

LETTER TO THE EDITOR

Problems in Erickson et al. 2009

In the course of pursuing research on dinosaur growth rates (Myhrvold, 2013), I attempted to replicate the findings of Erickson et al. for *Psittacosaurus lujiatunensis* (Erickson et al., 2009a). In their study, they use femoral length measurements and age estimates for 80 specimens of *P. lujiatunensis* to estimate body mass and age, fit a growth curve, and construct a life table. These results were then used to describe certain aspects of developmental timing and to compare with growth in other dinosaurs as well as extant birds and mammals.

There are statistical problems with the approach that Erickson et al. use for growth curve fitting (Myhrvold, 2013). This is evident in Figure 3 of their paper, which shows an ordinary linear regression using femur length as the independent variable and age as the dependent variable. Having thereby demonstrated that age is well fit by a linear relationship between femur length and age, Figure 6 of Erickson et al. proceeds to use age as the independent variable and fits a sigmoidal model to the cube of femur length. As discussed at length in Myhrvold (2013), the choice of independent variable in Figure 3 is inconsistent with that in Figure 6. The model is also inconsistent—if there is a linear relationship between age and femur length, then the cube of femur length should fit a cubic curve in age, not a sigmoidal one. Fitting a sigmoidal curve to cubic data gives nonsensical results (Myhrvold, 2013).

However, the main thrust of this letter is not these choices in statistical methodology, but instead to ask a far simpler question—can the growth curve results of Erickson et al. 2009 be replicated using their stated methods?

In Figure 6 of Erickson et al. 2009, they plot a two-parameter logistic growth model, $m(t) = \frac{ae^{bt}}{a + e^{bt} - 1}$, that has been fit to estimated mass-at-age data for *P. lujiatunensis*. According to the text, “Sigmoidal growth equations were fit using least squares regression and used to describe these data.” These are well known and robust statistical techniques and should be easily replicated. Erickson et al. report the best fit parameters $asa = 37.38$ and $b = 0.55$.

Unfortunately, the published regression equation cannot be replicated—the best fit that I get using the same function specified in the paper is radically different with $a = 88.4$ and $b = 0.345$ (for the full data set). Erickson et al. state that they use the R programming language (version 2.8.1) to do their curve

fits. I attempted replication with R (version 2.15.2) as well as with three commercial software packages: Microsoft Excel 2010 (version 14.0.7116.5000), Matlab (version R2013), and *Mathematica* (version 9.01). In no case can the results be replicated, even approximately—see Table 1. I also tried to replicate the regression equation published in the paper using a variety of other functions to test possible sources of error, including setting one parameter by fiat to the same value found by Erickson et al., or using different functions to anticipate possible typographical errors, all without success (Fig. 1, Table 1).

The data to be fit must first be transformed from age–femur length pairs in Table 1 to age–mass pairs using cubic scaling (developmental mass extrapolation), setting the mass of LPM R00117 to 26.8 kg. I tried this using both the full data set (all specimens in Table 1 of Erickson et al. 2009), as well as the subset of the data marked as having histologically determined ages, or subsets of unique points (ignoring multiplicity of identical points). I also used image processing techniques (in *Mathematica* using its `SelectComponents[]` and `ComponentMeasurements[]` functions, and subsequently adjusted by hand) to locate the center of the plotted points on a digital scan of Figure 6 of Erickson et al. to recover the as-plotted data set directly (see Myhrvold, 2013). The regression for each data set differed in parameters a , b but in no case was I able to even approximately replicate the Erickson et al. results—see Table 1.

However, for each data set, the results obtained were the approximately replicated across each of the different software systems, which agree to reasonable accuracy (Table 1). This shows that least squares regression is robust across different algorithms and software implementations for these data sets and the chosen model. Note further that some of these systems use a particular optimization algorithm (Levenberg–Marquardt for Matlab and R), others choose algorithm automatically (Excel), while *Mathematica* implements nine different algorithms. Each case gave the same results within reasonable accuracy.

A direct way to compare different least squares regressions is to sum the squared residuals. Table 2 shows the results for the regression of Erickson et al. 2009 using the full data set. The sum of squared residuals for the fits in this letter is more than 10-fold lower than for Erickson et al.’s fit. Similar

TABLE 1. Results from attempted replication of the curve fit of Figure 6 from Erickson et al. 2009

Data set	Origin	Regression source	Version	a	b
Unknown	Unknown	Erickson et al. 2009	2.8.1	37.38	0.55
Full data set	Table 1 of Erickson et al. 2009, converted to age-mass pairs	<i>Mathematica</i>	9.01	88.44	0.345
		Microsoft Excel 2010	14.0.7116.5000	88.45	0.345
		Matlab	R2013	88.44	0.345
		R	2.15.2	88.44	0.345
Full data set—unique age-mass pairs only	Duplicate points removed	<i>Mathematica</i>	9.01	79.17	0.349
		Microsoft Excel 2010	14.0.7116.5000	79.17	0.349
		Matlab	R2013	79.17	0.349
		R	2.15.2	79.17	0.349
Histologically aged subset	Labeled in Table 1 of Erickson et al. 2009, converted to age-mass pairs	<i>Mathematica</i>	9.01	86.57	0.346
		Microsoft Excel 2010	14.0.7116.5000	86.58	0.346
		Matlab	R2013	86.57	0.346
		R	2.15.2	86.57	0.346
Histologically aged subset—unique age-mass pairs only	Duplicate points removed	<i>Mathematica</i>	9.01	77.93	0.351
		Microsoft Excel 2010	14.0.7116.5000	77.93	0.351
		Matlab	R2013	77.92	0.351
		R	2.15.2	77.92	0.351
As-plotted data set	Recovered from digital scan of Figure 6 of Erickson et al. 2009	<i>Mathematica</i>	9.01	73.29	0.353
		Microsoft Excel 2010	14.0.7116.5000	73.26	0.353
		Matlab	R2013	73.26	0.353
		R	2.15.2	73.26	0.353

The first row is the original fit to a two-parameter (a , b) logistic function published in the caption of Figure 6 of Erickson et al. 2009. The following rows are the regression results from different software packages with the full data set (Table 1 of Erickson et al. 2009) or variations on that data set described in the text. In no case do they match the first row. Each case uses the two-parameter logistic function specified in Erickson et al. 2009. In the case of *Mathematica*, nine different optimization algorithms for computing the least squares regression were tried, including conjugate-gradient, gradient, Newton, quasi-Newton, Levenberg-Marquardt, Nelder-Mead, differential evolution, simulated annealing, and random search, all of which gave the result to within four significant digits using double precision arithmetic (64 bit) and a maximum of 1,000 iterations. Using extended precision (50 digit or 166 bit) arithmetic and up to 100,000 iterations did not change the results to four significant digits.

results hold for fits based on variations of the data set and model (Text S1 of Myhrvold, 2013).

Figure 6 of Erickson et al. 2009 also has graphical problems. The curve plotted in the figure does not match the published regression equation stated in the caption of the same figure. Figure 1 of this letter shows a digital scan of Figure 6 of Erickson et al. 2009, with the correct plot of the equation from the caption superimposed on top and plotted in red. The curve lies way to the left of the data points. The original plot (dark blue) goes through the points and it appears to have an asymptotic value that is quite different than that of the published equation. The curve plotted does not correspond to any of the best-fit regression attempts I made (solid and dashed curves in Fig. 1), and I was unable to identify any explanation for the origin of the plotted curve.

In addition, the data points plotted do not match the data set published in Table 1 of Erickson et al. 2009. Table 1 lists 80 specimens which have by 40 distinct age-size values. Of these, 20 are labeled as having their age determined by histology. There are also 20 distinct points plotted in Figure 6 of Erickson et al. 2009; 14 of the plotted data points match (approximately) the histologically aged data points from Table 1 when plotted as an overlay; six histologically aged points from Table 1 are missing. As an

example, Table 1 includes IVPP R00142, which is histologically aged at 2 years old; it is easy to see there is no data point at age 2 in Figure 6.

Five of the remaining data points plotted in Figure 6 do not correspond closely to any data point in Table 1 of Erickson et al. 2009, and one corresponds to a data point that is not labeled as histologically aged. It would appear that the data set plotted in Figure 6 of Erickson et al. 2009 is largely based on the histologically aged subset but I have no explanation for the points that do not match that subset. It is unknown whether the data set as plotted in Figure 6 was used in the regression analysis.

Figures are a primary way to communicate scientific results and it is troubling that Figure 6 does not plot the regression equation described in its own caption or appear to plot the referenced data set. Figures plotted by statistical software ought to be replicable to good accuracy (typically ± 1 pixel). The fact that the primary regression result also cannot be replicated is also a major issue. The growth rate results depend on the regression; as it cannot be replicated, all results depending on the regression are suspect. The correct (i.e., reproducible) regression has biological metrics like maximum growth rate that are very different than the correct, reproducible curve fit. This calls into question all conclusions based on these values.

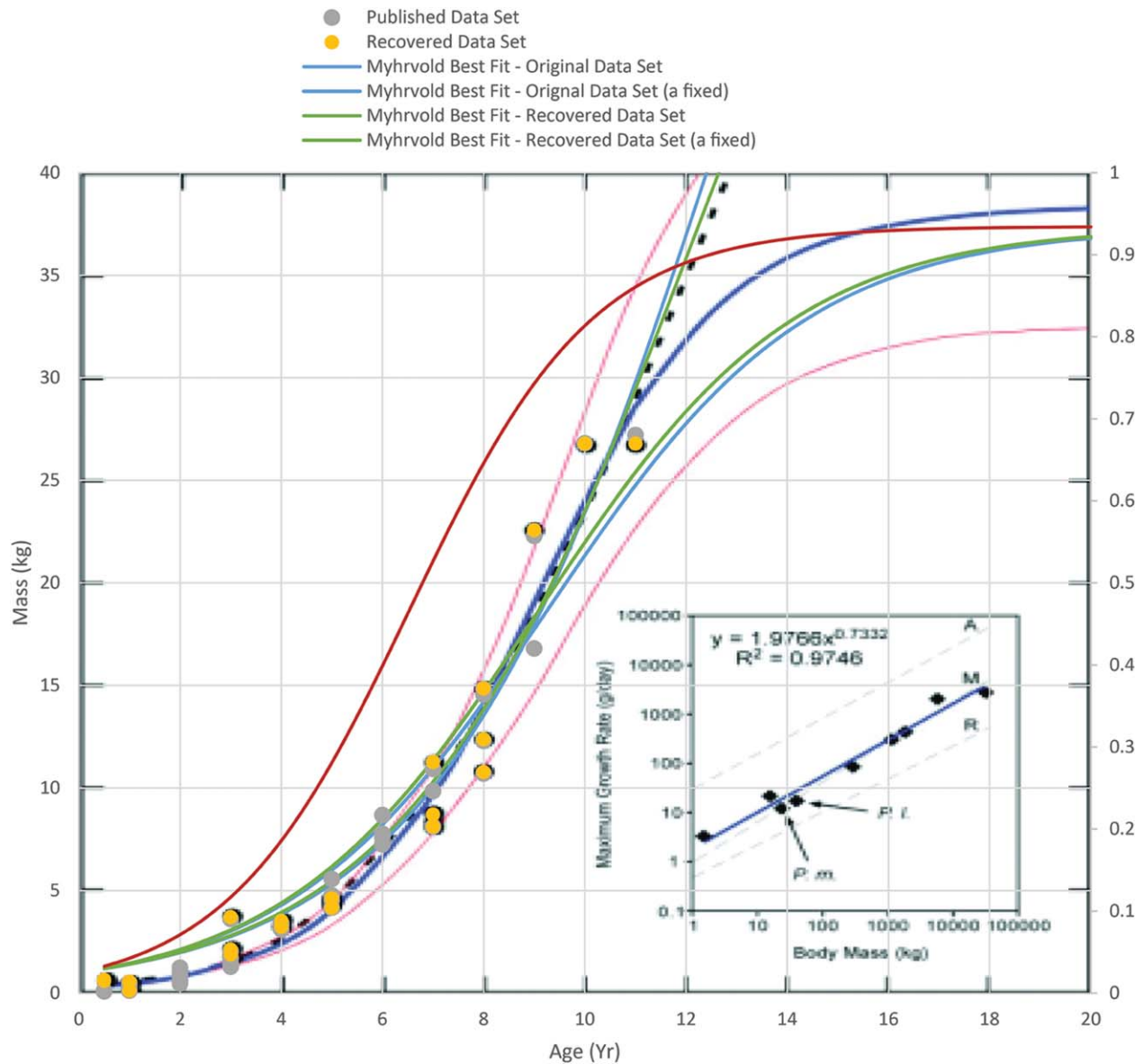


Fig. 1. A plot of Erickson et al.'s best fit equation, given in the caption of Figure 6 in Erickson et al. 2009, as well as the published and recovered data sets and two of my regression attempts, overlaid on top of the suspect figure. Erickson et al.'s best fit curve according to the caption published with Figure 6 matches neither the plotted best fit curve nor any of my attempted regressions. The recovered data set which I used in my reanalysis, plotted as yellow

points, shows close correspondence with the data set as plotted. Comparing the published data set with the recovered data set, it is clear that many of the original data points were not plotted in the published figure, and others appear to have been plotted inaccurately. The source of the data set as published, the best fit regression curve, and the regression curve as plotted are unknown.

The correct regression shows that the specimens are immature; the largest specimens are just over 30% of the maximum asymptotic size. As discussed by Myhrvold (2013), this must be treated with caution; with no data points in the asymptotic region all we can say with certainty is a self-consistency point that this data set cannot tell us much about the asymptotic properties of the chosen model. As a conse-

quence, it is not appropriate to treat the asymptotic size as a good estimate of the actual size achieved in life; it is an unsupported extrapolation far from the data points (Myhrvold, 2013).

This contrasts with the caption of Figure 6 of Erickson et al. 2009, which says "A Gompertz equation (black dashed line) shows similar fit for the empirical data but predicts an unreasonable

TABLE 2. Sum of squared residuals for Erickson's and two Myhrvold (2013) fits to the originally published *P. lujiatunensis* data set (Table 1 of Erickson et al. 2009). The Myhrvold (2013) results were obtained using *Mathematica*. The center column is a two parameter fit to the original logistic function in Erickson et al. 2009. The right column shows the results of fixing the parameter $a=37.38$, to match the value used by Erickson et al. 2009, and then performing a one-dimensional regression on the parameter b . Even when one parameter is fixed, one cannot recover the Erickson et al. value for the remaining parameter. In general, Myhrvold fits have more than a 10-fold lower sum of square error than the Erickson equation, demonstrating that they are much better fits.

Data set	Regression source	a	b	Sum of squared residuals
Full Data Set	Erickson <i>et al.</i> 2009	37.38	0.55	2197.57
	Best fit	88.44	0.345	133.52
	Best fit setting $a = 37.38$	37.38	0.388	181.102
Full Data Set – unique points only	Erickson <i>et al.</i> 2009	37.38	0.55	1928.01
	Best fit	79.17	0.349	81.64
	Best fit setting $a = 37.38$	37.38	0.390	121.39
Histologically aged subset	Erickson <i>et al.</i> 2009	37.38	0.55	1414.23
	Best fit	86.57	0.346	74.33
	Best fit setting $a = 37.38$	37.38	0.392	114.77
Histologically aged subset – unique points only	Erickson <i>et al.</i> 2009	37.38	0.55	1237.67
	Best fit	77.92	0.351	64.12
	Best fit setting $a = 37.38$	37.38	0.395	98.16
As-plotted in Figure 6 of Erickson <i>et al.</i> 2009	Erickson <i>et al.</i> 2009	37.38	0.55	1235.51
	Best fit	73.29	0.353	65.80
	Best fit setting $a = 37.38$	37.38	0.395	96.22

asymptotic size (107 kg).” This is problematic for two reasons. First, neither the Gompertz model referenced nor the correct fit to the logistic model discussed here can possibly yield a good estimate of the maximum size achieved during life because they both lack data points in the asymptotic region (Myhrvold, 2013 and references therein). Second, conventional statistical practice does not rely on the investigator’s opinion of “reasonable” size as a model selection criterion.

Fortunately, *P. lujiatunensis* is known from many specimens besides those studied by Erickson et al. 2009 (Hedrick and Dodson, 2013; Zhao et al., 2013) so future studies will have more data. They may also use a broader set of models, and use formal model selection criteria (Myhrvold, 2013).

Although the primary focus of this letter is the growth curve modeling, age determination and maturity are key inputs to the life table analysis done in other parts of Erickson et al. 2009. The problems found fundamentally change the interpretation of the regression, and potentially the interpretation of the life table as well. For example, a direct interpretation of the growth curve leads to the conclusion that the specimens are all immature (Myhrvold, 2013). In addition, the statistical methodology used by Erickson in other life table work has been shown to be invalid (Erickson et al., 2006; Steinsaltz and Orzack, 2011). That portion of the Erickson et al. 2009 paper, Table 2 and Figure 5, has been the subject of one correction already (Erickson et al., 2009b).

Proper scientific and statistical practice is to document all methods so that they can be replicated, including the details of the statistical analysis, any

data imputation or modification (American Statistical Association Committee on Professional Ethics, 1999). It is impossible for an external observer to say with any confidence what led to these problems, but similar errors, irregularities, and inconsistencies seemed to have occurred in several other papers published by the same first author (Erickson et al., 2001, 2004)—see Myhrvold (2013).

Nathan P. Myhrvold
Intellectual Ventures
3150 139th Ave SE Bellevue
Washington USA

LITERATURE CITED

- American Statistical Association Committee on Professional Ethics. 1999. Ethical guidelines for statistical practice. Available at: <http://www.amstat.org/about/ethicalguidelines.cfm>.
- Erickson GM, Currie PJ, Inouye BD, Winn AA. 2006. Tyrannosaur life tables: an example of nonavian dinosaur population biology. *Science* 313:213–217.
- Erickson GM, Makovicky PJ, Currie PJ, Norell MA, Yerby SA, Brochu CA. 2004. Gigantism and comparative life-history parameters of tyrannosaurid dinosaurs. *Nature* 430:772–775.
- Erickson GM, Makovicky PJ, Inouye BD, Zhou C-F, Gao K-Q. 2009a. A life table for *Psittacosaurus lujiatunensis*: initial insights into ornithischian dinosaur population biology. *Anat Rec* 292:1514–1521.
- Erickson GM, Makovicky PJ, Inouye BD, Zhou C-F, Gao K-Q. 2009b. Erratum. *Anat Rec* 292:1684–1684.
- Erickson GM, Rogers KC, Yerby SA. 2001. Dinosaurian growth patterns and rapid avian growth rates. *Nature* 412:429–433.

- Hedrick B, Dodson P. 2013. Lujiatun psittacosaurids: understanding individual and taphonomic variation using 3D geometric morphometrics. *PLoS One* 8:e69265.
- Myhrvold NP. 2013. Revisiting the estimation of dinosaur growth rates. Edited by Alistair Robert Evans. *PLoS One* 8:e81917.
- Steinsaltz D, Orzack SH. 2011. Statistical methods for paleodemography on fossil assemblages having small numbers of specimens: an investigation of dinosaur survival rates. *Paleobiology* 37:113–125.
- Zhao Q, Benton MJ, Sullivan C, Sander PM, Xu X. 2013. Histology and postural change during the growth of the ceratopsian dinosaur *Psittacosaurus lujiatunensis*. *Nature communication*, 4.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Revised: 17 July 2014; Accepted: 30 July 2014.

DOI 10.1002/ar.23101

Published online 18 November 2014 in Wiley Online Library (wileyonlinelibrary. com).