Your review report

Manuscript

Over 110 million mice and rats are bred and used in United States laboratories

Confidential feedback for the Editor

Your recommendation	• *Revise*
Is the study design appropriate to answer the research question (including the use of appropriate controls), and are the conclusions supported by the evidence presented?	• *Yes*
Are the methods sufficiently described to allow the study to be repeated?	• *Yes*
Is the use of statistics and treatment of uncertainties appropriate?	• *No*
Comments	*The assumption that the ratio of RM to AWA-covered animals is the same throughout all institutions must be named explicitly and should also be discussed in the section where the author addressed the potential that his estimates are too high or too low. Having a different ratio of RM to AWA-covered animal between the institutions for which the author has data on both "categories" of animals*

and the rest of the institutions in the USA can have a major impact on the overall estimated. An impact that is tremendous even for the slightest deviations in percentages. For example: Let us assume that the percentage of RM-to-total-animals-used in institutions for which there is no RM-data available is just 0.5% lower than in the 16 covered institutions (i.e. 98.8% instead of 99.3%). In that case, the estimate would be 170% of the "true" numbers. Conversely, if same percentage were just 0.5% higher, the estimate would only be roughly 30% of the "true" numbers. The author raises a valid point that this could cut both ways i.e. that it could lead to an over-estimation (if the ratio of RM-to-AWA-covered-animals is *higher* in the 16 covered institutions as compared to the rest of the institutions which conduct animal research) or an under-estimation of the animal number (if the ratio of RM-to-AWA-covered-animals is *lower* in the 16 covered institutions as compared to the rest of the institutions which conduct animal research). Given the data from Germany and Switzerland I linked to in the previous review, I deem the former more plausible than the latter but this up for debate or rather: up for additional scrutiny. It isn't something, however, that must be answered definitively in this manuscript. What must be addressed, however, is the inherent uncertainty of the estimate due to the difficulty to check the underlying assumption (similar ratio of RM-to-AWA-covered-animals between covered and non-covered institutions). This should not only be done by addressing the assumption as well as the potential effects on the estimate is violated (as described above), but also by being more cautious with regard to the wording and the presentation of the extrapolated numbers. Some examples (not exhaustive): - Wording: Switching the title to "Over 110 million mice and rats are bred and used in United States laboratories" provides the reader with a

false sense of certainty given the large uncertainties underlying the data and the assumptions used to extrapolate the numbers (after all, the number could be significantly lower or higher). I suggest to switch to the old title as this more accurately reflects the content. - Numbers: Similarly, providing estimates down to individual animal as for example done in Table 2 conveys a certainty of the estimate that is not warranted. Just changing the extrapolation factor from 0.007 (the rounded ratio of AWA-covered-animals-to-total-animals in the 16 covered institutions) to 0.007017665 (the exact ratio of AWA-covered-animals-to-total-animals in the 16 covered institutions) increases the estimate by almost 300'000 animals. Given the sensitivity of the estimate to such tiny changes in the extrapolation factor, it doesn't seem prudent to report estimates down to the individual animals as they are certainly wildly off. Estimates on the order of magnitude million of animals seems more reasonable (but of course one should mention why one does only report such rough estimates). - Numbers: Finally, I suggest reporting ranges of estimates which take into consideration the underlying uncertainties discussed above and in the previous review. For example, instead of only reporting a single point estimate, the author could report multiple point estimates assuming different ratios of AWA-covered-animals-to-total-animals (for example: 0.002, 0.007, 0.012, 0.017, 0.037, 0.057). This would immediately show the large effects these small changes would have on the overall estimates, thereby further emphasising the need for reliable and comparable data. Addressing the underlying uncertainties explicitly is especially warranted because the published numbers will certainly be used as a benchmark, as the author himself points out. As such, readers should be informed about corresponding uncertainties underlying this

benchmark. After all, the most valuable contribution of the manuscript is in my opinion not the estimate itself but the description of the process that brought about this estimate. After all, it will - hopefully - be the difficulties and uncertainties accompanying this process that will provide ideas for new paths to come up with more reliable estimates.

Is the presentation of the work clear?

- *Yes*

Are the images in this manuscript (including electrophoretic gels and blots) free from apparent manipulation?

- *Not applicable*

Confidential comments to the Editor

I still believe that the author's work is crucial contribution to the field of animal research and is worth a publication - *if* the underlying assumptions of the extrapolation and the effects that a deviation from these assumptions can have on the estimate are specifically addressed in text itself (see explanation in comment about statistics section above). In its current form, however, I cannot approve of the manuscript.

Feedback for the author(s)

Comments to the author(s)

I still believe that the author's work is crucial contribution to the field of animal research and is worth a publication *if* the underlying assumptions of the extrapolation and the effects that a deviation from these assumptions can have on the estimate are specifically addressed in text itself (see explanation below). The assumption that the ratio of RM to AWA-covered animals is the same throughout all institutions must be named explicitly and should also be discussed in the section where the author addressed the potential that his estimates are too high or too low. Having a different ratio of RM to AWA-covered animal between the institutions for which the author has data on both "categories" of animals and the rest of the institutions in the USA can have a major impact on the overall estimated. An impact that is tremendous even for the slightest deviations in percentages. For example: Let us assume that the percentage of RM-to-total-animals-used in institutions for which there is no RM-data available is just 0.5% lower than in the 16 covered institutions (i.e. 98.8% instead of 99.3%). In that case, the estimate would be 170% of the "true" numbers. Conversely, if same percentage were just 0.5% higher, the estimate would only be roughly 30% of the "true" numbers. The author raises a valid point that this could cut both ways i.e. that it could lead to an over-estimation (if the ratio of RM-to-AWA-covered-animals is *higher* in the 16 covered institutions as compared to the rest of the institutions which conduct animal research) or an under-estimation of the animal number (if the ratio of RM-to-AWA-covered-animals is *lower* in the 16 covered institutions as compared to the rest of the institutions which conduct animal research). Given the data from Germany and Switzerland I linked to in the previous review, I deem the former more plausible than the latter but this up for debate or rather: up for additional scrutiny. It isn't

something, however, that must be answered definitively in this manuscript. What must be addressed, however, is the inherent uncertainty of the estimate due to the difficulty to check the underlying assumption (similar ratio of RM-to-AWA-covered-animals between covered and non-covered institutions). This should not only be done by addressing the assumption as well as the potential effects on the estimate is violated (as described above), but also by being more cautious with regard to the wording and the presentation of the extrapolated numbers. Some examples (not exhaustive): - Wording: Switching the title to "Over 110 million mice and rats are bred and used in United States laboratories" provides the reader with a false sense of certainty given the large uncertainties underlying the data and the assumptions used to extrapolate the numbers (after all, the number could be significantly lower or higher). I suggest to switch to the old title as this more accurately reflects the content. - Numbers: Similarly, providing estimates down to individual animal as for example done in Table 2 conveys a certainty of the estimate that is not warranted. Just changing the extrapolation factor from 0.007 (the rounded ratio of AWA-covered-animals-to-total-animals in the 16 covered institutions) to 0.007017665 (the exact ratio of AWA-covered-animals-to-total-animals in the 16 covered institutions) increases the estimate by almost 300'000 animals. Given the sensitivity of the estimate to such tiny changes in the extrapolation factor, it doesn't seem prudent to report estimates down to the individual animals as they are certainly wildly off. Estimates on the order of magnitude million of animals seems more reasonable (but of course one should mention why one does only report such rough estimates). - Numbers: Finally, I suggest reporting ranges of estimates which take into consideration the underlying uncertainties discussed above and in the previous review. For example, instead of only reporting a single point estimate, the author could report multiple point estimates assuming different ratios of AWA-covered-animals-to-total-animals (for example: 0.002, 0.007, 0.012, 0.017, 0.037, 0.057). This would immediately show the large effects these small changes would have on the overall estimates, thereby further emphasising the need for reliable and comparable data. Addressing the underlying uncertainties explicitly is especially warranted because the published numbers will certainly be used as a benchmark, as the author himself points out. As such, readers should be informed about corresponding uncertainties underlying this benchmark. After all, the most valuable contribution of the manuscript is in my opinion not the estimate itself but the description of the process that brought about this estimate. After all, it will - hopefully - be the difficulties and uncertainties accompanying this process that will provide ideas for new paths to come up with more reliable estimates.