Estrogen Bioassay

Akshay Punwatkar, Andrew Patterson, Derek Wales, Xinwen Li,Tzu-Chun Hsieh

November 4, 2019

Summary

Analysis of a bioassay examining a possible estrogen agonist and antagonist. The agonist, Ethinylestradiol or EE, was confirmed in this analysis to have estrogen-like effects on the weight of the rat uterus. Additionally, ZM, a potential estrogen antagonist, was shown to possibly reduce the estrogenic effects of EE. However, without proper controls in the study, ZM may also operate on another mechanism of action than the one proposed here.

Introduction

Estrogen is an important hormone in mammals which controls numerous primary and secondary sex characteristics in the organism. Estrogen agonists and antagonists are classes of compounds which act on the estrogen receptor in place of estrogen to either activate or inhibit the receptor. In this study, several experiments were carried out to determine if the potential estrogen agonist Ethinylestradiol, or EE, had similar effects to estrogen on estrogen-free female rat uteruses. Additionally, the effects of a potential estrogen antagonist, ZM, were also examined in the study. Data from this project was then used to construct a model to determine if EE and ZM had estrogen agonistic or antagonistic effects on the rat uterus.

Data

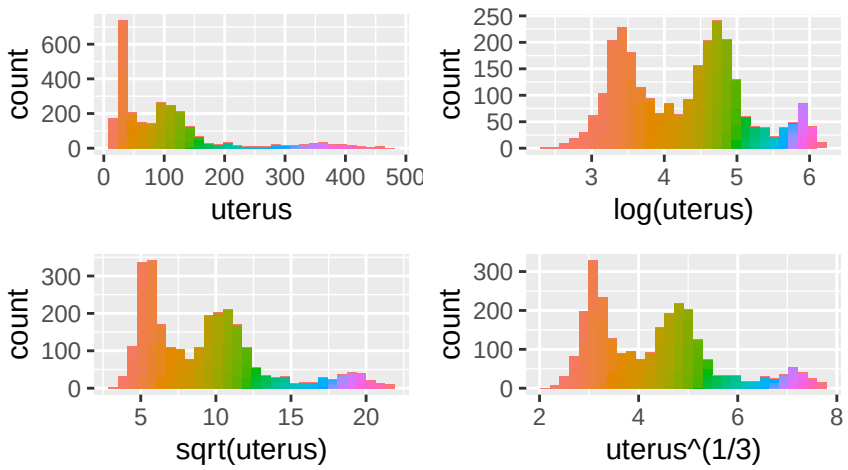
Data used in this analysis contained 2681 observations from different labs conducting research on whether the estrogen level would affect the uterus weights of rats. A cursory investigation of the data revealed there were 4 rows missing uterus weight values and 2 of rows missing weight values. Since these missing rows occurred randomly, which means these rows are in different groups, protocol types, or labs. Therefore, these rows were deleted.

For the variables in the data set, uterus, weight, EE, ZM were treated as numeric variables and protocal, lab, group were treated as categorical variables. In the research, there are only 3 kinds of dosage of ZM and 7 kinds of dosage of EE, however they were still treated as numerical variables because if treated as categorical variables, information would be lost between different dosages. A different dosage isn't just different from another dosage. For example, a 10 mg dose is 10 times a 1 mg dose. This information would be lost in a categorical variable. Another variable was added to the dataset, a new binary variable- mature, based on the value of protocol, to indicate whether the rats were mature. If the rat was categorized as protocol A or B, it would have value 0, and if the rat was categorized as protocol C or D, it would have value 1.

EDA

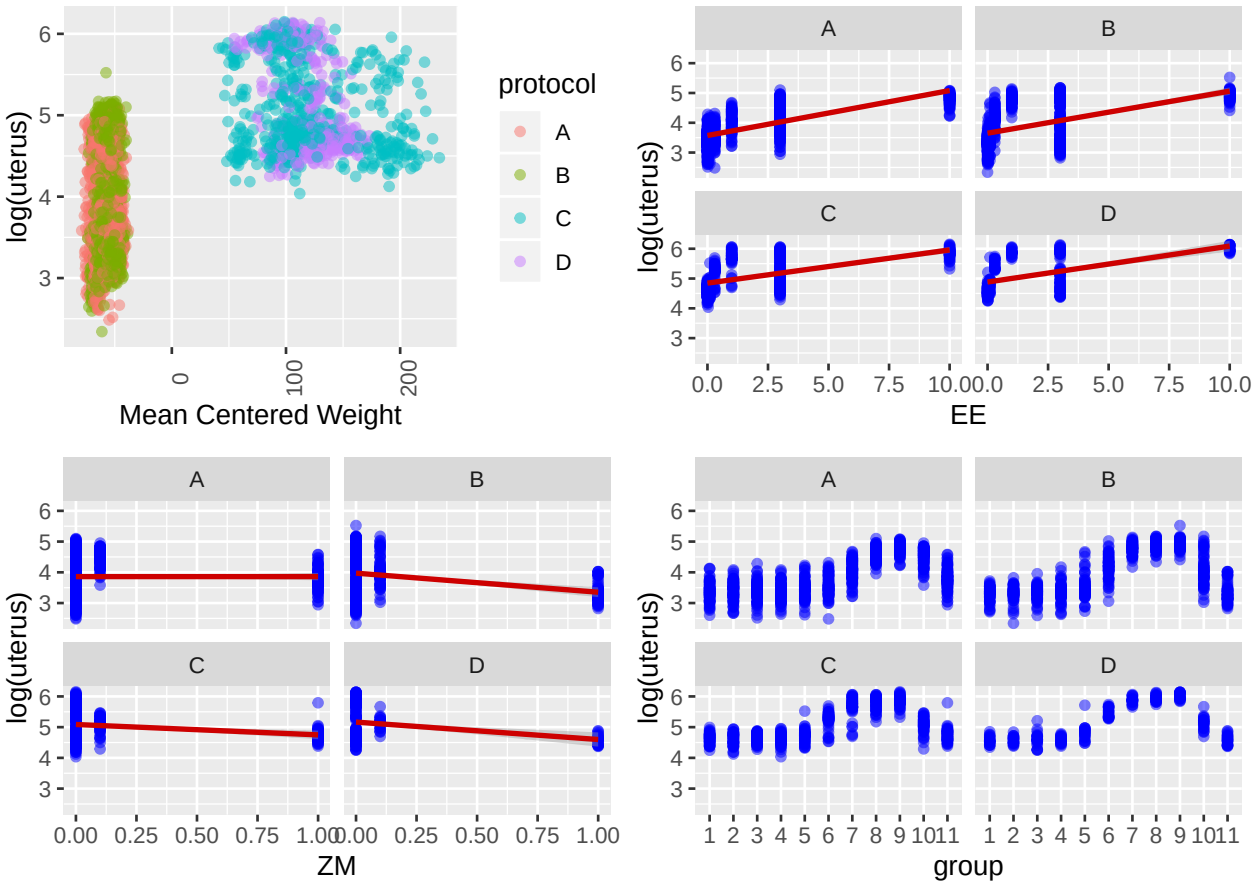
For the data set used, collinearity was investigated as the data was explored. Mature versus Protocol: Since mature is the new variable created based on protocol types, there is high collinearity between them. ZM and EE versus Group: Since groups is seperated depends on the different combination of dosage of ZM and EE, they are highly correlated with each other.

First, a histogram was plotted of the response variable- uterus weight. However, the distribution is skewed and is not a normal distribution. Out of the transformations square root, cube root, and log transformation, it seems that log transformation improves the distribution the most. As a result, even though it is still hard to say that the distribution of log uterus weight is a normal distribution, in the following analysis, log uterus weight was used as the response variable.



Moreover, plots were used to check the relatonship between log uterus weight with other variables. First, the difference of log uterus weight between labs was examined. By the distribution of log uterus weight of each lab, there appeared to be some differences. However, the data is further separated by different protocol types, in each protocol type group, each lab has a similar distribution of log uterus weight. This difference observed is caused by the fact that not all labs conduct experiment for every protocol type. Moreover, an apparent pattern seemed to indicate that the data points were clustered by group when plotting for log uterus weight and mean centered weight. There are four clusters in the plot. For the relation between log uterus rate and protocol, rates categorized as protocols C or D apparently have higher log uterus weight comparing to protocols A or B.

All these observations indicate that different protocol types would have different log uterus distributions. Therefore, a plot was made to examine the relationship between log uterus weight with each variable by different protocol types. This indicated that when comparing log uterus weight and mean centered weight, the mature rats have a negative pattern, while the immature rats do not appear to have a pattern. For the different dosages of ZM and EE, there appears to be a positive relationship between log uterus weight and ZM and a negative relationship between log uterus weight and EE. Lastly, it appeared that different groups of rats had different distribution of log uterus weights. However, since grouping is based on the dosage of ZM and EE and the experiment is concerned with studying the effect of ZM and EE, only ZM and EE will be included in the final model and the group variable will be excluded.



Model Selection Process

Based on the EDA, including lab and protocols as level predictors, EE, ZM, mean centered weight as numeric predictors, the first model constructed:

$$Log(Uterus) = (B_0 + \gamma_{lab} + \eta_{protocal}) + B_1 EE + B_2 ZM + B_3 MWeight + \varepsilon_{ijk}$$

The AIC for this model was 3314, and BIC was 3356. It was expected that both scores would be smaller for the final model. Mean centered weight was not very significant with a -0.72 t-value but the protocal variable does contain information of the weight of the rat and weight cannot help distinguish the difference between mature and inmature rats here. Instead of including the weight, another binary variable was created indicating either mature or immature rats. Because protocols A and B were done on immature rats, and protocls C and D were done on mature rats, the mature variable could help identify this difference between the two sets of protocols. Replacing weight with mature, The mature variable had a t-value of 40.33, and therefore seems to be a very significant predictor of uterus weight. However the normality issue was still not solved. There does not appear to be a multicollinearity issue in the model by checking the VIF scores, but the normality assumption seems to be violated from the qq-plot. Ignoring normality at this time, random slopes based on the protocols were created to test the sensitivity of different protocols detecting the effects ZM or EE. However, the model had a convergence issue. Due to this issue, protocol was instead included as a normal predictor instead of a level predictor. Interaction terms between protocol and EE and protocol and ZM were added to see the sensivity of the protocols in detecting ZM or EE as compared to the base protocol A. Weight was then added back into the model in hopes of distinguishing between the mature vs immature rats. However, mean centered weight was not very significant in the model and also caused a multicollinearity issue between weight and protocol. Thus, Weight was not included in the final model. The group variable was also added into the model as an extra predictor variable in the model once. This appeared to fix the normality and equal variance issues. However, by adding the group variable, the model loses information on

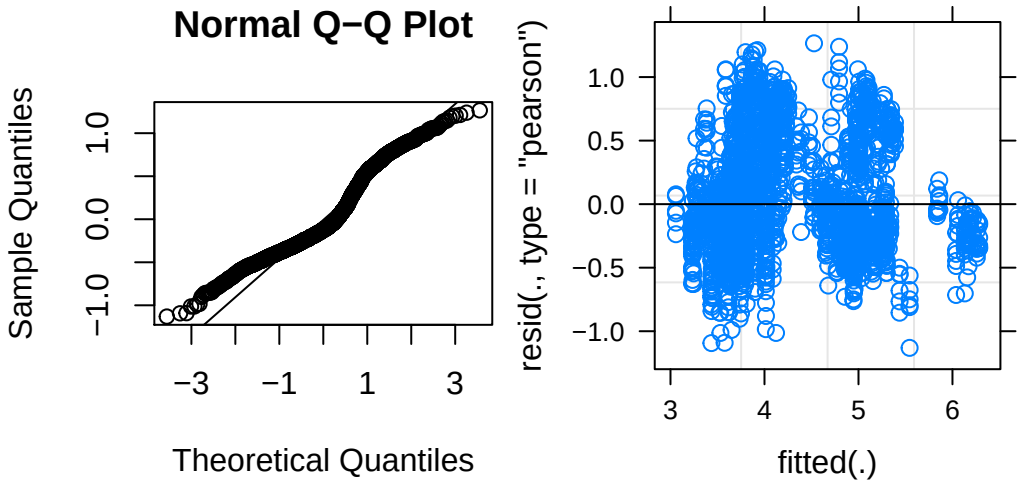
EE and ZM and the interpretation would not make sense to say different groups of rats can have different impacts. Adding group isn't a great idea here.

Model

The final model:

$$Log(Uterus) = (B_0 + \gamma_{lab}) + B_1 EE + B_2 ZM + B_3 Protocal + B_4 Protocol * EE + B_5 Protocol * ZM + \varepsilon_{ijk}$$

Model assumptions were examined by plotting residual vs fitted and qq-plot. As mentioned in the model selection process, the model does not satisify the normality assumption. There appears to be clustering in the residual plot, because the model cannot distinguish mature vs inmature rats, while Equal variance also appear to be violated.



In addition, multicollinearity was examined, and the VIF scores for each predictor in our final model were all between 2.6 to 5.6, indicating some multicollinearity but not enough to cause an issue for our interpretation. This is the best model considering all the questions asked of this data. The AIC score for this model is 3228 and BIC is 3310, which is lower than the first model above.

Table 1: Data Dictionary

Variable	Estimates (exp)	t-value
Intercept	35.202	79.785
EE	1.166	32.754
ZM	0.815	-4.366
Protocol B	1.123	4.255
Protocol C	3.878	44.457
Protocol D	3.887	34.308
EE: Protocol B	0.999	-0.126
EE: Protocol C	1.036	-4.545
EE: Protocol D	0.977	-2.186
ZM: Protocol B	0.531	-8.835
ZM: Protocol C	0.741	-3.830
ZM: Protocol D	0.580	-5.200

Results

There are three independent variables, a single hierarchical variable, and two interaction terms included in the model which are significant at predicting the weight of the rat uterus.

EE is a significant predictor of the weight of the uterus; as one unit of EE increases, the uterus weight will increase by a multiplicative effect of 1.17.The absolute t-value for this effect is 32.74, indicating this effect is significant at predicting uterus weight. This indicates that treating an estrogen free mouse with EE results in an increase in uterus weight.

For the independent variable ZM, as one unit of ZM increases, the uterus weight will decrease by a multiplicative effect of 0.81. The absolute t-value of this effect is 4.38, which is lower than EE but still significant at predicting the weight of the uterus. This indicates that the estrogen antagonist does have a negative effect on uterus weight. However, without proper controls as discussed in the conclusion, this study cannot conclusively say that this result is due to an estrogen antagonist effect.

For the random effects, each lab contained a different random intercept. The highest and lowest outliers were Chungkor and Poulenc labs, respectively at 0.337 and -0.330 intercepts.

The four protocols each had a significant effect on the weight of the uterus, with protocols C and D having an increased effect of as compared to A and B. Using protocol A as the baseline, protocol B had an multiplicative effect of on uterus weight of the rats by 1.12 with a t-value at 4.26. Protocols C and D had much higher effects, at 4.6 and 4.6 at t-values 44.46 and 34.31, respectively. This indicates protocols C and D had very significant effects on log uterus weight as compared to protocol A.

For the interaction terms, the interactions between the protocols and ZM and EE were significant. Of particular interest is the difference in the interaction term between ZM and protocol B versus the interaction term between EE and protocol B. The difference between these interaction terms, at 0.63, is larger than the other protocol interaction terms, indicating protocol B is the most sensitive protocol for determining the difference in effects between ZM and EE.

Conclusion

From these results, this study can conclude that EE has a positive effect on uterus weight while ZM has a negative effect on uterus weight. The different laboratories all have an effect on uterus weight, however this does not interfere with the overall conclusions of the study. Protocol B is the most sensitive protocol for detecting EE and ZM effects.

For further research, the point of this study was to see the effects of an estrogen agonist, EE, on uterus weight and to see the effects of a uterus antagonist, ZM, on uterus weight. From this study alone, ZM can be said to have an estrogen antagonistic effect. However, there were no controls for ZM as compared to EE. Without including a control ZM without EE, the actual mechanics of ZM on uterus weight cannot be determined. A further control group with ZM and without EE could determine if ZM is actually an estrogen antagonist, as expected, or is instead working by another mechanism on the uterus, such as a testosterone analog. The expectation of the estrogen antagonist is that, in the absence of all estrogen or estrogen analogs, there would be no decrease in uterus weight- only in the presence of EE would ZM show a reduction in uterus weight. If ZM were acting by another mechanism, then it would have a negative effect on uterus weight even in the absence of EE. Additional followups could investigate why there were differences in the laboratories' intercepts. A possible way to improve is to fit a mixture model to this dataset, since there is still bimodal pattern in the histogram of log(uterus).