PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Childbirth, morbidity, sickness absence, and disability pension: a population-based longitudinal cohort study in Sweden
AUTHORS	Wang, Mo; Laszlo, Krisztina; Svedberg, Pia; Nylén, Lotta; Alexanderson, Kristina

VERSION 1 – REVIEW

REVIEWER	Hanke Heun-Johnson University of Southern California, USA
REVIEW RETURNED	24-Mar-2020

GENERAL COMMENTS	The topic of whether childbirth and/or general morbidity is associated with receiving sickness or disability benefits is interesting and important, and the authors set out to research the associations between these factors in more detail in this paper. However, there are some issues with the approach and innovation that would need to be addressed, in particular the sample size, the presentation and interpretation of the results, and how the current study is adding to the existing literature (specifically in relation to the authors' own publications). Moreover, I believe the paper would benefit greatly from restricting the sample to those with only one childbirth, so that a second pregnancy (resulting in a generous parental leave period in the near future) does not affect the decision to apply for sickness absence or disability pension. The follow-up period of the analysis could start when parental leave ends, while also controlling for the duration of this parental leave period. Furthermore, grammar and style should be checked by a native English speaker as there are many minor errors, and words or whole sentences are sometimes vague. Abstract Objective – Second sentence is confusing, and "if exists" should be fixed. Not clear what "or inclusion year" refers to. Participants - do they need to have lived in Sweden during all the prior and follow up years as well? Measures – Does SA>14 mean 14 days? Regarding DP days, is this any number of days? Results – The results or the objective should be rephrased so that they fit together logically and the results make more of an impact: the second sentence of the Objective should be expanded to go deeper into the "knowledge regarding the role of morbidity", or the results should be worded in a way that highlights the departure from the status quo (if that is the case). The first sentence should also include "after controlling for prior morbidity and patient characteristics" since the unadjusted/crude model shows a higher risk. "Or July 2005 in the B0 group" can be left out. B0 is not defined until

	the methods section of the paper and the exact date does not add any information at this point. Strengths and limitations section – The fact that Sweden has a high employment rate among women does not automatically infer that health selection into paid work does not occur (but perhaps other regulations that are not mentioned can explain this). The reference used to support this claim later in the discussion is not in English and can't be used to check their statement. Regardless, relevancy of paid work for their interpretation of results is not directly discussed in the paper. Background "year" should be "years". Reference 13 is not related to the statement. The sentence containing "Strong positive associations with morbidity" is vague and could use some more specifics instead of words like "this". Reference 18 does not appear to be related to the statement. Reference 16 and 17 (prior studies by the authors) are relevant, but it leaves the reader with the question what is different about the current study. Only very late in the paper it is mentioned that this study is different because the authors expanded their previous findings to the general population – the question remains, however, whether the current contribution is significant and novel, since the prior studies, which utilize a twin cohort, arrive at a very similar conclusion, albeit with fewer details. It is unclear why it is "not surprising" and who is doing the questioning in the statement: "This may not seem surprising, however, this association is sometimes questioned". The references that follow the next clause consist of a non-peer-reviewed working paper (19), as well as a reference that is not related to the topic. Thus, the main motivation for this paper is not well-supported. It is not clear what the sentence that follows (on whether SA and DP are good measures) is adding to the authors' arguments, and one of the references related to bicycle accidents is not relevant. Methods Use the word "following" once in the last sentence of the first paragraph. Study population - Why did the authors decide on ages 18-39 years and not younger and/or older? Specific details on what happens with women in the analyses if they gave birth at these extremes e.g. age 18 or 39, especially considering the long lead-in and wash-out period, would be helpful. There is no mention how they dealt with deaths, women who leave the country or are in long-term care facilities. How are women who experienced events such as miscarriages, abortions, hysterectomies, stillbirth, fertilization treatments without results, etc etc categorized, if they're not excluded? Why is the final follow-up period 43 weeks and not just one extra year (Y+4) for simplicity? This adds to the confusion of all the time references (T0, Y+1-Y+3, Y-1) and the groups B0, B1, and B1+. It appears that the authors in most cases use the full explanation in most sentences throughout the paper (e.g. "indicating morbidity 1-3 years before T0 (Y-3 - Y-1), and/or 1 year after T0 (Y+1)"), and I see no reason to just describe the events and groups with words rather than T0, B1+ etc. throughout the paper to improve the readability. It is not clear what the date 2 July 2005 signifies. Randomization, or assignment of an index date by matching might be better to maintain some variability in the follow up period for the B0 group, since there

	is probably seasonality in applying for SA or DP. Morbidity - The clause that contains "operationalise" can be worded more specifically or be left out. The authors need to provide more details about specialized outpatient visits. Do they include all specialty health care, regardless of severity or recorded diagnosis (suspected or confirmed), and even for preventive visits or for disorders/diseases in remission? To label someone as having morbidity with one visit to any specialist regardless of the primary reason of the visit seems too broad/inclusive/heterogeneous. At the very least, using two or more outpatient events with the same diagnostic code (not in remission) rather than one event would be a better measure of actual morbidity. On page 9, the second and third paragraph appear to be duplicates. Is a hospitalization stay or outpatient visit excluded completely if there is a childbirth-related diagnosis observed, or only that particular diagnosis on the event? There were conflicting descriptions in the paper. Regardless, it may be more precise to use primary vs non-primary diagnoses to determine the reason for the visit (or other diagnostic codes rather than ICD10) and possible under/overcounting of either childbirth- or non-childbirth-related visits. The insurance section is useful for readers from other countries. More details on eligibility regarding prior work history (do people get benefits without a prior history of employment, and do immigrants get benefits immediately) would be helpful, as would an explanation of how the authors deal with variability in parental leave duration, i.e. if people return to work earlier. As mentioned before, including only at primiparous women could reduce overall confounding by future parental leave on SA or DP. Furthermore, matching on age and other characteristics would create a much cleaner comparison and balanced sample sizes. Outcomes – can use more details, or at least whole words and sentences. Included factors – Why is age not used as a continuous variable? Why are the other demographic characteristics besides age and education in table 1 not included as covariates? These all seem necessary and logical covariates. How reliable are these variables, especially for immigrants? What is the strategy for dealing with education levels of people who have not reached their final degree (i.e. at age 18-25)? Statistical analysis – The fact that anyone with DP prior to delivery is removed from the sample should be mentioned more prominently (e.g. in abstract); this changes the interpretation. It would be interesting to see some statistics on censoring (death and emigration). Results What statistical method is used to compare the results in table 1? The authors claim no difference between the three childbirth groups, but no statistical tests have been done or are reported (here or in the methods section). Some of the morbidity measures do appear different (specialized outpatient visits for B1 and B1+ vs B0) but this is not addressed nor supported with statistical tests. Using a consistent ordering of comparison groups in the text would make the last two sentences more readable. The small sample size in the SA and DP groups is a major issue, especially considering these subgroups are split up even further in subsequent analysis (morbidity vs no morbidity, short- and long-term SA). Is there a reason the analyses were restricted to childbirths in 2005 only? Adding more years could greatly improve the credibility

	of the final conclusions, as long as SA and DP policies did not change during the timeframe. The mean annual number of hospitalization days is not really a good metric, as the majority of people have no hospitalizations, and long stays (that also affect the probability of pregnancy in multiple ways) will skew the average. In the legend for figure 1d, "hospitalizations" should be "visits". The results in tables 2-4 are displayed in a complicated manner, and the text does not help in clarifying the results. It appears to be impossible to interpret the effect of prior morbidity (model 2 compared to model 1) for each subgroup as reference values are set to 1 across models. Issues of potential collinearity between morbidity prior to and after delivery are not addressed. The sample sizes of the subgroups are probably extremely small in some cases (e.g. for combination of SA spell > 90 days & 1 child & no morbidity prior & morbidity during follow up), but there are no details, only sample sizes for the overall groups. Is it not completely clear what is different in table 4 compared to table 2 and 3. Discussion References 40 and 41, related to women's roles in society and how that affects their health are from 1982 – considering the developments over the last 40 years in this area the authors should try to find more recent references. The conclusions regarding better health for those who gave birth appear stretched, at least considering the results of the unadjusted/crude model for SA, although it is hard to interpret the outcomes as they are currently presented. There is also a potential (voluntarily or involuntarily) self-selection into groups and there is confounding by parental leave and multiple births – these issues are only minimally discussed and not controlled for in the analyses. A discussion on why SA and DP outcomes are going in different directions for the childbirth group in the crude/unadjusted model would be useful. Reference 44 is not in English. Conclusions A strong association between morbidity and SA/DP (second sentence) does not necessarily negate the claim in the first sentence. Also, the conclusion that morbidity is associated with SA and DP should come as no surprise to the reader; it is the interaction with childbirths that makes this paper interesting. The following sentences assumingly address this, but should include "after controlling for morbidity", otherwise they are contradicting the unadjusted results. Also, if prior morbidity is indirectly caused by pregnancies (as stated in the discussion), that still means that childbirth is associated with SA and DP; controlling for it may attenuate the effect in these analyses on paper but does not reduce the original association. A discussion on the implications of these interactions on the health benefits system and the authors' point of view in this matter would be helpful to determine the overall significance of the authors' findings. The statement regarding "prioritizing domestic duties" is strongly worded and insufficiently supported by the articles cited by the authors (see comment on references 19 & 20 in Background).

REVIEWER	Norhayati Binti Mohd Noor Universiti Sains Malaysia, Malaysia
REVIEW RETURNED	12-Apr-2020

GENERAL COMMENTS	It is obvious that SA would have a strong association with DP. It is reported in the results that nulliparous women had the highest risk of SA and DP followed by one and more than one childbirth. This is the result of previous studies as stated in the Introduction section [12-15]. In the current study, the interest is to identify the extent of morbidity in these groups of women with SA. The explanation in the current study should be included in the previous papers to support their statement. It is a redundant publication. It still does not answer the question of WHY nulliparous women had the highest risk of SA and DP. ABSTRACT The results of this study are repeating the results in the previous study [12-15] but yet do not answer the objectives of the current study. Do not use abbreviations in the abstract. What does B0 mean? DISCUSSION The discussion is more of a statistical repetition of the results.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Hanke Heun-Johnson
Institution and Country: University of Southern California, USA

Comment:
The topic of whether childbirth and/or general morbidity is associated with receiving sickness or disability benefits is interesting and important, and the authors set out to research the associations between these factors in more detail in this paper. However, there are some issues with the approach and innovation that would need to be addressed, in particular the sample size, the presentation and interpretation of the results, and how the current study is adding to the existing literature (specifically in relation to the authors' own publications). Moreover, I believe the paper would benefit greatly from restricting the sample to those with only one childbirth, so that a second pregnancy (resulting in a generous parental leave period in the near future) does not affect the decision to apply for sickness absence or disability pension. The follow-up period of the analysis could start when parental leave ends, while also controlling for the duration of this parental leave period.

Authors' response: We thank the reviewer for the overall statement that this is an interesting and important research line. We also thank you for the helpful comments as well as for the nice suggestions for other studies, with other aims, regarding this theme. Such studies are indeed needed. Of the several points you raise we start with the most important one: how does this study add to the existing literature, specifically in relation to our previous studies. The major way this study adds to the previous ones is that in the present study we were able to include information about the women's morbidity. Actually, most studies on sickness absence (SA) in relation to childbirth do not include any aspects of morbidity although this could be considered of importance for both SA and disability pension (DP). The studies we know of that did this were from our group, however, based on a very selected and, compared to this study, small cohorts (n≈5000 women), based on data from the Swedish Twin Registry. It is not clear whether findings from twins about this may be generalized to women in the general population. Also, the larger cohort in these studies, allowed for more subgroup analyses.

In many Western countries the ill-health content of SA among women on SA, and especially among women with children, is questioned. It is rather assumed that women are on SA for other reasons (economic, to care for sick children or older relatives, because they are less interested in their paid work than in family life, etc.).

In our twin studies we showed that there was a strong association between morbidity and SA and/or disability pension (DP). However, in those studies we were only able to use information on hospitalization as a marker of morbidity. Here we were able to also use information about specialized out-patient healthcare.

We now revised the manuscript to better emphasize the novelty of our study; please see pages 5 and 6.

Here we study three groups of nulliparous women: those who did not give any birth in 2005 nor during the follow-up (B0), those who had one birth in 2005 and no more births during the follow-up (B1), and those who had one birth in 2005 and at least one more birth during the follow-up (B1+). You suggest to exclude the B1+ group. Actually, we find this group of extra importance to include as some have claimed that this group has very high SA; in Sweden this is even very often stated in the public debate. If you are not interested in that group, you can of course just ignore the results presented for them. The results for the B0 and the B1 groups are not altered by also presenting outcomes for the B1+ group. Maybe you are even suggesting that we exclude the B0 group? Again, in this exploratory study we see no reason for excluding any of these three groups. The B0 group is an important comparison group with relevance for the debate in welfare states about whether women who give birth, and especially those with more than one birth, have higher SA/DP than women who remain nulliparous. Future studies, with other aims, could probably further investigate these groups, and, as you suggested, maybe study the B1 group extra much, based on the findings from this study.

Also, you suggested that we start follow-up of SA after the parental leave period ends. However, that would indeed present many problems. As you say, Sweden has a generous system for this, and women can use 60 of the parental leave days before the expected delivery date. However, most women save these days till after the birth. Furthermore, both parents can use these days, that is, it is not always the women who gave birth who use most of them; of the 480 days of parental leave given for each child, During the study period, 60 of these days could only be used by the father and 60 days only by the mother, the rest of the days could be shared between the parents in any way the choose. Parental leave days may be and are often used part-time, i.e., parents work part-time and are on parental leave part-time during the same calendar period (even within the same day). Moreover, parental leave days may be saved and used later, up until the child turns eight years. For instance, some parents use some parental leave days each summer in order to prolong summer vacations. If it was a multiple pregnancy extra days per child are granted, i.e., a total of 660 days in case of twins, 840 days in case of triplets and 1020 days in case of quadruplets. Also, if the woman was on parental leave and due to disease or injury could not care for the child she could be granted SA benefits – or even DP, depending on the situation, during the period when parental leave often is taken. If the child does not survive the delivery, the mother is still granted 28 days of parental leave. In this project we have no information on days used for parental leave; our interest was on use of SA and DP days, irrespective of if the women were in paid work, on parental leave, studying, only engaged in domestic work, or unemployed the other days. Our aim was to elucidate to what extent this was associated with morbidity.

We have also, in line with what you asked for, improved the presentation of the results in Tables (Tables 1-5).

We are not sure what you mean with that the sample size is an issue; we have a large sample size (and also continuous outcomes, that is, not only the state at the end of follow-up, which is often the case in these type of studies), thus sufficient numbers even in our subcategories according to the independent variables, e.g., shown by the confidence intervals.

Comment:
Furthermore, grammar and style should be checked by a native English speaker as there are many minor errors, and words or whole sentences are sometimes vague.

Authors' response: The language has now been carefully checked in the revised manuscript.

Comment:
Abstract
Objective – Second sentence is confusing, and "if exists" should be fixed. Not clear what "or inclusion year" refers to.

Authors' response: We have now rewritten the abstract to increase clarity in the parts highlighted by the reviewer; please see page 2.

Comment:
Participants - do they need to have lived in Sweden during all the prior and follow up years as well?

Authors' response: No, only in the three prior years, not after inclusion. We now specify in the abstract that besides being nulliparous women had to be residents on 31 December 2004, as well as in the three years prior to this date, to be included in the study (page 2, paragraph 4). We specify in the Methods section that during the follow-up women were censored upon death or emigration (page 10, paragraph 1).

Comment:
Measures – Does SA>14 mean 14 days? Regarding DP days, is this any number of days?
Authors' response: As the Social Insurance Agency does not reimburse the first 14 days of SA spells for employees, we did not have information on those days. In order not to introduce bias regarding those unemployed who were reimbursed by the Agency also during the first two weeks of a SA spell, we excluded all information on the first SA 14 days of all SA spells. All DP days were included, as all DP benefits are reimbursed by the Social Insurance Agency. DP is usually granted for at least one year. This is now clearer in the abstract (page 2, paragraph 5).

Comment:
Results – The results or the objective should be rephrased so that they fit together logically and the results make more of an impact: the second sentence of the Objective should be expanded to go deeper into the "knowledge regarding the role of morbidity", or the results should be worded in a way that highlights the departure from the status quo (if that is the case). The first sentence should also include "after controlling for prior morbidity and patient characteristics" since the unadjusted/crude model shows a higher risk.

Authors response: We have now rewritten the objectives and results to be clearer; due to restrictions on the word count of the abstract we shortened the paragraph presenting the background and the objectives of the study. Nevertheless, we discuss the novelty of the study in more detail in the Introduction section. We have added "After controlling for study participants' prior morbidity and sociodemographic characteristics" to the first sentence in the Results section.

Comment:
"Or July 2005 in the B0 group" can be left out. B0 is not defined until the methods section of the paper and the exact date does not add any information at this point.

Authors' response: Thank you, we have now rewritten this sentence in light of this suggestion (page

2, paragraph 6).

Comment:
Strengths and limitations section – The fact that Sweden has a high employment rate among women does not automatically infer that health selection into paid work does not occur (but perhaps other regulations that are not mentioned can explain this). The reference used to support this claim later in the discussion is not in English and can't be used to check their statement. Regardless, relevancy of paid work for their interpretation of results is not directly discussed in the paper.

Authors' response: You are right - there is always a health selection in and out of the labour force. However, if a very large proportion of the population is in paid work, more people with different types of morbidity are in paid work; we now mention this in the "Strengths and limitations of this study" section , stating that a high employment rate "limits" the health selection (page 4). Indeed, our own results show that there is such a selection, that is, the higher risk of DP among those with the types of morbidity included here. We have now changed the reference to support this claim in the Discussion section; the new reference is Statistics Sweden, 2017. We discuss also the relevance of paid work for our findings (page 27, paragraph 3).

Comment:
Background
"year" should be "years".

Authors' response: We have rectified the text as suggested.

Comment:
Reference 13 is not related to the statement.

Authors' response: Thank you for highlighting the mistake in the insertion of this reference, we have now removed it and added the right reference (Björkenstam et al., 2020). We unfortunately had several problems with the used reference program in the version we last submitted, also regarding other references, we apologize for not seeing this before. All this is now corrected.

Comment:
The sentence containing "Strong positive associations with morbidity" is vague and could use some more specifics instead of words like "this".

Authors' response: We have now changed the text, based on your suggestion (page 5, paragraph 3).

Comment:
Reference 18 does not appear to be related to the statement.

Authors' response: Thank you for highlighting the mistake in the insertion of reference, we have now replaced this reference with a new one.

Comment:
Reference 16 and 17 (prior studies by the authors) are relevant, but it leaves the reader with the question what is different about the current study. Only very late in the paper it is mentioned that this study is different because the authors expanded their previous findings to the general population – the question remains, however, whether the current contribution is significant and novel, since the prior studies, which utilize a twin cohort, arrive at a very similar conclusion, albeit with fewer details.

Authors' response: Since the literature regarding the link between childbirth and SA is not consistent, it was not clear before we performed this study that findings would be similar to those in our previous studies that were conducted in a rather selected and smaller sample of twin sisters. We now clarify in the revised Introduction why we considered important to investigate the role of morbidity in the association between childbirth and SA/DP not only in twins but also in the general Swedish population (page 5, paragraphs 3 and page 6, paragraph 1).

Comment:
It is unclear why it is "not surprising" and who is doing the questioning in the statement: "This may not seem surprising, however, this association is sometimes questioned". The references that follow the next clause consist of a non-peer-reviewed working paper (19), as well as a reference that is not related to the topic. Thus, the main motivation for this paper is not well-supported. It is not clear what the sentence that follows (on whether SA and DP are good measures) is adding to the authors' arguments, and one of the references related to bicycle accidents is not relevant.

Authors' response: We agree and have now revised the sentences you refer to (page 6, paragraph 1). Being on SA more than 7 days in Sweden requests a medical certificate (page 9, paragraph 1), thus we believe an association between morbidity and being on SA is not surprising – however, this has repeatedly been questioned, especially regarding SA among women. We have now reformulated the sentences the reviewer refers to, to clarify the questions she raised and strengthen the rationale of the study (page 5, paragraphs 3 and page 6, paragraph 1-2). We also refer to a peer-reviewed publication (Angelov et al., 2008) instead of the previously cited non-peer-reviewed working paper. We have removed the reference related to the bicycle accidents.

Comment:
Methods
Use the word "following" once in the last sentence of the first paragraph.
Authors' response: We have revised the text as suggested (page 6, paragraph 3).

Comment:
Study population - Why did the authors decide on ages 18-39 years and not younger and/or older? Specific details on what happens with women in the analyses if they gave birth at these extremes e.g. age 18 or 39, especially considering the long lead-in and wash-out period, would be helpful. There is no mention how they dealt with deaths, women who leave the country or are in long-term care facilities. How are women who experienced events such as miscarriages, abortions, hysterectomies, stillbirth, fertilization treatments without results, etc etc categorized, if they're not excluded? Why is the final follow-up period 43 weeks and not just one extra year (Y+4) for simplicity? This adds to the confusion of all the time references (T0, Y+1-Y+3, Y-1) and the groups B0, B1, and B1+. It appears that the authors in most cases use the full explanation in most sentences throughout the paper (e.g. "indicating morbidity 1-3 years before T0 (Y-3 - Y-1), and/or 1 year after T0 (Y+1)"), and I see no reason to just describe the events and groups with words rather than T0, B1+ etc. throughout the paper to improve the readability.

Authors' response:
We now present in the manuscript the rationale for our choice of the age intervals and that women in the extremes were not treated differently in the analyses than women of other ages (page 7, paragraph 3). Regarding the age of 18, we wanted the women in the cohort to have at least some opportunity to have had SA, which can be granted to people aged 16 and above, in at least a couple of years before inclusion. The choice of 39 as upper age limit was mainly based on the very low number of women having a first birth after that age. We now more clearly specify that women who died or emigrated during the follow-up were censored regarding information on SA and DP during the year when this occurred (page 10, paragraph 1). Women with long-term hospitalization, e.g., due to mental diagnoses are followed with the registers similarly to women not in such care. We also clarify

that women with miscarriages, abortions, hysterectomies, stillbirth, fertilization treatments without results were kept in the analyses and could theoretically be in any of the three groups (page 10, paragraph 1). The 43 weeks was included to be sure that no SA in the third follow-up year would be due to problems related to a new pregnancy. We are well aware of that the number of weeks could have been set to 40 instead, or even less. However, here we wanted to be sure. Using 52 weeks, as you suggest, would mean that more women than necessary would have been excluded without any logical reason for this. You suggest this for simplicity – however, as all time is counted in relation to date giving birth, not calendar year, it would not have been easier to use 52 than 43 weeks. And, as stated, we would not have been able to justify the choice of 52 weeks.
Regarding stillbirths – they are treated as any other birth, that is, women who gave birth were included, irrespective of the outcome (dead or alive child, one or several children, etc.). Spontaneous abortions, induced abortions, and miscarriages before week 22 (i.e., the gestational age pregnancies need to reach to be included in the Medical Birth Register) were not categorized as giving birth. This is not a study of being a mother to small children. This is a study of SA and DP among women who gave birth or not, as stated in the manuscript. If she lived with the child afterwards or not (due to death, abortion, etc.), or if she lived with other children is not in focus here. Rather of morbidity before or after inclusion in the study was associated with morbidity in the three groups studied. It has in the laymen debate been claimed that childbirth does not include any health risks why it should not lead to any extra SA or DP. That debate is one of the reason for this study.
In longitudinal studies of this type it is customary to use the abbreviations of e.g., Y-3 in relation to T0 – that is T0 being the time of inclusion in the cohort.
We have now revised throughout the entire manuscript and the tables the texts where we used both the abbreviation and the full explanation in case of the study years or the childbirth groups.

Comment:
It is not clear what the date 2 July 2005 signifies. Randomization, or assignment of an index date by matching might be better to maintain some variability in the follow up period for the B0 group, since there is probably seasonality in applying for SA or DP.

Authors' response: We now clarify that 2 July 2005 represents the middle of the index year for the women who did not give birth (page 8, paragraph 2). We did not use matching, instead we choose included all the nulliparous women who had lived in Sweden in the past three years and who did not give birth in 2005, nor during follow-up. That is, all nulliparous women were included, and categorized into three groups, depending on future childbirths. Subsequently we calculated SA and DP during the three preceding and three following years, that is, e.g., the 365 days before (Y-1) and after (Y+1) 2 July 2005. We included all days for whole years for each group from Y-3 throgh Y+3, thus, it is not clear to us how you mean that seasonality of SA and DP would affect our findings. If we had calculated this for just a few months that could have been the case – and in that case, we would have used another study design.

Comment:
Morbidity - The clause that contains "operationalise" can be worded more specifically or be left out.

Authors' response: We have now left out this sentence, as suggested (page 8, paragraph 3).

Comment:
The authors need to provide more details about specialized outpatient visits. Do they include all specialty health care, regardless of severity or recorded diagnosis (suspected or confirmed), and even for preventive visits or for disorders/diseases in remission? To label someone as having morbidity with one visit to any specialist regardless of the primary reason of the visit seems too broad/inclusive/heterogeneous. At the very least, using two or more outpatient events with the same diagnostic code (not in remission) rather than one event would be a better measure of actual

morbidity.

Authors’ response: We used information on the main diagnoses given in inpatient and specialized outpatient healthcare as indicators of morbidity, i.e., morbidity requiring at least secondary healthcare (page 8, paragraph 3). In Sweden, as in many countries, the first-line healthcare is handled by primary healthcare, e.g., most of the treatments and investigations you specify are handled by primary healthcare. In this study we did not have any information about any primary healthcare visits. That is, we included the more severe morbidity, which can be seen both as a strength and a limitation. We now mention the lack of information on data from primary healthcare among the limitations of the study (page 28, paragraph 1).
In this explorative study we were interested in the occurrence of morbidity, not so much the type of morbidity, nor its severity, even if inpatient healthcare usually could be seen as more severe than outpatient healthcare and secondary as more severe than primary healthcare. However, the term ‘severity’ can apply to many aspects, one is in terms of risk for premature death, another is in terms of long-term associations with work incapacity. For many of the diagnoses requiring specialist healthcare, patients actually usually do not see their physician more than once/year, for instance when treated for multiple sclerosis, schizophrenia, etc.
Of course, many other ways to do this could have been used to obtain information about morbidity in the three groups. Here we choose to use the method used in the previous studies of smaller, more selected groups (e.g., data from the Swedish twin register), to investigate if the results could be replicated in this, larger and not so selected population; we were able to add also information on out-patient healthcare. If the results were in the same direction (as they are) one would think that this would be of interest to study in more detail. We agree that our findings show that it definitely would be of interest to pursue this line of research, focusing on different types of morbidity. However, that we will need to leave for future studies.

Comment:
On page 9, the second and third paragraph appear to be duplicates.

Authors’ response: We are afraid that we cannot see this – maybe there was a problem in how you were given the manuscript?

Comment:
Is a hospitalization stay or outpatient visit excluded completely if there is a childbirth-related diagnosis observed, or only that particular diagnosis on the event? There were conflicting descriptions in the paper. Regardless, it may be more precise to use primary vs non-primary diagnoses to determine the reason for the visit (or other diagnostic codes rather than ICD10) and possible under/overcounting of either childbirth- or non-childbirth-related visits.

Authors’ response: Regarding information on morbidity, we used the main diagnosis, as registered by the treating physician as the main reason for the visit or hospitalization. That is, if the woman was hospitalized or had specialized healthcare due to a disorder related to the pregnancy, childbirth or the postpartum period that was categorized as that. However, if she, e.g., had treatment for a fractured arm while being pregnant, it was the fracture of the arm that was used as the main diagnoses. That is, we did not include secondary diagnoses – related to pregnancy or to other comorbidity – in this explorative study. We now specify in the Methods section that we used information only on the main diagnoses (page 8, paragraph 3). However, when using the in-patient register to identify women giving birth, we used all diagnoses, not only the main diagnosis.

Comment:
The insurance section is useful for readers from other countries. More details on eligibility regarding prior work history (do people get benefits without a prior history of employment, and do immigrants

get benefits immediately) would be helpful, as would an explanation of how the authors deal with variability in parental leave duration, i.e. if people return to work earlier. As mentioned before, including only at primiparous women could reduce overall confounding by future parental leave on SA or DP. Furthermore, matching on age and other characteristics would create a much cleaner comparison and balanced sample sizes.

Authors' response: Regarding eligibility for SA benefits: as we stated in the paper, all people with income from work-related incomes (such as unemployment benefits, parental leave, etc.) are covered by this public SA insurance, regardless of if this is their first day of work. However, the person needed to be assumed to have an income of 9500 SEK that year (about 980 USD or 900 Euros/year). That is, most people are included. The same goes for students who due to long-term morbidity cannot conduct their studies. We have now included this information in the manuscript (page 9, paragraph 1). Moreover, as stated in the manuscript, all residents, irrespective of having had any income at all before, are eligible for DP benefits if having long-term or permanent work incapacity. If not having had any income before, a minimum standard level of benefit is provided. Otherwise about 65% of expected future income, up to a certain level, is provided.
That is, SA and DP benefits are not related to for how long one has worked (in many countries one needs to have been in work for a number of months in order to be covered) nor related to if one is born in Sweden or not.
Regarding parental leave – Sweden has a generous system for this; a total of 480 days of parental leave are granted for each child. Women can use up to 60 of these days before expected delivery date – however, most save these days till afterwards. Both parents can use the days, that is, it is not always the women who gave birth who used most of them; and each parent has to use at least 60 days. The days could be saved/used later, up until the child turned eight years. For instance, some parents use some days each summer in order to prolong summer vacations. If it was a multiple pregnancy 180 extra days per child was granted. Also, if the woman was on parental leave and due to disease or injury could not care for the child she could be granted SA benefits – or even DP, depending on the situation. We now clarify these aspects on page 9.
In this study we had no information on days used for parental leave; our interest was on use of SA and DP days, irrespective if the women were in paid work, on parental leave, studying or unemployed.
Thank you for these suggestions for other ways of designing studies in this area, studies with other aims. A matching procedure would imply another aim and other problems, especially as we were and are very interested in the B1+ group also.
With regard to the suggestion about including only the B1 group, please see our response to your overall assessment.

Comment:
Outcomes – can use more details, or at least whole words and sentences.

Authors' response: We have now revised the text in light of this suggestion (page 9, last paragraph and page 10, paragraph 1).

Comment:
Included factors – Why is age not used as a continuous variable? Why are the other demographic characteristics besides age and education in table 1 not included as covariates? These all seem necessary and logical covariates. How reliable are these variables, especially for immigrants? What is the strategy for dealing with education levels of people who have not reached their final degree (i.e. at age 18-25)?

Authors' response: The reason for not using age as a continuous variable was that we could not assume a clearly linear association between childbirth group and age and age and SA/DP in our cohort. We now re-run our multivariate analyses adjusting for age group, educational level, country of

birth, type of living area, and previous hospitalization and specialized outpatient visit (page 11, paragraph 1).
For those in the age group 18-25 years we used information on the highest attained education on 31 December 2004, in all three study groups. That means that they later might attain higher levels of education. The same actually goes for all age groups, as the possibilities for further education is high and promoted in different ways in Sweden – where, e.g., college and university education are free of charge as well as lower educational levels.
Many immigrants have had their highest education in Sweden, and that information is always sent to Statistics Sweden. Other immigrants are asked for such information when applying for work permit, asylum, immigration due to family reasons, etc. Those who want can have their education from other countries assessed regarding what type and level it would be in Sweden by a special Swedish authority working with this. If Statistics Sweden does not have information on the level of education for an immigrant, they try to obtain this through surveys sent to them by post. In this study we categorized those for whom we did not have information on educational level as elementary education at the most. This is due to the fact that some of the younger study participants might not have finished elementary school, due to different reasons – e.g. health reasons.

Comment:
Statistical analysis – The fact that anyone with DP prior to delivery is removed from the sample should be mentioned more prominently (e.g. in abstract); this changes the interpretation. It would be interesting to see some statistics on censoring (death and emigration).

Authors' response: We are now more clear about the fact that women with DP were excluded from some of the analyses (page 2, paragraph 6 and page 11, paragraph 1). We also mention that women who died or emigrated were censored in the analyses at the time of the event (page 10, paragraph 1).

Comment:
Results
What statistical method is used to compare the results in table 1? The authors claim no difference between the three childbirth groups, but no statistical tests have been done or are reported (here or in the methods section). Some of the morbidity measures do appear different (specialized outpatient visits for B1 and B1+ vs B0) but this is not addressed nor supported with statistical tests.

Authors' response: We now specify in the Methods section that the statistical tests used to compare our three childbirth groups on factors in table 1 were chi-square tests in case of categorical variables and Wilcoxon tests in case of continuous/count variables (page 10, last paragraph). We have now added the p-values corresponding to the comparisons of the three childbirth groups on the variables in Table 1. We have now also revised the text in the first paragraph of the Results section.

Comment:
Using a consistent ordering of comparison groups in the text would make the last two sentences more readable.
The small sample size in the SA and DP groups is a major issue, especially considering these subgroups are split up even further in subsequent analysis (morbidity vs no morbidity, short- and long-term SA). Is there a reason the analyses were restricted to childbirths in 2005 only? Adding more years could greatly improve the credibility of the final conclusions, as long as SA and DP policies did not change during the timeframe.

Authors' response: We have now rewritten the last two sentences in the Results section according to your suggestion.
Regarding the other issue you raise, we actually, see the large numbers we have as one of the main strengths of this study and are quite pleased that we have detailed data for such a large number of

nulliparous women – this is the first study with such a large dataset and the majority of our analyses are well-powered; please see the confidence intervals in our results. As we have such a large cohort, we did not consider adding more years. Also, it was for this year we had could have the best coverage of all information regarding baseline.

Comment:
The mean annual number of hospitalization days is not really a good metric, as the majority of people have no hospitalizations, and long stays (that also affect the probability of pregnancy in multiple ways) will skew the average.

Authors' response: One can use different measures regarding hospitalization. Here we chose to use four different such measures, each of them for three childbirth groups (B0, B1, B1+), i.e.: (1) mean number of days, (2) mean number of days excluding those related to pregnancy, childbirth or postpartum and (3) number of women with any hospitalization day, and (4) number of women with any hospitalization day excluding those related to pregnancy, childbirth or postpartum, respectively – for the years before and after childbirth. That is, we use both incidence and duration measures. Using the median would have been problematic since in case of some of these skewed measures the median in each childbirth group is 0 and thus would not have provided the possibility to meaningfully compare the groups.

Comment:
In the legend for figure 1d, "hospitalizations" should be "visits".

Authors' response: Thank you for alerting us, we have now revised the text.

Comment:
The results in tables 2-4 are displayed in a complicated manner, and the text does not help in clarifying the results. It appears to be impossible to interpret the effect of prior morbidity (model 2 compared to model 1) for each subgroup as reference values are set to 1 across models. Issues of potential collinearity between morbidity prior to and after delivery are not addressed.

Authors' response: We have now improved the presentation of Tables 2-5. We performed collinearity tests concerning morbidity prior to and after childbirth for all women and for women who had at least one childbirth. In the output shown below, the indicator Condition index does not show potential collinearity between morbidity before and after childbirth; a cutoff value of 30 is usually used for further investigation of collinearity. We now mention that there was no evidence of collinearity on page 11, paragraph 1)

Collinearity test for morbidity 1-3 years prior to and 1 year after childbirth in all women (n=470,656)
Collinearity Diagnostics
Number Eigenvalue Condition
Index Proportion of Variation
Intercept Morbidity before Morbidity after
1 1.37320 1.00000 0.31340 0.27049 0.05528
2 1.00000 1.17184 0 0.13692 0.82363
3 0.62680 1.48014 0.68660 0.59259 0.12110

Collinearity test for morbidity 1-3 years prior to and 1 year after childbirth in women who had at least one childbirth (n=38,413)
Collinearity Diagnostics
Number Eigenvalue Condition

Index Proportion of Variation
Intercept Morbidity before Morbidity after
1 1.68859 1.00000 0.15570 0.15401 0.00595
2 1.00000 1.29946 0 0.01087 0.96176
3 0.31141 2.32862 0.84430 0.83512 0.03228

Comment:
The sample sizes of the subgroups are probably extremely small in some cases (e.g. for combination of SA spell > 90 days & 1 child & no morbidity prior & morbidity during follow up), but there are no details, only sample sizes for the overall groups.

Authors' response: We now add the number of outcome events in Tables 2-5. To our knowledge this is the largest study to date in this field and the majority of our analyses are well-powered; please see the confidence intervals in our analyses.

Comment:
Is it not completely clear what is different in table 4 compared to table 2 and 3.

Authors' response: We have now improved the presentation of our tables. Tables 2 and 3 focus on the importance of any morbidity, whereas tables 4 and 5 focus on importance of cumulative morbidity.

Comment:
Discussion
References 40 and 41, related to women's roles in society and how that affects their health are from 1982 – considering the developments over the last 40 years in this area the authors should try to find more recent references.

Authors' response: You are right, as stated above, this and other such mistakes are due to a failure of the reference program we used. It is now corrected.

Comment:
The conclusions regarding better health for those who gave birth appear stretched, at least considering the results of the unadjusted/crude model for SA, although it is hard to interpret the outcomes as they are currently presented. There is also a potential (voluntarily or involuntarily) self-selection into groups and there is confounding by parental leave and multiple births – these issues are only minimally discussed and not controlled for in the analyses.

Authors' response: You are right in that issues of better health among those giving birth is not of interest in this study, and we have now deleted such text from the Conclusions. Of course there is a strong self-selection into all three groups. Very few give birth without wanting to do so nowadays, due to good contraception possibilities, good possibilities for legal and safe abortions, most women being self-supporting etc. There might be a higher risk of involuntarily belonging to the B0 and B1 groups. We cannot, of course, control for this here, as we have no such information. However, the question of self-selection into any of the groups is not of interest in this study. That could become issues for future studies, however, quite other types of data and data collection methods would be needed for such studies. We consider parental leave and multiple births rather as mediators than confounders of the association under study, thus adjusting for them would result in over-adjustments. We have now slightly revised the text of the Conclusions considering this comment and add that the conclusions are based on the multi-adjusted results.

Comment:
A discussion on why SA and DP outcomes are going in different directions for the childbirth group in

the crude/unadjusted model would be useful.

Authors' response:
The difference in the hazard ratios among the three models in table 2 and 3 that you refer to is due to differences in sociodemographic characteristics and especially in previous morbidity (a substantial proportion of which in the B1 or B1+ groups is likely to be pregnancy-related morbidity) and also in SA during the follow-up between the "B0, no morbidity in Y1" versus the groups "B1, no morbidity in Y1" and "B1+, no morbidity in Y1". In other words, when considering sociodemographic factors and previous health, women in B1 (and in case of SA>90 also women in B1+) are less likely to experience SA during the follow-up than those in B0.

Comment:
Reference 44 is not in English.

Authors' response: Again- references are now corrected.

Comment:
Conclusions
A strong association between morbidity and SA/DP (second sentence) does not necessarily negate the claim in the first sentence. Also, the conclusion that morbidity is associated with SA and DP should come as no surprise to the reader; it is the interaction with childbirths that makes this paper interesting. The following sentences assumingly address this, but should include "after controlling for morbidity", otherwise they are contradicting the unadjusted results. Also, if prior morbidity is indirectly caused by pregnancies (as stated in the discussion), that still means that childbirth is associated with SA and DP; controlling for it may attenuate the effect in these analyses on paper but does not reduce the original association. A discussion on the implications of these interactions on the health benefits system and the authors' point of view in this matter would be helpful to determine the overall significance of the authors' findings.

Authors' response: We have revised the conclusion in light of these suggestions.

Comment:
The statement regarding "prioritizing domestic duties" is strongly worded and insufficiently supported by the articles cited by the authors (see comment on references 19 & 20 in Background).

Authors' response: We have now reformulated this part of the conclusion and cite a reference published in a peer- reviewed publication (Angelov et al., 2018).

Reviewer: 2
Reviewer Name: Norhayati Binti Mohd Noor
Institution and Country: Universiti Sains Malaysia, Malaysia

Comment:
It is obvious that SA would have a strong association with DP. It is reported in the results that nulliparous women had the highest risk of SA and DP followed by one and more than one childbirth. This is the result of previous studies as stated in the Introduction section [12-15].
In the current study, the interest is to identify the extent of morbidity in these groups of women with SA.

Authors' response: Thank you for this comment. You are right that in most cases SA precedes DP. However, you can only be granted SA if you have some kind of income, e.g., from work or unemployment or parental-leave benefits. If you have/get long-term or permanent work incapacity due to injury of disease, you can still be granted DP, even if not having had any SA before. As some of your subsequent comments highlighted that the aims of the study were not clear, we have now rewritten the Introduction and better highlighted the rationale and aim of the study, i.e., to investigate whether SA/DP among nulliparous women of childbearing age, and especially among those who give birth, is associated with morbidity (pages 5-6).

Comment:
The explanation in the current study should be included in the previous papers to support their statement. It is a redundant publication. It still does not answer the question of WHY nulliparous women had the highest risk of SA and DP.

Authors' response: We are now clearer about the rationale and aim of the study, i.e., to investigate whether morbidity is associated with SA/DP among nulliparous women of childbearing age, and especially among those who give birth. We also highlight that this has been an important debate in Sweden and other welfare states in the past decade and that we have not studied this question in our previous population-based studies. Though a similar question has been investigated earlier in a small sample of twin sisters, it was not clear to what extent findings from this selected sample twin sisters (n≈5000) would be replicable in a larger, population-based cohort of all nulliparous women in a country (n>490,000), also using a wider indicator of morbidity (both hospitalization in specialized out-patient healthcare, compared to 'only' hospitalization' in the twin study) (page 5, paragraphs 3 and 4 and page 6, paragraph 1)..

Comment:
ABSTRACT
The results of this study are repeating the results in the previous study [12-15] but yet do not answer the objectives of the current study.

Authors' response: We are now clearer about the objectives of this study in the abstract, i.e., to investigate associations of morbidity with subsequent SA and DP among initially nulliparous women of childbearing age with no, one, or several childbirths during follow-up (page 2, paragraph 1).

Comment:
Do not use abbreviations in the abstract. What does B0 mean?

Authors' response: We have now removed abbreviations from the abstract and refer to women not giving birth instead of to B0.

Comment:
DISCUSSION
The discussion is more of a statistical repetition of the results.

Authors' response: There are very few studies in this field and thus more detailed comparisons with previous results were not possible. In this revised manuscript we now in the Discussion section better integrate our findings in the context of the study (page 26, paragraph 2).

VERSION 2 – REVIEW

REVIEWER	Hanke Heun-Johnson University of Southern California
REVIEW RETURNED	10-Jun-2020

GENERAL COMMENTS	The authors have clarified the objectives, background, and

methodology sufficiently. However, I am still questioning the overall novelty of the paper, and the authors did not (address) or were not able to markedly improve the issues related to the methods, outcomes, interpretation, and implications that were brought up in the previous review.

Remaining main issues:
The measures of morbidity are too crude (e.g. no measure of severity, using a dummy (one?) for hospitalization and specialized outpatient healthcare), especially for a paper that attempts to study the effect of morbidity on SA/DP.

There seems to be pregnancy-related confounding in the year prior to childbirth (fig 1d) for one of the main/explanatory covariates, something that wasn’t clear to me in the prior version, which I believe undermines a large part of the conclusion. It seems like -despite removing pregnancy/childbirth related diagnoses- there is still an uptick around T0 in specialized outpatient visits, thus it’s doesn’t appear to be an true indicator of underlying morbidity.

The tables remain hard to interpret; setting all models to the reference value of 1 makes comparison between models impossible. This has not been changed since the prior version.

The confounding due to the duration of parental leave, and the absence of information on this important variable make the results very hard to interpret.

The authors added the requested sample sizes to the tables, and as expected, many of these are very small. Having a large control group (as stated in the authors’ response) does not make the B1(+) subgroups much more reliable; reporting an n of 8 in the tables may even be violating the regulations of the data provider (an n of 11 or 12 is normally required, at least in the US). Considering the availability of data in Sweden, the authors may wish to combine several years to increase the reliability of their subgroup analyses.

Minor issues:
Background: The sentence: “It is, thus, of interest to study to what extent morbidity in these groups of women is associated with SA” does not make sense in its current context.

The n of the twin cohort study is mentioned twice in the same section.

I would advice against the word “often” at the start of the 4th paragraph; the statement still appears somewhat unfounded, but at least there is a better reference cited now. The word “rather” appears twice in one sentence.The last sentence of the 4th paragraph “Actually, also in …” seems unrelated to the goal of the paper since the authors are not testing people’s knowledge of the link between morbidity and SA/DP.

Morbidity: As mentioned before, mean nr of hospitalizations is not normally distributed and will be extremely affected by outliers, and differentially by group. The authors could present the percentage of people with a hospitalization (especially if that dummy is in the model) rather than the mean number in the results.

Statistical analyses: “<” character in 201<8.

	Table 1: It is not clear what the p-values are referring to for categorical variables: compared to B0? All cells compared? If the latter, does this make sense? Figure 1d: The statement “women in the B1 and B1+ groups had lower number of hospitalization days and specialized outpatient visits than women in B0, particularly the women in B1+” is not correct for outpatient visits in B1 group.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Hanke Heun-Johnson

Institution and Country: University of Southern California

Comment:
The authors have clarified the objectives, background, and methodology sufficiently.
Response: thank you.
However, I am still questioning the overall novelty of the paper, and the authors did not (address) or were not able to markedly improve the issues related to the methods, outcomes, interpretation, and implications that were brought up in the previous review.

Response:
Actually, there are no previous population-based studies of associations between morbidity of women in relation to giving birth (of one or more children) or not giving birth and their sickness absence (SA) or disability pension (DP). The only previous study on this, conducted by our group, was based on data for a, comparably, low number of twin-sisters born in Sweden (about 5000 women compared to about 490,000 in this study). Moreover, in the previous study only hospitalizations could be used as a proxy for morbidity. Here we had access to much wider data and for all individuals in Sweden fulfilling the inclusion criteria. That is, there is no question about that this is a very under-studied area and that this study provides results at a level not published before.
Maybe this comment more concerns the need for this type of study than the novelty?
In Sweden, as in many other countries, women have higher SA than men and especially during pregnancy and in many cases also in the years caring for young children. The latter has often been questioned, i.e. whether it is really due to work incapacity caused by morbidity or by other things. In the public debate in Sweden and in other countries it has been argued that women rather stay home in order to handle domestic work tasks. This ‘commonly accepted’ idea has been very prevalent, and probably affects employer’s willingness to hire women in fertile ages. Thus, in addition to contributing to scientific evidence, an important rationale for this study was to inform employers, policy makers and to contribute to this debate. We have now rewritten the Introduction to better highlight the rationale and the novelty of our study (pages 5 and 6).
In some countries it is assumed that the mothers also are staying home to care for sick children, something that is difficult to study. However, Sweden has a very generous benefit program to care for sick children who e.g. cannot be in ordinary daycare or school due to morbidity – a benefit system that actually is more generous than the SA benefit system. Therefore, we do not suspect that the women studied here were on SA in order to care for sick children.

Regarding the methods, we realized that we previously were not clear enough about the different measures of morbidity used, this has now been clarified in the revised method section. We have also included information on numbers not only of the full cohorts but also in the different sub groups analyzed in Tables 2, 3, 4 and 5. The n in the subgroups varies from 373 to 417,592 in the different analyzed groups. We do not regard that as small numbers, and realize that the reviewer mistook the out-come numbers for the 'sample size' numbers. We have clarified this in more detail in our responses below. We now add both number of outcomes and number of sample sizes corresponding to different exposure categories in Tables 2-5.
Regarding the outcomes, that is, the four different measures of SA and DP used: we are not sure what the reviewer finds problematic with them. In the literature at least 100 different measures of SA/DP are currently used, which of course is confusing in the research area of SA/DP/insurance medicine. Here we use aspects including both incidence and duration of SA and DP.
Regarding interpretations: as this is one of very few studies regarding associations between morbidity, childbirth, and SA/DP, we have been very careful with interpretations/conclusions in order not to be speculative. We merely state the findings, that is, the positive association between morbidity and SA/DP in all studied groups, and that those who gave birth had lower SA/DP than those we did not.
Regarding implications, we are not sure what you mean with this, we do not speculate about implications based on the results from this explorative epidemiological study.
We appreciate the reviewer's suggestions for future studies, using other cohorts, other types of data, other study designs and analyses and hope that this study will inspire others to conduct such studies.

Comment:
Remaining main issues:
The measures of morbidity are too crude (e.g. no measure of severity, using a dummy (one?) for hospitalization and specialized outpatient healthcare), especially for a paper that attempts to study the effect of morbidity on SA/DP.

Response:
Actually, we are glad that we had any data on morbidity, as such studies are lacking. In this explorative study we were interested in the occurrence of morbidity, not so much the type of morbidity, nor its severity, even if inpatient healthcare usually could be seen as more severe than outpatient healthcare, and secondary healthcare as more severe than primary healthcare –we had no information on primary healthcare. However, the term 'severity' can apply to many aspects, one is in terms of risk for premature death, another is in terms of long-term associations with work incapacity. For many of the diagnoses requiring specialist healthcare, patients actually usually do not see their physician more than once/year, for instance when treated for multiple sclerosis, schizophrenia, etc. The same goes for other very severe diagnosis. Hospitalization might also be more related to other things than severity, e.g., distance from home to the treating clinic. If an individual, as in parts of Sweden, live 600 km from the clinic, you might have to stay overnight before/after a treatment/surgery etc., but not if living in more densely populated areas.
In this observational epidemiological longitudinal cohort study we did not study 'effects of morbidity', but associations. For effect studies other types of study designs would be needed, study designs that are mainly not possible to conduct.
We have used different measures of morbidity, i.e., number of days in hospital or number of specialized outpatient visits – both with and without including such care related to pregnancy, childbirth and puerperium; specialized outpatient healthcare and hospitalization in Y+1, and cumulative morbidity in different time periods (as in Tables 4 and 5). Of course, many other ways could have been used to obtain information about morbidity in the three groups. Here we chose to apply the method used in the previous studies of smaller, more selected groups (e.g., data from the Swedish twin register), to investigate if the results could be replicated in this, larger and not so

selected population; we were able to add also information on out-patient specialized healthcare. If the results were in the same direction (as they are) one would think that this would be of interest to study in more detail in future studies and we hope that the results will inspire to such studies. We agree that our findings show that it definitely would be of interest to pursue this line of research, focusing on different types of morbidity as well as severity of that morbidity. However, that was not the aim of this study.
A comment of the reviewer (below) made us aware of that we had not been clear enough in defining the morbidity measures we used, why the text in the method section has now been revised (pages 8-9).

Comment:
There seems to be pregnancy-related confounding in the year prior to childbirth (fig 1d) for one of the main/explanatory covariates, something that wasn't clear to me in the prior version, which I believe undermines a large part of the conclusion. It seems like -despite removing pregnancy/childbirth related diagnoses- there is still an uptick around T0 in specialized outpatient visits, thus it's doesn't appear to be an true indicator of underlying morbidity.

Response:
As seen in figure 1, the number of hospitalization days and specialized out-patient visits increased much in the 12 months before childbirth among the two groups of women giving birth. When removing healthcare with diagnoses of pregnancy, birth, and puerperium, this was not the case regarding hospitalization days, however, regarding specialized out-patients visits, though at a low level. We are not surprised by this. Several diseases, e.g., autoimmune ones such as multiple sclerosis are diagnosed during a pregnancy or require more visits with the neurologist during pregnancy. The same goes for some other chronic conditions, e.g., rheumatoid arthritis and some mental or cardiovascular diseases. Other studies indicate a higher risk of domestic violence when pregnant, especially during a first pregnancy. It might also be so that women who are pregnant, especially if it is the first child, are not only more aware of different types of symptoms but also to a larger degree seek healthcare for them. We discuss these on pages 5 and 6. We do not see this as a confounder. Also, we had no concept of 'underlying' morbidity. We choose information about inpatient and specialized out-patient healthcare as one, of course crude, indicator of morbidity. Moreover, most people with different types of morbidity are not sickness absent due to the morbidity as it does not affect their work capacity to a level high enough to lead to SA. Nevertheless, we found a strong association between morbidity, as measured here, and SA/DP. Again, this is one of the first exploratory studies regarding these issues, and as many such studies lead to further questions. Even if the results in the d) part of the figure 1 are not surprising from a clinical view, they certainly inspire research questions regarding the specific diagnoses. We hope that such studies will be conducted.

Comment:
The tables remain hard to interpret; setting all models to the reference value of 1 makes comparison between models impossible. This has not been changed since the prior version.

Response:
This study and its result presentation are conducted within the epidemiological tradition. There are, of course, other ways those complex data could have been analyzed, using several different analytical methods and/or ways of presenting them – here we used an epidemiological study design and standard epidemiological analyses. As this is the first larger study of this subject, we chose to be very explorative, presenting several comparisons. As standard in epidemiology, we in the regressions used as the reference group the unexposed, i.e., the one that did not have any of the two "exposures" (i.e., childbirth and morbidity). Using the same reference group (parallel to having a common denominator) allows the interested reader also to compare the different groups to each other (by dividing their respective HRs with each other); one may, e.g., calculate the association between morbidity and

SA/DP within each childbirth group or the relation of giving versus not giving birth to SA/DP within each morbidity group. We included also analyses when we compared the two groups of parous women to each other, as that was a specific question of interest (that is, do the result differ between women who gave birth 'only' once during the study period compared to those who had more than one birth). We appreciate that the standard epidemiological ways of showing data might be a challenge if not used to that.

Comment:
The confounding due to the duration of parental leave, and the absence of information on this important variable make the results very hard to interpret.

Response:
You are right, we do not have information on parental leave. But, our aim was not to elucidate any associations with parental leave on SA/DP. We rather were glad that we had as much detailed information as we had for each of these >490,000 women, something relatively unique. Our interest was on their SA and DP days, irrespective of if the women were in paid work, on parental leave, studying, only engaged in domestic work, or unemployed the other days. Our aim was to elucidate to what extent morbidity was associated with SA/DP in these three groups of women. Thus, our aim was not to relate this to parental leave – if the women took this or not, nor for how long. We were here also not interested in the extent the women were in paid work – another important aspect.
What we studied is a very complex phenomenon, with factors at different structural levels being involved, and we, of course, do not aim to provide a full explanation. This exploratory study provides additional knowledge, that future studies can be inspired by, a little jigsaw piece, so to say.
We do not assume that morbidity, as measured here, was dependent on whether the women used parental leave days or not. Some women might not have taken any parental leave, some took it for a very long time, as we wrote quite extensively about in our previous review response to you. Also, those on parental leave who were hospitalized or due to morbidity could not care for the child, could claim SA, or even DP, depending on the situation; in this case she would not use parental leave (page 10, first paragraph). However, as clear from the results we presented, use of SA was much lower during Y+1. This was expected and has been shown in many other studies, as was our result regarding higher levels of SA during pregnancy was used. This is also why we in the regression models focused on Y+2 and Y+3 and not on Y+1.

Comment:
The authors added the requested sample sizes to the tables, and as expected, many of these are very small. Having a large control group (as stated in the authors' response) does not make the B1(+) subgroups much more reliable; reporting an n of 8 in the tables may even be violating the regulations of the data provider (an n of 11 or 12 is normally required, at least in the US). Considering the availability of data in Sweden, the authors may wish to combine several years to increase the reliability of their subgroup analyses.

Response:
The numbers the reviewer refers to – as indicated in the heading of the tables – are not the sample sizes, but the numbers with the respective "Outcomes"! The number of women in respective exposure group were of course many more. We have now in Tables 2-5 also added the 'sample sizes', that is, the women at risk in each of the analyses. The lowest number is 373 women, the highest 417,592 – that is, large groups.
We, of course, hope that this study will inspire others to conduct the same type of studies, based on other cohorts and other years. So far this is the largest study ever conducted in this area. Based on the results one in future studies might want to do as you suggest. We consider that our analyses have sufficient statistical power, for instance, the overwhelming majority of the confidence intervals in Tables 2-5 do not include 1.

We are, of course, extremely careful not to violate regulations of data providers. We use the standard of not presenting results involving numbers below 8, hence, it is not possible to identify the individuals based on the information given.

Comment:
Minor issues:
Background: The sentence: "It is, thus, of interest to study to what extent morbidity in these groups of women is associated with SA" does not make sense in its current context.

Response:
We have now deleted this sentence.

Comment:
The n of the twin cohort study is mentioned twice in the same section.

Response:
We now mention the n of the twin cohort only once.

Comment:
I would advice against the word "often" at the start of the 4th paragraph; the statement still appears somewhat unfounded, but at least there is a better reference cited now. The word "rather" appears twice in one sentence. The last sentence of the 4th paragraph "Actually, also in …" seems unrelated to the goal of the paper since the authors are not testing people's knowledge of the link between morbidity and SA/DP.

Response:
Thank you, we have revised the sentence considering also these suggestions.

Comment:
Morbidity: As mentioned before, mean nr of hospitalizations is not normally distributed and will be extremely affected by outliers, and differentially by group. The authors could present the percentage of people with a hospitalization (especially if that dummy is in the model) rather than the mean number in the results.

Response:
In the results you probably refer to, in figure 1, we present the number of hospitalization days, not the number of hospitalizations. However, in Table 1 we, as you suggest, presented the number of women in each of the three groups with at least one hospitalization (defined as having had at least one hospitalization day), both including and excluding those hospitalizations related to childbirth – both in the years before and after inclusion. We understand from your comment that we have not been clear enough in the method section on the two measures of morbidity we used, and have now revised the Methods section (pages 8-9).
There are different ways one can present these types of results. As there has only been one study about this subject before, we in this larger and population-based study choose to use some of the measures presented in that previous study, to increase comparability.

Comment:
Statistical analyses: "<" character in 201<8.

Response:
We have removed "<".

Comment:
Table 1: It is not clear what the p-values are referring to for categorical variables: compared to B0? All cells compared? If the latter, does this make sense?

Response:
We agree that p-values might be questioned here, which is why we had not included them in the first manuscript. We now present also in the footnote of Table 1 that the p-values – added based on a request during the previous review – corresponds to chi-square tests in case of categorical variables and Wilcoxon tests in case of continuous/count variables. These tests indicate whether there are differences across exposure groups in the variables under study. Since there are three childbirth groups and several variables have several categories, it would demand much space (but not be substantially informative) to present p-values at a more detailed level. With this large sample size in case of most comparisons differences would be expected to be statistically significant at the conventional p<0.05 level, both overall and if we compare specific groups; thus the actual values (n and %) will be more informative than individual p-values for the different comparisons.

Comment:
Figure 1d: The statement "women in the B1 and B1+ groups had lower number of hospitalization days and specialized outpatient visits than women in B0, particularly the women in B1+" is not correct for outpatient visits in B1 group.

Response:
Thank you, we have now revised the text.