Received: 6 August 2020 | Revised: 8 December 2020 | Accepted: 11 December 2020

DOI: 10.1111/joa.13390

Journal of Anatomy ANATOMICAL SOCIETY WILEY

METHODS

Bolstering geometric morphometrics sample sizes with damaged and pathologic specimens: Is near enough good enough?

D. Rex Mitchell[1,2] | Claire A. Kirchhoff[3] | Siobhán B. Cooke[4,5] | Claire E. Terhune[1]

[1]Department of Anthropology, University of Arkansas, Fayetteville, AR, USA

[2]Center for Anatomical Sciences, University of North Texas Health Science Center, Fort Worth, TX, USA

[3]Department of Biomedical Sciences, Marquette University, Milwaukee, WI, USA

[4]Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[5]New York Consortium in Evolutionary Primatology Morphometrics Group, New York, NY, USA

Correspondence
D. Rex Mitchell, Center for Anatomical Sciences, University of North Texas Health Science Center, Fort Worth, TX, USA.
Email: drexmitch311@gmail.com

Funding information
National Science Foundation, Grant/Award Number: NSF-BCS 1551669, NSF BCS-1551722 and NSF BCS-1551766

Abstract

Obtaining coordinate data for geometric morphometric studies often involves the sampling of dry skeletal specimens from museum collections. But many specimens exhibit damage and/or pathologic conditions. Such specimens can be considered inadequate for the analyses of shape and are excluded from study. However, the influences that damaged specimens may have on the assessment of normal shape variation have only been explored in two-dimensional coordinate data and no studies have addressed the inclusion of pathological specimens to date. We collected three-dimensional coordinate data from the cranium and mandible of 100 crab-eating macaques (*Macaca fascicularis*). Tests typically employed to analyze shape variation were performed on five datasets that included specimens with varying degrees of damage/pathology. We hypothesized that the inclusion of these specimens into larger datasets would strengthen statistical support for dominant biological predictors of shape, such as sex and size. However, we also anticipated that the analysis of only the most questionable specimens may confound statistical outputs. We then analyzed a small sample of good quality specimens bolstered by specimens that would generally be excluded due to damage or pathologic morphology and compared the results with previous analyses. The inclusion of damaged/pathologic specimens in a larger dataset resulted in increased variation linked to allometry, sexual dimorphism, and covariation, supporting our initial hypothesis. We found that analyzing the most questionable specimens alone gave consistent results for the most dominant aspects of shape but could affect outputs for less influential principal components and predictors. The small dataset bolstered with damaged/pathologic specimens provided an adequate assessment of the major components of shape, but finer scale differences were also identified. We suggest that normal and repeatable variation contributed by specimens exhibiting damage and/or pathology emphasize the dominant components and shape predictors in larger datasets, however, the various unique conditions may be more influential for limited sample sizes. Furthermore, we find that exclusion of damaged/pathologic specimens can, in some cases, omit important demographic-specific shape variation of groups of individuals more likely to exhibit these conditions. These findings provide a strong case for inclusion of these specimens into studies that focus on the dominant aspects of intraspecific

shape variation. However, they may present issues when testing hypotheses relating to more fine-scale aspects of morphology.

KEYWORDS
geometric morphometrics, normal variation, pathology, sampling, specimen

1 | INTRODUCTION

Geometric morphometrics (GM) is a powerful statistical methodology employed to study shape (geometry) with three-dimensional (3D) Cartesian coordinate data (Zelditch et al., 2004; Adams et al., 2013). This methodology is most often used to quantify morphological variation associated with environmental and/or evolutionary factors, and has surged forward over the last decade as a dominant feature in the toolkits of researchers who study form and function (Adams et al., 2013; Cooke & Terhune, 2015). However, a common problem often experienced by GM researchers is access to adequate numbers of specimens to reach viable sample sizes (Cardini & Elton, 2007; Cardini et al., 2015).

Recently, there has been a push for larger samples in morphometric analyses. Cardini et al. (2015) suggested a minimum of 15–20 specimens for a given sample to generate more consistent estimates of mean shape, centroid size variance, and shape variance. However, reaching such numbers of specimens is not always simple. Robust GM datasets for studies that examine vertebrate skeletal morphology often rely on dry bone specimens from museums. The number of specimens available from collections can be limited for some species. Furthermore, collections may often exhibit a range of conditions that can be considered deleterious to the reliability of shape data and its analysis. Typically, researchers will preferentially seek specimens with morphologies unaffected by damage, injury, or disease as these may be presumed to better represent normal variation. This can lead to many specimens being excluded, particularly those with missing landmark locations (Arbour & Brown, 2014; Strauss & Atanassov, 2006); and this can substantially reduce the size of datasets.

Specimens housed in museum collections can exhibit a range of conditions that could impact landmarking, including postmortem and/or perimortem damage, and antemortem pathologies. Postmortem damage includes damage or missing elements/features due to breakage (e.g., shelf damage), perimortem damage refers to unhealed injuries incurred either during or close to time of death (e.g., bullet wounds or shotgun pellets still *in situ*) (see Wheatley, 2008), and antemortem pathologies may include healed injuries and evidence of acute or chronic disease (Lovell, 1991). In a mammalian skull specimen, for example, a zygomatic arch may be cracked, broken, or missing (postmortem damage) (Figure 1a), bullet wounds (Figure 1b) or blunt-force trauma received close to or at the time of death may adversely affect the surrounding shape of the bone (perimortem damage), and antemortem pathologies, such as healed breaks, osteoarthritis, broken or lost teeth, dental caries and abscesses, and alveolar recession are often identified in representatives of many mammalian taxa (Figure 1b–f) (e.g. Cuozzo & Sauther, 2006; Dixon et al., 2000; Elgart, 2010; Fox, 1939; Fuss et al., 2018; Jurmain, 1989, 1997; Lovell, 1990; Sone et al., 2004; Van Valkenburgh, 1988, 2009). Notably, a single specimen could potentially exhibit all three of these conditions, and at different severities. The crab-eating or long-tailed macaque (*Macaca fascicularis*) is an ideal taxon to examine the influence of these types of specimens on shape analyses, because museums often have a sizeable range of specimens available, and many exhibit examples of these postmortem, perimortem, and antemortem conditions (Figure 1), which might often exclude them from GM analysis.

The influence that the inclusion of damaged/pathologic specimens may have on the analysis of 3D coordinate data has never been explicitly tested. Arbour and Brown (2014) performed a range of geometric morphometric analyses to test the impact of damaged specimens with missing landmarks, however, these analyses were carried out on 2D data. Furthermore, true morphologies of specimens with landmark coordinates impacted by antemortem damage and/or pathology (i.e., landmarks that are present but may be shifted in position from such conditions) have not yet been assessed for their influence on statistical outputs. Although increased sample sizes are often important to morphometric analyses, the inclusion of specimens with various forms of damage/pathology may alter the range of variation in shape, confound principal components (PCs), and obscure the influence of important correlations and predicting factors. It is therefore possible that the inclusion of damaged/pathologic specimens in GM datasets may incur both benefits and costs.

Here, we analyze landmark data from the cranium and mandible of a large sample of *M. fascicularis*. A range of standard GM tests was performed to assess normal variation across several datasets with differing degrees of damaged/pathologic specimen inclusion. First, we tested whether a larger dataset can be bolstered with such specimens. We hypothesized that the normal variation established by numerous undamaged, non-pathologic specimens, and also present in additional damaged/pathologic specimens, would overwhelm unique individual variation resulting from damage and/or pathology. If this hypothesis is supported, we expect that damaged/pathologic specimens will have little influence on known correlations associated with the allometry and sexual dimorphism previously established for *M. fascicularis* (Ito et al., 2011; Terhune et al., 2015; Yao, 2016), and that levels of covariation between the cranium and mandible would hold constant. We then performed these tests on a sample with only the most severely damaged and/or pathologic specimens. Here we hypothesized that these specimens, absent from the normal

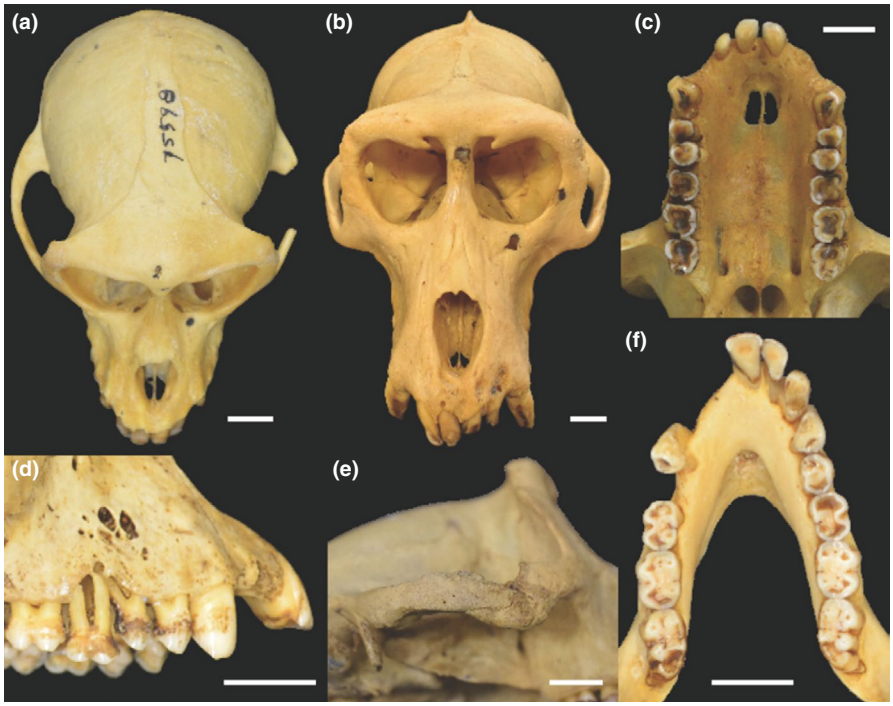

FIGURE 1 Examples of damaged/pathologic *Macaca fascicularis* specimens landmarked and analyzed in this study. (a) FMNH75598: postmortem damage (missing left zygomatic arch); (b) FMNH56493: perimortem damage (buckshot), and antemortem pathology (abscess fistulae associated with incisors); (c) NMNH114169: antemortem pathology (loss of an incisor, pulp cavity exposure of both maxillary canines, and heavy wear of the cheek teeth); (d) FMNH67720: antemortem pathology (abscess fistulae and alveolar recession of the cheek teeth); (e) FMNH56490: antemortem trauma (healed fracture of the zygomatic arch); (f) FMNH87428: antemortem pathology (loss of two incisors, a canine, and a premolar of the mandible with advanced alveolar resorption). Scale bar = 10 mm for all panels. Although specimen selection is always subjective, each specimen presented in this figure was considered, via consensus among the authors, to represent a severe case of pathology and/or damage and likely to be excluded from analysis. All are present in datasets 3, 4, and 5 (see Methods section for dataset selection criteria)

variation established by the undamaged, non-pathologic specimens in the previous tests, would demonstrate some deviation in statistical outputs from the initial datasets. Finally, we simulated a common scenario whereby a researcher is limited to a smaller number of good-quality specimens and must rely on the addition of damaged/pathologic specimens to analyze an adequate sample size.

2 | METHODS

Three-dimensional (3D) models were obtained from surface scanning the cranium and mandible of 100 (47 females, 53 males) largely intact adult *M. fascicularis*, housed at the Field Museum of Natural History and National Museum of Natural History at the Smithsonian Institution. Adults were identified via fully erupted third molars. Surface scanning was carried out using an HDI 120 blue-LED scanner (LMI Technologies) in the program FlexScan3D. At the time of scanning, any postmortem/perimortem damage (i.e., breaks, holes, missing regions) or antemortem conditions (i.e., antemortem tooth loss or breakage, pulp cavity exposure, dental abscess, skeletal evidence for periodontal disease, craniofacial trauma, TMJ osteoarthritis) were identified for each specimen (see Table S1 for specimen details).

Each cranium and mandible was scanned separately, and surface meshes exported as .ply files. Each mesh was then imported into Geomagic Studio (3D Systems) to be cleaned by filling in small sections of missing data with the "Mesh Doctor" and "Fill" functions. Cleaned meshes were then imported into Landmark Editor v. 3.6 for landmarking. A total of 188 landmarks (84 fixed and 104 semilandmarks) were placed on the cranium and 110 (36 fixed and 74 semilandmarks) on the mandible (Figure 2, Figure S1, Tables S2 and S3). Any landmarks representing postmortem/perimortem damage or missing regions were marked as missing data, while data for antemortem conditions were retained. For instances of antemortem tooth loss, the landmark representing the alveolus of the missing tooth was placed on the vacant diastema, but any landmarks representing the occlusal surface of the tooth (i.e., I1, P4, or M2) were marked as missing data. If alveolar recession was present, landmarks representing the alveolar margins were placed at the margin of receded bone regardless of severity.

From the compiled shape data of all specimens, five datasets were generated, each including both cranium and mandible data, which were analyzed separately. The first three datasets were used to test whether the inclusion of damaged/pathologic specimens alters statistical outputs of shape analyses. Dataset 1 included only specimens with all landmarks present and with no identified signs of

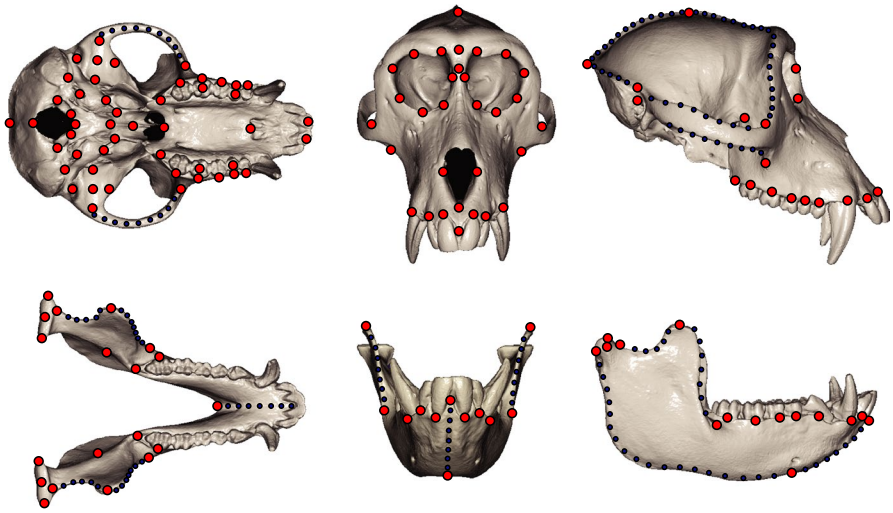

FIGURE 2 Landmark configurations for the cranium and mandible of *Macaca fascicularis*: red = fixed landmarks, blue = semi-sliding landmarks (see Tables S2 and S3 for landmark descriptions)

TABLE 1 Sample size (*n*) for each dataset (cranium/mandible), and for the two-block partial least-squares (2B-PLS). 2B-PLS sample sizes represent only individuals with both cranium and mandible.

Dataset	Description	Female *n*	Male *n*	Total *n*	2B-PLS n
1	Only specimens with no damage/pathology)	9/15	26/29	35/44	32
2	No or only mild damage/pathology	31/33	39/40	70/73	69
3	No or only mild damage/pathology, and severe cases (all specimens)	40/45	53/53	93/98	92
4	Severe cases of damage/pathology only	10/12	12/11	22/23	22
5	20 specimens with no damage/pathology bolstered with 20 specimens with severe damage/pathology	20/20	20/20	40/40	40

damage or pathology; therefore "ideal" for landmarking to examine normal (i.e., non-pathologic) variation. For Dataset 2, two researchers (authors CET and CAK) independently assessed the quality of each specimen and designated which individuals had abnormalities minor enough to be adequate for landmarking (e.g., missing few teeth, minor osteoarthritis of the TMJ, etc.). All specimens independently considered adequate by both researchers, including those of Dataset 1, were included in Dataset 2. This represents a subjectively selected dataset most likely sampled by the typical GM researcher. Dataset 3 included all specimens, regardless of the presence and severity of damage and/or pathology. To identify whether the analysis of severely damaged and/or pathologic specimens alone resulted in divergent statistical outputs, Dataset 4 included only the severely damaged/pathologic specimens that were excluded from Dataset 2 but included in Dataset 3. Finally, Dataset 5 included 10 females and 10 males from Dataset 1 (i.e., specimens with no damage/pathology) and an additional 10 females and 10 males from Dataset 4 (i.e., severely damaged and/or pathologic). These were randomly selected using the 'sample' R base function. Because there were only nine females in Dataset 1, one additional female was randomly selected from dataset 2 to make a total of 10. This represented a hypothetical dataset with male and female numbers below the recommended number per group of 15–20 individuals (Cardini et al., 2015), which is bolstered by more severely damaged/pathologic specimens to reach 20 specimens for each sex.

Landmark data were imported into R (R Core Team, 2018) and analyzed using the *geomorph* package (v. 3.2.1) (Adams and Otárola-Castillo, 2013; Adams et al., 2020). Any missing landmarks (see Table S1 for details on missing landmarks for each specimen) were estimated via thin-plate spline interpolation using the "estimate.missing" function (Adams & Otárola-Castillo, 2013; Gunz et al., 2009). A generalized Procrustes analysis was performed using the "gpagen" function, to remove variation in the shape data attributable to scale, position, and orientation (Rohlf & Slice, 1990). Semilandmarks were slid by minimizing bending energy and size was represented by centroid size for each specimen. Statistical outliers were identified using the "plotOutliers" function. This function plots each specimen's Procrustes distance from the consensus shape. Specimens that fell above the upper quartile were subsequently removed from the dataset, and the superimpositions performed again on the raw data of the remaining specimens. These outliers were female specimens with exceedingly small temporal fossae and were not associated with any damage/pathologic conditions of interest in this study. We therefore removed these individuals to limit potential confounding factors associated with larger

Procrustes distances and to maintain focus on the damaged/pathologic specimens. The definitions and final sample sizes for each dataset are presented in Table 1.

For each dataset, principal component analysis (PCA) was performed on the Procrustes residuals and the contribution of variation from the first five PCs and their corresponding shape associations were identified. Shape associations for each PC were investigated using the "plotRefToTarget" function. We then visually inspected vector displacements between corresponding landmarks of PC maxima and minima. Shape variation is only minimally described to identify consistencies among the datasets, since these details are specific to the taxon and arbitrary in the context of the hypotheses here.

Next, multivariate regressions were performed using the "procD.lm" function to identify associations between shape (Procrustes residuals), cranial or mandibular size (as represented by centroid size), and sex. Specifically, we examined the relationship between shape and size (~ln[centroid size]), then sex (~sex), and we also employed a model using sequential (Type I) Sum of Squares to examine the influence of sex when variation attributable to size is accounted for (Klingenberg, 2016) (~ln[centroid size] + sex). Significance was assessed using permutation tests with 1000 iterations.

Lastly, a two-block partial least-squares (2B-PLS) analysis (Rohlf & Corti, 2000) was performed on the cranial and mandibular Procrustes shape data for each dataset using the "two.b.pls" function. This analysis allowed us to examine the degree of covariation between cranial and mandibular shape. The sample sizes of some datasets were smaller for this test (Table 1), because this analysis only operates on specimens for which there were both a cranium and a mandible present from each individual and some had been removed as outliers. From these tests, we obtained the correlation (r) between the two blocks (cranium and mandible) and significance was assessed using a permutation test with 1000 iterations. The proportion of variation of each block associated with covariation was then determined by dividing the variance (squared standard deviation) of PLS1 by the total variance, for both Block 1 (cranium) and Block 2 (mandible). The RV coefficient was obtained by performing the 2B-PLS in the MorphoJ software (Klingenberg, 2011).

3 | RESULTS

3.1 | PC Analysis

The descriptions of the PCA results for each dataset are presented in Table 2. For all PCs, we briefly describe the most important anatomical region driving shape variation along that axis; representative images of shape definitions and corresponding vector displacements are available in Tables S4 and S5. Although we provide these descriptions and briefly discuss below, the most important result of this analysis is whether consistent components of shape variation are captured by the same PC axes across datasets.

Shape variation represented by PC1 of the cranium was consistent across datasets and primarily reflected size of the temporalis attachment area. The proportion of variance for this PC axis ranged from 36.66% to 48.14%. All subsequent PCs for the cranium contributed 11.90% or less and often differed in their shape associations and order of importance (Table 2) depending on the dataset in question. The shape variation represented by the first four PCs of the cranial data were mostly consistent for Dataset 1 (specimens with no damage/pathology), Dataset 2 (selected specimens), and Dataset 3 (all specimens). PC1 represented similar shape variation across each of these three cranial datasets, however, the additional individuals in Dataset 2 resulted in a rearrangement of the subsequent three PCs and an increase in the proportion of variation contributed by PC1. With the addition of the most questionable specimens, Dataset 3 presented only very minor differences to Dataset 2.

For Dataset 4 (the most severely damaged/pathologic specimens) the first five PC axes of the cranial data had broadly similar shape associations to the previous datasets, with PCs 1, 2, 3, and 5 having been identified in one or more prior datasets. However, PC4 was defined by a novel shape association, the morphology of the temporalis attachment area. Dataset 5, consisting of a small number of good-condition specimens bolstered by damaged/pathologic specimens, was also largely consistent with the results of the other datasets but differed slightly. Of the first five PCs, four were consistent with Dataset 2 and Dataset 3, however, PC4 instead matched the novel shape association introduced by Dataset 4, representing temporalis attachment shape.

For the mandible data, the proportion of variation (between 21.31% and 32.02%) and shape associations represented by PC1 was mostly consistent across all datasets. All subsequent PCs contributed less than 15.76% variation. The first four PCs often demonstrated similar shape associations, albeit in different order, across the datasets, but began to diverge more clearly at PC4–PC5.

The shape variation represented by the first five PCs was partly consistent between Datasets 1 and 2 for the mandibular data, but with some differences. For Dataset 1, PC1 primarily reflected gonial angle morphology as well as anteroposterior projection of the incisors/canines. In contrast, for Datasets 2 and 3, PC1 largely reflected the mediolateral distance between the left and right ramus and gonial angles, as well as mandibular symphysis angulation. Overall, Dataset 1 was less similar than Datasets 2 and 3 were to one another. For Dataset 4, the shape variation represented by PC1 of the mandible was more consistent with Dataset 2 and Dataset 3, while the subsequent PCs were more aligned with Dataset 1 in having associations with bigonial breadth width and anteroposterior ramus breadth. Lastly, the results of Dataset 5 for the mandibular data were mostly consistent with Datasets 1 and 4, though PC3, PC4, and PC5 varied in the aspects of shape variation they represented.

3.2 | Procrustes Regressions

Results of the regression analyses are presented in Table 3. Both the cranial and mandibular data showed significant ($p \leq 0.005$) relationships between shape and both size and sex under

TABLE 2 Results of the principal component analysis showing shape variation associated with the first five PC axes for each dataset, including proportions of total variation (%). AP =anteroposterior, SI =superoinferior, ML =mediolateral.

	%	Cranium	%	Mandible
Dataset 1				
PC1	36.66	Temporalis attachment area	25.19	ML distance between L and R ramus, Gonial angulation, incisor/canine AP projection
PC2	11.90	Bizygomatic breadth	13.14	Coronoid process SI projection relative to condyle, AP length of inferior margin of corpus
PC3	10.09	Facial prognathism	10.17	Development of gonial angle, genioglossal fossa depth
PC4	5.29	Premaxillary SI angulation	8.43	Ramus AP breadth
PC5	4.72	Premaxilla AP position/ projection	6.59	Bigonial ML breadth
Total	68.66		63.52	
Dataset 2				
PC1	48.14	Temporalis attachment area	26.77	ML distance between L and R ramus, mandibular symphysis angulation
PC2	7.62	Facial prognathism	11.37	Coronoid process SI projection relative to condyle
PC3	5.62	Premaxilla AP position/ projection	9.81	Development of gonial angle
PC4	5.51	Bizygomatic breadth	7.28	AP angulation of the ramus in the parasagittal plane
PC5	4.12	Short and anteriorly positioned postcanine toothrow/ zygomatic arch SI height	6.17	Genioglossal fossa depth, toothrow length
Total	71.01		61.40	
Dataset 3				
PC1	47.21	Temporalis attachment area	26.78	ML distance between L and R ramus, mandibular symphysis angulation
PC2	7.31	Facial prognathism	11.93	Coronoid process SI projection relative to condyle
PC3	5.89	Premaxilla AP position/ projection	8.96	Development of gonial angle
PC4	5.41	Bizygomatic breadth	6.78	AP angulation of the ramus in the parasagittal plane
PC5	3.89	Zygomatic arch SI height	5.80	Genioglossal fossa depth
Total	69.71		60.25	
Dataset 4				
PC1	45.83	Temporalis attachment area	32.02	ML distance between L and R ramus, incisor/ canine AP projection, ML width of posterior toothrow
PC2	9.66	Premaxillary SI angulation	13.65	Coronoid process SI projection relative to condyle
PC3	7.18	Bizygomatic breadth	9.38	Bigonial ML breadth
PC4	5.54	Temporalis attachment shape	8.22	Ramus AP breadth, genioglossal fossa depth
PC5	4.88	Premaxilla AP position/ projection	6.35	Development of gonial angle
Total	73.09		69.62	

(Continues)

TABLE 2 (Continued)

	%	Cranium	%	Mandible
Dataset 5				
PC1	43.32	Temporalis attachment area/approximation of the temporal lines	21.31	ML distance between L and R ramus, incisor/canine AP projection, ML width of posterior toothrow
PC2	9.20	Premaxillary SI angulation	15.76	Coronoid process SI projection relative to condyle
PC3	8.75	Bizygomatic breadth	11.15	Ramus AP breadth
PC4	5.74	Temporalis attachment shape	8.76	Development of gonial angle
PC5	4.51	Premaxilla AP position/projection	5.91	Condylar SI projection, genial fossa development
Total	71.52		62.89	

independent tests, but R^2 values differed considerably. For Datasets 1, 2, and 3, increased sample sizes resulted in increases to the amount of variation linked to both size and sex. This result was consistent regardless of the severity of damage/pathology exhibited by the included specimens and coincides with more dense clustering of individuals along the allometric trajectories with the inclusion of damaged/pathologic specimens (Figure 3). The influence of sex in the multivariate models, where size was also included as a factor, was consistently marginal in significance in the cranial data for the first three datasets, with p-values near to 0.05. Only Dataset 3 demonstrated a significant (p = 0.014) size-adjusted effect of sex in the cranium. Sex was a significant factor in the multivariate model for all mandible datasets ($p \leq 0.012$).

Dataset 4, by contrast, did not follow the trends of the previous three datasets, as it was the dataset with the smallest sample size and yet demonstrated the highest R^2 for size and sex. This was also the only dataset where there was no overlap between sexes in the regression plots (Figure 3) and a comparison of the centroid size distributions between datasets indicate that this is because the males of Dataset 4 consist of larger individuals with smaller standard deviation (Figure 4).

The proportion of variation linked to size and sex in Dataset 5 were mostly consistent with Datasets 2 and 3 for the cranial data and most similar to Dataset 1 for the mandibular data. In addition, the multivariate regression indicated that the effect of sex was highly significant when adjusted for size (R^2 = 0.078, p = 0.001), consistent with the first three datasets. The allometric plots also show greater overlap between the sexes compared to Dataset 4 (Figure 3). The mean centroid size for males for both the cranium and mandible is smaller than Dataset 4 and has a larger standard deviation, more similar to Datasets 1–3 (Figure 4).

TABLE 3 Results of the Procrustes regressions of shape versus ln(centroid size) and sex showing samples size (n = cranium/mandible), the coefficient of determination (R^2), and the significance of the regression (p).

	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Dataset 5	
	n = 35/44		n = 70/73		n = 93/22		n = 21/23		n = 40/40	
	R^2	p	R^2	p	R^2	p	R^2	p	R^2	p
Cranium										
~Size	0.190	0.001	0.329	0.001	0.338	0.001	0.375	0.001	0.334	0.001
~Sex	0.104	0.005	0.243	0.001	0.256	0.001	0.326	0.001	0.262	0.001
~Size+Sex										
Size	0.190	0.001	0.329	0.001	0.338	0.001	0.375	0.001	0.334	0.001
Size-adjusted Sex	0.036	0.100	0.017	0.070	0.016	0.014	0.026	0.565	0.022	0.209
Mandible										
~Size	0.154	0.001	0.160	0.001	0.175	0.001	0.260	0.001	0.132	0.001
~Sex	0.119	0.001	0.139	0.001	0.151	0.001	0.230	0.001	0.120	0.001
~Size+Sex										
Size	0.154	0.001	0.160	0.001	0.175	0.001	0.260	0.001	0.132	0.001
Size-adjusted Sex	0.060	0.002	0.054	0.001	0.052	0.001	0.064	0.012	0.078	0.001

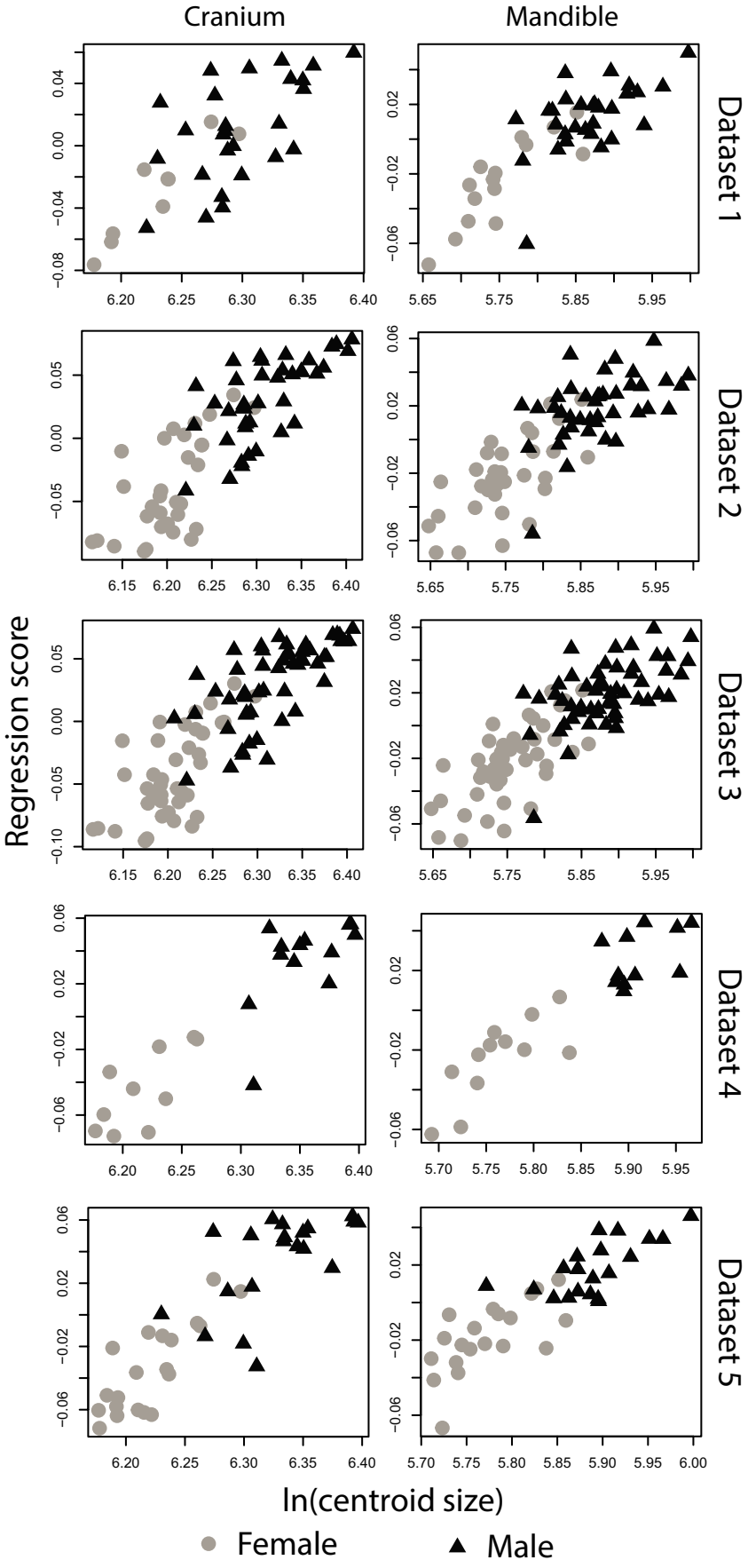

FIGURE 3 Allometry plots (shape ~ln[centroid size]) of the cranium and mandible for each dataset

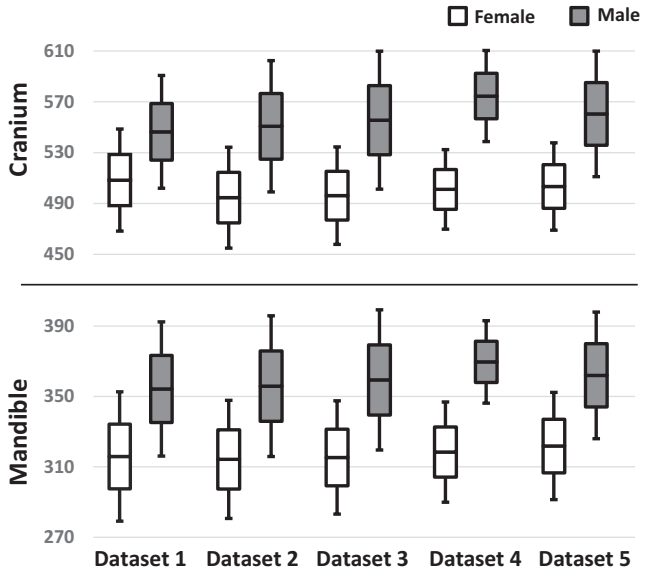

FIGURE 4 Boxplots of mean centroid sizes and SDs from each dataset

TABLE 4 Two-block partial least-squares (2B-PLS) results, including sample size (*n*), correlation coefficient (*r*), *p*-value (*p*), proportion of variation contributed by covariation in Block 1 for cranial data (B1) and Block 2 for mandibular data (B2), and the RV coefficient (RV)

Dataset	*n*	*r*	*p*	B1	B2	RV
1	32	0.779	0.007	0.302	0.129	0.437
2	69	0.796	0.001	0.447	0.240	0.502
3	92	0.818	0.001	0.446	0.243	0.536
4	22	0.913	0.001	0.445	0.265	0.691
5	40	0.868	0.001	0.422	0.193	0.565

3.3 | Two-block partial least-squares

The results for the 2B-PLS between the cranium and mandible are presented in Table 4. For the first three datasets, the p-values remained consistent (p = 0.001), and the correlation (r) was high, ranging from 0.779 to 0.818. As with the regression models, the degree of covariation (r) between the cranium and mandible increased with sample size. The proportion of variation contributed by each block (B1-cranium and B2-mandible) also increased alongside sample size. A similarly steady increase relative to sample size is found for the RV coefficient (Table 4). However, Dataset 4 did not follow these trends, with the highest correlation (r = 0.913) and the highest proportions of variation for all datasets found alongside the smallest sample size (Table 3). Dataset 5 also had a greater correlation between blocks (r = 0.868) than Datasets 1, 2, and 3, but the proportions of variation contributed by each block are consistent with those expected for sample size. In other words, all datasets except Dataset 4 demonstrate increases in covariation with an increase in sample size.

4 | DISCUSSION

This study examined multiple datasets of *M. fascicularis* with differing degrees of inclusion of damaged/pathologic specimens. Our goal was to identify the effects that the inclusion of such specimens has on the analysis of shape variation. We addressed three main questions: (a) Does the inclusion of damaged/pathologic specimens into larger datasets appreciably impact statistical outputs? (b) Do specimens with the most severe damage/pathology represent "normal" shape variation if tested alone? and (c) Can a dataset composed of a small number of typically ideal specimens bolstered by damaged/pathologic specimens offer findings consistent with datasets that exclude damaged/pathologic specimens? Our findings present a strong case that morphometric analyses can benefit from the inclusion of these specimens for hypotheses interested in the dominant aspects of intraspecific normal variation; however, there are some caveats to consider.

4.1 | Including damaged/pathologic specimens in larger datasets

An obvious benefit to including damaged/pathologic specimens is the larger pool of potential specimens available for analysis, because larger sample sizes typically have higher statistical power and a lower chance of Type II errors (Smith, 2018). However, the introduction of new specimens into a dataset may also shift the orientation and influence of PCs (Adams et al., 2011), thus potentially altering interpretations of the data. Specimens distorted by damage and/or pathology may therefore be expected to contribute deviations to normal shape variation that could generate new PC shape associations, or at least shift the importance of shape associations.

We found that the first three PCs were mostly consistent across our Dataset 1 (good condition specimens), Dataset 2 (selected specimens), and Dataset 3 (all specimens) with regard to both their proportions of total variation and their shape associations. This suggests that the first few PCs largely represent morphological changes in regions not affected by the damage/pathology. However, some differences in how shape variation was distributed among the PC axes were found between our Datasets 1 and 2 and more pronounced differences were observed from approximately PC4 and into lower order PCs. It is therefore possible that only considering specimens without any pathology or damage may not as reliably represent population variation and could potentially skew shape variation unrealistically. In contrast, we found only very slight differences between Dataset 2 and Dataset 3. Because Dataset 2 was intended to represent the most likely specimens sampled by the average researcher (i.e., normal specimens plus those with relatively minor damage or pathology), these results further indicate that the inclusion of specimens with the most severe conditions had little deleterious effect on shape associations or proportions of variation for the first four PCs, at least for larger datasets of these sample sizes.

The regressions and 2B-PLS analysis showed that the proportions of variation and correlation coefficients (R^2 and r values, respectively) attributable to dominant predictors of shape (size and sex) and covariation between the cranium and mandible increased with the addition of damaged/pathologic specimens and overall increasing sample sizes. The greater proportions of variation found in both the Procrustes regressions and 2B-PLS for Dataset 2 and Dataset 3 suggest that the predominantly normal variation imparted by the most questionable specimens strengthens these relationships, rather than the analyses being hindered by their damage and pathologies.

One caveat to this may be influences on shape that are statistically marginal in significance. Craniofacial size and sex are strongly correlated in the species studied here. Female shape and male shape share similar allometric trajectories, with most separation between sexes occurring due to differences in centroid sizes. The influence of sex after adjusting for size in the cranium was consistently marginal in significance across datasets, with p-values near to 0.05. However, the inclusion of the most questionable specimens in Dataset 3 represented the only case of significance when the p-value shifted from >0.05 to <0.05. The inclusion of these specimens was able to discern significance where Datasets 1 and 2 could not. Therefore, for marginal p-values that dwell around 0.05, interpretations may vary between researchers who are particularly stringent on the meaning of this arbitrary threshold for significance (see Smith, 2018).

The results of these initial tests demonstrate that including damaged/pathologic specimens into an already sizeable dataset retained clearly significant relationships and increased the proportion of variation attributable to known biological predictors of shape, thereby supporting our initial hypothesis. The findings suggest that, rather than impeding tests on normal shape variation, these specimens can enhance the statistical merit of dominant shape predictors and more clearly define their influences on shape variation. However, marginal influences on shape, with p-values near to 0.05, may vary in significance depending on sample composition and size.

4.2 | Sampling damaged/pathologic specimens alone to examine shape variation

Dataset 4 included only the most severely damaged and/or pathologic specimens and identified a somewhat different pattern of results than the other datasets. The PCA found shape associations previously identified by the other datasets; however, lower order PCs demonstrated different shape associations, including a novel PC4 for the cranium. The R^2 values of predictor variables were greater than all previous datasets, and covariation was also greater in the 2B-PLS. Given that this dataset had the smallest sample size, this finding countered the previously identified trend of increasing R^2 values with increased sample size, indicating that analysis of these specimens alone, and in smaller numbers, may deviate R^2 values from more expected estimates for the population as a whole. There was also greater sex disparity along the allometric trajectory in Dataset 4, indicated by a lack of overlap between the sexes. An important consideration for Dataset 4 is that, despite adequate total sample size, numbers for each sex (n = 12) are lower than suggested for consistent estimates of size and shape parameters (Cardini & Elton, 2007; Cardini et al., 2015). This can bring into question the validity of sex-related findings for this dataset and so these are minimally discussed. However, what can certainly be noted from Dataset 4 is an inherent bias towards larger males, which isolated the sexes along the allometric curve. This bias is evidenced by the smaller standard deviations for male centroid sizes in Dataset 4, since randomly selected datasets with sample sizes <30 should have relatively larger standard deviations (Cardini & Elton, 2007).

There are aspects of behavioral ecology for the species studied here that may have influenced the likelihood that these specimens would be considered inappropriate for sampling. *Macaca fascicularis* is known to exhibit considerable agonistic behaviors among males, including physical assault and biting (De Waal, 1977). Injuries are known to increase in number and severity with age among primates (Bramblett, 1967; Jurmain, 1997; Lovell, 1990), although prevalence can differ substantially between species (Jurmain, 1997; Lovell, 1991). Larger and/or older individuals may therefore be more likely to exhibit more severe cases of antemortem tooth loss, extreme dental wear, and other trauma. Furthermore, larger, more aggressive individuals may be preferentially shot or otherwise dispatched by museum collectors, resulting in perimortem damage. Damaged/pathologic specimens may therefore be more common in particular groups of individuals and raises two important points for consideration: (a) in potentially consisting of mostly larger individuals, the analysis of severely damaged/pathologic specimens alone could limit the representative range of variation and potentially obscure the influence of size-independent predictors; and (b) the omission of such specimens may inadvertently ignore important demographic contributions to intraspecific variation. Thus, it is possible that the inclusion of these larger male specimens was responsible for the shift in significance of size-adjusted sex differences in Dataset 3 crania.

4.3 | Bolstering smaller datasets with damaged/pathologic specimens

Dataset 5 was generated to simulate a scenario that can often occur when not enough undamaged, non-pathologic specimens are available in museum collections to produce acceptable sample sizes (Cardini & Elton, 2007; Cardini et al., 2015), while other specimens available may be considered too poor in quality and inadequate for representation of normal shape variation. This dataset simulated an extreme case by including only 10 specimens from each sex from Dataset 1 and bolstering the sample to 20 of each sex with specimens randomly selected from Dataset 4. The first three PCs of this dataset were consistent with the initial three datasets, but PC4 was instead consistent with the novel shape association found for PC4 of Dataset 4. This indicates that, at least in a small dataset, less influential PCs representing smaller scale shape associations can be

influenced by damage and/or pathology, or possibly morphologies more common to certain population demographics that exhibit these conditions. The regression analyses gave proportions of variation linked to size and sex more in line with the first three datasets and similar results were found for covariation in the 2B-PLS. Therefore, bolstering small sample sizes with damaged/pathologic specimens can provide an adequate means of assessing the dominant components and highly significant predictors of shape, but lower-order PCs and marginal trends should be interpreted with caution.

Although these findings may be true for *M. fascicularis*, there are almost certainly differing degrees of the prevalence and severity of damage and lesions in other primate and mammalian species. This will likely result in differing degrees of their influence in geometric morphometric analyses and may also produce additional impacts in other taxa not observed here. Therefore, future studies would benefit from considering damage and pathology when collecting data and interpreting results.

5 | CONCLUSION

The collection of reliable landmark data is one of the most fundamental aspects of the GM methodology. However, specimen selection will always be subjective, and researchers must consider what morphologies constitute "normal" shape variation when choosing individuals to be included in their samples. By incorporating specimen condition into shape analysis of normal variation, our findings demonstrate four important points:

1. The inclusion of specimens with relatively mild pathologies and/or damage into larger datasets can further increase sample size and provide a more thorough representation of shape variation.
2. When sample sizes are larger with many ideal specimens also present, aspects of shape associated with damage and/or pathology will likely be obscured by the dominant, consistent trends in shape variation. Damaged/pathologic conditions are often unique in morphology and, therefore, less likely to be repeated in coordinate data to a high enough degree to represent one of the major axes of shape variation.
3. The inclusion of damaged/pathologic specimens into smaller datasets may confound lower-order PCs and evidence for less influential variables, but the most influential factors will likely remain mostly consistent.
4. In some cases, specimens with evidence of damage/pathology could represent important aspects of variation across a species that may be missed if exempt from analysis. This is particularly true if certain population demographics, such as individuals of a given age, sex, or size, are predisposed to increased risks of damage, injury, or pathology that may see them excluded from sampling. Including specimens such as these may therefore be important for more accurately capturing shape variation within a species. However, this is probably less critical for studies of interspecific or intergeneric variation.

In summary, we suggest that the inclusion of damaged/pathologic specimens is likely adequate for assessing the dominant influences on normal shape variation, but results may vary for less influential factors. Therefore, careful consideration should be taken when selecting specimens that will adequately address a hypothesis, particularly for smaller sample sizes of a single species. In all cases, it is important for researchers to outline their sampling rationale and be clear about whether specimens with pathology and/or damage were included. The documentation and assessment of various pathologies and/or damage may be especially important for understanding the limitations inherent in and interpretation of analyses of shape variation.

ACKNOWLEDGEMENTS

Research carried out under the funding of the National Science Foundation (NSF-BCS 1551669 [SBC], NSF BCS-1551722 [CAK], NSF BCS-1551766 [CET]). We would like to thank Darrin Lunde and Paula Bohaska of the National Museum of Natural History and Lawrence Heaney and Adam Ferguson of the Field Museum for access to the specimens used in this research. We thank two anonymous referees for their helpful comments.

AUTHOR CONTRIBUTIONS

DRM conceived the study, designed the methodology, carried out analyses, and led the writing of the manuscript; CAK, SBC, and CET collected the data and advised the study design; and all authors contributed substantially to the drafts and gave final approval for publication.

DATA AVAILABILITY

All models used in this research were created by the authors and are available at morphosource.org. Digital object identifiers (DOIs) for each specimen are presented in Table S1. Landmark configuration data used in this study may be made available upon request to author Claire Terhune.

ORCID

D. Rex Mitchell https://orcid.org/0000-0003-1495-4879
Claire A. Kirchhoff https://orcid.org/0000-0002-9375-6282
Siobhán B. Cooke https://orcid.org/0000-0001-8265-6705
Claire E. Terhune https://orcid.org/0000-0002-5381-2497

REFERENCES

Adams, D.C., Cardini, A., Monteiro, L.R., O'Higgins, P. & Rohlf, F.J. (2011) Morphometrics and phylogenetics: Principal components of shape from cranial modules are neither appropriate nor effective cladistic characters. *Journal of Human Evolution*, 60(2), 240–243. https://doi.org/10.1016/j.jhevol.2010.02.003

Adams, D.C., Collyer, M. & Kaliontzopoulou, A. (2020) Geomorph: Software for geometric morphometric analyses. R package version 3.2.1. https://CRAN.R-project.org/package=geomorph

Adams, D.C. & Otárola-Castillo, E. (2013) Geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4(4), 393–399. https://doi.org/10.1111/2041-210X.12035

Adams, D.C., Rohlf, F.J. & Slice, D.E. (2013) A field comes of age: geometric morphometrics in the 21st century. Hystrix. *The Italian Journal of Mammalogy*, 24(1), 7–14. https://doi.org/10.4404/hystrix-24.1-6283

Arbour, J.H. & Brown, C.M. (2014) Incomplete specimens in geometric morphometric analyses. *Methods in Ecology and Evolution*, 5(1), 16–26. https://doi.org/10.1111/2041-210X.12128

Bramblett, C.A. (1967) Pathology in the Darajani baboon. *American Journal of Physical Anthropology*, 26(3), 331–340.

Cardini, A. & Elton, S. (2007) Sample size and sampling error in geometric morphometric studies of size and shape. *Zoomorphology*, 126(2), 121–134. https://doi.org/10.1007/s00435-007-0036-2

Cardini, A., Seetah, K. & Barker, G. (2015) How many specimens do I need? Sampling error in geometric morphometrics: testing the sensitivity of means and variances in simple randomized selection experiments. *Zoomorphology*, 134(2), 149–163. https://doi.org/10.1007/s00435-015-0253-z

Cooke, S.B. & Terhune, C.E. (2015) Form, function, and geometric morphometrics. *The Anatomical Record*, 298(1), 5–28. https://doi.org/10.1002/ar.23065

Cuozzo, F.P. & Sauther, M.L. (2006) Severe wear and tooth loss in wild ring-tailed lemurs (*Lemur catta*): a function of feeding ecology, dental structure, and individual life history. *Journal of Human Evolution*, 51(5), 490–505. https://doi.org/10.1016/j.jhevol.2006.07.001

De Waal, F.B. (1977) The organization of agonistic relations within two captive groups of Java-monkeys (*Macaca fascicularis*). *Zeitschrift für Tierpsychologie*, 44(3), 225–282.

Dixon, P.M., Tremaine, W.H., Pickles, K. et al. (2000) Equine dental disease Part 3: a long-term study of 400 cases: disorders of wear, traumatic damage and idiopathic fractures, tumours and miscellaneous disorders of the cheek teeth. *Equine Veterinary Journal*, 32(1), 9–18. https://doi.org/10.2746/042516400777612099

Elgart, A.A. (2010) Dental wear, wear rate, and dental disease in the African apes. *American Journal of Primatology*, 72(6), 481–491. https://doi.org/10.1002/ajp.20797

Fox, H. (1939) Chronic arthritis in wild mammals. Being a description of lesions found in the collections of several museums and from a pathologic service. *Transactions of the American Philosophical Society*, 31(2), 73–148.

Fuss, J., Uhlig, G. & Böhme, M. (2018) Earliest evidence of caries lesion in hominids reveal sugar-rich diet for a Middle Miocene dryopithecine from Europe. *PLoS One*, 13(8), e0203307. https://doi.org/10.1371/journal.pone.0203307

Gunz, P., Mitteroecker, P., Neubauer, S., Weber, G.W. & Bookstein, F.L. (2009) Principles for the virtual reconstruction of hominin crania. *Journal of Human Evolution*, 57(1), 48–62. https://doi.org/10.1016/j.jhevol.2009.04.004

Ito, T., Nishimura, T. & Takai, M. (2011) Allometry and interspecific differences in the facial cranium of two closely related macaque species. *Anatomy Research International*, 2011, 849751. https://doi.org/10.1155/2011/849751

Jurmain, R. (1989) Trauma, degenerative disease, and other pathologies among the Gombe chimpanzees. *American Journal of Physical Anthropology*, 80(2), 229–237.

Jurmain, R. (1997) Skeletal evidence of trauma in African apes, with special reference to the Gombe chimpanzees. *Primates*, 38(1), 1–14.

Klingenberg, C.P. (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11, 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Klingenberg, C.P. (2016) Size, shape, and form: concepts of allometry in geometric morphometrics. *Development genes and evolution*, 226(3), 113–137. https://doi.org/10.1007/s00427-016-0539-2

Lovell, N.C. (1990) Skeletal and dental pathology of free-ranging mountain gorillas. *American Journal of Physical Anthropology*, 81(3), 399–412.

Lovell, N.C. (1991) An evolutionary framework for assessing illness and injury in nonhuman primates. *American Journal of Physical Anthropology*, 34(S13), 117–155.

R Core Team (2018) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/

Rohlf, F.J. & Corti, M. (2000) Use of two-block partial least-squares to study covariation in shape. *Systematic Biology*, 49(4), 740–753. https://doi.org/10.1080/106351500750049806

Rohlf, F.J. & Slice, D. (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Biology*, 39(1), 40–59.

Smith, R.J. (2018) The continuing misuse of null hypothesis significance testing in biological anthropology. *American Journal of Physical Anthropology*, 166(1), 236–245. https://doi.org/10.1002/ajpa.23399

Sone, K., Koyasu, K. & Oda, S.I. (2004) Dental and skull anomalies in feral coypu, *Myocastor coypus*. *Archives of Oral Biology*, 49(10), 849–854. https://doi.org/10.1016/j.archoralbio.2004.02.015

Strauss, R.E. & Atanassov, M.N. (2006) Determining best complete subsets of specimens and characters for multivariate morphometric studies in the presence of large amounts of missing data. *Biological Journal of the Linnean Society*, 88(2), 309–328. https://doi.org/10.1111/j.1095-8312.2006.00671.x

Terhune, C.E., Hylander, W.L., Vinyard, C.J. & Taylor, A.B. (2015) Jaw-muscle architecture and mandibular morphology influence relative maximum jaw gapes in the sexually dimorphic *Macaca fascicularis*. *Journal of Human Evolution*, 82, 145–158. https://doi.org/10.1016/j.jhevol.2015.02.006

Van Valkenburgh, B. (1988) Incidence of tooth breakage among large, predatory mammals. *The American Naturalist*, 131(2), 291–302.

Van Valkenburgh, B. (2009) Costs of carnivory: tooth fracture in Pleistocene and Recent carnivorans. *Biological Journal of the Linnean Society*, 96(1), 68–81. https://doi.org/10.1111/j.1095-8312.2008.01108.x

Wheatley, B.P. (2008) Perimortem or postmortem bone fractures? An experimental study of fracture patterns in deer femora. *Journal of Forensic Sciences*, 53(1), 69–72. https://doi.org/10.1111/j.1556-4029.2008.00593.x

Yao, L. (2016) Phylogeography and phenotypic evolution of *Macaca fascicularis* in Southeast Asia. Doctoral dissertation, The University of Chicago.

Zelditch, M.L., Swiderski, D.L. & Sheets, H.D. (2004) *Geometric morphometrics for biologists: A primer*. Waltham, MA: Academic Press.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Mitchell DR, Kirchhoff CA, Cooke SB, Terhune CE. Bolstering geometric morphometrics sample sizes with damaged and pathologic specimens: Is near enough good enough?. *J Anat*. 2021;238:1444–1455. https://doi.org/10.1111/joa.13390