Author’s response to reviews

Title: Is any job better than no job at all? Studying the relations between employment types, unemployment and subjective health in Belgium.

Authors:

Karen Van Aerden (kvaerden@vub.ac.be; karenvanaerden@hotmail.com)
Sylvie Gadeyne (sgadeyne@vub.ac.be)
Christophe Vanroelen (cvroelen@vub.ac.be)

Version: 1 Date: 29 May 2017

Author’s response to reviews:

The tables that are included in our 'response to the reviewers' document cannot be added in this box. They are available upon request.

We would like to thank the reviewer for his/her valuable comments on our manuscript “Is any job better than no job at all? Studying the relations between employment types, unemployment and subjective health in Belgium.”. In this document, we will go through the comments one by one to explain the changes we made in the manuscript following the suggestions of the reviewer. The page and line numbers mentioned in our responses refer to the revised version of the manuscript.

General comments

However, there are some major weaknesses of the article that needs to be improved as especially the magnitude of the association between health measures and types of employment (as well as unemployment then) is not trustworthy in current manuscript. The main problem is that the analyses does not well explain how the estimates can be considered to how low bias. This is especially obvious from the different of some of the estimates between model 2 and model 3 and 4. The authors must be able to explain which estimates are trustworthy and why. Current text are written in such a way that both the estimate from model 2 and 3 are considered trustworthy and considering that the odds ratio can go from 3.41 to 1.97 this is not a reasonable conclusion.

We understand the reviewer’s concern with regard to the trustworthiness of the estimates presented in the manuscript. In answer to the reviewer’s specific comments, we have double-checked the bivariate correlations between the different labour market positions and social situation indicators (see tables 1, 2 and 3 in this document). Based on these results, we have decided to eliminate the ‘material deprivation’ variable from the analyses. The large reduction in
the odds ratios from model 2 to model 3 are due to the important role of respondents’ household situation in explaining the relationship between labour market position and health. This is also clearer in tables 2 and 3 in the manuscript as coefficients for the controlling variables have been added. On top of this, the large reduction in the odds ratios can also be explained by the fact that maximum effect sizes (from 0 to 1 for each labour market position) are reported for the labour market typology variables. Since the reported effects are large, the potential effect of including control variables is also large.

The second most important issue to deal with is the composition of the four precarious employment groups (one of them the reference group “Standard jobs”). It needs to be better explained and reasoned on why they are good definitions and also a reference for why it is reasonable to use the approach from the authors, i.e. to create variables based on a cluster analysis and to use these variables in the logistic regression based on probabilities.

The typological approach towards the measurement of employment quality follows from the need for a multidimensional approach, in order to transcend the mere comparison of respondents in employment with those out of employment and looking for a job (the unemployed). This need for a multidimensional approach towards employment quality emerges from the literature.


We have clarified the use and the definitions of the waged employment groups in the methods section (lines 238-253) and the discussion section of the manuscript (lines 419-430).

For most of the above mentioned issues are collinearity issues, i.e. correlated exposure variables, a very important thing to deal with. I suggest that the authors are looking into papers that have evaluated important criteria that need to be fulfilled for logistic regression, e.g. Bagley et al.
mentioned in one of my comments below, and evaluate whether these criteria are well enough dealt with in the article besides collinearity.

All correlations between the independent variables in the analyses were checked, in order to make sure that no collinearity issues were present (see tables 1, 2 and 3 in this document). As a result, the ‘material deprivation’ variable was removed from the analyses. We also read the article by Bagley and colleagues. The criteria mentioned in this article are dealt with, apart from those that were impossible to implement because of sparseness of the data (e.g. interaction effects).

Major Compulsory Revisions

Abstract:

1. Results from the study should be presented in the abstract. The background part of the abstract can be much reduced and give space for this. It would also be valuable to have a clearer description of what the health measures are in the method part of the abstract.

The results part of the abstract was elaborated, in order to include more information about the results of our study (lines 49-55). A clearer description of the health outcomes was provided in the methods part of the abstract, as was suggested by the reviewer (lines 45-48).

Background:

2. Row 154-156: The statement at either needs to be clearer so that the uniqueness in the manuscript is obvious or it needs to be revised. There are previous studies that have include precarious employment and unemployment in the analyses, e.g. Virtanen et al. "Health inequalities in the workforce: the labour market core-periphery structure" and Flint et al. "Do labour market status transitions predict changes in psychological well-being?". My interpretation from current text is that the submitted manuscript consider itself to be the first to include both precarious employment and unemployment in their analyses, which is not the case. Maybe is the solution to be clearer with that employment quality is referring to a combination of 5 different work-related properties instead of the traditional definition of precarious employment and that this is unique.

We agree with the reviewer that this statement does not entirely reflect the uniqueness of our manuscript. We have removed the statement from the manuscript and replaced it with an explanation of the added value of our approach (lines 135-140). Also, we have included the two studies mentioned by the reviewer in our manuscript (lines 604-610).
3. Row 177: It is written "the second main objective". The first objective is not obvious from the paragraph. I strongly recommend to write two-three sentences in an own paragraph where the objectives are clearly written. If not, it must at least be obvious what the first objective is.

We followed the suggestion of the reviewer. The original statement was removed from the manuscript and replaced by a short paragraph focusing only on the objectives of the study (lines 161-165).

Methods:

4. There is no information on how self-perceived general health is measured.

The missing information was added to the manuscript (lines 289-293).

5. The "Center for Epidemiological Studies Depression Scale" is used for self-rated mental health. Seven of twenty items are used in the manuscript but the rationale behind this reduction is not explained. Such explanation should be added to the manuscript.

The reduction from twenty to seven CESDS-items occurred because only these 7 items were included in the GGS. We have tried to make this clearer by adapting the wording in the manuscript (lines 296-298).

6. The probabilities to be in each of the four groups are used for four variables that will be highly correlated. Why this unorthodox way of defining variables is working in a logistic regression needs to be explained as the collinearity between the newly created variables could highly bias estimates. A reference to a manuscript that is advising this creation of variables and their use in a logistic regression (or similar statistical method) might handle the issue. It intuitively feels like a reasonable choice to create variables but that does not mean that it will work well in the analysis and an explanation is therefore required.

Given the fact that the latent classes are created based on respondents’ answering patterns considering their job features – so exactly aimed at maximizing the distinction between groups of respondents – the correlations between the four job type variables should be fairly low. The table below (table 1) shows the bivariate correlations between the four types of waged employment. Although correlations are statistically significant, they are not problematic for our analyses. Therefore, it is possible to simultaneously include these variables into a binary logistic regression analysis without risking bias to the estimates.

Table 1. Bivariate Pearson Correlations between the four job types
The use of latent class variables as independent variables in (logistic) regression analyses is a technique that is known as the 3-step approach for latent class modelling with covariates (Bray, Lanza & Tan, 2015; Bakk, Tekle & Vermunt, 2013). Usually, the 3-step approach implies that a Latent Class Cluster Analysis is performed and that the latent class probabilities are used to assign each respondent to a single cluster (modal assignment). We choose not to use modal assignment, but to model the latent class probabilities per respondent as independent variables. In other words: each respondent’s probability to belong to a certain cluster is expressed in a decimal value between 0 and 1. Modeling latent class probabilities has the advantage (compared to modal assignment) of not incorporating the uncertainty of misclassification as a result of classifying respondents in a particular cluster. Classification errors are thus minimized. This is explained in lines 305-313 of the manuscript.

The same technique to create job types based on employment quality indicators was also used on data from the European Working Conditions Survey (EU27). The job types found in that study are similar to the job types that are discerned in this manuscript, with that difference that the European data yielded five instead of four job types because two types of precarious employment were discerned (Van Aerden, Puig-Barrachina, Bosmans & Vanroelen, 2016). The striking similarities between the results on European data and those on Belgian data – despite the use of a different survey and different variables – serve as a strong validation for the latent class approach.


Results:

7. The presentation of results need to be improved. To start with is the focus on the "uncontrolled" model too large. You are controlling for other variables because you consider there to be important confounders that needs to be included in the analysis. The crude odds ratio is therefore of very low, if any, value for your results. The crude odds ratio would benefit from being presented as crude odds ratio or crude model in Table 2 and 3. I suggest that you rename other models from model 2-4 to model 1-3 and change Model 1 to either crude odds ratio or crude model. It is important to show the crude odds ratio in the table to
explain how the odds ratio is affected by the confounding variables. It gives an indication of how important the more advanced models are. It is still, of course, very important to evaluate whether the extended models are improving the estimates of the odds ratio, i.e. decreasing the bias, but this should be dealt with in the discussion where the analysis method are recommended to be commented on. Keep in mind that the crude odds ratio could potentially have the lowest bias even if this unlikely to be the case and when performing the analysis it needs to be understood also if model 2 are presenting better estimates than the crude odds ratios. The result section would also benefit a lot from a more compact reporting. I suggest that the result section concludes for what labour market positions that there is a significantly poorer health. As it is significant results for all or no model for all labour market positions it might be sufficient to only specify the range for the ORs of current model 2-4.

We followed the suggestions of the review, which means that we changed the names of the models in table 2 and table 3 to “basic model”, “model 1”, “model 2” and “model 3”. We have also adapted the text of the results section (lines 342-351, lines 356-358 and lines 360-364) in order to reduce the focus on the crude odds ratios. We have also tried to report the results in a more compact way.

Discussion:

8. Row 387-394: The discussion should not summarize what has been done. It is sufficient to mention this information in the method part (and the aim). This information should be removed from the discussion section.

The summary has been removed from the manuscript.

9. Row 396-397: When concluding results in the discussion you should rely on your analysis method and it should not be commented in detail how variables where controlled for unless there are some important message that need to be given. Now the same message is repeated when it is written that it holds also for the more extensive model.

The statement has been removed from the manuscript.

10. Row 398-402: This information are not relevant for this part of the article. It should either be mentioned in the method section why these analysis are performed or in the discussion of methodological concerns why your choice of analysis is relevant. Now it is presented in relation to your results where it is surplus.
This information was already mentioned in the background section of the manuscript (lines 157-159), in order to explain why the introduction of the household context and social support variables in the model is necessary to obtain a good understanding of the relationship between labour market position and health. As suggested by the reviewer, it was removed from this part of the manuscript.

11. Row 404-409: As in previous comment, you need to rely on your analysis. Now you are arguing about how additional variables are affecting the estimates and not about which estimate is most reliable. You must not be able to judge which one of your estimates is the most reliable one but you cannot assume that both are somewhat correct, which current phrasing are. The true estimates of the effect from different labour market positions is not affected by which variables are in the model. It is simply a parameter value that our statistical model to the best of its ability tries to solve. How small the bias is depend on the set-up of variables in the model and the lowest bias is of course for the best set-up of variables which we cannot confirm as the true parameter remains unknown. The different model set-ups need to be validated as far as possible and there should be a good explanation. If the estimate is strongly affected by collinear variables then you need to make sure that collinear variables are not part of the analysis. That some estimates are so highly affected by the addition of household variables needs to be better taken care of in the analyses and the discussion of them. An alternative way of dealing with this is to present stratified estimates and to check whether the estimates are the same for all outcomes for the controlling variable. If not it is a sign of the household variables explaining that the effect on health from e.g. unemployment are different on individual level and it might for such reason be needed to present things differently. However, the possibility to present results on group level are often difficult to do in a trustworthy way due to small sample sizes and might not work for your study for such reason.

We have adapted this part of the manuscript and shifted the focus from the effect on the estimates towards the validation of our model and the effect on the relationship under study, namely that between labour market position and health (lines 392-396). Presenting stratified estimates is unfortunately not possible due to the limited size of our sample.

12. Row 423-427: In line with the comments above, this needs to better elaborated. My interpretation of this is that the estimate is mainly affected by collinearity between household variables and labour market status variables where some of these are partly derived from similar information in the cluster analysis.

The wording was changed in accordance with the previous comments of the reviewer (lines 408-412). Table 2 in this document presents the bivariate correlations between the labour market
status variables on the one hand and the social context variables on the other hand. The table shows that none of the social context variables is highly correlated with a labour market status variable. We also want to stress that no similar information was used to construct the labour market position typology and the social context variables. Of course, certain of the employment quality indicators and social context indicators are related to each other. For example, a respondent’s income partly determines the financial situation and level of material deprivation of the household. Showing these relations is one of the important intentions of our study. However, also other factors determine the situation at the household level. The fact that the employment quality indicators were used to construct latent classes (that were later included as probability scores for individuals to belong to each class) also means that problems due to multicollinearity are avoided. As a consequence, the relationship between the labour market position variables and the social context variables is not too strong.

Table 2. Bivariate Pearson Correlations between the six labour market positions and the five social situation indicators

13. The cluster analysis created four groups of individuals. Why these should be considered as valid groups needs to be discussed as this is a very central part of the content of the article.

We agree with the reviewer that the use of our typological approach is an essential part of our study and should therefore be discussed in this section of the manuscript. We have added a paragraph to the manuscript (lines 419-430).

Minor Essential Revisions

Abstract:

1. Row 30-32: Health inequalities are brought up first in abstract while it is nowhere mentioned in the remaining part of the article. It feels like a good idea to bring it up in the remaining part of the article. Either do so or delete this argumentation from the abstract as it cannot be important in only the abstract.

The statement about health inequalities was removed from the abstract.

Background:

2. Row 124: "Find" does not work in this sentence as you are referring to the results of their result.
The wording was changed (line 104).

3. Row 127-129: Grammar needs to be improved. The scale is not what is related to poor health. It is employment quality that is related to poor health.

The wording was changed (lines 108-109).

4. Row 179-181: The text has no value for the article. It only informs about the standard practice in research articles in Archives of Public Health and should therefore in my opinion be deleted. This explanation of what to expect next in the text also feels surplus on row 111-113.

Both passages were deleted, as suggested by the reviewer.

Methods:

5. There is no information about the number of invited in the GGS (only that 7163 participated). This information should be added.

The information about the number of individuals that were invited to participate in the Belgian GGS was added to the manuscript (lines 176-179).

6. On row 207 it is written: "assuming missing at random (MAR)". It is unclear what this means. Are values imputed randomly based on other values or is the best predicted value replacing missing value? The rationale behind the procedure must be rewritten so that it is understandable how values are imputed.

The MAR-procedure is no imputation procedure. It is not necessary to impute missing values, because the classification of cases can proceed based on those variables that are observed for the respondent under consideration. This explanation was added to the manuscript (lines 197-199).

7. Row 233: AIC, BIC and CAIC are mentioned. These short forms have not been explained previously in the text. Revise the manuscript accordingly.

The short forms were written in full (lines 224-225).
8. Row 274-279: Material deprivation is mentioned. It is unclear if this definition has been used before and whether the questions are part of a validated instrument. Especially item 5 (eating fish, meat or chicken) feels a bit controversial as this might not be due to the lack of possibility to eat these food dishes, e.g. a vegetarian would skip them for obvious reason. Also in the method section, a clarification of the definition of material deprivation is required.

Due to our response to the reviewer’s next comment (9), the material deprivation indicator is no longer included in the manuscript.

9. Material deprivation and perceived financial situation are likely to measure very similar things. Please explain how they could both be used in your analysis. It feels obvious that there is a collinearity problem between variables.

Table 3 shows the bivariate correlations between the five social situation indicators. The highest correlation is indeed seen between the material deprivation variable and the financial situation variable. Because of this, we have checked what happens to the model if we leave out either material deprivation or financial situation. The results from these analyses are presented in tables 4 (general health) and 5 (mental health). From these results, it is clear that including material deprivation in a model which already contains financial situation is not very useful. The models containing only the financial situation are almost identical to the model containing both financial situation and material deprivation. Therefore, we have decided to leave the material deprivation variable out of the analyses presented in the manuscript.

Table 3. Bivariate Pearson Correlations between the five social situation indicators

Table 4. Results for three different models relating labour market positions to poor self-perceived general health

Table 5. Results for three different models relating labour market positions to poor self-rated mental health

10. The social support cut-off value surprises me. From the description, the scale seems to be from 0-60 (6 items with scale 0-10). Is it really as few as 10% that scores 5 or lower or have I misunderstood the calculation of scores for social support? Even if this is correct it needs to be specified what a high score corresponds to. It is confusing that the first question (availability of someone to talk with) is indicating a positive situation while the second question (emptiness) is indicating a negative situation. Please clarify how the score for this measurement is calculated and also check if the cut-off value is correct.
The procedure for the construction of the social support indicator is clarified here and also in the corresponding part of the manuscript (lines 280-286). First of all, three of the six original items were recoded because they indicate a positive situation instead of a negative situation such as the remaining three items. Then, the six items were summed and the resulting scale ranging from 0 until 6 was transformed in a scale ranging from 0 until 10. We agree with the reviewer that the original cut-off – isolating 10% of the sample – is rather strict. Therefore, a new cut-off value of 7 (out of 10) was applied to dichotomise the social support scale. The adapted cut-off value means that 20.2% of the sample is now classified as reporting ‘low social support’.

11. Row 330-331: Write "household situation variables" because if I understood it correctly are you using four variables that are indicators of household situation.

We followed the suggestion from the reviewer. The wording was adapted to make clear that three household situation variables were included to represent respondents’ household situation (lines 327-329).

Results:

12. The coefficients for controlling variables should according to experts (see e.g. Bagley et al. “Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain”) be presented in articles. They should therefore be added in Table 2 and 3 as they give highly relevant information for your statistical model.

We followed the suggestion from the reviewer and added the coefficients for the controlling variables in Tables 2 and 3. We have also commented on significant effects in the text of the manuscript (lines 350-351 and lines 363-364).

13. "Odds ratios and 95% confidence intervals" should not be part of the title of a table. They should be given either in the table or as a footnote to the tables. Also the number of participants should not be given in the title. As you are using a sample, "n" should be used and not "N". "N" is referring to the population size according to the commonly agreed standard.

The information with regard to the number of participants and the content of the tables was deleted from the title of the tables and given as a footnote to the tables instead (lines 368, 373, 378 and 383).
Discussion:

14. Policy implications from the article needs to be discussed in the article. The authors are expected to have ideas on how the content of the article can be used in decision making or at least if there is a requirement of further analyses before there can be some advices to give to the decision makers. It would also be valuable to have a suggestion on what future research to recommend based on current study.

The authors believe that further research is needed before making firm statements about policy implications. We have added our suggestions considering future research to the discussion section of the manuscript (lines 432-447).

References:

15. Reference 13 and 23 are referring to "Self-reference removed". This needs to be corrected.

The two references were added to the manuscript (lines 558-560 and 612-614).

16. The surname is Paul in reference 17, i.e. author name is Karsten Paul and not Paul Karsten.

We thank the reviewer for bringing this mistake to our attention. The reference was adapted (line 582).

Discretionary Revisions

1. The background would be improved if it was more compact and clearer written with a focus on precarious employment in regard to the definition you have decided to use. The background would also benefit from being shortened. However, I don't consider it necessary to rewrite it to fulfil this suggestion.

We have tried to improve the readability and the focus of the background by shortening the text (redundant parts were removed) and by changing the order of certain paragraphs.

2. The manuscript would benefit from more references in peer-reviewed journals. At least the references 2, 4, 5, 6, 10, 14, 15, 16, 24, 26, 27, 28, 30 and 31 constitutes of references to books, reports and similar.
References 2 and 6 were deleted from the manuscript, due to the shortening of the background section. The three references added to the manuscript are publications in peer-reviewed journals.