PEER REVIEW HISTORY

This paper was submitted to a another journal from BMJ but declined for publication following peer review. The authors addressed the reviewers' comments and submitted the revised paper to BMJ Open. The paper was subsequently accepted for publication at BMJ Open.

(This paper received three reviews from its previous journal but only two reviewers agreed to published their review.)

ARTICLE DETAILS

TITLE (PROVISIONAL)	Precarious employment, business performance and occupational injuries – a study protocol of a register-based Swedish project.
AUTHORS	Orellana, Cecilia; Kreshpaj, Bertina; Johansson, Gun; Burström, Bo; Kjellberg, Katarina; Hemmingsson, Tomas; Axén, Magnus; Davis, Letitia; Wegman, David; Bodin, Theo

VERSION 1 – REVIEW

REVIEWER	N. Dragano University Clinic Düsseldorf,Germany
REVIEW RETURNED	16-Sep-2018

GENERAL COMMENTS	The research topic is interesting and has a high public health relevance as the rise of non-standard forms of employment calls for a precise assessments of the related health risks. Occupational injuries have been related to employment status in some studies but there is a lack of large scale longitudinal studies. Register based approaches such as the one presented in the protocol have the potential to generate new knowledge in the field. My remarks mainly focus on the description of the methods and the databases. I would like to ask for some clarifications: #Research question / theory: It is clear to me that it is relevant to investigate a) the effect of precarious employment on occupational injuries (OI) and b) the influence of business performance on injuries. But I do not understand how both questions relate? Research cited in the introduction suggests a differential underreporting of OI by employment status – with less reports from precariously employed workers. On the other hand, it may well be that precarious workers have more injuries than others (which they do not report). Do you see any risk of bias resulting from both contradictory trends? Research question 3 must be answered first in my view. If the datasets are not appropriate for measuring OI, questions 1 and 2 are obsolete. #Methods / operationalization of precarious work Standard definitions are type of contract, self-reported job insecurity, irregular working hours and others. In the discussion section you describe that important variables like type of contract

	are not included in the datasets. Can you further specify which kind of proxies you plan to use instead? #Methods / outcomes Authors point at a pronounced underreporting of OI in registries in the introduction section. Are there any indications that the datasets used in this study provide an appropriate coverage of cases? I realized that this question will be researched in objective 3 but have you made any feasibility analyses in advance to ensure that your study will work? Do you expect that AFA has a more precise measurement of OI than ISA? Do the records in NPR allow to identify OI? And if yes: do you plan to use this register for the validation study too? If you use several datasets to identify OI: how do you define your primary outcome (best off? combined?) #Methods / business indicators I wonder why an ecological design will be applied. As far as I understand, it is possible to link all datasets. Thus, it would be possible to conduct a multilevel study with workers as level 1 and companies and their characteristics as level 2. #Methods / assess underreporting RQ6 on page 6 suggests that the validation study will rely on 2013 data only. Why do you restrict this analyses to one year? #Methods / analyses To investigate RQ4 a time-lag analyses with one exposure measurement will be performed. Will it also be possible to measure the exposure more than once at the beginning of the observation period? Insecure employment is volatile and it makes a difference if a person spends several years in insecure employment or just a short period. Can you take repeated exposure measurement into account? #Methods / analyses I am not a statistician but a sample size of roundabout 7 million persons means that you will find a lot of highly significant associations. How do you make sure that findings are relevant and replicable – and not the result of multiple-testing with a large number of observations? #additional minor remarks: Wording of the bullets on page 4 is a bit strange, English should be checked. E.g. incomplete sentences in bullets 4 and 5. Table 1: information in the last column is very detailed and can be shortened

REVIEWER	Fernando G. Benavides Center for Research in Occupational Health, Universitat Pompeu Fabra (Spain)
REVIEW RETURNED	12-Oct-2018

GENERAL COMMENTS	This manuscript presents an interesting protocol to study the impact of precarious employment and business performance on occupational injuries. The study will be based on a national–wide register data from all residents aged 18-70 in Sweden with any registered income during at least one year, between 2003 and

	2015. It includes around 7 000 000 workers, and they can be tracked between companies and workplaces. This will once again show the potential of the studies based on the connection of administrative records. No doubt, interesting data may be missing, but these kinds of studies are, besides possible, necessary. However, there are aspects related to objective 2, the main objective of the study, which are not defined in this protocol, leaving for a systematic review (objective 1) for the operational definition (according to available data) of precariousness and a workshop to finally (some have been identified in preliminary discussions) define indicators of business performance. The analysis according to business performance will possibly provide the most original results, since precariousness, measured through temporality (an important limitation of this study is not having information on type of contract), have already been widely studied (see Kivimäki et al. AJE 2003, Benavides et al OEM 2006, among others). Likewise, the results of the analysis according to the characteristics of the workplace will be of great interest (specially by workplace sex distribution). An important limitation, pointed out by the authors, is to leave informality out, since it is not included in the administrative records used in the study, so it would be necessary to add to the manuscript an estimate of what this population represents in Sweden. A recent publication by Jùlia et al J Public Health (Oxf) 2018 estimates a prevalence of informal employment of around 5% in the EU27. Another limitation not indicated by the authors, although described in methods, is the exclusion of injuries in itinere (during transit to/from work), it would be advisable to estimate their proportion in the set of occupational injuries. In itinere injuries, unlike those occurring in the workplace, are increasing in some countries (see Lopez-Ruiz et al AAP 2014). I would recommend stratifying all analyses by sex and age groups, because the difference between men and women, on the one hand, and between young and old, on the other, are important and known. (see Eurostats work accident statistics). Finally, although the authors mention the realization of a PhD thesis among the planned products, in my opinion this study can lead to more than one PhD thesis, besides being the seed for a possible European project.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
General answer: We thank the reviewer for this positive overall comment. We will provide a response for the questions, one by one.

1) Research question / theory:
a) It is clear to me that it is relevant to investigate a) the effect of precarious employment on occupational injuries (OI) and b) the influence of business performance on injuries. But I do not understand how both questions relate?

b) Research cited in the introduction suggests a differential underreporting of OI by employment status – with less reports from precariously employed workers. On the other hand, it may well be that precarious workers have more injuries than others (which they do not report). Do you see any risk of bias resulting from both contradictory trends?
c) Research question 3 must be answered first in my view. If the datasets are not appropriate for measuring OI, questions 1 and 2 are obsolete.

Answer regarding research question / theory
a) One possible explanation for the relationship between occupational injuries (OI) and business performance is that companies in economic deterioration may also, to a higher extent, apply precarious employment (PE) conditions to remain profitable. The opposite is also plausible, i.e. an association between companies using precarious employment conditions and a lower productivity and profitability over time. These relations are largely unstudied, and not within the scope of this protocol. Changes have been made on page 5 which hopefully bridges the gap between the two main exposures and makes is clearer why these are combined in the same study.

b) It is true that underreporting of OI may be higher among precarious workers, and conversely, these workers may have a higher rate of injuries compared to non-precarious workers. The potential bias from these contradictory trends cannot be mitigated as we don't have the data to conduct a longitudinal analysis of underreporting. Identifying "lower-than-expected" injury rates is an exploratory approach that we are contemplating, but haven't finished discussing.
We believe however that the best way to observe differences in the underreporting for precarious workers compared to non-precarious, is to conduct a stratified analysis. This will be possible once we have obtained results from RQ1 and are able to use this in RQ2 (of the new version of the study protocol). We have made changes in the manuscript, both in the Analysis plan (page 16, under RQ2) and added a brief paragraph at the end of the discussion (page 23).

c) We assume the reviewer wishes that objective 3 comes before objective 1 and 2. Regarding the order of the objectives/research questions, we agree that the most logic would be to have objective 3 before objective 2 and have changed accordingly (pages 6-7 in the Introduction, and pages 15-17 in the Analysis plan). We respectfully disagree to the second part of the comment: "If the datasets are not appropriate for measuring OI, questions 1 and 2 are obsolete". ISA is used to produce the official injury statistics of Sweden and the results are likely to be relevant. Under-reporting will be reported and discussed as limitation in these etiological studies and thanks to our work on under-reporting in the second objective, this discussion will be better informed than previously.

2) Methods / operationalization of precarious work
Standard definitions are type of contract, self-reported job insecurity, irregular working hours and others. In the discussion section you describe that important variables like type of contract are not included in the datasets. Can you further specify which kind of proxies you plan to use instead?

Answer regarding methods / operationalization of precarious work
Although PE has been studied from different perspectives and in different disciplines, there is still a lack of a standard definition. Evidence is mounting that temporary contracts per se are not a very useful exposure (Koranyi et al 2018), rather employment security could be measured in other ways. We are likely to use variables such as length of employment, multiple jobs and fragmented careers. However, we do not want to discuss this in the protocol as it would mean forgoing the results from RQ1 – systematic review and operationalization.

3) Methods / outcomes

a) Authors point at a pronounced underreporting of OI in registries in the introduction section. Are there any indications that the datasets used in this study provide an appropriate coverage of cases? I realized that this question will be researched in objective 3 but have you made any feasibility analyses in advance to ensure that your study will work?

b) Do you expect that AFA has a more precise measurement of OI than ISA?

c) Do the records in NPR allow to identify OI? And if yes: do you plan to use this register for the validation study too?

d) If you use several datasets to identify OI: how do you define your primary outcome (best off? combined?)

Answer regarding methods /outcomes

a) We strongly believe that our two available sources of occupational injuries (ISA and AFA) have an appropriate coverage of cases. In order to make these populations comparable and decrease source-dependence that may arise in the capture-recapture analysis, we may initially restrict the analysis to companies in the public sector and large companies in the private sector, which are most likely to be covered by collective bargain. We have run preliminary analysis and have a good size population to be included that are similar in their characteristics.
We have succeeded in conducting the linkage of all databases that will be used for the estimation of underreporting. To ensure clarity and provide the readers with some preliminary results, we have decided to include a table showing the number of occupational injuries in Sweden for the year 2013 being part of ISA, AFA and the overlap. We have included it as Table 3, under “Preliminary findings” in page 18.

b) We do not expect differences in the OI definition, as the definition of OI being used by AFA insurance is the same as the one used by ISA (stated in Table 1, under “Injury reporting and definition”). However, there are some other variables complementing the OI definition (regarding description of the accident that led to the injury, severity, etc.) that could be more thoroughly described in one register than the other.

c) NPR cannot be used to identify occupational injuries. It will be used to classify injuries in AFA and ISA into different levels of severity.

d) Finally, regarding our definition of our primary outcome (best-off, combined), we realize that this may not have been completely clear in the submitted version of the study protocol. We have added a paragraph entitled “Linkage of data sources” in the Methods section (pages 13-14). The only RQ (or objective) in which data from both OI registers will be used is the one using capture-recapture method. Due to the availability of data from the AFA register only for the year 2013, for all other research questions, only ISA will be used.

4) Methods / business indicators
I wonder why an ecological design will be applied. As far as I understand, it is possible to link all datasets. Thus, it would be possible to conduct a multilevel study with workers as level 1 and companies and their characteristics as level 2.

Answer regarding methods /business indicators
The reviewer is right: we are in fact able to link workers (individual level) within companies and their characteristics (company level), and even workplaces within companies. This may not have been clearly stated in the submitted version and we have therefore moved this from the “Description of data sources” into the new sub-heading “Linkage of data sources”, as part of the methods section (pages 13-14). We have here included minor changes, which hopefully can make this clearer.

5) Methods / assess underreporting
RQ6 on page 6 suggests that the validation study will rely on 2013 data only. Why do you restrict this analyses to one year?

Answer regarding methods / assess underreporting
The reason for this restriction is the availability of data provided by one of the sources, AFA (page 10, lines 2-4). To conduct this analysis, we must restrict the time period in ISA to the same year, and make our calculations based on this year only. We do, however, count with a reasonable number of OI’s for that year (n=103 589, combining both OI registers) which will enable us to run both crude and adjusted analysis, including restricted analyses and/or stratifications.
We have included a table with preliminary results of the linkage of datasets (Table 3, page 18), to give the readers a better idea of the population size.

6) Methods / analyses
a) To investigate RQ4 a time-lag analyses with one exposure measurement will be performed. Will it also be possible to measure the exposure more than once at the beginning of the observation period? Insecure employment is volatile and it makes a difference if a person spends several years in insecure employment or just a short period. Can you take repeated exposure measurement into account?

b) I am not a statistician but a sample size of roundabout 7 million persons means that you will find a lot of highly significant associations. How do you make sure that findings are relevant and replicable – and not the result of multiple-testing with a large number of observations?

Answer regarding method / analyses

a) Although changes in PE may affect certain outcomes (for example, mental health), we do not believe that this is the case for the association with OI. As part of the Analysis Plan (page 17), we have written: “It is unlikely that there is an accumulation of risk or latency in the precarious employment–occupational injuries relationship. We will therefore measure risk and outcome at the same point in time, i.e. precarious employment and injury in the same year.” We hypothesize that PE has an effect on the occurrence of OI’s, but within a relatively short time-frame.
In spite of that, it would be interesting to investigate if there is such association with changes in PE over time, and we could look deeper into this in the form of a sensitivity analysis.

b) When starting the project, we soon realized that the big amount of data and numerous possibilities to conduct the analyses called for a published study protocol, which forced us to put everything into context and adhere to a plan instead of venturing on a “fishing expedition”.
We thank the reviewer for pointing to the risk of mass-significance, and therefore we have decided to include the population attributable fraction as part of the main outcomes to be reported apart from relative risks. Changes have been made in the text of the protocol, as follows: in RQ5, page 17, we have added “Apart from this, we will provide the population attributable fraction as part of our results”. It has also been added as part of RQ6 (page 17-18).

7) Additional minor remarks:
Wording of the bullets on page 4 is a bit strange, English should be checked. E.g. incomplete sentences in bullets 4 and 5.
Table 1: information in the last column is very detailed and can be shortened

These changes have been made according to the reviewer's suggestions.

Reviewer: 2
General answer: We thank the reviewer for the interest in our protocol. We do share the interest in the area, and its relevance. We will provide replies for each one of the comments, which we have grouped and numbered for simplicity.

1) Question about objective 2

However, there are aspects related to objective 2, the main objective of the study, which are not defined in this protocol, leaving for a systematic review (objective 1) for the operational definition (according to available data) of precariousness and a workshop to finally (some have been identified in preliminary discussions) define indicators of business performance.

The analysis according to business performance will possibly provide the most original results, since precariousness, measured through temporality (an important limitation of this study is not having information on type of contract), have already been widely studied (see Kivimäki et al. AJE 2003, Benavides et al OEM 2006, among others). Likewise, the results of the analysis according to the characteristics of the workplace will be of great interest (specially by workplace sex distribution).

Answer regarding objective 2
We believe temporality can be constructed from the register data since we have information on number of employers per year. We believe this is the real issue, rather than fixed-term contract (which would be a proxy for temporary employment).

2) Question about informality

An important limitation, pointed out by the authors, is to leave informality out, since it is not included in the administrative records used in the study, so it would be necessary to add to the manuscript an estimate of what this population represents in Sweden. A recent publication by Jùlia et al J Public Health (Oxf) 2018 estimates a prevalence of informal employment of around 5% in the EU27.

Answer regarding informality
We thank the reviewer for the comment. We have included an estimate of the prevalence of informal employment, using data from the latest version of the European Working Conditions Survey. We have included this in the discussion (page 22), and added this source into our reference list.

3) Injuries in itinere
Another limitation not indicated by the authors, although described in methods, is the exclusion of injuries in itinere (during transit to/from work), it would be advisable to estimate their proportion in the set of occupational injuries. In itinere injuries, unlike those occurring in the workplace, are increasing in some countries (see Lopez-Ruiz et al AAP 2014).

Answer regarding injuries in itinere
Although the occurrence of traffic-related injuries is certainly of interest, we have decided to exclude these. The reason behind this is that injuries that occurred during transit to/from work may be covered by car insurances and may therefore not appear in the OI registers. We have added a brief paragraph in the discussion section (page 20) and added the suggested reference.

4) Stratified analysis
I would recommend stratifying all analyses by sex and age groups, because the difference between men and women, on the one hand, and between young and old, on the other, are important and known. (see Eurostats work accident statistics).

Answer regarding stratified analysis
Although not clearly stated, we had planned to include stratified analysis by sex and age for all of our associations. We are thankful for this comment, so we can make the clarification, adding this in the Covariates/Confounders section, in page 14: “All analyses will be stratified by sex and age.”

5) Planned products
Finally, although the authors mention the realization of a PhD thesis among the planned products, in my opinion this study can lead to more than one PhD thesis, besides being the seed for a possible European project.

Answer regarding planned products
The reviewer has a very good point, and we may have been too modest about the potential of the project. We have rephrased this in the Dissemination section, starting on page 23.

VERSION 2 – REVIEW

REVIEWER	Nico Dragano University of Düsseldorf
REVIEW RETURNED	27-Nov-2018

GENERAL COMMENTS	Thanks to the authors for their response! I follow most of your arguments (and thanks for the several clarifications) with one important exception. It relates to the following question / answer from the response letter: “ 2) Methods / operationalization of precarious work Standard definitions are type of contract, self-reported job insecurity, irregular working hours and others. In the discussion section you describe that important variables like type of contract are not included in the datasets. Can you further specify which kind of proxies you plan to use instead? Answer regarding methods / operationalization of precarious work Although PE has been studied from different perspectives and in different disciplines, there is still a lack of a standard definition. Evidence is mounting that temporary contracts per se are not a very useful exposure (Koranyi et al 2018), rather employment security could be measured in other ways. We are likely to use variables such as length of employment, multiple jobs and fragmented careers. However, we do not want to discuss this in the protocol as it would mean forgoing the results from RQ1 – systematic review and operationalization.„ This does not convince me. If you want to study precarious work you should be sure that it is possible to operationalize it. Otherwise

	you cant do any of the suggested analyses as you miss the main exposure. An appropriate definition is the prerequisite for a study protocol in this case or do you see this different? And one minor personal remark (which can be ignored): you introduce population attributable fraction in response to my answer 6b. I personaly dont like the PAF in occupational epidemiology because it suggests that it is possible to fraction out effects of a quite complex exposure with many different links (some direct some indirect) to the outcome – which is not possible in my view.

REVIEWER	Fernando G. Benavides Center for Research in Occupational Research, Universitat Pompeu Fabra, Barcelona, Spain
REVIEW RETURNED	12-Dec-2018

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.

VERSION 2 – AUTHOR RESPONSE

Response for Reviewer 1

We realize that our previous response was not entirely satisfactory. Although type of contract can actually be obtained from both occupational injury registers, the quality and completeness of the data may be limited. This is why we have planned to obtain variables to construct precarious employment mainly from the LISA register. These variables include number of employers, number of sources of income, etc. With this, we plan to construct our own variables such as length of employment, multiple jobs and fragmented careers, to name a few.
We have made some changes in the description of Main variables (page 10), as well as in the Discussion section (starting at the bottom of page 18).

VERSION 3 – REVIEW

REVIEWER	Nico Dragano University Hospital Düsseldorf, Germany
REVIEW RETURNED	17-Dec-2018

GENERAL COMMENTS	The reviewer completed the checklist but made no further comments.