitfort MJ, Roberts-Smith N, et al. 2000. The influence of life history and diet on the distribution of catarrhine primates during the Pleistocene in eastern Asia. J Hum Evol 39:131–157.

**33** Patterson SK, Sandel AA, Miller JA, et al. 2014. Data quality and the comparative method: the case of primate group size. Int J Primatol 35:990–1003.

**34** Hassel-Finnegan HM, Borries C, Larney E, et al. 2008. How reliable are density estimates for diurnal primates? Int J Primatol 29:1175–1187.

**35** Martin P, Bateson P. 2007. Measuring behaviour: an introductory guide. Cambridge: Cambridge University Press.

**36** National Academy of Sciences, National Academy of Engineering, Institute of Medicine of the National Academies. 2009. On being a scientist: a guide to responsible conduct in research. Washington, DC: National Academies Press.

**37** Martin RD, Genoud M, Hemelrijk CK. 2005. Problems of allometric scaling analysis: examples from mammalian reproductive biology. J Exp Biol 208:1731–1747.

**38** Deaner RO, Nunn CL, van Schaik CP. 2000. Comparative tests of primate cognition: different scaling methods produce different results. Brain Behav Evol 55:44–52.

**39** Smith RJ, Jungers WL. 1997. Body mass in comparative primatology. J Hum Evol 32:523–559.

**40** Gordon AD, Marcus E, Wood BA. 2013. Great ape skeletal collections: making the most of scarce and irreplaceable resources in the digital age. Yearbk Phys Anthropol 152:2–32.

**41** Borries C, Lu A, Ossi-Lupo K, et al. 2014. The meaning of weaning in wild Phayre's leaf monkeys: last nipple contact, survival, and independence. Am J Phys Anthropol 154:291–301.

**42** Lee PC. 1996. The meanings of weaning: growth, lactation, and life history. Evol Anthropol 5:87–96.

**43** van Noordwijk MA, Willems EP, Utami-Atmoko SS, et al. 2013. Multi-year lactation and its consequences in Bornean orangutans (*Pongo pygmaeus wurmbii*). Behav Ecol Sociobiol 67:805–814.

**44** Harvey PH, Clutton-Brock TH. 1985. Life history variation in primates. Evolution 39:559–581.

**45** Dubman E, Collard M, Mooers AO. 2012. Evidence that gestation duration and lactation duration are coupled traits in primates. Biol Lett 8:998–1001.

**46** Asquith PJ. 1989. Provisioning and the study of free-ranging primates: history, effects, and prospects. Yearbk Phys Anthropol 32:129–158.

**47** Whitlock MC. 2011. Data archiving in ecology and evolution: best practices. Trends Ecol Evol 26:61–65.

**48** Nosek BA, Alter G, Banks GC, et al. 2015. Promoting an open research culture: author guidelines for journals could help to promote transparency, openness, and reproducibility. Science 348:1422–1425.

**49** Parker TH, Nakagawa S, Gurevitch J. 2016. Promoting transparency in evolutionary biology and ecology. Ecol Lett 19:726–728.

**50** Strier KB, Altmann J, Brockman DK, et al. 2010. The primate life history database: a unique shared ecological data resource. Methods Ecol Evol 1:199–211.

**51** Redmond E, Wilson JR. 2012. Seven databases in seven weeks: a guide to modern databases and the NoSQL movement. Dallas: Pragmatic Bookshelf.

**52** Wilson G, Aruliah DA, Brown CT, et al. 2014. Best practices for scientific computing. PLoS Biol 12:e1001745.

**53** Wieczorek J, Bloom D, Guralnick RP et al. 2012. Darwin core: an evolving community-developed biodiversity data standard. PLoS One 7:e29715.

**54** Sadleir RMFS. 1969. The ecology of reproduction in wild and domestic mammals. London: Methuen.

**55** Bronson FH. 1989. Mammalian reproductive biology. Chicago: University of Chicago Press.

**56** Leigh SR. 1994. Relations between captive and noncaptive weights in anthropoid primates. Zoo Biol 13:21–43.

**57** Rafferty AR, Wong BBM, Chapple DG. 2015. An increasing citation black hole in ecology and evolution. Ecol Evol 5:196–199.

**58** Takahata Y, Suzuki S, Agetsuma N, et al. 1998. Reproduction of wild Japanese macaque females at Yakushima and Kinkazan Islands: a preliminary report. Primates 39:339–349.

**59** Altmann J, Alberts SC. 2005. Growth rates in a wild primate population: ecological influences and maternal effects. Behav Ecol Sociobiol 57:490–501.

**60** Borries C, Koenig A, Winkler P. 2001. Variation of life history traits and mating patterns in female langur monkeys (*Semnopithecus entellus*). Behav Ecol Sociobiol 50:391–402.

**61** Sandel AA, Miller JA, Mitani JC, et al. 2016. Assessing sources of error in comparative analyses of primate behavior: intraspecific variation in group size and the social brain hypothesis. J Hum Evol 94:126–133.



## *Articles in Forthcoming Issues*

- **Cooperation, Collective Action, and the Archaeology of Large-Scale Societies**
  *David M. Carballo & Gary M. Feinman*
- **The Psychology of Cooperation: Insights from Chimpanzees and Children**
  *Alicia P. Melis and Felix Warneken*
- **Gang warfare: What are the Criteria for Deriving Assertions from Facts in Science?**
  *Kenneth M. Weiss*
- **Less of a Bird's Song than a Hard Rock Ensemble**
  *Robert Hosfield, James Cole & John McNabb*
- **Explanations for Adaptations, Just-So Stories, and Limitations On Evidence in Evolutionary Biology**
  *Richard J. Smith*
- **Tropical Forests and the Genus *Homo***
  *Patrick Roberts, Nicole Boivin, Julia Lee-Thorp, Michael Petraglia, & Jay Stock*
- **The Crown Joules: Energetics, Ecology, and Evolution in Humans and other Primates**
  *Herman Pontzer*