




# Peer Community In Ecology

## In the quest for estimating true survival

**Matthieu Paquet**  based on peer reviews by **Rémi Fay** and 1 anonymous reviewer

Maria C. Dzul, Charles B. Yackulic, William L. Kendall (2024) The importance of sampling design for unbiased estimation of survival using joint live-recapture and live resight models. arXiv, ver. 3, peer-reviewed and recommended by Peer Community in Ecology.

<https://doi.org/10.48550/arXiv.2312.13414>

Submitted: 31 December 2023, Recommended: 30 October 2024

**Cite this recommendation as:**

Paquet, M. (2024) In the quest for estimating true survival. *Peer Community in Ecology*, 100637. [10.24072/pci.ecology.100637](https://doi.org/10.24072/pci.ecology.100637)

Published: 30 October 2024

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

---

Accurately estimating survival rate and identifying the drivers of its variation is essential for our understanding of population dynamics and life history strategies (Sæther and Bakke 2000), as well as for population management and conservation (Francis et al. 1998, Doherty et al. 2014). Many studies estimate survival from capture–recapture data using the Cormack–Jolly–Seber (CJS) model (Lebreton et al. 1992). However, survival estimates are confounded with permanent emigration from the study area, which can be particularly problematic for mobile species. This is problematic, not only because CJS models under estimate true survival in populations where permanent emigration occurs (i.e. they estimate “apparent” survival), but also because some factors of interest may affect both survival and emigration (e.g., habitat quality, Paquet et al. 2020), leaving the interpretation of results challenging, for example in terms of management decisions.

Several methods have been developed to account for permanent emigration when estimating survival, for example by jointly analyzing CMR data with data on individuals’ locations at each capture/resighting site (to estimate their dispersal distances; Schaub and Royle 2013, Badia Boher et al. 2023), with telemetry data (Powell et al. 2000), mark recovery data (Burnham 1993, Fay et al. 2019), or with live-resight data (Barker 1997).

The Barker joint live-recapture/live-resight (JLRLR) model can estimate survival when resight data are continuous over a long interval and from a larger area than the capture recapture data. This model becomes particularly promising with the growing collection of data from citizen science, or remote detection tools (Dzul et al. 2023). However, as pointed out by Dzul et al., this model assumes that resight probability is homogeneous across the area where individuals can move, and this assumption is likely violated for example because of non-random movements or because of non-random location of resighting sites.

In their manuscript, Dzul et al. performed a thorough simulation study to evaluate the accuracy of survival estimates from JLRLR models under various study designs regarding the location of resight sites (global, random, fixed including the capture site, and fixed excluding the capture site). They simulated data with varying survival and movement values, varying recapture and resight probabilities, and varying sample sizes. Finally, they also

developed and fitted a multi state version of the JLRLR model. They show that JLRLR models performed better than CJS models. Survival estimates were still often biased (either positively or negatively) but they were less biased when resight sites were randomly located (rather than at fixed locations), when recapture sites were included in the resighting design, and when using the multi state JLRLR model they developed.

This study highlights (multistate) JLRLR models as an alternative to CJS models one should consider when auxiliary resight data can be collected. Moreover, it shows the importance of evaluating not only model performance, but also the efficiency of alternative sampling designs before choosing one for our studies. Hopefully, this study will help the authors and other researchers making a more informed and efficient choice of model and design to estimate survival in their study populations.

### **References:**

- Jaume A. Badia-Boher, Joan Real, Joan Lluís Riera, Frederic Bartumeus, Francesc Parés, Josep Maria Bas, and Antonio Hernández-Matías. Joint estimation of survival and dispersal effectively corrects the permanent emigration bias in mark-recapture analyses. (2023) *Scientific reports* 13, no. 1: 6970. <https://doi.org/10.1038/s41598-023-32866-0>
- Richard J Barker (1997) Joint modeling of live-recapture, tag-resight, and tag-recovery data. *Biometrics*: 666-677. <https://doi.org/10.2307/2533966>
- Kenneth P. Burnham (1993) *Marked Individuals in the Study of Bird Populations* (ed. J.D. Lebreton), pp. 199–213. Birkhäuser, Basel
- Kevin E. Doherty, David E. Naugle, Jason D. Tack, Brett L. Walker, Jon M. Graham, Jeffrey L. Beck (2014) Linking conservation actions to demography: grass height explains variation in greater sage-grouse nest survival. *Wildlife biology* 20, no. 6 : 320-325. <https://doi.org/10.2981/wlb.00004>
- Maria C. Dzul, Charles B. Yackulic, William L. Kendall (2023) The importance of sampling design for unbiased estimation of survival using joint live-recapture and live resight models. arXiv, ver.3 peer-reviewed and recommended by PCI Ecology <https://doi.org/10.48550/arXiv.2312.13414>
- Rémi Fay, Stephanie Michler, Jacques Laesser, and Michael Schaub (2019) Integrated population model reveals that kestrels breeding in nest boxes operate as a source population. *Ecography* 42, no. 12: 2122-2131. <https://doi.org/10.1111/ecog.04559>
- Charles M. Francis, John R. Sauer, Jerome R. Serie (1998) Effect of restrictive harvest regulations on survival and recovery rates of American black ducks. *The Journal of Wildlife Management* : 1544-1557. <https://doi.org/10.2307/3802021>
- Jean-Dominique Lebreton, Kenneth P. Burnham, Jean Clobert, David R. Anderson (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs* 62.1: 67-118. <https://doi.org/10.2307/2937171>
- Matthieu Paquet, Debora Arlt, Jonas Knape, Matthew Low, Pär Forslund, and Tomas Pärt (2020) Why we should care about movements: Using spatially explicit integrated population models to assess habitat source-sink dynamics. *Journal of Animal Ecology* 89, no. 12: 2922-2933. <https://doi.org/10.1111/1365-2656.13357>
- Larkin A. Powell, Michael J. Conroy, James E. Hines, James D. Nichols, and David G. Krentz. Simultaneous use of mark-recapture and radiotelemetry to estimate survival, movement, and capture rates. (2000) *The Journal of Wildlife Management* : 302-313. <https://doi.org/10.2307/3803003>
- Bernt-Erik Sæther, Øyvind Bakke (2000) Avian life history variation and contribution of demographic traits to the population growth rate. *Ecology* 81.3 : 642-653. [https://doi.org/10.1890/0012-9658\(2000\)081\[0642:ALHVAC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[0642:ALHVAC]2.0.CO;2)

Michael Schaub, J. Andrew Royle. Estimating true instead of apparent survival using spatial Cormack–Jolly–Seber models (2014) *Methods in Ecology and Evolution* 5, no. 12: 1316-1326.  
<https://doi.org/10.1111/2041-210X.12134>

## Reviews

### Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.48550/arXiv.2312.13414>

Version of the preprint: 2

### Authors' reply, 16 October 2024

[Download author's reply](#)

[Download tracked changes file](#)

### Decision by [Matthieu Paquet](#) , posted 04 October 2024, validated 07 October 2024

Dear Authors,

First I really want to apologize for the time it has taken me to get back to you with this decision.

Your preprint entitled “The importance of sampling design for unbiased estimation of survival using joint live-recapture and live resight models” has now been reviewed, by one of the previous reviewers. As you will see, the reviewer was positive about the revision and had no more major comments, but still highlight a few points that you should carefully address before I can recommend your pre-print.

In addition to the reviewers comments I have the following feedback:

1) Regarding my previous comment on the use of STAN for one model vs MARK for another, thank you for your explanation. Please add it in the main text so that the reader (who may not look into the review history) can read it too.

2) Also you still didn't answer whether STAN was used in a Bayesian framework or not (but I guess so from the code). Please provide brief information on which priors were used, (even if in the appendix), and which diagnostics tests were performed for the convergence of the 3 chains (Gelman Rubin R hat statistics?) and assessing the fit of the model (e.g., any posterior predictive checks?).

3) Same regarding the answer to Reviewer 1's previous comment: “Table 1: why select 26 sites in the random + fixed but only 24 in the fixed and random only designs?”. Please justify in the text as well (including for the case with the updated number of sites).

I look forward to reading the revised version of this preprint.

Best wishes,

Matthieu

### Reviewed by [Rémi Fay](#), 11 September 2024

It is the second time I review this manuscript that assesses the influence of resight sampling design of a CMR monitoring on survival probability estimated by a Barker joint live-recapture/live-resight model. I appreciated the efforts the authors have done to improve the manuscript, especially rewriting of the methods, the clarification the conceptual figure 1 and the new complementary analyses. Overall, the manuscript is now

clearer. After reading this second version, I have no major concern but I find several points that could be improve.

While improvement have been made for the method section, I think slightly more could be done. Readable structure is key for manuscript clarity.

The first paragraph of the motivation section provides information that justify the simulation design. I suggest to rename this section "Motivation of the simulation design", because overall motivations are expected to be given in the introduction. The second paragraph is a mixed of introduction (l. 146-150), and simulation design (l. 151-156) which is fully developed in the following section. I suggest the move these information where they are expected to be (i.e. introduction and simulation design respectively). Additionally, the "Simulating animal movement" seems to appears out of nowhere to me. I think the second sentence could be merged with the motivation section, and the other could be reduced and merged with the paragraph explaining how data were simulated. The mention of the alternative scenario with lower recapture probability could be introduced along with the alternative scenario investigating the influence of sample size (l. 218).

During my previous review I expressed concern with the confusion between probability and percentage. These are two different things. I note that the other reviewer had the same concern. I think it is fair to make changes when reviewer concerns converge.

In my previous review, I suggested to specify which data have been used to fit the CJS model. The author response is "Done". But I was not able to find this information in the manuscript...

Abstract: I suggest to simplify "is often confounded with emigration" instead of "is often confounded with permanent and temporary emigration"

Note 1 (line 208): specify "from the capture site" after "emigration probability"

l. 219 "Resight sampling design" instead of "Resight process" ?

l. 220 Table 1 has been moved to the appendix. It should be Appendix C. Please check the correspondence between figure/table names throughout the text.

"Three" instead of "four" scenarios shown with colors on figures?

Please specify somewhere what is the number of simulations that have been done for each scenario.

l. 225: "inherently had higher resight events" instead of "inherently had higher resight probability" ? l. 226 Similarly could be "resight number" instead of "resight probability" ? I found these formulations with probability confusing because resight probability is fixed (0.8).

l. 299 remove "observed". I think this is not needed.

Would be nice to add a few sentences in the results section to mention the influence of sample size and detection probability on model performances with references to the corresponding appendix.

l. 454-455: "animals located into the capture site were more likely to be resight" instead of "animals resighted in the capture site were more likely to be resighted". To be crystal clear, would be useful to recall that individuals outside the capture site could be in site where there is no resight effort and so that the actual resight probability at the individual level is lower. Something trivial for you could be less for someone which read the manuscript for the first time.

Appendix A: I suggest to remove the figure A1 and the table A1 which are just repetition of figure 2 and appendix C. Please refer directly to these figure and table.

## Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.48550/arXiv.2312.13414>

Version of the preprint: 1

### Authors' reply, 01 August 2024

[Download author's reply](#)

[Download tracked changes file](#)

Decision by [Matthieu Paquet](#) , posted 27 March 2024, validated 28 March 2024

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.

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.

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.

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.

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

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?

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.

I look forward to reading the revised version of this preprint.

Best wishes,  
Matthieu

## Reviewed by anonymous reviewer 1, 26 March 2024

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.

**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.

**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.

**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.

**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.

**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.

**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?

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.

**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.

**Intro:**

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

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”?

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.

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

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.

**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?

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?

2nd paragraph: “and continuous capture-only (resighting) data from autonomous antenna arrays (refer-

ences).” What does “references” mean here?

2nd paragraph: Methods are typically 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

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.

#### *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.

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?

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.

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.

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.

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

Figure 3: Is there a reason to define  $m_{A,BB}$  and  $m_{B,BB}$  as opposed to  $m_{A,AA}$  and  $m_{B,AA}$  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.

#### *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.

## **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.

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

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?

Figures 6/7 - could probably be a panel in 1 figure

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



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

$\hat{c}$  paragraph: You say “suggesting that as capture probability increased, lack of fit may be more diagnosable through  $\hat{c}$  simulations.” I’m not following the reasoning here, why 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?

“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  $\hat{c}$  values, which aren’t that variable.

#### *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.

#### **Discussion**

1st paragraph: Add percent sign to relative bias amounts

2nd paragraph: This is picky, but you are saying resight probabilities and then using 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).

3rd paragraph: “the multistate model includes more nuisance parameters 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. nuisance parameters 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.

### **Reviewed by Rémi Fay, 29 January 2024**

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.

#### Majors comments

The method section gives all needed information but it could be better structured. Adding further subsections 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.

I have some concern with the GOF comparison, but note that I am not a specialist of GOF, nor familiar with the median  $\hat{c}$  methodology. Thus, the authors will correct me if I am wrong. Median  $\hat{c}$  methodology provides a poor estimation since it corresponds to the average value of the distribution of all possible  $\hat{c}$ . It seems to me that this estimation should be strongly affected by the range of  $\hat{c}$  that is considered as possible. While  $\hat{c}$  estimate from the median  $\hat{c}$  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  $\hat{c}$  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).

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.

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.

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"

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

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

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

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.

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).

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

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

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

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

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.

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.

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).

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

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 of the bias computation could be move from the label of the axis to the legend. In this way, you may provide clear definition without abbreviation.

Page 19: Again, bias of what?

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

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.

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.