**Peer Review and Communication History**

**MS Title**: Poisson Regressions: A Little Fishy

**Author Names**: Ryan, William H., Evers, Ellen R. K., Moore, Don A.

**Submitted**: July 24, 2020

**Editor First Decision—Revise & Resubmit**

Nov 5, 2020

Dear William H. Ryan,

I am writing in regard to your manuscript “Poisson Regressions: A Little Fishy” (MS 1130596), which you submitted for consideration at Collabra: Psychology. The reviews that I obtained were provided by experts in areas related to your work, and I found their comments to be detailed, thoughtful, and constructive. Both reviewers thought that topic was important. However, these reviewers also identified some limitations of the work and raised some concerns that you would need to address before the paper could be published here. Therefore, I invite you to submit a revised version of this paper for further consideration.

I think the biggest issue is that the paper does not go far enough in explaining when poisson regressions should be used, what the underlying assumptions are, why these assumptions do not hold in many cases, and how this leads to the problematic results. As a result, readers might come away from the paper with an increased suspicion of this method, but I think it will be very hard for the to come away with a clear understanding of why the approach can be problematic or how to decide on the appropriate way to analyze data for which a poisson regression might be tempting. As Reviewer 2 notes, a clear overview of these issues could be useful, but as currently written, the paper is not effective in accomplishing this goal. Almost every section needs quite a bit of additional explanation and clarification. Reviewer 2 made a number of very good and very detailed suggestions about places where your discussion could be expanded and clarified, and I strongly agree that such a thorough overhaul would be needed for this paper to be publishable at Collabra: Psychology. So although I will allow for a revision, it will need to be a very thorough revision to address the issues that the reviewers raised.

I think that Reviewer 2 hit the main issues that need to be addressed so please pay careful attention to these, but I did also have some additional minor concerns and suggestions (described below). You should also address additional issues raised by Reviewer 1, who was more critical about the contribution that the paper can make.

You occasionally have incomplete comparisons that make the points you raise hard to follow. For instance, in the second paragraph on p. 3, you mention a “better” method and “higher” rates of false positive without making it clear what the comparison is. There are other examples in the paper.

That entire paragraph (the second on p. 3) could use a rewrite. The tense in the fourth sentence doesn’t seem appropriate, the “this” in the sentence is an unclear referent, and the transition between this clause and the next is jarring because you switch from general ideas to a summary of an investigation that you have not yet introduced.

The discussion of your literature search on p. 4 is confusing. You mention details that are not explained clearly enough. You mention that you found 9 papers that incorrectly used poisson, but you don’t say how many used it correctly, how you know that the use was incorrect, or how you determined that there was a time trend.

The footnote that accompanies this sentence is also confusing. You refer to a task and your selection of this task (the voodoo doll task) that you have not yet described to the reader and seems irrelevant to the sentence to which the footnote refers.

The next sentence mentions a specific false positive rate for a set of studies, but you have yet to explain how one can calculate false positive rates from the data that are available, so this sentence will be hard for readers to understand or evaluate.

In the last sentence in the first paragraph of p. 5, you mention a comment from an editor to a colleague. I’m not categorically opposed to providing examples like this, as they can reveal things that go on behind the scenes; but this example did not seem to make an important point.

Related to this, I think that extreme descriptors can sometimes be appropriate, there were a couple places (e.g., the “outlandish” on p. 6) that seemed hyperbolic given the issues you raised up to that point.

Your description of the preregistration is incomplete and confusing. From the description on p. 6, the reader gets the impression that you correctly predicted the direction of your effect, which seemed implausible to me, even in cases where false positive rates are inflated. Looking at the pre-registration document, I see that you did preregister both directional hypotheses, which makes more sense to me. So please add more details to the in-text description and be clear that you preregistered both directional hypotheses.

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Richard Lucas

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Overall, this paper is interesting in that I appreciate its main point (people make critical analyses errors in areas in which they are unfamiliar), and in taking both a simulated and experimental approach. However, this paper is very limited in scope and incremental contribution. Broadly speaking:

* there are already a large number of simulations comparing Poisson to overdispersed models
* plotting results for their practical significance often does not lead to entirely inconsistent findings
* when there are problems in the simulated data, it is often in areas that are not typically found in actual data
* authors might usefully obtain the original data from published papers to explore both practical and statistical significance
Again, I think the overall message is important, but it has been repeated elsewhere. Sincerely, I hope the authors will continue in the revision of this paper, because I they can clarify/accentuate the problem, as well as provide stronger guidance/steps for future authors to follow, to adopt appropriate analyses.
(Side note on a side point made by authors on p. 7: calories and grams aren’t count data, not appropriate for Poisson)

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

First, let me apologize for the delay submitting my review. Adapting to the Covid-19 crisis, online teaching, and the political situation in the US has slowed my review responsiveness. I’m deeply sorry.

In general, this is the beginnings of a useful pedagogical paper. It clearly shows that blindly using a Poisson model for count data is problematic. However, as a teaching tool, the paper can be much improved in terms (1) clearly describing the assumptions and model forms for both the Poisson model and proposed alternatives and (2) clearly describing the statistical analyses and results reported.

1. The paper doesn’t do a great job of explaining clearly what the assumptions of the different models discussed are or why these models make these assumptions. As the authors note, Poisson regression is often introduced to students in the simplistic form as “the model you use for count data”. Accordingly, to be an effective pedagogical piece, I think the paper should more carefully help readers to understand the workings and design of the various models.
a. For the Poisson model, the paper states that the model assumes the variance of the data equals its mean. It would be helpful to unpack why this assumption is made. A Poisson distribution is not just a distribution of any count variable, but specifically a variable that is the count of events that occur after a specific period of time from a process with a known constant rate (λ). The reason the mean and variance are equal is because the distribution is modeling this constant-rate type process. Making this clearer might help readers to see why this model is a poor assumption for many psychological problems.
b. For the negative binomial model, the paper doesn’t describe the assumptions at all or explain why these might be more reasonable. This is a problem because the negative binomial also makes assumptions about the relationship between the mean and the variance, though ones that are often more reasonable. Violating these can be problematic, as the paper notes on page 9. The sentence “Even negative binomial regression begins to yield the same false positive ratio as Poisson regression at very high variance to mean ratios” illustrates the issue, but the reader is not guided toward understanding it. At present, the paper doesn’t give a clear idea what the negative binomial model is and seems to mostly be suggesting that readers replace a blind use of the Poisson model with a blind use of a negative binomial model.
c. For the OLS model, it would similarly be useful to clearly state the assumptions of the OLS model and why these are often violated by count data. Relatedly, it would be useful to discuss cases when an OLS model might still be reasonable for count data (when the assumption violates are likely to be of minor consequence).
d. As the t-test is just an OLS model with a single categorical predictor, I suggest dropping this particular analysis entirely, perhaps retaining Footnote 7 making this point.
e. Several other alternative models, such as overdispersed/quasi-Poisson and zero-inflated models, are shown in the R code, but not described in the text. I suggest adding brief sections to the paper describing these methods, their assumptions, and when they might be appropriate.
f. Testing for overdispersion is illustrated in the R code but not discussed in text. In general, a clear tutorial on choosing between alternative model specifications and testing the chosen model’s assumptions would be very helpful.
2. The presentation of the example statistical results is not complete, and it’s not clear what some of the numbers are supposed to mean in the results section. For example, the first set of results includes a confidence interval, but no point estimate. What is the CI for? What effect size? If it is Cohen’s d, I suggest instead doing a raw mean difference or the slope coefficient for group from the Poisson regression model. Please also interpret the effect size, rather than merely stating that participants in one condition ate “more”—how many more? The omission of the effect size point estimates and interpretations is repeated for the transformed model results later in the paragraph.
a. To make the mean = variance assumption violation clearer, please report the variance of the response variable as well as the mean.
b. A table showing all of these results would also be much easier to digest than just a list of statistics in running text.
3. The paper spends a fair amount of space on the point that preregistering a bad analysis doesn’t lead to good inference. I’m not sure how useful a point this is. It isn’t really built up at all—is false claiming of preregistration as an arbiter of validity common in this literature or something? At present, it reads as though it were an out of left field pet peeve of the author, rather than a point that is clearly connected to the rest of the paper.

Minor points/typos:

1. In footnote 1, you have a small typo. You refer to Linear, Poisson, and Negative Binomial models as cases of the “general linear model”. It should be the “*generalized* linear model”.
2. On page 9, you refer to the “false positive ratio”. This should be the “false positive rate”.
3. Footnote 8 is virtually identical to Footnote 1. You can probably omit one of them.
4. The following are some general comments about R code style best practices to facilitate reproducibility and understandability for other users of your code

a. library() should be preferred over require() in most cases, such as in your example scripts. The major reason is that library() will error immediately if the package is not available, whereas require() doesn’t error until an exported object is called. This can make it harder for users to debug. See <https://yihui.org/en/2014/07/library-vs-require/>
b. You should generally avoid using setwd() at all in R scripts. This makes the script specific to a specific machine and requires users to always edit the script before they are able to use it themselves. This can be a big barrier to new users of R. Instead, a project/folder-oriented workflow is more reproducible, wherein scripts are written assuming that they are run in a fresh new R session where all file paths refer to directories within the directory holding the script. See <https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects> and <https://here.r-lib.org/> for discussions. For this paper, I suggest simply instructing the reader to import or create a data frame with their data, then include output showing what the data in the example look like.

I hope you find these comments helpful as you revise your paper.

Best,
Brenton Wiernik

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |

**Author Response**

**June 29, 2021**

Dear Dr. Lucas,

Please accept our apology for the delay between your decision letter and this resubmission. We are grateful for the time you and the Collabra reviewers have spent reading our manuscript, thinking about how to make it better, and providing us with helpful guidance. Thank you for the insightful and useful reviews – we have used them as the basis for an extensive revision covering almost every section of the paper.

We believe the fact that researchers are still misusing Poisson analyses underscores the importance of our paper. We seek to fill a gap in the current literature by providing guidance to the broad mainstream of researchers who are not reading the statistical literature on Poisson analyses. In our revision, we have endeavored to provide a clear and simple explanation of when to avoid Poisson analyses and what alternatives exist. So, while adding additional explanations we tried to keep our main text accessible to these readers, while providing an appendix with more detail. We hope this balance will help readers remember the simple lesson while also giving those who want to engage more deeply the opportunity and tools to do so.

We have reworked the paper to include substantially more explanations of the Poisson distribution and how its assumptions may be violated (including new illustrative figures), as well as more detailed explanations of Negative Binomial and Linear Regression and how to decide which alternative to use. We also have greatly expanded the Appendix. It now contains explanations of all the methods used in the tutorial code, including tests for overdispersion, additional alternative methods of analyzing count data such as zero-inflated Poisson, and discussions of how to choose between these methods, as suggested by Reviewer 2.

The changes we have made in response to all the suggestions and comments in the reviews are enumerated point-by-point on the addendum to this resubmission letter. Our responses are in lettered (A., B., C., etc.) paragraphs.

We are delighted to be able to accept your invitation to resubmit the revised manuscript now, to be considered for publication in *Collabra: Psychology*.

Sincerely,

William Ryan

Editor Feedback:

I am writing in regard to your manuscript "Poisson Regressions: A Little Fishy" (MS 1130596), which you submitted for consideration at Collabra: Psychology. The reviews that I obtained were provided by experts in areas related to your work, and I found their comments to be detailed, thoughtful, and constructive. Both reviewers thought that topic was important. However, these reviewers also identified some limitations of the work and raised some concerns that you would need to address before the paper could be published here. Therefore, I invite you to submit a revised version of this paper for further consideration.

I think the biggest issue is that the paper does not go far enough in explaining when poisson regressions should be used, what the underlying assumptions are, why these assumptions do not hold in many cases, and how this leads to the problematic results. As a result, readers might come away from the paper with an increased suspicion of this method, but I think it will be very hard for them to come away with a clear understanding of why the approach can be problematic or how to decide on the appropriate way to analyze data for which a poisson regression might be tempting. As Reviewer 2 notes, a clear overview of these issues could be useful, but as currently written, the paper is not effective in accomplishing this goal. Almost every section needs quite a bit of additional explanation and clarification. Reviewer 2 made a number of very good and very detailed suggestions about places where your discussion could be expanded and clarified, and I strongly agree that such a thorough overhaul would be needed for this paper to be publishable at Collabra: Psychology. So although I will allow for a revision, it will need to be a very thorough revision to address the issues that the reviewers raised.

I think that Reviewer 2 hit the main issues that need to be addressed so please pay careful attention to these, but I did also have some additional minor concerns and suggestions (described below). You should also address additional issues raised by Reviewer 1, who was more critical about the contribution that the paper can make.

You occasionally have incomplete comparisons that make the points you raise hard to follow. For instance, in the second paragraph on p. 3, you mention a "better" method and "higher" rates of false positive without making it clear what the comparison is. There are other examples in the paper.

1. Thank you for this clear guidance. We have revised this passage to clarify the comparisons as you suggest.

That entire paragraph (the second on p. 3) could use a rewrite. The tense in the fourth sentence doesn't seem appropriate, the "this" in the sentence is an unclear referent, and the transition between this clause and the next is jarring because you switch from general ideas to a summary of an investigation that you have not yet introduced.

1. We have done our best to rewrite the paragraph, simplifying and clarifying as you suggest.

The discussion of your literature search on p. 4 is confusing. You mention details that are not explained clearly enough. You mention that you found 9 papers that incorrectly used poisson, but you don't say how many used it correctly, how you know that the use was incorrect, or how you determined that there was a time trend.

1. We appreciate your guidance here. We removed discussion of the time trend. We have revised the sentence describing the results of our search to describe the overall number of papers we found, and why we believe the 9 did not use it correctly (lines 44-47): “We find evidence that incorrect use of Poisson is widespread. A review of the *Journal of Personality and Social Psychology* found 18 papers using Poisson regression to analyze count data in the past 10 years; of these 18 papers, 9 appear to have used it incorrectly; using Poisson on data in which the variance is not equal to the mean”.

The footnote that accompanies this sentence is also confusing. You refer to a task and your selection of this task (the voodoo doll task) that you have not yet described to the reader and seems irrelevant to the sentence to which the footnote refers.

1. We revised this footnote (1) to only describe the literature search within the four journals mentioned here, and discuss the voodoo doll task only later in footnote 2 when it has been introduced in the main text.

The next sentence mentions a specific false positive rate for a set of studies, but you have yet to explain how one can calculate false positive rates from the data that are available, so this sentence will be hard for readers to understand or evaluate.

1. We have added an explanation of how this rate is calculated on lines 48-50: “Based on the simulations presented later, data with overdispersion equivalent to that found in several of these papers can be expected to lead to false positive rates of up to 60%.”

In the last sentence in the first paragraph of p. 5, you mention a comment from an editor to a colleague. I'm not categorically opposed to providing examples like this, as they can reveal things that go on behind the scenes; but this example did not seem to make an important point.

1. We have deleted this example.

Related to this, I think that extreme descriptors can sometimes be appropriate, there were a couple places (e.g., the "outlandish" on p. 6) that seemed hyperbolic given the issues you raised up to that point.

1. We have revised the wording to delete many of the more extreme descriptors.

Your description of the preregistration is incomplete and confusing. From the description on p. 6, the reader gets the impression that you correctly predicted the direction of your effect, which seemed implausible to me, even in cases where false positive rates are inflated. Looking at the pre-registration document, I see that you did preregister both directional hypotheses, which makes more sense to me. So please add more details to the in-text description and be clear that you preregistered both directional hypotheses.

1. We have revised our description of the preregistration to explicitly make it clear that we preregistered both directional hypotheses, on lines 181-188: “To demonstrate how the improper use of Poisson regression can lead to misleading results, we set out to test the highly implausible theory that seeing a blue shirt might prime thoughts of water, thereby affecting their consumption of Swedish Fish gummy candies. We pre-registered opposing hypotheses that the color of the experimenter’s shirt would either increase or decrease Swedish fish consumption (the preregistration can be found at http://aspredicted.org/blind.php?x=di2sg4).”

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This may be the last opportunity for major editing, therefore please fully check your file prior to re-submission.

Reviewer 1:

Overall, this paper is interesting in that I appreciate its main point (people make critical analyses errors in areas in which they are unfamiliar), and in taking both a simulated and experimental approach. However, this paper is very limited in scope and incremental contribution. Broadly speaking:

* there are already a large number of simulations comparing Poisson to overdispersed models
1. While savvy statisticians may know the dangers of Poisson regressions, their inappropriate usage remains too common. While (in principle) everyone knew about the risks of undisclosed researcher degrees of freedom, it took researchers repeat their dangers before the research community came to see the dangers of p-hacking and address them. Even if the reviewer would never consider using Poisson with overdispersed data, too many others continue to do so. If there are published simulations the reviewer thinks we should be citing, we welcome specific suggestions.
* plotting results for their practical significance often does not lead to entirely inconsistent findings
1. We hope that our figures illustrate our claims in a way that makes the practical implication readily apparent.
* when there are problems in the simulated data, it is often in areas that are not typically found in actual data
1. To help address this concern and better demonstrate the practical significance of these findings, we have created Appendix Figure 1, which shows a plot of the variance to mean ratios from papers we looked at in our literature review of the Voodoo Doll task on the same graph as our simulation results. We also added a short discussion of this question on lines 450-463. This figure demonstrates that many papers do fall into the range of variance to mean ratios where false positive rates can be expected to be high.
* authors might usefully obtain the original data from published papers to explore both practical and statistical significance
1. We are not opposed to taking this step, but we are dubious that it would add much to the persuasive impact of the argument put forward in our paper.

Again, I think the overall message is important, but it has been repeated elsewhere. Sincerely, I hope the authors will continue in the revision of this paper, because I they can clarify/accentuate the problem, as well as provide stronger guidance/steps for future authors to follow, to adopt appropriate analyses.

1. Thank you for this encouragement. We have attempted to clarify and accentuate the problem in our revision, as you suggest.

(Side note on a side point made by authors on p. 7: calories and grams aren’t count data, not appropriate for Poisson)

1. The revised manuscript clarifies our point: that arbitrary decisions made in design and analysis can affect the dispersion of the resulting distribution, producing non-Poisson distributions of count data.

Reviewer 2:

First, let me apologize for the delay submitting my review. Adapting to the Covid-19 crisis, online teaching, and the political situation in the US has slowed my review responsiveness. I’m deeply sorry.

1. We understand. These things have also slowed our research productivity and response to your feedback.

In general, this is the beginnings of a useful pedagogical paper. It clearly shows that blindly using a Poisson model for count data is problematic. However, as a teaching tool, the paper can be much improved in terms (1) clearly describing the assumptions and model forms for both the Poisson model and proposed alternatives and (2) clearly describing the statistical analyses and results reported.

1. The paper doesn’t do a great job of explaining clearly what the assumptions of the different models discussed are or why these models make these assumptions. As the authors note, Poisson regression is often introduced to students in the simplistic form as “the model you use for count data”. Accordingly, to be an effective pedagogical piece, I think the paper should more carefully help readers to understand the workings and design of the various models.
2. We have added discussions of the assumptions of each model – specific line numbers for each are given below. In addition, while it was not asked for explicitly, we also added an expanded explanation of permutation tests’ operation and assumptions on lines 156-164, following our expanded discussions of OLS, Poisson, and Negative Binomial.

a. For the Poisson model, the paper states that the model assumes the variance of the data equals its mean. It would be helpful to unpack why this assumption is made. A Poisson distribution is not just a distribution of any count variable, but specifically a variable that is the count of events that occur after a specific period of time from a process with a known constant rate (λ). The reason the mean and variance are equal is because the distribution is modeling this constant-rate type process. Making this clearer might help readers to see why this model is a poor assumption for many psychological problems.

1. We have added a significantly longer explanation of the Poisson distribution in a paragraph found on lines 106–127, which explains what it is intended to model and why this results in errors. This includes the addition of Figure 1, which demonstrates how Poisson regression’s distributional assumptions lead to incorrect fitting by showing its fit to datasets with increasingly high variance to mean ratios.

b. For the negative binomial model, the paper doesn’t describe the assumptions at all or explain why these might be more reasonable. This is a problem because the negative binomial also makes assumptions about the relationship between the mean and the variance, though ones that are often more reasonable. Violating these can be problematic, as the paper notes on page 9. The sentence “Even negative binomial regression begins to yield the same false positive ratio as Poisson regression at very high variance to mean ratios” illustrates the issue, but the reader is not guided toward understanding it. At present, the paper doesn’t give a clear idea what the negative binomial model is and seems to mostly be suggesting that readers replace a blind use of the Poisson model with a blind use of a negative binomial model.

1. We have added a significantly larger explanation of the Negative Binomial model on lines 131-149 which explain the distribution it comes from and how this translates to a regression model. In addition, we added Figure 2, which shows a Negative Binomial distribution fit to the same data as in Figure 1, and demonstrates how this distribution can better account for increased variance to mean ratios.

C. For the OLS model, it would similarly be useful to clearly state the assumptions of the OLS model and why these are often violated by count data. Relatedly, it would be useful to discuss cases when an OLS model might still be reasonable for count data (when the assumption violates are likely to be of minor consequence).

1. We now discuss in more detail the assumptions of the OLS model on lines 89-104, as a way of both explaining these assumptions and helping contrast this model to Poisson/Negative Binomial.

d. As the t-test is just an OLS model with a single categorical predictor, I suggest dropping this particular analysis entirely, perhaps retaining Footnote 7 making this point.

1. We removed this analysis from the paper and refer to it as OLS throughout, retaining a footnote (footnote 9) which notes that this is equivalent to a T-test.

 e. Several other alternative models, such as overdispersed/quasi-Poisson and zero-inflated models, are shown in the R code, but not described in the text. I suggest adding brief sections to the paper describing these methods, their assumptions, and when they might be appropriate.

1. We have added extensive appendix sections for quasi-Poisson and zero-inflated models on lines 389-413 which explain how these function in more detail.

f. Testing for overdispersion is illustrated in the R code but not discussed in text. In general, a clear tutorial on choosing between alternative model specifications and testing the chosen model’s assumptions would be very helpful.

1. Several tests of overdispersion are presented, along with how to interpret them and the intuition behind how they function, are now discussed in the appendix section “Tests of Overdispersion”, lines 414-436, as well as how to use them in practice. In addition, we added a section to the appendix “Deciding which model to use” on lines 437-449 which discusses in more detail which models to use, and what to consider when deciding between them, as well as citations to relevant literature for further, in-depth reading.
2. The presentation of the example statistical results is not complete, and it’s not clear what some of the numbers are supposed to mean in the results section. For example, the first set of results includes a confidence interval, but no point estimate. What is the CI for? What effect size? If it is Cohen’s d, I suggest instead doing a raw mean difference or the slope coefficient for group from the Poisson regression model. Please also interpret the effect size, rather than merely stating that participants in one condition ate “more”—how many more? The omission of the effect size point estimates and interpretations is repeated for the transformed model results later in the paragraph.
3. We have revised the reporting of statistical results for the experiment – only the initial Poisson regression results are displayed in the text. The remainder of the results are now presented in the newly-created Table 1, where the means, variances, and regression p-values for all analyses are presented. In addition, we have created Figure 3, which shows the distribution of the number of fish chosen in each condition, as well as the Poisson distribution fit to the data, to more intuitively illustrate how this distribution is failing to accurately reflect the data in a way that statistics alone may not show as clearly. This is paired with Figure 4, which shows the same data with the Negative Binomial distributions fit to the data overlaid, to demonstrate how this distribution can better fit the data.

a. To make the mean = variance assumption violation clearer, please report the variance of the response variable as well as the mean. b. A table showing all of these results would also be much easier to digest than just a list of statistics in running text.

1. We have added this to Table 1, where we have moved statistical results instead of putting them in the text.
2. The paper spends a fair amount of space on the point that preregistering a bad analysis doesn’t lead to good inference. I’m not sure how useful a point this is. It isn’t really built up at all—is false claiming of preregistration as an arbiter of validity common in this literature or something? At present, it reads as though it were an out of left field pet peeve of the author, rather than a point that is clearly connected to the rest of the paper.
3. We significantly shortened the section discussing preregistration – in full it is now on lines 57-59: “In addition, unlike other, more familiar, forms of p-hacking such as conducting multiple tests, pre-registration does not prevent inflated rates of false positives when Poisson regressions are used with overdispersed data.” Our intent with this section is just to make it clear that misuse of Poisson is still a concern even if you are already carrying out other good and useful open science practices.

Minor points/typos:

1. In footnote 1, you have a small typo. You refer to Linear, Poisson, and Negative Binomial models as cases of the “general linear model”. It should be the “generalized linear model”.
2. This typo has been fixed.
3. On page 9, you refer to the “false positive ratio”. This should be the “false positive rate”
4. This typo has been fixed.
5. Footnote 8 is virtually identical to Footnote 1. You can probably omit one of them.
6. We removed the first instance of this footnote.
7. The following are some general comments about R code style best practices to facilitate reproducibility and understandability for other users of your code

a. library() should be preferred over require() in most cases, such as in your example scripts. The major reason is that library() will error immediately if the package is not available, whereas require() doesn’t error until an exported object is called. This can make it harder for users to debug. See <https://yihui.org/en/2014/07/library-vs-require/>

1. We have updated the code to use library() – thank you for the tip!

b. You should generally avoid using setwd() at all in R scripts. This makes the script specific to a specific machine and requires users to always edit the script before they are able to use it themselves. This can be a big barrier to new users of R. Instead, a project/folder-oriented workflow is more reproducible, wherein scripts are written assuming that they are run in a fresh new R session where all file paths refer to directories within the directory holding the script. See <https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects> and <https://here.r-lib.org/> for discussions. For this paper, I suggest simply instructing the reader to import or create a data frame with their data, then include output showing what the data in the example look like.

1. We have updated the tutorial to not use setwd(), instead readers are instructed what form the data.frame should take when used and shown output.

I hope you find these comments helpful as you revise your paper.

Best, Brenton Wiernik

**Editor Final Decision—Accept**

July 25, 2021

Dear William H. Ryan,

I have now had a chance to read over your manuscript “Poisson Regressions: A little fishy”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Richard Lucas