List of changes: addressing the comments

We would like to thank again the reviewers for this last round of comments. We addressed this time the remarks on reviewer-by-reviewer basis.

Reviewer 1:

- *The evaluation of model performance is the biggest shortfall of this paper. The authors note that only 2 out of 70 observations of mosquito captures are outside the 90% credible interval of the posterior predictive expected number of catches (pg. 12 lines 40-45). However, the 90% credible intervals on the predictions are very large. The authors should include additional information on the estimated error in the predicted number of catches (e.g. RMSE etc) and provide a more in-depth discussion of model performance.*

  - Adrien:

- *Please add further details clarifying why dispersal of WT mosquitoes might differ from the Sterile GM males that the model is estimating and whether using WT MRR experiments might negatively affect predictions of GM mosquito dispersal.*

  - Adrien and Geoff:

- *The methods would benefit from additional discussion of the specific expertise of the experts that were surveyed. Generally, what qualified them to make these predictions/on what basis were the predictions made? Why was this approach preferable to acquiring priors from lab-based studies currently in the literature?*

  - Geoff:

- *Were surveys of experts conducted in a standardized way? It would be helpful if you could supply in a supplement the specific questions that each expert was asked and their responses.*

  - Adrien:

- *The discussion and/or introduction would benefit if the authors specifically discussed how the results of this study were used to assess risk and how they informed the release of the GM mosquitoes.*

  - Keith:

- *The supplemental document appended at the end of the manuscript appears to just be the manuscript in pdf format.*

  - Adrien:

- *Typo pg. 13, line 61: 'define the bounds of accuracy that are regulators and stakeholders believe are adequate'*

  - Adrien

Reviewer 2:

- *P2L28: The authors introduce genetic containment (male sterility) as a convenient means to meet the recommendations of a "phased-release" strategy, to collect the data required for evaluating risks on GDMMs. Of course the authors are correct here. However, the way this is phrased suggests that genetic containment is a convenient alternative to physical containment. This makes the assumption that the data gathered for, say a DSM male, are equivalent to the risks associated with alternative designs, for example gene drive replacement strategies.*

  Keith:

- *P2L14: "force effector genes" is inaccurate because suppression gene drives do have to carry an "effector". I would suggest rephrasing to "gene drives to drive engineered alleles or chromosomes through vector...".*

  Keith:

- *P2L38: Males of these mosquitoes are reproductively contained. Currently the way it reads suggests that there is full genetic containment. This is not the case here, DSM females are actually fertile - which is how the strain is maintained. Perhaps this needs to be emphasized more strongly also in the discussion because "containment" here only refers to the released population. The population in the lab, an accidental release of females, or contamination of females in the released population - all of these have to be taken into account when arguing that this strain has the benefit of being "contained".*

  Keith:

- *P2L59 vecotr should be vector.*

  Adrien:

- *P12L50 What does it mean that the authors "held back a proportion of the observation data". The point of this was lost on me - although admittedly with very limited experience with the methods used here that is probably true for many sections. This is not a criticism of the work, just an honest take from a researcher in the field that does not analyse modeling papers in depth on a regular basis.*

  Adrien: