Dosing Donkeys: Replies to Comments

Comments are highlighted in yellow and our responses are highlighted in green to make it easier to read.

**Reviewer 1:**

Comment: “Yes, there was a very clear goal of providing a model that had better interpretability and accounted for proportionate adjustments. The conclusion provides two situations in which each of the two proposed models may be used. My only criticism would be to state your overall goals in the introduction rather than at the end of section 2 which discusses previous work to make it clearer/more digestible for someone that may be skimming your work. “

Response: Thanks for the helpful suggestion! We previously had a sentence discussing our goals in the introduction, but your comment made us realize it wasn’t super clear/explicit. We’ve changed the structure of the introduction a bit to incorporate the goals in a more concise way.

Comment: “Yes, plenty of equations along with their rationale and specific tests are provided to give the reader a clear understanding of the exact steps that the authors took. The subsections are clearly labeled and follow a logical order that guides the reader through a natural flow of the analysis.”

Response: Thank you!

Comment: “Yes, the two models discussed in the conclusion are given a discussion about which may be more appropriate to use in a situation. The authors also aimed to provide better interpretability with comparable performance to the Milner and Rougier model and discussed the relative error of their models with a table of Beta values with their confidence intervals.”

Response: Thank you! We are glad that made sense!

Comment: “All of the tables help illustrate important points that the authors discuss in the paper and validate results or confirm findings. Though the authors are limited by the page length, I would suggest adding a plot with a graphical representation of the author’s loss function overlaid so that readers can compare it with what was presented in Milner and Rougier. I would also argue for removing Figure 1 which shows the log transformations; this was a preliminary check and was not necessary to include; the Appendix may have been a more appropriate place. I would have also liked captions on the plots; the tables don’t really need captions but for people that are skimming the paper, captions on plots would bring out the main takeaways from those diagrams.”

Response: Thank you for the feedback! We did not overlay our loss function with the paper’s since their loss function is on relative error whereas ours is on the log scale. We also decided to keep Figure 1 because we felt that it would be nice for the reader to visually confirm that our model was promising and that it didn’t violate assumptions. We will attempt to add captions to plots if space allows!

Comment: “Yes, there was a substantial discussion of how the authors validated their work. They used cross validation on their final model to evaluate performance. They also discussed appropriateness of a linear model and provided a plot to support their claims. They also discussed normality of residuals and constant variance assumptions and conducted a Breusch-Pagan test to confirm whether their model violated any assumptions. They’ve also conducted sensitivity analysis for height. Though the authors have discussed how sensitive the predictions can be for changes in beta, I would have liked to see a graphical interpretation of this or some sort of table illustrating this.”

Response: We did not end up displaying this graphically, but we did incorporate further discussion based off of this comment in the section where we interpret our Beta values! We did include graphs for the effects of deviations to Length, Girth, Height, or a misclassification.

Comment: “There were quite a few functions that I didn’t automatically have and had to figure out what packages to install (pracmath, olsrr, lmtest). I also had considerable trouble using the paranomo package that Milner and Rougier provided; I would exclude that package in the code and load in the data differently because it took a bit of time to figure out that I wasn’t going to get it to work.”

Response: These are good things to note- we fixed our code to use the donkeys.Rda file as opposed to the paranomo package. We added a comment in the code chunk where we load our libraries that one has to download the libraries listed.

Comment: “Yes, the flow of the paper is very easy to follow and there are sufficient subheaders that it makes it easy to understand. The paper is generally free of typos and grammatical errors and easy to read.”

Response: This is good feedback for us and good to hear!

Comment: “The authors discussed how this model provides better results than Mliner and Rougier and has several disadvantages including interpretability of coefficients, interpretability of intercept, and applies proportional adjustments for BCS and Age. They discussed their limitations of not knowing the exact loss function used in Milner and Rougier and how predictions could be improved if this were known. I would have liked more discussion of limitations of their approach in the field (Shiny app vs. nomogram).”

Response: Thanks for this feedback. We have now hosted our Shiny App so that it is accessible on mobile devices. We added a discussion addressing the convenience and simplicity of our app in the last paragraph.

Comment: “Would it be possible to remove the colored boxes used as links to the figures? I found this to be a bit distracting and it may be better to underline in a different color instead.”

Response: We couldn’t find a way to do so and the links might make it easier for a reader to refer to the tables and plots. But if it is distracting one can open the pdf on a web browser (Chrome works for us).

Comment: “Did you consider recoding the age variable like Milner and Rougier to reduce the number of levels?”

Response: This is a great point and something we haven’t really considered in our first draft. We ended up recoding the Age variable and binning all donkeys above the age of 5 together.

Comment: “Why did you choose to define relative error differently?”

Response: We felt that predicted/actual makes more sense than actual/predicted. Predicted/actual gives you an idea of how well your model perform proportionally, and conventionally used. Actual/predicted is the flipped interpretation and is less intuitive.

**Reviewer 2:**

Comment: “The case study is pretty clear in articulating the research goals. They could write a little more on the background motivating the problem. I think they could have also highlighted that comparison to Model 1 (the Milner and Rougier model) was also one of their goals, but their overall focus of optimizing and validating a model is stated and evaluated clearly. If I’m being nitpicky then a sentence clearly stating that the goal of the case study is “…” would have really solidified the objective overall.”

Response: Thanks for this comment! We added more to the introduction and restated our sentence to make it more clear that we would be discussing Model 1. We also reformatted our sentence to make our goals more clear.

Comment:

Overall, the analyses match the research questions and available data. Through the analyses, they identify important variables in estimating donkey weight, validate the model, and generate a tool (the Shiny App) for prediction of donkey weight. As I was reading through the paper, I had the following comments on the analyses:

Pros:

1. Do a good job of explaining issues with the previous model and motivating certain choices for the formulation of a new model (eg., proportionate adjustments for AGE and BCS)
2. Good comparisons to Model 1 (the MR model) throughout the paper
3. The Approach does a good job explaining how exactly you will validate the identified model
4. Effective preliminary checks to see if log transformation of variables affects the linearity relationship between variables

Cons:

1. What does applying the model mean? The preliminary checks section could do a better job to explain how this is relevant to the model being promising.
2. Having an output giving the MSE and mean relative squared error in section 4.3 (or how you calculated it) could be potentially useful Also is an MSE of 87.5 considered high? The report could do a better job of explaining the relevance of these statistics.
3. What do model diagnostics do? Including a sentence to explain the relevance of model diagnostics to validate your model in the results section could be useful (Edit: I see that there is a section in the discussion, but I still think adding a sentence in the result/more information in the caption could be clearer)
4. I think interpretation of beta values could do better in actually explaining what the beta coefficients mean to vets in the field since this case study is meant to be accessible to them

Response:

1. Thank you for such a detailed response! We realize “applying” the model may have been a big unclear and have clarified what we mean in that section. We added additional discussion in the preliminary checks section, including discussions about binning through looking at VIF to check multicollinearity.
2. We’ve included definitions for MSE and MRSE to clarify this in our final draft. We also discussed the relevance of these statistics in the context of donkey weight and relative error.
3. We ended up deleting our model diagnostics graphs due to space constraints, so this comment is no longer relevant.
4. We’re glad you brought this up- we have rewritten this section to address this!

Comment: “I think most of the methods are described in enough detail, but some parts of the methods section are lacking in clarity. Just reading section 4.5, I don’t fully understand what they did to generate table 2. Additionally, It would have been helpful if they had shown the equation/visualized the asymmetric loss function in some way. It is clear what approach and model were used to evaluate the hypothesis. I think including and Approach section gave a useful roadmap for how they were approaching their analyses.”

Response: We have modified our Section 4, so hopefully our methods are more clear now! We have added a visualization of the asymmetric loss function- thanks for bringing that up!

Comment: “I think the report includes a correct interpretation of results provided. Regarding effectiveness, I had some questions about how they defined interpretability (included in 2 and 10). My main clarity issues were in section 5.1. Overall, they did a really good job presenting and explaining their findings. Sensitivity analyses helped substantiate their conclusions. They included confidence intervals and p-values for their coefficients which was helpful.”

Response: Addressed in above response, but we have modified our sections discussing interpretability. Glad to hear our tables and sensitivity analysis was helpful!

Comment: “All tables and figures are clear. I do not think any need to be eliminated/new ones are needed, but I do think that figure captions should be more informative and titles should be on graphs, to stay consistent with general scientific writing. Right now, the captions do not give much information on what we can actually discern from the figures.”

Response: We will try to add titles and captions if space allows! Thanks!

Comment: “The model is appropriately and thoroughly validated (through k-fold cross validation, predicted vs. actual plots, loss analysis, etc.) They went above and beyond to check model assumptions and confirm that they are valid (like using the Breusch-Pagan test). They conducted sensitivity analysis to check the model against other models to validate that their predictors were important. I do not think they needed any more validation.”

Response: Great!

Comment:

* + 1. Difficulties in loading the paranomo package for my version of R 4.0.0 (had to work around this issue by directly downloading donkeys dataset on my laptop and then loading the rda file into R)
       1. Could specify what version of R they used to figure out how to load the donkeys package
    2. Due to having to load the donkeys package differently from the authors, I had to change a few calls to variables (eg., Weight 🡪 donkeys$Weight)
    3. Work was pretty reproducible but it would have been nice if the sections in Sensitivity Analysis and Final Betas had comments in it

Response:

i: We have fixed this problem- paranomo package is no longer needed. ii: One should be able to run given code now. iii: We’ve added comments- thanks!

Comment: “Yes. the paper is professionally presented and executed very well One nitpicky comment would be to center their title for Table 2.”

Response: This is now Table 3 in our report. The “title” was actually a column title, so it wasn’t supposed to be centered. We realize what we previously had was a big visually unappealing, so we’ve modified it a bit to balance the table.

Comment: “The strengths of the analysis are that it does as effective job of proving why their model is better than the proposed model and validating it. They clearly present the strengths of ther model: prediction model is more interpretable, has proportional adjustments rather than additive adjustments to match the cylindrical criterion, and has better performance. I do think they can do a more thorough job discussing either what they mean by interpretability, or making it clearer for readers (in comparison with Model 1). They do a good job of also acknowledging the limitations. They identify that their model is more sensitive to beta changes and measurement error; they note that it could run into problems practically due to how age is assessed and how data is gathered (room for error). I think their analysis could have also incorporated an asymmetric loss function (as in the paper) for more thorough loss comparison to Model 1.”

Response: We have modified our interpretation sections so hopefully it is more clear what we mean now. We incorporated a loss function, but it may not have been super clear- we address this by adding a new graph to display the asymmetric loss function we use.

Comment:

* Why is predicted/actual more accurate than actual/predicted?
* Why did you choose different baseline categories for your model, as compared to model 1 (the Milner-Rougier model?)
* Why weren’t over predictions considered more harmful in your loss functions? Doesn’t this deviate from the Milner-Rougier paper where overdosing was penalized more than underdosing?
* Why is the idealized cylindrical model where beta(length)=1 and beta(girth)=2?
* Why exactly is your proposed model more interpretable? Could you explain more why this model is more accessible to vets in the field?

Response:

* We felt that Predicted/Actual would be more interpretable since it’s the relative error to the true value.
* Baseline categories don’t affect predictions, it just determines what the adjustment is based on. The paper fits to BCS = 3 and Age > 5 years, but we fit on all data so we are free to choose our baseline category . We will address binning our BCS and Age categories.
* We addressed this, though it may not have been too clear, so we’ve added more to this discussion. We use the loss function in our report for the sake of simplicity and using ordinary least squares. However, we later discuss how our model performs under an asymmetric loss, and also train our model on asymmetric loss.
* We can discuss this more.
* We modified our interpretation of Beta values to address this.

**Reviewer 3:**

Comment: “The analyses match the question and data. The project aims to generate an interpretable model yielding good predictions based on the original study, and the authors did a good job in identifying in what way they could improve the model and proving that the model was indeed more effective.”

Response: Great!

Comment: The authors provided a clear description in why they proposed their Model 2 and how they proceeded to their final model 3. The steps were explained with enough details and all hypotheses were checked with corresponding techniques. However, I was a little confused by the section 3.3 loss function and how it is integrated in the model selection process. In section 3.7, the authors mentioned comparing model 3 with model 1 “across a variety of loss functions,” and I would love to see a more detailed explanation on this phrase.

Response: Thank you for the comments! We decided to proceed with our explanation of the loss function since we felt that it was decently clear, and we add that we’re performing ordinary least squares regression. Section 3.7 is more of a road map, and the variety of loss functions is explained later.

Comment: “The authors provided an effective interpretation of the final model and its coefficients. The conclusion that their model works better than the original model was also substantiated with evidence from different tests.”

Response: Great! Thanks!

Comment: “The majority of the tables and figures are clear and informative. However, for Figure 1, which plots the collinearity between the three numerical log transformed variables, I did not see enough elaboration on that and why the authors kept the log transformation if it does not help with improving the model’s assumptions.”

Response: We kept the log transformation due to the assumption of cylindrical shape. We have modified our explanation in our report to make it more clear!

Comment: “Residual plots were used to check the modeling assumptions, and they are valid for the model. Sensitivity Analysis was used to confirm the significance of the variable Height, and the results are reasonable. The current analyses are sufficient, and no additional ones need to be conducted.”

Response: Thanks!

Comment: “The only challenge was that my R version is 4.0.0, which does not support the package “paranomo.” I just deleted the line library(paranomo) and the rmd file knitted smoothly without any error.”

Response: We have changed our code so that it doesn’t require paranomo anymore. Instead, we load in the Rda file and the code given should run fine. Thanks!

Comment: “Overall, the paper is professionally presented without distracting errors. One tiny error I spotted was on the bottom of page 1, where Length should be squared.”

Response: Thanks! It was intentional for Length to not be squared.

Comment: “The strengths of the analysis include providing a model that is easier to interpret and checking the model’s performance and validity through variety of methods. The limitations include the proposed model being more sensitive to changes in β and errors in measurements, as well as the multicollinearity in the variables. All are clearly stated by the authors.”

Response: We didn’t discuss multicollinearity in our first report, but we’ve incorporated that into our new report!

Comment: “Do you think removing the influential points will improve the model’s performance?”

Response: Our model diagnostics indicate there are no influential points. We also removed outliers! Thanks!

**Amy/Raphael:**

**(Comments on Report)**

Comment: “Don’t start sentence with reference #”

Response: Fixed! Thank you!

Comment: “Age + BCS are not in this model”

Response: We have modified equation (1) to include Age and BCS. Thank you!

Comment: “Be careful about mixing up parameters and estimates”

Response: We have fixed the marked Betas and checked the Betas in the other parts of the report to make sure we didn’t make that mistake.

Comment: “They evaluated each loss function separately.”

Response: We removed the sentence in our previous report that states they sum the loss function and discuss it in a later paragraph. In the paper, they use separate loss functions to obtain their lambda value of 0.5 to determine the transformation of the Weight variable. However, we were unsure of what their loss function was later on when they optimize for one set of coefficients, so we assume that they sum their separate loss functions to obtain those values. We added discussion on this approach and show that the paper’s model actually has a lower minimum relative error, giving us confidence that our loss function is taking the errors for overdose seriously, if not even more seriously. So even if the Milner and Rougier paper did not sum the two loss functions, training a model with our proposed asymmetric loss should not pose much danger for donkeys.

Comment: “Note that there no 1 cm donkey! This is extrapolating beyond range of data so be careful”

Response: We were too caught up in arguing that our model provides better estimates as length and girth tend to 0, but we realize now that predicted weight is only really meaningful in the range of the data. We’ve taken out the interpretation with a 1 cm donkey. It is interesting to note though that our model predicts better for a baby donkey! We have left this detail out from the report due to space constraints.

Comment: “Did you consider any other forms, e.g. polynomials on the log scale? Sensitivity analysis to this specification?

Response: This is a great suggestion, and we have incorporated this into our final report. This is in Section 4.4 and the results are summarized in Table 3.

Comment: Circled typos (Beta\_10 and the Betas that look like B’s)

Response: Fixed!

Comment: error\_i? distribution?

Response: We have added the error term to our model and show that it is normally distributed.

Comment: fix labels in Table 1 (add referent)

Response: Fixed! Thanks!

Comment: “If you sum, you would kill a donkey with anesthesia if you worm it correctly- best to keep separate when one consequence is severe.”

Response: Addressed in an above response that discusses the loss function!

Comment: “How do you predictions compare with nomograms and the truth? Can you show error of your prediction vs. error of nomogram in the same figure?

Response: We didn’t put this in the same figure, but we’ve incorporated side by side histograms (Figure 3) in our final report to address this.

Comment: “Be precise. What is 1 unit? Cm? Be more specific. OK to interpret on log scale. Need a more comprehensive interpretation of the model.

Response: We were looking at an increase in Betas here, so we didn’t include a unit since Beta is unitless. We realized that this may have been slightly confusing, so we went back and added in a change in the variables (with units), then went on to discusses the impact that Betas could have on the predicted weight.

Comment: “I love your app! You may wish to add range checks so that if you input a value too far outside the range of the data (e.g. 1.4 cm), you are prompted to re-enter.

Response: This was a great suggestion- we looked into this, and R Shiny currently does not enforce bounds on numerical inputs (it allows you to specify the min and max that you want your input values to be in, but it doesn’t actually enforce it and it’s currently a git issue). We also looked into the validate function but it was giving us an error message no matter what every time. We instead decided to just print warning messages that specify a reasonable range for Length, Girth, and Height. We got the min and max for the range of each variable by looking at the minimum and maximum measurements in the dataset. However, the user is allowed to input any numbers they want.

Comment:

+) Shiny app; report is well edited; figures and tables are well edited; show good understanding of nomograms; models are very clearly described; model is well motivated

-) In section 5.1, intepretation of \beta\_k is incorrect (but almost correct); homoscedasticity does not seem to be violated (the variance of the errors seems constant over the range of the predictions)

Response: We have fixed our interpretation of \beta\_k . We have fixed the homoscedasticity discussion!

Comment: “Tips: use slides for the presentation instead of scrolling up and down the report; do not used arrows to denote vectors, simply writing "let $\beta = (\beta\_1, \beta\_2)$ be the vector of coefficients where $\beta\_1$ corresponds to the intercept and $\beta\_2$ to the coefficient of the variable Length" is sufficient for us toknow that beta is a vector; MRSE instead of MSRE; did you consider minimizing the asymmetric loss function instead of the symmetric squared loss? It would yield a more performant model wrt the asymmetric loss”

Response: Thanks for the comments! Yes, we have re-recorded our video with slides. We took out the arrows in the vectors. We have changed MSRE to MRSE.

We use a symmetric loss function for the sake of simplicity and using the ordinary least squares regression, which gives us standard errors useful for interpretation. However, we later perform tuning to find an asymmetric loss function that is close to the papers’ asymmetric loss function. We then show that our model performs better than the Milner and Rougier model given the asymmetric loss. We also additionally trained the model with this asymmetric loss and a reader could instead use this Asymmetric Model. We did not propose this as the final model because as expected, this skewed predictions more toward under-predictions, which might be undesirable without more information on the level of asymmetry.