**Comments from the Editor, Prof. Loseke Kruuk**

This looks like an excellent contribution to the Proc B special edition on gene drive; the referees are very positive, but have also made some very useful suggestions.

A quick point about the abstract. This reads very well, until the last sentence, which puzzled me: if the aim is population control, why is it a problem if the drive allele spreads rapidly? I realise you explain this in the Discussion, but it's puzzling seeing it presented without explanation in the abstract - I certainly got to the last sentence and assumed I'd misunderstood something. So this needs some clarification/modification. 

Thank you, that’s helpful! I have changed this sentence to something that makes sense without the Discussion.

**Comments from the Associate Editor, Dr. Tom Price**

Two expert reviewers have read your MS, and both like it and recommend it be accepted with minor revisions, which I am happy to recommend. Both have some questions that need to be dealt with, and also suggestions to improve the breadth and clarity of the MS. I particularly agree with the point that your results on resistance are very interesting, but you do not discuss it much in the Discussion, which seems a missed opportunity.

I have expanded the section on resistance as suggested.

NB One thing you might think about for the future: I really like your github repository with the code. But if a reviewer forks your repository to explore the simulations, my understanding is you will likely be able to identify them, breaking reviewer anonymity (e.g. my github name is TomARPrice), potentially with the reviewer not realising this. It might be worth either having a recommendation on the front page to reviewers to clone it (which I think can't be tracked), or to set up a new github account with a non-informative name. Alternatively you could have a "code for reviewers" appendix of the code included in the review document so reviewers don't have to use github. But I don't use github much so I might be misunderstanding this.  
  
Thanks, that is an interesting point! I have never used the “fork” feature so it hadn’t occurred to me. You’re right that one can download the code and data anonymously using the ‘clone’ feature, or the direct download button associated with individual files in the Github repo, so I think it’s workable but does raise some anonymity issues.

I also recently came across this page, which I will look into for the future: <https://github.com/tdurieux/anonymous_github>

**Comments from Reviewer 1**

This manuscript formally examines the ecological and evolutionary dynamics of various forms of Z linked gene drives by means of simulations. Z-linked drive constructs could be used to control/extirpate pest species with a ZW sex determination by turning them into males. Holman explores a impressive range of potential drive construct designs, ecological conditions (local and global competition within and between the sexes, population structure), and potential resistance mechanisms. The simulations demonstrate that Z linked drives can be powerful tools to eradicate populations within relatively short time periods and with very few animals released. At the same time, and in line with research on other homing based drive constructs, the results suggest that drive resistance will evolve quickly if resistance is possible. Creating constructs that make resistance evolution difficult should therefore be a priority in their design.

My general impression of the paper was very positive. The scope for type suppression drives is large, as ZW species include major agricultural and disease pests. I was particularly impressed with the broad range of scenarios addressed. As a result, I think this fits very well within the scope of the special issue, and will be a valuable contribution to the field in general. That said, I think there are still a few (fairly minor) issues that need addressing / clarifying. Interestingly, most of them seem to arise from arguably one of the main strengths of the MS, namely the sheer volume of issues addressed, and the challenges this poses with respect to summarising of the results and their presentation.

Thank you for the helpful comments and encouragement.  
  
## General Comments  
  
1) Structure. In terms of presentation/accessibility for the reader- given that you consider so many different processes, I was wondering whether the whole MS would benefit if the goals / methods / results were explicitly grouped according to different processes that are consistent / recognisable to reader throughout the paper. E.g. drive design/type (drive type, strength and cost), drive resistance, ecological factors (competition, density dependence and structure), release strategy, interactions. You could then clearly mention those as goals in the last paragraph of the introduction, and they would then reappear as sections in the methods and results. Some aspects (drive design, release strategy) seem more easily manipulated than others (ecology), so it may help assess the reader what the scope there is for successful drive design. By the way, I like the first results paragraph that shows a representative example for the each of the qualitative outcomes, showing the actual evolutionary dynamics, so would leave that section as is. 

Thank you for the feedback. I have re-written the manuscript with this in mind. The main changes are to the structure of the results, which uses similar sub-headings to those proposed here, though I also adopt that structure in the other sections while keeping it in prose form.

2) GLM approach. I was interested to see the binomial GLM used as a means to explore / summarise the large / multidimensional parameter space. This is the first time I've seen it used in this way. At the same time, it raised a couple of questions I wasn't quite sure what to make of, so I would appreciate if you could maybe briefly clarify.

TL;DR: I still think the glm approach is valid, but it’s not optimal in several ways. I have replaced it with Latin hypercube sampling, and the new results are similar but much richer.

1) A slightly philosophical point- It seems interesting that you create simulations that are quite complex, and potentially non-linear (-monotonic), but they are then analysed / looked at through the lens of an analysis that 'only' allows for logit-shape effects / two-way interactions, and as a result, a lot of the previously implemented complexity is potentially lost again. I guess there is no way to analyse this large space without some kind of information loss. Secondly, I was wondering how much your GLM conclusions depend on your choice of parameter values / combinations? In what way are the parameter combinations a representative sample of the space (or what might be observed in nature)?

Thanks for this thought-provoking comment. My original aim was to be able to measure the **marginal effect** of each of the model’s parameters on the main outcome variable (the occurrence of extinction), i.e. the average effect across a range of realistic/informative values for the other model parameters.

However since the original submission, I learned about Latin hypercube sampling (LHS), which is an efficient way to evenly sample from the full range of a high-dimensional parameter space. I think this method provides a better way to achieve my original aim (indeed, that is what LHS is mostly used for), because I am not constrained to look at particular parameter combinations selected *a priori* (which begs the obvious question “what about other parameter spaces, and did my choices dictate the conclusions?”), and I can sample more of the parameter space with less CPU time. Also, this method gets around having to assume linearity in the relationship between each parameter and the extinction probability, which I agree is a problem because it discards information.

**I have therefore completely re-run the model and re-made all figures to address this comment**. In the revision, it is now apparent that many parameters do have a non-linear relationship with the two response variables (i.e. extinction probability, or time-to-extinction). Also, because I now sample more-or-less the entire biologically plausible parameter space, there is much less scope to worry about whether the parameter space selected is representative.

3) Modelling. I was a bit confused about the various ways by which you modelled density-dependent fecundity. First, why does equation S1 have read as F\_i=1+r\*w(Nglobal,Nlocal)? Would that not imply F\_i=1 if female fitness w\_i=0, hence expected reproduction under a Poisson distribution of 1 (rather than 0)? Or do you subtract 1 at the end (as the Poisson is not defined for lambda=0)?

### Thanks for spotting this. This was actually an error left over from an earlier iteration of the model, so I am glad you caught it. The equation is now fixed (in the manuscript, and in the code for the model), and the revised equation ensures that females with a fitness of 0 indeed have an expected fecundity of zero.

### There was one major change in the results once I fixed this error: the female-sterilising Z\* allele no longer succeeds in causing extinction, ever (previously, it occasionally caused extinction). This is because Z\* females now correctly have zero fitness, instead of having a small but non-zero fitness as before. The other results were qualitatively unaffected.

Moreover, I couldn't quite wrap my head around your use of carrying capacity K and how exactly it relates to the various processes of density dependence. Should K not be related/bound by the values D\_i can possibly take on? Put differently, how do you currently avoid D\_i>K, which potentially produces negative fertilities? Finding a way to relate scale K with respect to D\_i may also avoid your "reverse"/trucation problem where overall fecundity rates sum(Fi) exceed K.

Another good point – this was documented in my code, but I had forgotten to also describe it properly in the methods (corrected in the revision). I dealt with the possibility D > K by adding an “if” statement to the code, to set D = K if D > K. This stops fecundity from going negative: the lowest possible fecundity per female is now 2 offspring (for female genotypes with normal fitness, at least): this means that a ‘full’ population of wild type individuals will tend to replace itself, since females produce an average of 1 son and 1 daughter each when D = K.

4) Resistance evolution. Given both how detailed you model the various forms of resistance, and how common resistance evolution was as an outcome (if it was possible), it seemed to me as if you dedicated quite little space to actually talking about the results, e.g. which form of resistance is particularly likely, and therefore which ones could be prioritised to reduce the risk of it occurring.

Thanks for this. I have expanded this part of the Discussion as suggested (starts line 292).

## Minor points / Details  
  
L84: Maybe specify that you only release them once.

Done, thanks

Also, it appeared that you didn't really explore the effect of different n\_release on dynamics. Is there a particular reason for this? Maybe at least be worth mentioning somewhere, given that this is a parameter that can be quite easily tweaked with rather easily by someone planning an intervention (compared to some of other parameters you consider). I guess there would also be the option of examining different temporal release strategies (e.g. repeated release), or the release of different genotypes, but I wonder whether this would not be too much given that you explore a lot of factors already.

I originally fixed the release size at 20 individuals simply to minimise the number of variable parameters, because I figured that the effect of increasing the release size could be guessed reliably (bigger release = more likelihood of extinction, though with diminishing returns). Similarly, I felt that multiple releases could only help the gene drive allele to spread.

In the revised version, I now allow the release size to vary as suggested. As expected, the release size has little effect on the probability of extinction provided that the release is not too tiny. This is because a gene drive that is able to invade the population when it is very rare does not require a large release to get started. The only obstacle to invasion is the possibility that every individual in the release stochastically fails to breed, but having a release of 20 or 30 seems to ensure that this never happens (because the chance the whole release fails is related to p^n, where p is the probability per individual of reproductive failure and n is the release size, so it quickly goes to zero as n increases).  
  
LL98-99: In terms of readability, would it make sense to use the terms female and male drive throughout the MS, rather than switching between shredding / gene conversion and female / male drive. Maybe it's just me, but it requires that extra step of remembering which drive type happens in which sex. Along a similar line, would it make sense to give the autosomal resistance loci a name that immediately tells the reader it creates resistance at male or female drive, respectively (M^r and F^r or something). It took me a while to figure out what they do exactly, and what the difference between them is.

Thanks, that’s a helpful suggestion. I have kept the original terminology, because the W-shredder is not necessarily a gene drive (it depends on whether the missing daughters are replaced with Z\*-bearing sons). But I have been careful to make the manuscript even more accessible by not jumping between these terms so often.  
  
L110: I would either not mention the parameters here at all, or briefly mention what they do (as done in the Figures- "shape of density dependence", "Maximum fecundity" etc.)

As suggested, I now write out the parameters’ meanings as well as their symbols. Note that there is also Table 1, which matches symbol names to their meanings, and the labels on the facets of Figures 3 and 4, so once the MS is formatted with the figures in place, there will be a Glossary close at hand.

LL236ff: Could an argument for combining male and female drive be that it makes resistance less likely (as the target species has to come up with a resistance mechanism against both)? Do you find this to be the case?

That’s an interesting idea, though I don’t think it’s the case for the type of gene drive considered in this MS. Essentially the female part of the drive (a Z-linked allele that destroys the W chromosome) is the part that suppresses the population: if there is any resistance to the female part (e.g. a trans-acting autosomal allele that protects the W) then the population will not go extinct. The male-acting part mostly serves to help the shredding allele spread faster, by ensuring that it is positively selected in males as well as females, or to help it spread at all (in cases where the drive allele has below-wildtype fitness in females). There are lots of parameter spaces where we can either leave out the male component, or make the male part easy to resist, and the population still goes extinct.

Figure 2. I couldn't find Table S3 that shows the parameter values used for the three panels. Also, would it make sense to only show dynamics after the burn-in (or dedicate ~10% of the figure to the burn-in)? In particular, panel A now mainly shows the burn-in, and not the interesting, drive-related dynamics at the end, which is even slightly cut off.

I had mistakenly omitted that table, but I’ve added it now (now Table S1).

I have adjusted the x-axis of Figure 2A so that the burn-in period takes up a similar amount of space to the other two, as suggested. I also fixed the end point getting slightly cut off, so it’s clear that the population went extinct.  
  
Figure 3. Would it make sense to standardise the y axes here? It would give the reader already an impression of the strength of the effects. Fair enough if you want to dedicate Figure 4 to this question.

Thanks, I did have the same thought! I agree that it is a bit misleading to have different y-axes, although the advantage of varying the scales is that readers can then see the patterns for all the weaker effects clearly (they’d just appear like flat lines, otherwise). I have re-worded the legend to make it even more clear that the axis scales are different.

Supplementary Methods/Calculating Fitness. I thought it was a bit confusing that you had a separate paragraph on fitness, when all it does (if I understood correctly) is create differences in fertility (females) and mating success (males), which you elaborate in separate paragraphs later (now under the term 'fertility'). It might make sense to just use fertility (rather than fitness) throughout, and explain those costs in the paragraphs were you actually explain how they were implemented?

I have merged the fitness section into the male mating success and female fecundity sections as suggested. I don’t think I used the term ‘fertility’ anywhere, and I don’t think it makes sense to sub it in for ‘fitness’. However I have re-written this part of the methods to anticipate this issue a bit.

## Typos  
L82: Living in \*k\* discrete habitat patches?

Thanks, this typo has now been fixed

L135: spread -> spreads?

Thanks – changed “spread” to “invaded” to make the intended meaning more clear

Supplementary equation (4): p\_k rather than p\_j?  
  
Thanks – I fixed this typo, and also changed the index k to l, since the letter k was already being used to index patches.   
  
  
Referee: 2  
  
Comments to the Author(s)  
In this manuscript, Holman provides simulation results for a W-shredder gene drive considering several parameters and finds conditions optimal for the driver to spread. Simulations are likely appropriate (vs analytical solutions) in this case because the number of parameters likely make analytical solutions intractable.

Thank you for the helpful comments.   
  
Lines 71-73 - this is unclear - are you suggesting that Fisherian SR selection would act on other loci in the genome? If so, would that selection be to restore optimal SR or to suppress the driver? Can these two be lumped together as the same thing?

The sentence in question was:

*For example, the altered sex ratio might intensify the fitness advantage accruing to any resistant W chromosomes or autosomal modifiers that prevent W-shredding (due to Fisherian selection for an even sex ratio; [26])*

I have tried to make this clearer as follows:

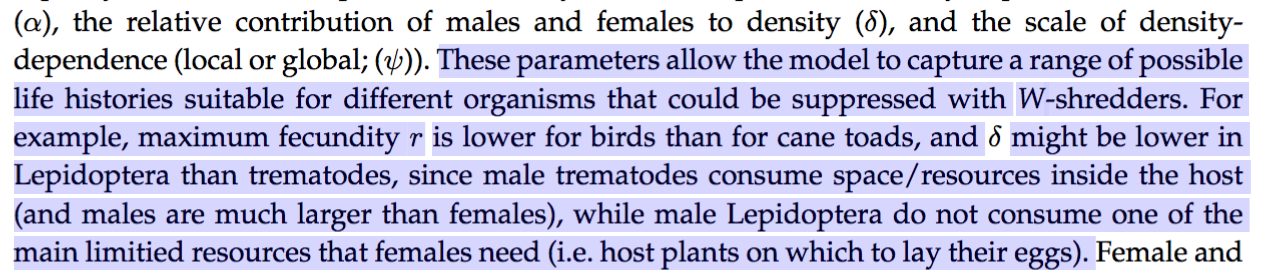
*For example, alleles that prevent W-shredding might have an especially large fitness advantage relative to resistance alleles against ‘standard’ CRISPR drives [see 17,18], since the former restore a normal sex ratio as well as defusing the harmful effect of the driver on individual fitness.*

Line 100 - What are the implications of resistance alleles being cost free? Is that likely in this scenario?

If resistance alleles lowered fitness, they would be less likely to invade, and so extinction would be more likely, all else equal. I now say this explicitly in the manuscript.

Line 113-114 - I appreciate the incorporation of population structure, but it seems like the three different systems (trematodes, leps and birds) would have extremely different demographic models. It is probably worth a much more detailed discussion about how each system would fit the different models tested.

I slightly expanded this section, which now reads:



Given the space limitations, and the hypothetical nature of W-shredders at present, I elected not to elaborate further. The model does not aim to capture every ecological detail, or to compare the likely success of a shredder for e.g. toads vs trematodes – rather, its aim (quoting from the Introduction) is “to test which properties of the gene drive and the ecology of the target species are critical to determining the likelihood and speed of extinction”

Figure 1 - I can think of clearer ways to present the data. Maybe show the actual shredded chromosomes (or sterilizing somehow).

There is no data in Figure 1. Figure 1 already shows sterilising through the offspring being absent. I don’t think it’s essential to draw chromosomes to get the point across, especially since the molecular mechanisms of W-shredders are currently hypothetical, are not the focus of this paper.

Figure 2 - Why not plot counts instead of frequency? Aren't the counts what actually matters? (see comments below too)

I’m not sure I’ve ever seen a population genetics model that plots allele counts instead of allele frequencies, so I think this change would confuse more than clarify.

Also, one can infer the numbers from this plot. The legend explains that the maximum possible population size was 10,000, so an allele frequency of 10% means there are 1,000 copies of the allele when the population is at its maximum value, or 0.1 \* [current population size] otherwise.

Figure 3 - Is there a way to present this more simply? It was a lot of work to understand what was going on? A couple ideas: a) switch Figure 4 to Figure 3 and either put figure 3 as 4 or move it to the appendix - Figure 4 is a better overall summary. b) Ignore gene conversion in the main figure (or add it as an additional parameter panel) and put the interaction (current Fig 3) in the appendix. c) Maybe (god forbid) use bars instead of dots.

I totally re-designed the figures in the revision. I took your suggestion to not plot the results separately by gene conversion and I think it’s much improved, thanks!

I chose to use dots instead of bars because I’d like to vary the y-axis scale, and it’s considered misleading to do that with bar charts since one can make differences look bigger or smaller by changing the y-axis scale. With dots, I think people are less inclined to assume that the y-axis starts at zero. There are some bars in the revised figures for the qualitative parameters, and I start all the bars at y=0 to avoid this issue.  
  
Line 151 - Isn't the most important parameter the initial frequency? In fact, I'd argue that it is likely the initial count, not the frequency. In other words, if populations started at 10^4, 10^6 or 10^8, I bet that the absolute count is more important than the frequency. This is a point made in some of the other gene drive modeling papers as well as the evolutionary rescue field in general. After all, this is a special case of evolutionary rescue. The absolute count vs. frequency point is alluded to in line 287.

Here I meant “most important parameter out of the ones I manipulated and plotted”, which did not include release size in the original. I now model release size too, but you are right that having a release size close to zero (vs having an adequately large release) would have a strong effect on the probability of extinction.

You’re also correct that this generalises to evolutionary rescue, but actually I think it generalises to the population genetics of any positively-selected allele (e.g. a gene drive, or any ‘normal’ allele that improves individual fitness). According to classical pop gen (e.g. Kimura, Gillespie, Moran etc), the fixation probability of a positively-selected allele is

*π* ≈ 2*sN*e/*N*.

This means that the fixation probability of the positively-selected allele only depends on the strength of selection and population structure, but it is independent of the frequency or copy number of the allele (the key assumptions used to derive this expression are that *N* is not tiny – over 100 is plenty – and *s* is not too small).