A peer-reviewed version of this preprint was published in PeerJ on 11 June 2015.

View the peer-reviewed version (peerj.com/articles/1000), which is the preferred citable publication unless you specifically need to cite this preprint.

Logan CJ, Palmstrom CR. 2015. Can endocranial volume be estimated accurately from external skull measurements in great-tailed grackles (Quiscalus mexicanus)? PeerJ 3:e1000 https://doi.org/10.7717/peerj.1000
Is there a field proxy for brain size in great-tailed grackles \textit{(Quiscalus mexicanus)}? 

Corina Logan, Christin Palmstrom

There is an increasing need to validate and collect data approximating brain size on individuals in the field to understand what evolutionary factors drive brain size variation within and across species. We investigated whether we could accurately estimate endocranial volume (a proxy for brain size) as measured by computerized tomography (CT) scans, using external skull measurements and/or by filling skulls with beads and pouring them out into a graduated cylinder for male and female great-tailed grackles. We found that while females had much stronger correlations than males, estimations of endocranial volume from external skull measurements or beads did not correlate with CT volumes at a standard that surpassed our strict criteria. We found no accuracy in the ability of external skull measures to predict CT volumes because prediction intervals from data points overlapped extensively. We conclude that we are unable to detect individual differences in endocranial volume using external skull measurements. These results emphasize the importance of validating and explicitly quantifying the predictive accuracy of brain size proxies for each species, and each sex, under consideration.
Is there a field proxy for brain size in great-tailed grackles (*Quiscalus mexicanus*)?

Corina J. Logan*1, Christin Palmstrom2

1SAGE Center for the Study of the Mind and 2College of Creative Studies, University of California Santa Barbara, Santa Barbara, California USA 93106

*Corresponding author: Corina J Logan, SAGE Center for the Study of the Mind, Building 551, University of California Santa Barbara, Santa Barbara, California 93106-9660, USA, 805-570-6821, logan@sagecenter.ucsb.edu

Abstract

There is an increasing need to validate and collect data approximating brain size on individuals in the field to understand what evolutionary factors drive brain size variation within and across species. We investigated whether we could accurately estimate endocranial volume (a proxy for brain size) as measured by computerized tomography (CT) scans, using external skull measurements and/or by filling skulls with beads and pouring them out into a graduated cylinder for male and female great-tailed grackles. We found that while females had much stronger correlations than males, estimations of endocranial volume from external skull measurements or beads did not correlate with CT volumes at a standard that surpassed our strict criteria. We found no accuracy in the ability of external skull measures to predict CT volumes because prediction intervals from data points overlapped extensively. We conclude that we are unable to detect individual differences in endocranial volume using external skull measurements. These results emphasize the importance of validating and explicitly quantifying the predictive accuracy of brain size proxies for each species, and each sex, under consideration.
Introduction

While comparing brain sizes across species has led to a greater understanding of the evolutionary factors correlated with brain size variation at a broad scale (e.g., Iwaniuk & Nelson 2003, Sakai et al. 2011, Sol et al. 2005), little is known about the within species causes and consequences of variation in brain sizes (see Gonda et al. 2013, Thornton & Lukas 2012). Additionally, the accuracy of brain size proxies, which are frequently used in such comparisons, are not often validated. Therefore, the accuracy of these measures and how they compare to measures in other species is questionable (Healy & Rowe 2007). Intraspecies brain size comparisons are rare perhaps due to the difficulty of obtaining data on a number of factors (e.g., biometric measurements, reproductive success, dominance rank, position in the social network, cognitive abilities, etc.) for the same individuals. Acquiring such data is key for understanding what contributes to the evolution of brain size among individuals, as well as across species (Gonda et al. 2013, Logan & Clutton-Brock 2013, Thornton & Lukas 2012).

We investigated whether endocranial volume, a proxy for brain size (Iwaniuk & Nelson 2002), can be approximated using measurements of the external skull in great-tailed grackles (*Quiscalus mexicanus*, JF Gmelin, 1788). Finding such a proxy would greatly ease the collection of data on brain sizes since head measurements can be taken on live birds, thus allowing for correlations with any number of other factors on which data are gathered on this species in the field and the lab. Great-tailed grackle body sizes are sexually dimorphic (Johnson et al. 2000), therefore we expect sex differences in brain sizes and we investigate proxies for each sex independently. We used endocranial volumes calculated from computerized tomography (CT) scans to represent actual endocranial volumes since this measure is the most precise. We compared CT volumes to skull length, width, and height measurements to determine whether the correlation between these two methods and the accuracy of external measures in predicting CT volumes warrants their use as a proxy for endocranial volume. We also evaluated the bead method of generating endocranial volume, where glass beads are poured into
the skull and then out into a graduated cylinder, to increase the value of our research by determining
whether this widely used method also accurately predicts actual endocranial volume as estimated by
CT scans in this species.

Methods

Specimens

We collected data from February through September 2014 on 40 great-tailed grackle skulls (Table S1),
20 female and 20 male (some analyses have 19 males because on one of their skulls the bill was broken
off, thus we could not acquire its skull length measurement), obtained from the Museum of
Southwestern Biology (n=24, Albuquerque, NM), the Ornithology Division of the University of
Kansas (KU) Biodiversity Institute (n=15, Lawrence, KS), and the Santa Barbara Museum of Natural
History (n=1, Santa Barbara, CA). Skulls of unknown age were aged by Andy Johnson if they were
from the Museum of Southwestern Biology or by us if they were from KU. Skulls were aged using the
percentage of ossification to classify each as adult (100% ossified unless it was collected in February-
May because this would mark the start of that individual’s first breeding season after having hatched
June-August in the previous year) or immature (<100% ossified when collected September-December
indicating it had hatched that year; del Hoyo et al. 1992, Winker 2000, Pyle 1997).

Collecting endocranial volume measurements

Linear measurements: Linear measurements of skulls were collected using calipers as would be
measured on a live bird in the field. We recorded skull length from the base of the bill to the back of
the skull along the occipital crest (Figure 1), height from the posterior edge of the foramen magnum to
the top of the skull along the frontal region (Figure 2), and width at the widest part of the brain case
along the squamosal bones (Figures 3). All measurements were taken to the nearest 0.1mm. We
estimated endocranial volume using a number of volumetric shapes and data transformations to determine which best correlated with actual endocranial volumes from CT scans. The volumetric shapes included were: cube (Length x Width x Height), sphere \( \frac{4}{3}\pi r^3 \), where \( r = \frac{2L}{3} \) or \( \frac{2W}{3} \) or \( \frac{2H}{3} \),

ellipsoid \( \frac{4}{3}\pi abc \), where \( a = \frac{2L}{3} \), \( b = \frac{2W}{3} \), \( c = \frac{2H}{3} \), and cone/pyramid \( \frac{1}{3}bh \), where \( b = W \), \( h = H \). We included log, natural log, and exponential transformations of the data, and also allowed polynomial terms.

CT scans: Skulls were CT scanned at the Pueblo Radiology Medical Group in Santa Barbara, California using a Siemens 16-slice Somatom Sensation 16 (1mm slices, 100Kv, 150MAs, 380mm FOV, soft tissue window, analyzed with bone algorithm on). Endocranial volume (cm\(^3\)) was calculated using the DICOM viewer OsiriX v5.8.5 (32-bit, Pixmeo SARL, Switzerland; Figure 4) for 1x1mm slices (regular) and for 1x1mm slices that were taken with the CT scanner bed moved 0.5mm forward (offset), using the average endocranial volume \( \frac{1}{2} \) in analyses. The offset was added to increase the precision of the endocranial volume measurements since grackle craniums are small (approximately 20mm in length), resulting in about 20 slices per scan (one slice every 1mm).

The offset allowed us to measure more area (one slice every 0.5mm) by increasing the number of slices to approximately 40 per skull.

Beads: Endocranial volume was measured by pouring 1mm diameter glass beads (BioSpec Products, catalog number 11079110) into the cranium through the foramen magnum until full. The skull was repeatedly shaken to settle the beads and then filled again until the beads reached the posterior foramen...
magnum without falling out (Figure 5). The volume was calculated by pouring the beads out of the skull and into a graduated cylinder (5ml in 0.1ml graduations, World Precision Instruments, Inc., catalog number CG-0160; note that 1ml=1cm$^3$). In cross-species comparisons, there is mixed evidence about whether pouring the beads into a graduated cylinder introduces error when compared with pouring the beads onto a scale and converting their mass into volume (4% difference: Miller 1997, 0% difference: Isler et al. 2008). However, addressing this issue in an intraspecies study means ensuring that the same amount of error, if any, is introduced for each skull, thus making sure that the relative differences between skulls remains unaffected. This measurement error was controlled in our study because we used the same methods on every skull.

Statistical analyses
The female and male data (analyzed separately) were normally distributed (Anderson Darling normality test $p>0.05$).

We used generalized linear models (GLMs) to determine how well linear and bead measurements correlated with volumes from CT scans, while examining whether the year the skull was collected improved the model fit. GLMs were carried out in R v3.0.2 (R Core Team 2014) using the MCMCglmm function (MCMCglmm package, Hadfield 2010), while applying the dredge function (MuMIn package, Barton 2012) to select the top model using the Akaike weight (Akaike 1981). Females and males were evaluated in separate models. Full models included endocranial volumes from CT scans as the response variable with the following explanatory variables: volume of a cube or sphere or ellipsoid or cone + age * year collected, or skull length + skull width + skull height + age * year collected. GLMs were conducted on the top model for each sex to explore whether the adjusted coefficient of determination (adjusted $r^2$) improved by transforming the explanatory variable endocranial volume proxy in the following ways: squared, cubed, quadratic, exponential, square root,
log, log base 10, and a polynomial with a degree of two or three. Of these, the model with the highest adjusted $r^2$ was chosen as the final top model for that sex and included in the results below.

Since we want to predict CT volumes from linear measures, we validated whether this was possible by generating prediction intervals: the interval in which new observations would occur with 95% probability. We applied the predict function in the MCMCglmm package to the top model for each sex and evaluated whether fitted values (predicted CT volumes) had credible intervals small enough such that there was little to no overlap with other fitted values, thus allowing the discrimination of individual differences.

Data availability
The data from skull measurements and intraobserver reliability, and the R code are available at the knb repository (will be posted soon).

Results
Intraobserver reliability
CP had a very high degree of intraobserver reliability for volume measurements when data included both sexes (Pearson’s product moment correlation: linear method $r^2=0.86$, $p=0.0003$, $n=9$; bead method $r^2=0.90$, $p=0.0001$, $n=9$; CT scans $r^2=0.94$, $p=0.005$, $n=5$) and a substantial amount of reliability for individual linear measurements (skull height $r^2=0.71$, $p=0.004$, $n=9$; skull width $r^2=0.69$, $p=0.003$, $n=9$; skull length $r^2=0.83$, $p=0.001$, $n=9$; Landis & Koch 1977). Intraobserver reliability was also very high when evaluating males independently, though the sample size for the CT volumes was too small to be significantly correlated (linear method $r^2=0.84$, $p=0.02$, $n=7$; bead method $r^2=0.89$, $p=0.008$, $n=7$; CT scans $r^2=0.96$, $p=0.12$, $n=3$). Female data could not be analyzed separately due to the small sample size ($n=2$ for CT scan volumes, $n=1$ for other measures). We can rule out that males and females were
measured with different levels of accuracy, which might have caused the poor correlations between bead volumes/linear measures and CT volumes for males in the analyses below. Male skulls had narrower length and width ranges for individual linear measurements than females, making accuracy more difficult in males (see data at the knb repository). Therefore, it is likely that female intraobserver reliability would be at least as high for female skulls as it was for male skulls.

Correlations between methods

Volumes from CT scans were poorly (for males) to moderately (for females) correlated with volumes from linear measurements, with the sphere being the best fitting shape for both sexes according to the Akaike weights (the radius was based on skull width for males and skull height for females). The top female model showed a positive relationship between CT volumes and volumes from using the skull height as the radius for a sphere, volumes were larger for immatures than for adults, and volumes slightly decreased over the years collected (Akaike weight=0.60, 

\[ y = 0.00002 \times VolumeSph + 1.10 \times Age - 0.007 \times Year + 1.44, \]  
adjusted \( r^2=0.80, \) \( p<0.0001, \) model 1; Figure 6a). The top male model showed a positive correlation between CT volumes and volumes using a quadratic polynomial of the skull width as the radius for a sphere, volumes were slightly larger for immatures than for adults, and volumes decreased slightly over the years collected (Akaike weight=0.26, 

\[ y = 0.47 \times VolumeSph + 0.19 \times VolumeSph^2 + 0.12 \times Age - 0.003 \times Year + 5.25, \]  
adjusted \( r^2=0.39, \) \( p=0.02, \) model 2; Figure 6b). Transformations of the explanatory volume variables or substituting volume for individual linear measurements (length, width, height, or some combination of these) did not improve the adjusted \( r^2 \) for females.
Volumes from CT scans were only moderately positively correlated with volumes from the bead method for both sexes: the top female model showed that endocranial volumes decreased slightly over time ($\text{Akaike weight}=0.74$, adjusted $r^2=0.77$, $p<0.0001$, $y = 0.37 \times \text{VolumeBead} - 0.0005 \times \text{Year} + 11.24$, model 3; Figure 7a), while the top male model included age, with immatures having smaller volumes than adults ($\text{Akaike weight}=0.45$, adjusted $r^2=0.68$, $p<0.0001$, $y = 0.66 \times \text{VolumeBead} - 0.09 \times \text{Age} + 0.66$, model 4; Figure 7b). None of the models from the bead method or linear measurements had high enough Akaike weights to make strong inferences about the data (Burnham & Anderson 2002).

None of the correlations between CT volumes and linear measures met our subjective minimum criteria ($r^2>0.88$) for a strong enough relationship to predict endocranial volumes from linear measurements of live birds in the field. Since we want to predict CT volumes from linear measures, we determined whether this was possible by generating prediction intervals for the top female and male models for the linear measurements (models 1 and 2) and bead method (models 3 and 4). We found that the lower and upper limits of the 95% confidence intervals of the predicted values for both sexes show extensive overlap such that individual differences would not be able to be resolved if a new, unvalidated data point was obtained (Table 1).

**Comparing method means**

Endocranial volume means were significantly different from each other when comparing across methods (mean±standard deviation: females: $\text{volumeCT } 2.29\text{cm}^3±0.20$, $\text{volumeSphere } 32459.1\text{mm}^3±4344.7$, $\text{volumeBead } 2.60\text{ml}±0.28$; males: $\text{volumeCT } 2.54\text{cm}^3±0.15$, $\text{volumeSphere } 59292.1\text{mm}^3±2360.4$, $\text{volumeBead } 2.91\text{ml}±0.21$; Welch two sample t-test: females: $\text{CT x Sphere } t=31$, $p<0.0001$, df=19; $\text{Sphere x Bead } t=-31$, $p<0.0001$, df=19; $\text{Bead x CT } t=4$, $p=0.0003$, df=34; males: $\text{CT x Sphere } t=31$, $p<0.0001$, df=19; $\text{Sphere x Bead } t=-31$, $p<0.0001$, df=19; $\text{Bead x CT } t=4$, $p=0.0003$, df=34;
Discussion

While female great-tailed grackle endocranial volumes from linear measurements were moderately correlated with volumes from CT scans, which we consider a more accurate proxy for brain size, the correlation did not meet our criteria of having a coefficient of determination ($r^2$) greater than 0.88—a level of correlation that might allow the resolution of individual differences in endocranial volumes. This correlation was weak in males, which is likely due to the sexual dimorphism in this species and potentially influenced by traits that correlate with male reproductive success (tail length and iridescence; Johnson et al. 2000). Perhaps additional biometric measurements would explain more of the variation in their endocranial volumes from CT scans, however we only had access to skulls for most of the specimens and therefore could not test this hypothesis.

We were more interested in whether a given value of $x$ (some external skull measurement) could accurately predict $y$ (actual endocranial volume from CT scans), rather than setting a subjective criterion about how high $r^2$ should be, especially given the extensive debate around the latter approach (e.g., Legates & McCabe Jr. 1999, Müller & Büttner 1994). In particular, $r^2$ does not allow one to investigate differences in the variance of individual data points because it “…describes the proportion of the total variance in the observed data that can be explained by the model” (Legates & McCabe Jr. 1999, p. 233, emphasis added). Our predictive analyses showed that prediction intervals for new data points overlapped to such a degree (within 95% credible intervals) that it was not possible to distinguish among individuals, as we would need to when collecting linear measurements on new individuals in the field. Therefore, we must conclude that there is not a field proxy accurate enough yet to estimate endocranial volume, and thus brain size, in great-tailed grackles.
Predictive analyses are crucial for determining the accuracy of predicting individual data points by a particular method and should be applied extensively in future research, rather than relying solely on correlation coefficients (r) or coefficients of determination (r^2). The omission of such an analysis leaves data uninterpretable for its purported use of discerning intraspecies differences in a morphological feature. Additionally, we caution against using a proxy validated in one species as evidence that the same proxy will apply to other species (e.g., great tits: Dreyer 2012). Until intraspecies validations of brain size proxies using skull or head measurements have been validated across species, we cannot assume that what works (or not) for one species will work (or not) for another.

The bead method was only moderately correlated with CT volumes in both sexes and prediction intervals also extensively overlapped for individual data points. Iwaniuk and Nelson (2002) validated the strong relationship between the endocranial volumes from beads and actual brain masses in 81 bird species (r^2=0.98, p<0.01). Great-tailed grackles appear to be an anomaly since this relationship does not hold for them. However, great-tailed grackles and common grackles are among the species with the largest ranges in endocranial volumes (as measured using the bead method) when compared with the other species in Iwaniuk & Nelson’s (2002) study (common grackles: mean ± SD=2.59ml ± 0.37; Iwaniuk & Nelson 2002; great-tailed grackles: female 2.60ml ± 0.28, male 2.91ml ± 0.21; this study). It appears that grackle skulls are more variable than skulls in other species and it is not clear how this variation relates to brain size.

To infer differences in brain size among individuals of the same species, and of the same sex, there must be a high degree of accuracy to have the ability to detect actual individual differences (Legates & McCabe Jr. 1999, Logan & Clutton-Brock 2013). Our results highlight the need to validate brain size proxies and their predictive power for each species under investigation, and for each sex if they are sexually dimorphic. It is unfortunate that there is not an easier, more accurate way to
approximate brain size in the field where we have the potential to understand how evolutionary factors
drive brain size variation within species. However, this study accentuates the importance of knowing
how accurate brain size measures are when including such data in analyses.

Acknowledgements

We thank Scott Grafton at the University of California Santa Barbara (UCSB) and Lawrence Harter
and Chuck Scudelari at the Pueblo Radiology Medical Group for providing access to the CT scanner;
Mario Mendoza at UCSB for scan planning; Sabrina Cruz and Kerri Moore at the Pueblo Radiology
Medical Group for scanning; Steve Rothstein for advice and sponsoring CP’s grant; and Mireia Beas-
Moix and the Cheadle Center for Biodiversity and Ecological Restoration at UCSB for hosting the
skull loans, Dieter Lukas for analysis assistance, and Jay Taylor for feedback on a previous version of
this manuscript. We are grateful for skull loans from the Museum of Southwestern Biology (and to
Andy Johnson for aging skulls), Krista Fahy and the Santa Barbara Museum of Natural History, and
the Ornithology Division of the KU Biodiversity Institute.

References

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-
theoretic approach. New York: Springer.
Edicions.


1

Skull length

Measuring skull length as it is measured on live birds
Skull height

Measuring skull height, replicating the height that can be measured on live birds
Skull width

Measuring skull width at the widest part of the braincase as it would be measured on live birds
Figure 4 (on next page)

CT scan of grackle skulls

CT scan showing five grackle skulls using OsiriX
Bead method

Skull holes are plugged with cotton and then the cranium is filled with glass beads
Plots of the correlations between Volume Sphere and Volume CT for females and males.

Correlations between CT volumes and the volume of a sphere as calculated from linear measurements for female (A) and male (B) adults (small circles) and immatures (large circles), with the year the skull was collected represented by a red-blue spectrum (earlier years are redder and recent years are bluer). Note that regression lines only reflect VolumeSphere\textendash VolumeCT and do not correct for age or year (factors in the top model for both sexes) as in the GLMs.
Figure A shows the relationship between CT Volume (cm³) and Volume Sphere (mm³) for different samples, with points marked in red and blue. Figure B displays another set of data points, also indicating a linear relationship between the variables, with a different color scheme for clarity.
Plots of the correlations between Volume Bead and Volume CT for females and males.

Correlations between CT volumes and bead volumes for female (A) and male (B) adults (small circles) and immatures (large circles), with the year the skull was collected represented by a red-blue spectrum (earlier years are redder and recent years are bluer). Note that regression lines only reflect VolumeBead~VolumeCT and do not correct for age (in the top male model) or year (in the top female model) as in the GLMs.
Grackle skulls on the CT scanner
Table 1 (on next page)

Predicted CT volumes and their prediction intervals

Predicted CT volume values (fitted value) and the predicted intervals in which these new data points would occur with 95% credible intervals based on inputs from linear measures or the bead method in the top female and male models for each method.
Table 1. Predicted CT volume values (fitted value) and the predicted intervals in which these new data points would occur with 95% credible intervals based on inputs from linear measures or the bead method in the top female and male models for each method.

<table>
<thead>
<tr>
<th>Linear Measurements</th>
<th>Bead Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
</tr>
<tr>
<td></td>
<td>Fitted value</td>
</tr>
<tr>
<td>Males</td>
<td>Fitted value</td>
</tr>
<tr>
<td>2.72</td>
<td>2.37</td>
</tr>
<tr>
<td>2.41</td>
<td>2.17</td>
</tr>
<tr>
<td>2.40</td>
<td>2.12</td>
</tr>
<tr>
<td>2.60</td>
<td>2.33</td>
</tr>
<tr>
<td>2.42</td>
<td>2.10</td>
</tr>
<tr>
<td>2.52</td>
<td>2.22</td>
</tr>
<tr>
<td>2.77</td>
<td>2.45</td>
</tr>
<tr>
<td>2.67</td>
<td>2.41</td>
</tr>
<tr>
<td>2.35</td>
<td>2.06</td>
</tr>
<tr>
<td>2.55</td>
<td>2.30</td>
</tr>
<tr>
<td>2.46</td>
<td>2.21</td>
</tr>
<tr>
<td>2.64</td>
<td>2.40</td>
</tr>
<tr>
<td>2.53</td>
<td>2.24</td>
</tr>
<tr>
<td>2.58</td>
<td>2.31</td>
</tr>
<tr>
<td>2.55</td>
<td>2.31</td>
</tr>
<tr>
<td>2.64</td>
<td>2.39</td>
</tr>
<tr>
<td>2.54</td>
<td>2.29</td>
</tr>
<tr>
<td>2.55</td>
<td>2.24</td>
</tr>
<tr>
<td>2.52</td>
<td>2.26</td>
</tr>
<tr>
<td>2.51</td>
<td>2.24</td>
</tr>
</tbody>
</table>