Revision round #1

Decision for round #1 : *Revision needed*

Dear Authors,
Your preprint entitled "Can "true" survival be estimated without global resights: it depends on "true" movement" has now been reviewed, and the reviewers' comments are appended below. As you will see, both reviewers are positive about the study, notably regarding the way it is written, and its relevance for optimizing capture recapture/resight designs. Yet they have several insightful and constructive comments that all need to be addressed carefully before your preprint can be recommended.

One main comment shared by both reviewers is that the manuscript requires important clarifications of the model description. Both reviewers provide several comments and suggestions in that regard. I was also surprised to see no equations. I appreciate that the model is quite complex, but would it be possible to guide us through the model and model parameters with a couple of very simple examples of capture recapture/resight histories (e.g. 2-3 time intervals)? That could also help readers implementing the model with other software than MARK/Stan.

Author response: We have now included an example encounter history and likelihood equation in the Methods section.

On that note, why was Stan only used for the multi state version of the model? Wouldn't the use of Stan for both models help with model comparison? You could still keep the results from MARK e.g. in appendix to confirm the results are nearly identical when using MARK vs Stan. Also, it wasn't clear to me whether Stan was used in a Bayesian framework or not. If so, please provide the prior used for each parameters and the chains' settings.

Author response: The multistate version of the Barker model implemented in Program MARK differs from the formulation we are suggesting (i.e., MARK formulation only allows for one transition between capture events, whereas our model includes the possibility of two transitions), so we cannot implement our multistate version in Program MARK. We used Program MARK for the single-state simulations, as this version is more commonly used by researchers and also computationally much faster than Stan.

I also agree with one reviewer that it feels important, at least for one or two most relevant sampling designs, to see how model performance changes with sample size.

Author response: We have included Appendix B, that addresses the impact of sample size.

Importantly, please provide R files (or text files) of the codes and not a PDF following the manuscript. Also include the R code that was used to run Stan, not only the Stan code. In order to improve re-usability of the code, try to not "hard code" numbers throughout. For example replace 12 by nOcc / 2, or 25 by nOcc + 1 (if it is indeed the reason for using 12 and 25 here). Also annotate the code more (e.g. what is tt? what is tt2 ?). We need to be able to easily review the code and re-run/re-use codes and models.

Author: We have done this in the revision.

Finally please also provide a version of the manuscript with line numbers to help us review the revised manuscript.

Author: Yes, sorry about there being no line numbers. The original MS had line numbers, but ArXiv wouldn't allow me to upload a document with line numbers.

I also have some more specific comments:
Abstract: "One popular method for estimating true survival", perhaps delete "popular" as it was not a familiar method for us at least?

Author response: Done

Discussion: "Another alternative for species with one-time ontogenetic movement dispersal (e.g., natal dispersal in birds) is the mark-recapture approach with natal dispersal described by Badia-Boher et al. (2023)." I am not sure why a reference here, focusing on one time movement, is better than referring to more general spatial capture recapture methods? See for example more general, earlier work:
-Ergon, Torbjørn, and Beth Gardner. "Separating mortality and emigration: modelling space use, dispersal and survival with robust-design spatial capture–recapture data." Methods in Ecology and Evolution 5.12 (2014): 1327-1336 (which you cite already elsewhere).
-Schaub, Michael, and J. Andrew Royle. "Estimating true instead of apparent survival using spatial Cormack–Jolly–Seber models." Methods in Ecology and Evolution 5.12 (2014): 1316-1326.

Author response: Note that we do cite Ergon & Gardner (2014) elsewhere. Most of the focus of this study is differentiating permanent emigration from survival. Since permanent emigration only occurs once (by definition), the Badia-Boher model would likely be more appropriate as Ergon & Gardner focus on temporary emigration and Schaub & Royle assume a normal dispersal kernal, which focuses more on short-distance movements and may be less appropriate for permanent emigration.

I look forward to reading the revised version of this preprint.
Best wishes,
Matthieu
by ***Matthieu Paquet****, 27 Mar 2024 10:39*
Manuscript: **https://doi.org/10.48550/arXiv.2312.13414**
version: 1

Review by anonymous reviewer 1, 26 Mar 2024 06:46

Overall, I think this is a nice, straightforward contribution to the literature. It is generally well-written and rather concise given the amount of simulations and the model to be described. I would like to see more clarity in the model description and better links between the figures (particularly the diagrams) and text. I think the number of figures and tables could be reduced (the # figures to text ratio is pretty high) – I'm suggesting moving some to a supplement but not removing them totally. So, it looks like there is a lot of text below but I don't think the manuscript necessarily needs a ton of work, most of my comments are asking clarifying questions or trying to work out the details of the simulation and that is because I'm just not that familiar with the Barker model. I appreciate your time and effort in putting this manuscript together, I hope that my comments and suggestions are somewhat useful for you.

Questions from the journal:
Does the title clearly reflect the content of the article? I would say that while I think the title is kind of catchy, I had no idea what "global resights" meant until I read the paper and didn't understand what "true" movement meant in this context. I think a more descriptive title would be helpful.

Author response: We have changed the title to the following: ***The importance of sampling design for estimation of survival using joint live-recapture and live resight models***

Does the abstract present the main findings of the study? My answer to this is not really, but I think it won't be much work to update it. I found it a bit difficult to follow some parts of the abstract, I think a bit of context would help (I understand the need to keep it short though). I am not sure what the global and random model designs are? I'm going to guess that global means you sample inside and outside the capture site, but I'm not sure how that is different from random – maybe you literally sample everywhere in the global but only random locations in the latter? You then mention the results with bias in the inside vs outside models followed by a statement that you developed a multistate version of the JLRLR, but I'm not sure what states are included or why this is included here. Did you develop a new model that reduces the bias after you tested the original model? --Those were my thoughts before reading the paper – I can see now what these terms mean, but I think to get the reader interested, you need to have more clarity in the abstract. There's a

lot going on in this paper so it's a challenge to highlight the right things, but I think just adding in a few details on the main take aways will really enhance the readability of the abstract. I have no idea what the word limit is on the abstract but you are under 300 words right now so I'm hoping there is some more space.

Author response: Yes- the reviewer is correct in their interpretation of global and random. We have defined global in the abstract as such: *"This model assumes resightings are global, or spatially widespread to encompass the entire area where animals can disperse to, an assumption that may not be realistic for widely dispersing species." We have also changed a sentence to state that the MS Barker formulation was developed to help improve/reduce bias associated with the single-state version: "In effort to help minimize bias, we developed a multistate version of the JLRLR and demonstrated reductions in survival bias compared to the single-state version for most designs."*

Are the research questions/hypotheses/predictions clearly presented? In general, the ideas are clearly presented. I think the main thing to build on in the introduction is a need for consistency in text and figures (see my notes below on the intro). In addition, it would be good to add a few hypotheses or expectations to the introduction – you must have had some idea what would happen because you developed another version of the model to account for differences in detection between sites.

Author response: We moved our hypotheses from the Methods to the end of the Introduction.

Does the introduction build on relevant research in the field? I think so, but I have to admit I'm more familiar with CJS and JS models and did not know anything about the JLRLR model prior to reading this.

Are the methods and analyses sufficiently detailed to allow replication by other researchers? My lack of familiarity with the Barker JLRLR model had me struggling a bit with details in the methods section. I think overall the authors have done a good job describing the necessary model details but there are some areas where the clarity could be improved – please see my notes below. Also, the code is submitted as a pdf, and I did run it because of this. I highly recommend that the authors make it available as an R script or other file type which can be more easily viewed and run.

Author response: We include code appendices as R files in the revision, as requested.

Are the methods and statistical analyses appropriate and well described? Overall, yes. Again, there are some areas where the authors could be a little clearer in describing the terms and model details, see below.

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? Not applicable.

Are the results described and interpreted correctly? Overall, yes! I think maybe a figure or two could be moved to an appendix as the figure to text ratio is pretty high.

Author response: In the revision, we have moved one table to an appendix and combined two figures into one double-panel figure.

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? Overall, I think the discussion is good. However, I would like to see some more information on why the high movement case leads to more bias in the multistate model especially for the permanent emigration scenario. And why does it get reduced when immigration and emigration are allowed? What does movement look like under the two scenarios (ref, high) relative to the study area? It would be great to relate this back to the case study you are interested in as well – like how much are fish moving relative to the size of your plots? See my other notes below.

Author response: One potential reason why the high movement case has higher bias in the multistate Barker is that the high movement scenario leads to lower estimates of resight probability (R) and this could affect survival estimation. We have added descriptors of the emigration and movement rates in the main body of the paper to help readers understand what the scale parameter on the Cauchy distribtuion actually means in terms of movement.

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? Yes, I think the authors do a good job of not overstating the implications. That said, I think they could take a few of their findings and make a few bigger statements/recommendations based on what they found (particularly what they were seeking information on for their own study in terms of the use of fixed resight locations)

Larger concerns/questions:

1. I'm not sure what is reasonable, but I imagine that it would be hard to survey half of the stream you are interested in for your fish (in fact you say that large sections of the river are never sampled). But the study design you implemented surveys about half the stream section (unless I'm misunderstanding) and you don't vary this at all in the simulations. I also noted below that even your low-p simulation is likely to have a high overall detection probability because of the number of occasions. So, I guess my question is, are your simulations all based on large amounts of sampling with pretty high detection that wouldn't be achievable in your own system?

Author response: Yes- the reviewer is correct that some of the sampling designs suggested are not easily achievable/realistic for fishes in large rivers. Because we wanted to focus more on bias associated with sampling design, and not bias associated with low values of 'p', 'R', and % of area sampled, we chose to simulate with fairly high values of R & p. However, we have

included a more realistic sampling scenario that we term 'low observability' as part of Appendix A. These simulations provide additional insight as to biases that can arise under low R & p. Additionally, we have deleted the low-p simulation from the main paper, because it was misleading (i.e., while p was low, R was still high so that there was little difference between ref & low-p).

2. There didn't seem to me to be a way to understand what high movement means in your simulation. Yes the scale parameter of the Cauchy is 5, but what does that mean in your stream? What are the units of the stream? How many sites could a fish cover between occasions? Is this realistic with your actual study species? How often do the animals move (at one point you say you simulate survival and movement as continuous processes, but then say they can move and die between capture and resight periods which sounds discrete to me – am I missing something?) And are the animals considered inside a stream reach or not (binary, as opposed to having a "home range" center which you talk about how movement about that center can impact detection). In addition to better understanding this scenario, I would like you to discuss more why your multi-state model performed worse than the single-state model for the high move scenario.

Author response: Yes- it is difficult translate the Cauchy distribution into how animals move and it is unitless, so 1 can mean 1cm or 1km or 100km. The only movement relevant to this model is movement into and out of the sample and resight sites. We calculated the true emigration probability for our simulations where the scale parameter was 1 versus 5 and report this in the study and the proportion of animals that moved 1, 5, and 10 spatial units. This level of movement may be too high for our study species, so we have included a more realistic spatial unit for simulations in Appendix A.

Other comments/suggestions: Without page numbers or line numbers, it was really hard to write specific comments so I often just copy/pasted the authors' text and I did my best to sign post where comments are coming from but it could be a beast to read and for that I apologize.

Author response: Sorry about that, arXiv didn't allow line numbers.

Intro:
2nd paragraph – what are auxiliary detection data? Is that a specific term for the JLRLR data collection?

Author response: we have changed this to 'resight data'

Figure 1: why isn't the first period capture occasion (1) since you say animals are released at that time….what is that first occasion on the figure really supposed to be? Is the "Ideal design" the "Global design"?

Author response: The model is similar to a CJS model in that it conditions on first capture- so the first time where capture probability can be estimated is the second occasion. We have changed the title on the figure to Global design to be consistent with the rest of the manuscript.

3rd paragraph – first sentence: the figure doesn't show the resight probability is the same (the assumption), in fact, it shows under the second scenario that this is clearly violated. Be clear what you are referencing with the figure at the end of this sentence.

Author response: We have changed this reference from (Figure 1) to (Figure 1, top panel).

Last paragraph: define "memory behavior", or add a citation. Is "home ranging" a word?

Author response: We have changed this to "movement around a home range" and "breeding site fidelity"

Last paragraph: maybe I'm just not familiar enough but fixed versus random for design is just not clear to me at this point. Is random synonymous with ideal design in Figure 1, if so, then use I suggest using ideal here instead.

Author response: No- the ideal design is the global design. We have changed Figure 1 to reach 'global design' instead of 'ideal design'. Figure 2 is a useful illustrative example for the different designs.

Methods

1st paragraph: "Adult survival estimates from models of humpback chub in a fixed site in western Grand " - what does fixed site mean here, is it just 1 location for both marking, release, and resighting?

Author response: We have added '(capture site)' here to clarify.

1st paragraph: For the next part, I like the framing here, I'm just confused on if you are actually only doing resights during these visits to the site or if you are capturing and releasing? Maybe you could add a few details here on how you actually collect the data – are resights from PIT tags or from physical capture?

Author response: We have rearranged this section to help improve clarity. The trips that visit 'resight sites' outside the sample site do mark and release fish, however, the Barker model only describes fish after they are first released in the capture site.

2nd paragraph: "and continuous capture-only (resighting) data from autonomous antenna arrays (references)." What does "references" mean here?

Author response: That was a mistake 😊 – I forgot to fill in references. I have added some.

2nd paragraph: Methods are typically written in past tense.

Author response: Yes- our Methods are written in past tense.

Table 1: this could maybe go in a supplement and add the relevant information to the text and/or caption of Fig.2

Author response: We have moved this table into an appendix

Figure 2 caption: can you add the number of sites? Also, this red/green combo is probably not great for people that are red/green colorblind….I would suggest using colorblind friendly palette.

Author response: We have changed from red to orange. The number of sites is in the main text and in the Appendix, so we did not add it to the caption.

Single state Barker
1st paragraph: "Note that occasion 1 corresponds to release-only"….I'm just not following the design here. When would I have a situation where I release animals but it's not a capture occasion (the first one on the Figure 1) and then why is the first capture occasion 'release-only' if there are marked animals out there? I'm afraid this just isn't intuitive to me.

Author response: The model is similar to a CJS model in that it conditions on first capture- so the first time where capture probability can be estimated is the second occasion.

1st paragraph: "so that 11 capture probabilities are estimable" - are resight probabilities separate from capture and recapture probabilities? This indicates, unless I'm reading it wrong, that you are calling all of these the same?

Author response: Yes- this is described earlier in that paragraph "Parameters in the single-state Barker model include S (survival), p (capture probability for animals in the capture site), R (resight probabilities for live animals inside and outside the capture site), F (probability of emigration from the capture site), F' (probability of immigration into the capture site), and r

(probability a dead tag is reported)." We have changed some of the wording in this paragraph to help improve clarity.

2nd paragraph: did you divide the river reach into 50 equal but contiguous "sites" such that the global model would then sample all 50 for resights? A few more details would be helpful here.

Author response: We have added '(i.e., in all 50 segments)' to help with this. The sites where animals can be resighted is also described in table (Appendix C) and illustrative (Figure 2) formats.

2nd paragraph: If the resighting is continuous, then how can you estimate survive from capture to resight and then resight to capture? Again, I think maybe Figure 1 is not consistent with what you are describing here….it sounds more here like you have discrete resight occasions (maybe they are longer or something). I'm just struggling with how this looks in terms of the sentence here – which indicates that there is a capture-to-resight survival, but Figure 1 indicates the resight period starts right at the end of the capture occasion.

Author response: The time step for the simulation does not match that of the model. This was intentional to create a scenario where animals could be resighted but die between capture occasions. We have reworded this section as follows – "Animals could move among the 50 sites that occurred along a river (i.e., line). In the Barker JLRLR model, mortality and movement occur during the continuous resight interval so that animals can be resighted but die before the next capture occasion. To simulate this, we 'broke up' each capture-to-capture interval into two sub-intervals (capture-to-resight and resight-to-capture). This allowed animals to be resighted and die before the next capture event, thus imitating a continuous survival process across capture-to-capture intervals. Note that the survival probabilities estimated by the model correspond to the square of the simulated survival probabilities from sub-intervals (i.e., 80% survival over capture-to-resight and from resight-to-capture, corresponding to 64% survival from capture to capture)."

4th paragraph: I like the figure and table, but I think it would be really helpful to describe things like – what does random mean? It looks like you perhaps randomly select 24 sites each time period for resights but this is just not spelled out anywhere, you have to infer it from the figure. In table 1, it looks like you select the random sites from those that are not already selected for in the fixed sites…again, I'm inferring that but I think just a couple sentences spelling it out would be useful.

Author response: We have added some text to help clarify.

Table 1: why select 26 sites in the random + fixed but only 24 in the fixed and random only designs?

Author response: This is because we wanted to achieve a better balance of fixed vs random sites and also maximize spatial spread for fixed sites in designs that included both fixed and random sites. Despite including two extra resight sites and some random sites, the designs with fixed and random sites didn't really perform better than just fixed sites.

Figure 3: Is there a reason to define m_A,BB and m_B,BB as opposed to m_A,AA and m_B,BB for example or does it just not matter which is the relative category? Also this Figure has capture and resight interval numbered different from how you state it above…"During each capture occasion (i.e., odd-numbered time periods 1,3,5,…23)," this figure indicates that capture occasions would be 1 – 12 and resight 1-12.

Author response: The main point is that there is one parameter that is estimated by subtraction, which we happened to choose as m_A,BB and m_B,BB. The multistate Barker simulations in this model were exploratory. If there was interest in using the MS Barker model for a real-world application, more work could be done to determine if(how) survival estimates are affected by which movement parameter is estimated by subtraction. We have changed to read 't+2' for the last capture occasion, based on the reviewer's suggestion.

Multi-state JLRLR model

2nd paragraph. Since you have essentially created discrete intervals for capture and resight (I think), and you estimate survival between those states and allow movement, I don't really understand why you need these "secondary" states….isn't an animal just surviving with Phi_AA if it's captured in A at time t and then resighted in A at t+1 and has "transition probability" or movement of say theta_AA, This feels like a complicated way to write the probabilities for standard multi-state model.

Author response: We wanted to be able to model animals that could be captured at time t in state A, resighted ant time t+1 in state B, and captured at time t+2 in state A. This also allows for animals to be captured in A at time t, resighted in A at t+1, and move into state B (i.e., move between resight and capture) before capture occasion t+2 OR captured in A at time t, resighted in B at t+1, and remain in B (i.e., move between capture and resight). Simpler versions of this model could be constructed, for sure.

Results

Single-state Barker JLRLR model

1st paragraph: I'm pretty sure this is the first time we are seeing the terms high-S and high-move and quite honestly, I forgot what these scenarios are because it's been a while so maybe a quick reminder.

*Author response: Sorry- we realized we didn't abbreviate these appropriately earlier in the Methods. We have corrected this to read: **In addition to the above-described reference simulation set (ref), we also simulate the following scenarios: high survival (high-S) with survival = 90.25%, and high movement (high-move) where the scale parameter is set to 5 on the Cauchy distribution[1].**" Hopefully this is helpful now in the results section.*

Figure 4/5: all of the in text discussion is percentage, so why not put the figures in percentages too?

Author: Good point -we have switched to percentages in these figures.

Figure 4/5: It seems like the low-p scenario is really similar to the reference. Do you think this is possibly because even in the low p scenario, the overall detection is still quite high? What would you expect if detection were even lower, like 0.05? What is the detection of your fish in your study?

Author: We agree with the reviewer that the detection was still quite high for low-p and that it was misleading to have that as a scenario, so we have deleted it form the main paper. Instead, we have included a more realistic scenario for our study species as part of Appendix A that includes lower p, R, and fewer resight sites.

Figures 6/7 – could probably be a panel in 1 figure
Author: We have made this change.

3rd paragraph: here you switch to relative bias in decimals.

Author: In the revision, we now report percent relative bias and have changed this throughout the paper.

For the c-hat paragraph, it says values ranged from 1.0-1.5, but the table indicates a much narrower window, 1.24-1.40.

Author response: The c-hats in the table correspond to the mean c-hats across all five replicates. We have added wording to clarify.

c-hat paragraph: You say "suggesting that as capture probability increased, lack of fit may be more diagnosable through ĉ simulations." I'm not following the reasoning here, why

[1] This translates to an emigration probability of ~92-94% (conditional on survival) across capture occasions, and where ~85% of animals in the central site (i.e., can move no more than 25 units) move at least 1 spatial unit, ~43% move at least 5 spatial units, and ~19% move at least 10 spatial units.

would a higher detection make it easier to detect lack of fit? There doesn't seem to be a pattern here with the other scenarios between the low-p and reference p, and this statement suggests that you think all of the cases should have a lack of fit but you just can't detect it – is that correct?

Author response: We have deleted this because we no longer report the low-p scenario in the main paper (instead- we have Appendix A which is a better reflection of low-p/low-R).

"Table 1" – this Table should be table 2. I think you can move this to a supplement as it's really a whole table just to show the c-hat values, which aren't that variable.

Author response: Good catch, but since the old Table 1 is now an Appendix, this is actually Table 1 in the revision.

Multi-state Barker JLRLR model
Figure 9 – you could probably move this one to a supplement and just say the results were similar but had less bias.

Author response: In an effort not to have too many supplements, we have retained this in the main text.

Discussion
1st paragraph: Add percent sign to relative bias amounts
Author response: Done.

2nd paragraph: This is picky, but you are saying resight probabilities and then using percents
Author response: Probabilities are often expressed as percents.

3rd paragraph: Is the ideal design the global design (just using 1 term throughout would be helpful, unless they are different, then clarifying the difference is needed).

Author response: Yes- we have changed 'ideal' to 'global'

3rd paragraph: "the multistate model includes more nuisanceparameters associated with movement and resight and, while our simulations illustrated that it reduced bias compared to the single state version, both positive and negative biases were still present." 1. nuisanceparameters should be 2 words, and 2. Unless I'm reading the figures wrong, you really only saw a small amount of positive bias in 1 scenario but this sentence makes it seem like positive and negative bias were similar.

Author response: We have fixed 'nuisance parameters'. The sentence is saying both positive and negative biases are present, not that they are equal in magnitude.

Review by Rémi Fay, 29 Jan 2024 13:15

In the current study, the authors explore the consequences of the violations of model assumption on survival estimate for the Barker joint live-recapture/live-resight model. They focus on the spatial heterogeneity in the resight pressure, something that is likely to be common in practice but which is not expected by the standard version of this model.
I'd like to thank the authors for this work, I enjoyed reading it. Overall, I think this is a useful study that is tackling an important potential bias in CMR monitoring. I have several comments and suggestions to improve the study's clarity, and potentially to strengthen the results. My major concerns are about the clarity of the method section and the lack of consideration for the influence of sample size.

Author response: Thanks – to address these concerns, we have included an analysis of sample size as an appendix in the revision (Appendix B). We found no difference in directionality of biases across designs, though scenarios with lower sample size had higher uncertainty in survival estimates. We did not evaluate sample size effects in the multistate Barker model, but we do mention how the large number of parameters would likely require large sample size : "However, the high number of parameters in the multistate version will likely limit its use to studies with high sample sizes (e.g., lots of animals released, recaptured, and resighted)."

Majors comments
The method section gives all needed information but it could be better structured. Adding further sub-sections would be helpful. The first paragraph is about motivations. The second gives an overview but is a bit confusing. Then, the following section gives more details about the Barker joint live resight model, but this information is mixed with the description of the virtual study site, how individuals life-histories have been simulated. But this is the same for both models (classical and multi-state JLRLR model). Thus, I would expect that this is explained once and for all at the beginning of the section.
More generally, it would be nice to see the successive steps more clearly: (i) description of the virtual study site, (ii) how individual survival and move, (iii) how capture histories are simulated based on these data, (iv) models are used to estimates survival.

Author response: We have added subheadings and restructured the Methods section to be more in line with the Reviewer's suggestion. Think it reads more clearly now (hopefully) – thanks!

I have some concern with the GOF comparison but note that I am not a specialist of GOF, nor familiar with the median c-hat methodology. Thus, the authors will correct me if I am

wrong. Median c-hat methodology provides a poor estimation since it corresponds to the average value of the distribution of all possible c-hat. It seems to me that this estimation should be strongly affected by the range of c-hat that is considered as possible. While c-hat estimate from the median c-hat methodology could be better than nothing in a specific application of a CMR model, I am not convinced that this technic is appropriate to compare c-hat among models. It seems to me that is asking too much to this rough methodology. Furthermore, looking at their values in table 1, I found that they were all very similar (range from 1.24 to 1.40).

Author response: Yes- goodness of fit for these models is confusing. We could have done some GOF tests for CJS models, but I don't think that would have told us much as there is no violation in CJS model structure for fish in the capture site. McCrea et al. (2014) mention GOF testing, but this seems to be only for models with dead recoveries (I don't see any info on F and R parameters that are typically in the Barker model). McCrea et al. (2014) also mention that fitting only to the live recaptures and ignoring dead recoveries can be problematic.

We have elaborated on the c-simulation tool in MARK by adding the following: "To our knowledge, while there have been some GOF tests developed for the Barker model with dead recoveries (McCrea, Morgan, & Pradel 2014), there are no GOF tests specifically designed for the Barker JLRLR model with live resights. Because no contingency-table GOF tests were available, we evaluated GOF for simulated models using the median-ĉ simulation tool in Program MARK (White & Burnham 1999). According to Cooch (2008), this median-ĉ simulation tool is still a 'work in progress' that is useful for diagnosing lack of fit that is due to extra-binomial noise but may not be able to diagnose all different types of lack of fit."
Yes- the c-hat estimates are all very similar and all fairly close to 1, indicating no major violations in model fit that can be diagnosed by the c-hat simulation tool.

One thing that is missing in these simulations is the effect of sample size on model performance. This could be critical for the performance of the Barker joint live resight model, and even more for the multi-state version. These models are very interesting but their advantages come with the cost of higher data requirement. Thus, it is unfortunate that sample size received no consideration in this manuscript which aim to assess the robustness of these models. This may also affect the comparison with the CJS model. It would be very insightful to see, at least for one or two sampling designs, how model performance changes with the number of individuals released at each capture occasion, all other thing being equal.

Author response: Thanks – to address these concerns, we have included an analysis of sample size as an appendix in the revision (Appendix B). We found no difference in directionality of biases across designs, though scenarios with lower sample size had higher uncertainty in survival estimates. Increasing sample size (from 100 releases per site to 400 releases per site), did not improve estimation so that the comparison with CJS should be similar.

The figure 1 is very important and I think it can be rework to improve the readability. I am very familiar with CMR models but it takes me some time to understand this figure. Some suggestions: keep the minimal number of individuals needed to explain the different possibilities. There is a lot of text, are all information equally important? The X in the legend let me believe that there were cross on the figure, while they are illustrating letters. Use Y? or a,b,c…? The time could be shown more explicitly. It is not super easy to understand the meaning of the colored dotted and solid lines below the circles. Are the colors really informative? Use gray instead of yellow for individual dying during the interval? May be use a square instead of a circle for dead individuals. Use circles in the legend for the meaning of white, blue and orange since they are circles on the figure.

Author response: We have simplified Figure 1 and taken many of the Reviewer's suggestions

Others comments

Because there was no line number on the manuscript, my location indications are limited to the page. Please double check that line number are provided before any submission. Throughout the manuscript, it is not always easy to understand which sites the authors are talking about (capture or resight), as for instance in the last sentence of the introduction. I think this could be easily solved using a specific term for the capture site (e.g. capture area). Page 3: add permantly after "did not move out"

Author response: Done

Page 6: remove "(mostly)", close to the end of the paragraph

Author response: Done

Page 9: use 0.8 instead of 80% since it is a probability. The same for all probabilities.

Author response: We have expressed all probabilities as percentages, which should be acceptable.

Page 9: I suggest rephrasing: low capture probability with p = 0.2 high survival between capture occasions with S = 0.9025,

Author response: We have expressed all probabilities as percentages, which should be acceptable.

Page 9: The opposite way seems more logical to me: designs that included resight sites in the capture site inherently had higher resight probability than designs that excluded resight sites from the capture site.

Author response: I think we are saying the same thing? I have re-worded it in attempt to make it easier to understand: "Because most animals did not move out of the capture site, designs where resights could occur in the capture site (e.g., designs 3 & 5) inherently had higher resight probability than designs where resights could not occur in the capture site (i.e., designs 4 & 6)."

Figure 2 : Remove the « 5) » within the top right panel. And be careful, some yellow resight sites which are excluded from the capture site are actually slightly overlapping the capture site on the figure (e.g. panel 4).

Author response: Corrected.

Page 11: remove "(e.g 9&10, 45&46)". This is more confusing than helpful.

Author response: Done

Page 12: Please specific which data have been used to fit the CJS model (only capture data I guess).

Author response: Done

Page 12: Missing comas after "randomly chosen)", "fixed resight sites", "(otherwise 0)", "3 & 4 for the reference"

Author response: We have added after randomly chosen and (otherwise 0), but we disagree on the other commas

Page 13: "Bayesian framework" instead of "Bayesian models"

Author response: Done

Page 13: found strange to say that "p […] can differ inside and outside the capture sites" since this parameter is specific to the capture site.

Author response: we have deleted 'p' here.

Page 14: Scenarios with permanent emigration or immigration/emigration have to be better introduce in the method section. They are too briefly mentioned so that reading the result

we may not understand what is the permanent emigration model. It could be useful to better summarize which scenarios have been used to simulate the data. See my major comment about the structuration of the methods section.

Author response: We have restructured the Methods section. We have also added more description of the permanent emigration and immigration/emigration scenario.

Figure 4 & 5: Merge these two figures in a single panel? Would help for comparison. The legend is similar for both. Please specify from the first sentence of the legend that it is the bias in survival estimates (the same for the others figures).

Author response: We did make a version where we tried to put both graphs on the same figure, but the figure was difficult to read so we have left them as separate figures in the revision.

Page 16: Bias of what? To be crystal clear, specify which parameter is referred to.

Author response: Done

Figure 6: There are two purple points which are far from the core distribution of the estimates. I wonder if the number of simulations use is high enough to reliably depict the distribution of all potential estimates. Please, specify what is the vertical line. The detail ofthe bias computation could be move from the label of the axis to the legend. In this way, you may provide clear definition without abbreviation.

Author response: There are sometimes outliers in these analyses – the general point of the paper was to look mostly at the general trend and not specifically to evaluate the properties of the outliers. We have added text describing the vertical and horizontal lines, and we've added a description of relative bias to the figure caption.

Page 19: Again, bias of what?

Author response: Done

Page 22: "Our simulation results demonstrate that heterogeneity" may be specify that it could be spatial or temporal ?

Author response: we have added 'spatial'

Page 22: "Our results suggest that substantial biases in survival may arise when" Specify the type of model you are now talking about to make the distinction with previous sentences crystal clear.

Author response: We have changed wording

Page 22: remover "that were" in "(i.e., animals that were in the capture site were more" and add a coma after closing the bracket.

Author response: Done