**General comments**

**R1 -** 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.

**R2 -** 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.

--------------------------------------------------------------------------------------------------------------------------------------

**Abstract (p1-2)**

**R1** - The Abstract reads well and includes background, methods, results and conclusions.

**Introduction (P3-4)**

**R1** - The Introduction chapter is well written and informative.

**Materials and Methods (P4-9)**

**R1** - The Materials and methods chapter is clearly written, I have only minor comments and questions:

**Trial design (P5)**

***R1*** *-* ***P5-L9****: Mesh gauge measurements, how many meshes measured, the force used and SD or SE should be standard information for cod-end measurements.*

**Model development (P6-9)**

**R1** - **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.

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)**.

**R2** - **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.

**R2 -** . **Equations 2, 3, 4, 7**. Should turn subscripts round on the πs to make them consistent with the subscripts on Y.

**Covariates (P6-7)**

**R2 - p7, l4-9.** Didn’t understand the modelling of catch weights at all. What is *Wi*? 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.

**Subsampling offset (P7)**

**Multinomial random effects (P7-8)**

R1 - **P8-L10**: 'Dirichlet' needs a reference (there is a reference to Thorsén 2014, but it is in the Discussion chapter on page 12).

**Inference (P8-9)**

**Estimation (P9)**

**Results (P9-11)**

**R1**-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.

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).

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

**Discussion (P12-15)**

**Model developments (P12-13)**

**R2 - Model interpretation**

I get the feeling that some interpretational issues have been ducked. For example, there is a strongcatch 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’.

There are strong correlations in the random effects. Does this mean that the 80, 90 and 100 mm codends 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.

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.

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.

**R2 - Model adequacy and structure**

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.

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 aware 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.

**Covariate effects (P13-14)**

R1 - **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.

**Fishery implications (P14-15)**

**R1** - 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.

(Covered in discussion but needs work)