

On the use of relative brain size

Simeon Q. Smeele^{1,2,3}

March 2022

¹Cognitive & Cultural Ecology Research Group, Max Planck Institute of Animal Behavior, Radolfzell, Germany

²Department of Human Behavior, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

³Department of Biology, University of Konstanz, Konstanz, Germany

Subject areas: evolution

Keywords:: relative brain size, comparative analysis, structural equation model, Bayesian

Author for correspondence: Simeon Q. Smeele, ssmeele@ab.mpg.de

Abstract

There is a long standing interest in the effect of relative brain size on other life history variables in a comparative context. Historically, residuals have been used to calculate these effects, but more recently it has been recognised that regression on residuals is not good practice. Instead, absolute brain size and body size are included in a single regression, with the idea that this controls for allometry. This practice leads to a second, potentially even greater estimation error when both relative brain size and body size are a cause of the third variable in question. In this study I use a simple simulation to show how the effect of body mass is masked if absolute brain size is included. I propose the use of structural equation models to simultaneously estimate relative brain size and its effect on the third variable, leaving the effect of body mass unconfounded.

Introduction

Brain size is often theorised to be important in the life history of species and has been linked to everything from innovation rates (Lefebvre, Reader, and Sol 2004), to sociality (Dunbar and Shultz 2007) and longevity (Minias and Podlaszczuk 2017). It is important to note that it is often the *relative* and not the *absolute* size of the brain that is of interest. When talking about relative brain size, it is always implicit that brain size is relative to body size or a specific brain region relative to the whole brain. This to control for allometric scaling. There exists a large diversity of relative brain size measures (Healy and Rowe 2007), but some decades ago the use of residual brain size was proposed and quickly became the most popular (Jerison 1973). This measure is simply the residual from a model that has log body size as predictor and log brain size as response variable. It is an intuitively attractive approach, since it seems to get at *that extra bit of brain* a species has. Multiple challenges were recognised early on. Probably the most debated is the effect of phylogeny (Armstrong 1983). When comparing across multiple species nested at different levels of taxonomy, how does one estimate the ‘true’ intercept and slope of the model? Several techniques have been proposed to mitigate this, but the debate is ongoing.

More recently another statistical caveat has been highlighted (for an overview see Freckleton (2002)). The use of residuals is great for visualisation, but cannot be used in statistical models. The proposed solution is to simply include body size as a predictor variable in the final model, which is very similar to the use of

residual brain size, if - and only if - relative brain size is the response variable. This approach takes care of both the phylogenetic signal (since models can include the phylogenetic variance-covariance matrix) and correct estimation of uncertainty. However, one major issue has been overlooked: if relative brain size is a predictor variable, including body size as a second predictor can lead to incorrect inference, because body size is not controlling for variation in brain size, as brain size is not the response variable. Brain size does, however, contain information about body size as body size is a cause of brain size. This will result in a reduction of the estimated effect of the body size. To properly include both relative brain size and body size one needs a system that contains regressions with brain size as response (of body size) and as predictor (of the third variable).

A structural equation model is such a system. It contains regressions for each variable and allows brain size to be response and outcome variable simultaneously (Bowen and Guo 2011). Structural equation models have been used before to include relative brain size in a comparative study in combination with phylogeny, measurement error and imputation [REF own paper]. The aim of this current paper is to show the estimation error in ordinary linear regressions using a simulation and propose a simple Bayesian structural equation model as a solution that can be easily adapted to most comparative studies.

Methods

To show the difference between a case where relative brain size is the *response* variable and where it is a *predictor* variable I simulated two simple cases with three variables. To avoid discussion about directionality I have simply named the third variable z . To further simplify the simulation I have not attempted to simulate realistic values for body size and relative brain size, but just used values with mean = 0 and standard deviation = 1. This allows me to draw general conclusions that are not sensitive to the scale of the variables.

I simulated 20 datasets per case and present parameter estimates for all datasets. Frequentist linear models were fitted with the *lm* function from base R (R Core Team 2021). Bayesian models were fitted using the *cmstanr* package (Gabry and Češnovar 2021), which runs the No U-turn Sampler in Stan (Gelman, Lee, and Guo 2015) with four chains and default sampler settings.

Case I: relative brain size as response variable

In the first case absolute brain size is caused by body size and z . The interest of the study is to what extent z causes additional increase in brain size. In other words, to what extent z correlates with relative brain size. I simulated 20 data sets with 100 species with the following structure:

$$\begin{aligned} \text{body size} &\sim \text{normal}(0, 1) \\ z &\sim \text{normal}(0, 1) \\ \text{brain size} &\sim \text{normal}(\mu_{\text{brain}}, 1) \\ \mu_{\text{brain}} &= 1 * \text{body size} + 1 * z \end{aligned}$$

I analysed the resulting data with a frequentist linear model and with the Bayesian equivalent. Then I plotted the estimated effect of body size and z to show how well parameters were retrieved.

Case II: relative brain size as predictor variable

In the second case both body size and relative brain size are a predictor of z . The interest of the study is to what extent relative brain size causes z . I simulated 20 data sets with 100 species with the following structure:

$$\begin{aligned}
&\text{body size} \sim \text{normal}(0, 1) \\
&\text{relative brain size} \sim \text{normal}(0, 1) \\
&\text{brain size} = \text{relative brain size} + 1 * \text{body size} \\
&z \sim \text{normal}(\mu_z, 1) \\
&\mu_z = 1 * \text{body size} + 1 * \text{relative brain size}
\end{aligned}$$

I analysed the resulting data with both a frequentist and Bayesian linear model where brain size and body size were included as predictor variables. Additionally I analysed the data with a Bayesian structural equation model that included sub-models for all causal paths:

$$\begin{aligned}
&\text{body size} \sim \text{normal}(\alpha_{\text{body}}, \sigma_{\text{body}}) \\
&\text{brain size} \sim \text{normal}(\mu_{\text{brain}}, \sigma_{\text{brain}}) \\
&\mu_{\text{brain}[i]} = \alpha_{\text{brain}} + \beta_{\text{body}} * \text{body size}_{[i]} \\
&z \sim \text{normal}(\mu_z, \sigma_z) \\
&\mu_{z[i]} = \alpha_z + \zeta_{\text{body}} * \text{body size}_{[i]} + \zeta_{\text{brain}} * (\text{brain size}_{[i]} - \mu_{\text{brain}[i]}) \\
&\alpha_{\text{body}}, \alpha_{\text{brain}}, \alpha_z \sim \text{normal}(0, 1) \\
&\beta_{\text{body}}, \zeta_{\text{body}}, \zeta_{\text{brain}} \sim \text{normal}(0, 1) \\
&\sigma_{\text{body}}, \sigma_{\text{brain}}, \sigma_z \sim \text{exponential}(1)
\end{aligned}$$

The model includes a regression for each variable. Body size is not a function of any variable. Brain size is a function of body size. z is a function of body size and relative brain size (where relative brain size is the difference between the actual and expected brain size). The last three lines are the priors for all parameters. Note that the priors for the slopes (β and ζ) are set to $\text{normal}(0, 1)$, which regularises them slightly and assumes no effect of the predictors. For empirical studies theory might provide more informative priors, which would further increase the accuracy of the model.

Results

For the Case I, where relative brain size was the response variable, both models estimated the parameters very well (see Figure 1). For Case II, where relative brain size was a predictor variable, the effect of brain size was estimated well by all models, but the effect of body size was only estimated correctly by the structural equation model (see Figure 2). Both the frequentist and Bayesian linear model estimated the effect of body mass to be essentially 0.

Discussion

Several recent studies have claimed to study the effect of relative brain size by including both absolute brain size and body size as predictors (Street et al. 2017; González-Lagos, Sol, and Reader 2010; Isler and Van Schaik 2009). These studies also reported the direct effect of body size and sometimes drew conclusions based on the sign of this effect. Since all these studies used some version of a linear model (be it phylogenetic and/or Bayesian), they actually tested the effect of absolute brain size, as including body size only accounts for allometry in the response variable. The simulations in this paper showed that including body mass as additional predictor to control for allometric scaling of brain size works well if relative brain size is the response variable, but not when relative brain size is a predictor variable. Perhaps counter-intuitively, the effect of brain size was still estimated correctly by all models. It was the body size effect that was biased in

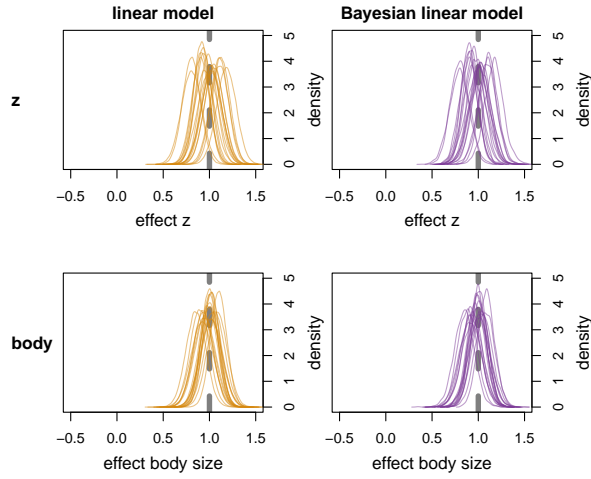


Figure 1: Parameter estimates from the linear model and Bayesian linear model with brain size as response variable. Dashed grey line is the true value. Orange density plots are normal distributions based on the mean and SE from the linear model. Purple and green density plots are the posterior distributions for the Bayesian models.

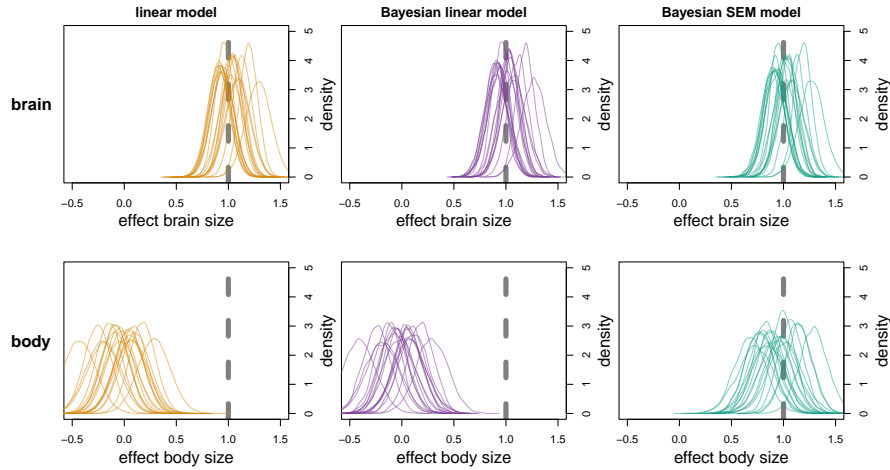


Figure 2: Parameter estimates from the linear model, Bayesian linear model and Bayesian structural equation model (SEM) with relative brain size as predictor variable. Dashed grey line is the true value. Orange density plots are normal distributions based on the mean and SE from the linear model. Purple and green density plots are the posterior distributions for the Bayesian models.

the linear models. The use of such models would lead to the incorrect inference that there was no effect of body size at all.

One way to create some intuition about what is going on is that absolute brain size (which was the actual predictor variable included in the linear models) contains information about both body size and relative brain size. In a sense this variable controls for the body size effect already, leaving the direct effect of body size 0. In a case where body size itself does not have an effect on z , it does not actually need to be included at all. Using absolute brain size would be fine. The variation in brain size due to allometric scaling would just create noise. Using absolute brain size would of course lead to a less precise estimate of the effect of brain size, so using relative brain size from a structural equation model would still be preferable. By presenting a simple version of such a model, I hope to encourage future work to adopt this practice to ensure that all parameter estimates, and the inference drawn based on these, are statistically sound.

Data accessibility. All data is generated in the simulation. Code is publicly available at: https://github.com/simeonqs/On_the_use_of_relative_brain_size.

Competing interests. I declare I have no competing interests.

Funding. I received funding from the International Max Planck Research School for Quantitative Behaviour, Ecology and Evolution.

Acknowledgements. XXX

References

- Armstrong, Este. 1983. “Relative Brain Size and Metabolism in Mammals.” *Science* 220 (4603): 1302–4.
- Bowen, Natasha K, and Shenyang Guo. 2011. *Structural Equation Modeling*. Oxford University Press.
- Dunbar, Robin IM, and Susanne Shultz. 2007. “Evolution in the Social Brain.” *Science* 317 (5843): 1344–47.
- Freckleton, Robert P. 2002. “On the Misuse of Residuals in Ecology: Regression of Residuals Vs. Multiple Regression.” *Journal of Animal Ecology*, 542–45.
- Gabry, Jonah, and Rok Češnovar. 2021. *Cmdstanr: R Interface to 'CmdStan'*.
- Gelman, Andrew, Daniel Lee, and Jiqiang Guo. 2015. “Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization.” *Journal of Educational and Behavioral Statistics* 40 (5): 530–43.
- González-Lagos, C, Daniel Sol, and SM Reader. 2010. “Large-Brained Mammals Live Longer.” *Journal of Evolutionary Biology* 23 (5): 1064–74.
- Healy, Susan D, and Candy Rowe. 2007. “A Critique of Comparative Studies of Brain Size.” *Proceedings of the Royal Society B: Biological Sciences* 274 (1609): 453–64.
- Isler, Karin, and Carel P Van Schaik. 2009. “Why Are There so Few Smart Mammals (but so Many Smart Birds)?” *Biology Letters* 5 (1): 125–29.
- Jerison, Harry J. 1973. “Evolution of the Brain and Intelligence.”
- Lefebvre, Louis, Simon M Reader, and Daniel Sol. 2004. “Brains, Innovations and Evolution in Birds and Primates.” *Brain, Behavior and Evolution* 63 (4): 233–46.
- Minias, Piotr, and Patrycja Podlaszczuk. 2017. “Longevity Is Associated with Relative Brain Size in Birds.” *Ecology and Evolution* 7 (10): 3558–66.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Street, Sally E, Ana F Navarrete, Simon M Reader, and Kevin N Laland. 2017. “Coevolution of Cultural Intelligence, Extended Life History, Sociality, and Brain Size in Primates.” *Proceedings of the National Academy of Sciences* 114 (30): 7908–14.