**Cohort specific disability trajectories among older women and men in Europe 2004-2017.**

**Manuscript ID: EJOA-D-21-00095**

|  |  |  |
| --- | --- | --- |
| **Comments and suggestions** | **Authors’ reply** | **Pages** |
| **Editors comments** |  |  |
| I think the paper is really well-structured and presents some novel findings (i.e. cohort specific analyses). However, after reading the Discussion, I thought it would also be good to have more ideas and a better framing of your main findings. What are possible reasons for country difference (e.g. higher among younger people in Eastern)? Why are sex-differences greater in eastern and southern countries? Given that SHARE allows you to study such differences I think a broader conceptual framing and discussion (beside the empirical analysis) would make your paper stronger. What was the rational (beside the fact that this is how is was done before) to regroup countries into the four categories you have used? Do these country-group have similar policies? Do they differ in terms of gender-related aspects or family-policies? This not necessarily needs to be mentioned in the Introduction, but I think the Discussion needs to come up with some ideas. Maybe also because of different socioeconomic circumstances? There is much literature on this (based on SHARE as well) that could be useful and incorporated here. | We thank the editor for the kind words and for the thoughtful comments. We have now tried to address the comments and suggestions in the following fashion: * The four country groups are based on an adapted version of the Ferrera (1996) typology of welfare state models. We have now added an extended motivation of the grouping to the data and methods section.
* We have also added a paragraph in the discussion, where we discuss potential causes of the greater sex differences in disabilities observed in Eastern and Southern Europe. In addition, we have also included new models in the paper, adjusting for education and income. This adjustment partly, but not wholly, attenuate the sex-gap in disabilities.
 | * 3-4
* 8-9
* Figures 1 & 2 (and the corresponding tables in the appendix)
 |
| The results section is rather long and could be shortened by highlighting the most important findings from tables and figures. Also, I wondered why you have decided to present findings of pooled analyses in Figure 1 (since this is not really related to your main research question). Why not focussing on the country-specific analyses? | We appreciate this comment. We have now shortened the results section substantially, mainly by omitting the pooled analyses (as suggested). | * 5-6
 |
| Please briefly specify how estimated probabilities (denoted “estimated probabilities” in the figure) were exactly predicted? Are these average marginal predictions (based on “margins” in Stata)?  | We have now clarified in the methods section that we are (as guessed) estimating the probabilities using the ‘margins’ command in Stata. | * 5
 |
| I generally liked the fact that your models do not include too many variables, but I nevertheless wondered if you should not include additional variables to strengthen your finding on country differences and assuring that differences are not due to country-compositions (e.g. different educational background or different workforce composition?). These indicators are available in SHARE and could be included.  | We appreciate this suggestion. We have now included models adjusted for education and income in order to analyze to what extent the observed patterns can be attributed differences and changes in the distribution of socioeconomic conditions. | * 5
* 7
* 8-9
* Figures 1 and 2
 |
| Given that IADL and ADL are also available as count-variables (number of limitations), I think the reader may wonder why you have decided to use a binary indicator in both cases (1+ limitations). This includes loss of information and in case you have run the same analyses with count-variables you may mention this as an additional test of robustness? Or justify your decision more explicitly.  | The rationale for dichotomizing the variables is to make the distinctions between disability/autonomy and - on an aggregated level – to assess the accumulation of non-independent individuals, rather than to assess the average severity of disabilities. We have now added a passage to the methods section explaining this. | * 4
 |
| Please carefully check and follow the guidelines for authors, including change of abstract (unstructured) and change of reference style (incl. references in the text by name and year in parentheses). | We thank the editor for drawing our attention to this. We have now revised the manuscript in line with the author guidelines. | Throughout the entire manuscript |
|  |  |  |
| **Reviewer 2** |  |  |
| It is would be good to describe more why the authors decided to use mixed effect logistic regression (and perhaps compare with other methods, such as mixed effect probit regression or latent growth model). | We appreciate this comment. The main reason why we use the mixed effect model to start with is that it has been used successfully in previous studies to track cohort-specific trajectories (cf. Marshall et al. 2015). Since we wanted to use mixed effect regression models and our outcomes were binary, we opted for the logit link function. We could, just as well, have used a mixed effect probit regression model – as suggested by the reviewer. Yet, as the estimates are ultimately transformed into estimated probabilities – it is unlikely that a probit link function would yield substantially different results.  |  |
| What does the "robustness of the estimates" mean? | We agree that this formulation was unnecessarily ambiguous. What we meant to say was that the statistical power for analysing specific countries was too low, and such analyses would yield volatile and erratic patterns. Therefore we collapse several countries into regional clusters to “power up” and yield more stable patterns. We have now added a passage to the methods section to explain this. | * 3
 |
| Related to point 2 above, what is its connection with four regions-based groups? | The four region-based groups were used to garner enough observations in each, even in the oldest age groups, to get stable and credible results. This is now explained in the methods section. | * 3-4
 |
| It would be great to describe reason to create the cohort birth year group 1920-1929, 1930-1939, 1940-1949, etc. |  We appreciate this comment. We have now clarified, in the methods section, that the rationale for dividing the sample into these birth cohorts is to allow us to study the cohort-specific trajectories across subsequent birth cohorts.  | * 3
 |
| Is it possible to conduct sensitivity analysis using complete cases only? Or others? | We have conducted several sensitivity analyses (which are not presented in the manuscript, as the appendix is quite lengthy already). We have run the models without the weights, and we have run mixed effect linear models with the outcomes as ordinal variables. The results of both of these exercised show yield similar patterns – and have reassured us that the observed patterns are quite robust.  |  |