**Reviewer: 1**

Comments to the Author

I find the subject relevant for the journal and after minor revision i can recommend it for publication.

The experimental setup is well conducted, the statistical methods clearly explained and in general the paper is well written.

The Abstract reads well and includes background, methods, results and conclusions.

The Introduction chapter is well written and informative.

The Materials and methods chapter is clearly written, I have only minor comments and questions.

*We thank this reviewer for a very helpful review that has clarified a number of important points within the revised manuscript.*

P5-L9: Mesh gauge measurements, how many meshes measured, the force used and SD or SE should be standard information for cod-end measurements.

*Forty meshes were measured at 125 Newtons per cod-end. Details of these measurements and the standard errors are now included on Lines 88-89.*

P6-L10: Each cod-end is set as a category in the model, could it somehow be dealt with as an ordinal/continous variable? Instead of categorising by cod-end type, can you categorise trawls as A, B, C, D from starboard to port and add mesh size as a choice-specific attribute? The differences between hauls would then be included in the random effects I guess.

*We thank the reviewer for this thought-provoking comment, however we affirm that the choice of nominal mesh-size categories as the response is appropriate because it reflects the experimental design in that mesh size is what differs among the cod-ends and we are most interested in the retention by gear type as opposed to retention from port to starboard. We address port-starboard differences via a position effect, which has been further developed in line with both reviewer’s comments on position effects.*

*An ordinal response would likely make an a-priori assumption on the direction of the proportion retained across increasing mesh sizes (proportional odds assumption). Treating the response as nominal categories relaxes this assumption and is more generally applicable, e.g., when sorting grids are trialled. As we are presenting the method as a general approach we think nominal response categories are most appropriate.*

What about including interaction terms, e.g. length-catch weight or mesh size to investigate effects of the slope (length) parameter? Information on adding interaction in the model would be of interest, different slope parameters are mentioned in the Results chapter (P10-L25).

*Thanks for highlighting this important point. In the revised manuscript, we now include two-way interactions between carapace length and bulk weight and carapace length and net position. These considerably improved the model fit (Table 2). We now present an effects plot to assist in interpreting the influence of the interactions on the proportions retained (Figure 7).*

P8-L10: 'Dirichlet' needs a reference (there is a reference to Thorsén 2014, but it is in the Discussion chapter on page 12).

*We now include the original derivation reference on Line 167.*

*Mosimann, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. Biometrika 49 (1-2), 65–82.*

The Results chapter is clear and well written, but catch composition and quantity by haul is missing. After all, the main objective, as I read it in the Abstract, is to provide statistical framework to investigate influences on catch composition.

*RONÁN TO UPDATE. Catch composition is statistically modelled via the proportions retained at length (Figures 2, 4, and 5).*

Discussion

P13-L5: "Net position, total cod-end weight and carapace length significantly affected the numbers of Nephrops retained..."? If the cod-end catch is mostly Nephrops, then the sentence reads like 'catch quantity is amongst factors that affected catch quantity'.

Here we need to consider what we are investigating; the main interest is the effect of explanatory variables mesh size and catch quantity on length dependent efficiency (selectivity).

*Catch weight per cod-end (see response to naming below) included other species of which Nephrops comprised approximately half the bulk weight across all hauls (noted on Lines 200-202). We are focussed here particularly on how the catch weight per cod-end affects Nephrops retention across length classes in different diamond mesh sizes. In this paper we focus on improving size retention. RONÁN AND DARAGH TO UPDATE.*

Also, catch weight would be more correct than cod-end weight, as the weight of the cod-end per se is a constant.

*Thanks for highlighting this important naming mistake! We now use “catch weight per cod-end” as the variable name; “cod-end” is retained to highlight that the weights are cod-end specific (four per haul).*

P13-L9: The statement that mesh opening angles affect selectivity may be true, but it does not explain the effect of mesh size on size selectivity.

*We have removed the original sentence and replaced with references on the effects of mesh size on Nephrops retention.*

For the fishery implication section, it would be logical to mention factors other than mesh sizes like twine thickness/stiffness, cod-end circumference and square meshes to compensate for 'poorer' size selectivity when using quad rig. These factors are covered in eg Frandsen (2010) which you refer to in other context.

*Lines 319-321 now includes other measures to improve Nephrops size retention in line with this comment.*

**Reviewer 2**

Review of ICESJMS-2015-625

A general catch comparison method for multi-gear trials: application to a quad-rig trawling fishery for Nephrops by Daragh Browne, Cóilín Minto, Ronan Cosgrove, Brian Burke, Daniel McDonald, Rickard Officer, Michael Keatinge

This is a really interesting methodological paper which develops a multinomial mixed model framework for investigating catch comparison data from multiple codends. The methods are sound, the availability of code means that they are likely to be used, and the paper should certainly be published. However, work is still needed to improve the model development. Also, multinomial models can be really difficult to interpret and, if the method is to be used properly by others, it is incumbent on the authors to describe and interpret their own results more fully. Finally, I would like to see more discussion of how further work might incorporate the data structures that often arise in catch comparison trials.

*We thank Dr Fryer for a very helpful review. We hope to have addressed the points below and generally improved the model and interpretation for users.*

Model development

1. p6, l1-6. It took me ages to get my head around the notation, and the problem is simply that the matrix Y isn’t well described. What is an observation? I eventually worked out that it is a combination of length class and haul, and n is the total number of length-haul combinations. But my natural inclination was to think of the haul as the observation, and I was looking for an h subscript on the Y matrices.

*We have updated the definition of Y to describe that each observation/row is a length class in a given haul and now use a haul and length-class subscript (h,i) throughout.*

1. p7, l4-9. Didn’t understand the modelling of catch weights at all. What is W i? Is it the total catch weight for haul hi across all four cod-ends? Why would this have a common effect on all four cod-ends? Much more explanation needed.

*Wi is now Wh,j throughout, that is cod-end specific total catch weights in a given haul. We have given serious consideration to the effect of catch-weight on the proportions. We agree with Dr Fryer that the effect should not be considered common across cod-ends. Choice-specific effects are common where the variable differs by choice, however, the restriction of a common effect here is unrealistic. What we’ve now implemented is a cod-end-specific weight effect. This allows for the retention per cod-end to differ depending on the cod-end weight.* Treating weight as such resulted in an improved model fit (Table 2). For interpretation of the effects, we now include an additional figure (Figure 7).

1. eqns 2, 3, 4, 7. Should turn subscripts round on the πs to make them consistent with the subscripts on Y.

*Changed in the manuscript.*

Model interpretation

1. I get the feeling that some interpretational issues have been ducked. For example, there is a strong catch weight effect. Looking at Table 2, it is arguably the most important individual term. Yet the results only describe how AIC changes with catch weight and the discussion (p13, l11-15) only reiterates that catch weight was significant and affects selection. What was the sign of γ, and what does that mean for the relative catch rates of the four nets? I worry about models where the inclusion of terms is rather ‘black-box’.

*We thank Dr Fryer for this important issue with our original presentation. We now two features that we hope resolves interpretation difficulties: a table of parameter estimates (Table 3) and a fixed effects Figure 3 and discuss them on lines 215-226.*

1. There are strong correlations in the random effects. Does this mean that the 80, 90 and 100 mm cod-ends are behaving similarly and, in particular, differently to the (reference) 70 mm cod-end? Why would that be? I presume the correlation matrix would look rather different if the 100 mm cod-end was used as the reference.

*From the updated Figure 3, in 4 out of 12 hauls the sign of the random effects differs among the three odds ratios, indicating that they are not behaving similarly with respect to the sign of the effect across all hauls. However, it is clear that the random effects are strongly correlated, potential reasons for this include the order of hauling the quad-rig aboard. We discuss this on lines 272-274. With regards to the baseline, Hartzel et al. note that with an arbitrary/unconstrained covariance matrix, the model is structurally the same irrespective of the baseline category. We ran a model with 100mm as the baseline and recovered the same fixed effects parameters and likelihood, we discuss how the covariance matrices can be transformed on lines 274-282.*

1. Do all the four cod-ends differ from each other? Do e.g. the 90 and 100 mm cod-ends have similar catch rates? Some plots of the differences between the fitted proportions, or reporting the results of adjusted pairwise comparisons would help.

*We have included a figure containing the odds ratio between the six combinations with Bonferroni corrections to the pointwise confidence intervals (Figure 6). The 70/80 odds ratio decreases over carapace length but the confidence intervals span one throughout. The 90 and 100mm don’t differ whereas the others do. Additional work on the coverage of these confidence intervals would be required (e.g., fitting a model omitting non-significant parameters in Table 3) to draw stronger inference, as noted on lines 228-233.*

1. It wasn’t until I looked at Table 2 that I realised that the net-configuration term was an ‘interaction’. I initially thought it was a ‘main effect’ with four levels (inner and outer port and inner and outer starboard). But the results (p10, l1) do suggest such a main effect (‘inner port position typically fishing worst, and the other starboard or port fishing better’), so why not fit one and then add the interaction and see how important it is. Indeed, with more days fishing, you might treat this interaction as a higher-level random effect.







*Thanks for this extremely interesting point. We've now fit models with position main effects (Table 3) and position effects differing by mesh. Indeed, the main effects model containing just position fit better than one including the interaction. We think this is a very valuable result reflecting different fishing power.*

Model adequacy and structure

1. The random effects are clearly important, but how do you know that they are adequate? In selection and catch comparison studies it is common for both the intercept and the slope to vary between hauls. Here, for example, the carapace effects might vary between hauls. I don’t think the paper needs to try to fit such a correlation structure – it has progressed things plenty as it is – but I think there should be some more discussion about how correlation structures more complex than an ‘additive’ haul effect are often required and how they might be incorporated.

*We have included a discussion on this on lines 287-298.*

1. The proportions might be expected to converge to 25% at large lengths. I know the quadratic terms in length did not improve the model fit (at least based on AIC), but one of the problems of linear length relationships is that they can be driven by the proportions retained at smaller and intermediate lengths (where there are usually more individuals) with the fitted values at larger lengths then constrained to follow the linear relationships wherever they may lead. I have often found when fitting loess smoothers to catch comparison data that the relative catch rates asymptote to 50% for large lengths, even though the smoothers are not a ‘significant’ improvement on a linear model (just not enough power in the data). So there is a need to be award of the dangers of over-interpreting the fits at large lengths. One possibility might be to put some constraints on the length relationships (e.g. so that the proportions are all equal to 25% above some ‘large’ length) and see how that compares to the unconstrained model.

*We fully agree and include a section on future model developments that would allow for non-linearity to be more fully explored lines 297-303.*

Rob Fryer

1 February 2016