Reviewer's report

Title: Long term unemployment, income, poverty, and social public expenditure, and their relationship with self-perceived health in Spain (2007-2011)

Version: 0  Date: 13 Jun 2017

Reviewer: William Hulme

Reviewer's report:

This study uses two waves of the Survey on Living Conditions (2007-10 and 2008-11) to investigate the impact of unemployment length, poverty and regional social public expenditure with self-perceived poverty in Spain. Overall the methods and results, while suitable for showing an association between perceived ill health and unemployment length, do not support the conclusion of the 'casual' hypothesis over the 'selection' hypothesis.

**Major comments**

The decision to dichotomise perceived health results in a high loss of information. Consider retaining original groups.

Unemployment length is certainly shown to be associated with perceived health, but this model is not sufficient to establish a direct causal effect of unemployment length on perceived health, however plausible it may be. These are longitudinal observational data (with repeated measures each year), and although year has been included in the model the longitudinal aspect is not captured (relabel years 2007-2011 as A, B, C, D, E and reorder, and the model will be the same). See eg http://isites.harvard.edu/fs/docs/icb.topic156289.files/General/chapter_12jun07-1.pdf. An alternative hypothesis that perceived bad health leads to longer term unemployment is as compatible with the data as the hypothesis preferred by the authors. The reality is likely to be more complex than either hypothesis, with both factors influencing each other.

Is unemployment length counted from the start of the survey, or from the start of the when the person was unemployed? For instance, for data in 2007 for a person who has been unemployed since 2005, are they in the 'unemployed 0-12 months' or the 'unemployed 24-36 months' group?
If the latter, then what of people who were unemployed for longer than 48 months? There is no category for this. If the former, then you are conflating the long-term unemployed with those who are recently unemployed. This has strong implications for interpretation.

You include chronic illness to "control for bidirectionality in the relationship between unemployment and health". It's not clear how this works. For one thing, chronic illness may be a cause of unemployment (or inactivity) or itself may be caused by unemployment. Clearly, chronic illness is very highly correlated with perceived health (OR=23.52), and in many ways it captures the same thing. Such a high odds ratio suggests there are very few individuals with chronic illness who rate themselves as having good health. All this is doing is causing instability in your model because you have near-separation, and the estimates of the remaining odds ratios are drawing almost all their information from the no chronic illness group. You could remove this variable - yes, you're removing the variable that best predicts the response, but this prediction is unhelpful if you consider that chronic illness is (almost) a subset of those with bad perceived health. Further you could exclude all individuals who begin the period with a chronic illness. Thus you start only with 'objectively healthy' individuals and you can observe whether the likelihood that they perceive themselves to be less healthy is associated with extended periods of unemployment.

**Minor comments**

The distinction between "inactive" and "other inactivity" is not clear.

Table 2 should include the number of individuals in each group.

"Out of the total number of unemployed, 11.18% have been so for less than 12 months, 11.93% between 12 and 23 months, 5.25% between 24 and 35 months, and the remaining 2.49% for between 36 and 48 months in the period under consideration." Shouldn't these percentages add up to 100%? or else another group 'unemployed for longer than 48 months' makes up the rest?
"Fixed effects multilevel model" is an oxymoron as multilevel implies the inclusion of random effects, making it a "mixed-effects" model.

The description of multi-level models is unnecessary, but also very unclear. It's enough to describe the linear component of the model (b_0 + b_1 + b_2*Z_ik + u_jk + nu_k) so that the multilevel structure is clear, and drop the rest with suitable reference to relevant literature.

Not clear how 9003 individuals with "continuous presence in the data during the four years" can add up to on 34692 observations. Shouldn't this be 9003*4=36012? Unless there is an overlap of individuals in the two waves, or some individuals were not available in the first or last year of their wave. This should be clarified.

How many individuals who did not have a continuous presence were excluded? How might these bias results? For example if they were not included due to death (more likely to be in ill health), or because they moved abroad (more likely to be in work), this is a case of informative censoring.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal