PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Suicide Risk in Relation to Air Pollen Counts: A Study Based on Data from Danish Registers
AUTHORS	Qin, Ping; Waltoft, Berit; Mortensen, Preben; Postolache, Teodor

VERSION 1 - REVIEW

REVIEWER	Besancenot, Jean-Pierre 1Réseau National de Surveillance Aérobiologique
REVIEW RETURNED	19-Jan-2013

THE STUDY	The fourth point is unjustified
GENERAL COMMENTS	The aim of this paper was to test the hypothesis that airborne pollen, and more especially tree pollen, may be a significant immune trigger for suicidal behavior. The study was about a population of 2.86 million (what account for more than half the national population of Denmark) from January, 1982 to December, 2006 (1304 consecutive weeks). A so extensive sample size has the advantage ensuring a strong statistical power in the analysis; however, it may find significant connections that are not dominantly present in the real world. According to the authors, which are most likely right, it is the first attempt to test the association between atmospheric pollen and suicide risk using empirical and individually linked data. The statistical analysis used a generalized additive Poisson model and was rigorously conducted. The data were correctly adjusted for possible confounding effects of location, calendar time and weather conditions as temperature, cloud cover and humidity – factors that have been suggested to be possible bioclimatic mediators of seasonal peaks of suicide through changes in neurotransmitters and hormones. Lastly, the stratified analyses were wisely performed by sex, age group and history of mood disorder of the subjects. Results indicate a significant association between suicide risk and air pollen counts. In the main the observed association remained after controlling for effects of region, calendar time, and weather conditions. The influence of pollen counts on suicide differed significantly by sex. So, males appeared to be more sensitive to lower concentrations of pollen, while females tended to show a stronger effect but only when the pollen counts reach to a certain level. Interestingly, the authors did not note any additional effect of a very high level of air pollen (more than 100 grains per cubic meter of air and per day). The personal history of mood disorders only had slight influence on suicide rate in the population, regardless of sex and age. The discussion is rather relevant and well managed. The paper is supported by a substantial bibliography, which includes 44 references; 35 of them are after 2000, and 23 after 2005, the most recent ones being dated 2012. I only have a couple of quick points. First of all, only data of grand pollen counts, i.e. all taxa confounded,

	were taken into account. The authors try to justify this practice on the basis of the limited availability of the data. Such an assertion is surprising: if the Danmarks Meteorologiske Institut has not correctly stored all the files, these could probably be obtained from the European Aeroallergen Network Pollen Database. Additional information on that point would be welcome. In particular, birch pollen really deserved to be considered separately. Second, it would have been helpful to indicate why the study closes in 2006, while the paper is only submitted in December, 2012. Finally, among the possible mechanisms underlying the link between pollen and suicide, the authors emphasize, and rightly so, that aeroallergens may precipitate suicide via mediators of allergic inflammation. But it might not have been too much to devote a few lines to the side effects of some antiallergic drugs (pseudoephedrine, antihistamines and corticosteroids), which may worsen prosuicidal factors such as night-insomnia, day-somnolence, agitation, anxiety, impulsivity, and cognitive disturbance. In my view, with these minor additions, the paper could be published.

REVIEWER	Ho Kim, Ph.D. PROFESSOR of BIOSTATISTICS & EPIDEMIOLOGY Associate Dean, SCHOOL OF PUBLIC HEALTH, SEOUL NATIONAL UNIVERSITY 1 KWANAK-RO KWANAK-GU, SEOUL 151-742, KOREA
REVIEW RETURNED	11-Feb-2013

THE STUDY	This manuscript aimed to report an association between suicide and pollen exposure with individual level data from Danish registers. I think, as they said, this is the first paper successfully showing the association. The paper is well written in general. My concern is about the statistical model. It is well kwon among air pollution epidemiologists that the estimated effects are very sensitive to the choices of variables and other options of the model. I don't think the authors provided enough information on their models. Detailed information of the statistical model should be included to show the validity of this study. I strongly recommend to include a professional statistician as a co-author of the manuscript. They used "none-parametric" which is an unusual terminology. "non-parametric" or "nonparametric" is a proper word. Please google "none parametric" and see what happens. Please refer to the comments below. 1. I'm not quite sure that the final model is valid. At Fig 1, CI's for large value of total pollen covers 0 to 1 which means 100% uncertainty. And I'm not sure about the meaning of "Pr>Chisq: 0.0126". Please specify the exact model at the bottom of Fig 1. 2. It is not quite sure how they control for other factors like temperature, cloud cover? I guess they were added linearly into the model. It is well kwon that temperature shows nonlinear association with mortality. Please be more specific on describing adjustment methods for other variables. 3. The authors should compare the effect of pollen to other variables like temperature, season, etc. This comparisons will be helpful to discuss the effect of pollen with known risk factors. 4. How to control time is not that simple. We have to control long-term trend, year-to-year variation, seasonal and monthly trend.

	Please be more specific on time trend adjustment.

REVIEWER	Fotios Papadopoulos, MD Associate Professor Dep of Neuroscience, Psychiatry Uppsala University
REVIEW RETURNED	17-Feb-2013

THE STUDY	Only completed suicides are included in this study. There is no control group. Results cannot be generalized for the whole population or even the population of patients with depression or allergy.
GENERAL COMMENTS	This is a well-written paper with an interesting research question, interesting statistical approach and engaging discussion of the results. The combination of individual-level data in a time-series analysis is undoubtedly advantageous and superior over simpler ecological approaches. Sub-analyses by gender, age groups and history of mood disorders are also adding value to the paper. **Major concerns** 1. pollen counts as the exposure measure are obtained from two stations covering the two included regions, the Middle Jutland Region with an area of 13.000 km² and Copenhagen. No data is provided on the variation of pollen counts within regions. I am not so confident with Denmark's geography, but I believe that the including regions have a quite large area and include both rural and urban/semiurban areas, something that may also influence pollen counts. Thus these pollen counts should be seen as rough approximations of the actual exposure in individual level. This should be discussed in the paper. 2. Please provide description of the raw data, preferably as a time-series graph of suicide and pollen counts, temperature etc. 3. Suicide is a very complex phenomenon. It would be very interesting to quantify the potential of pollen-associated pathophysiological mechanisms in suicide prevention. How much of the suicide variance can be explained by pollen counts versus by other climatic variables? 4. The authors control for climatic variables in their analyses. I wonder how cloud cover is defined. If it takes into account day length or not. If not then sunlight would be a better measure of sunlight exposure, which is associated with suicide risk in the literature. 5. Please provide more information on missing values for pollen counts and how those were managed. It is stated eg that if there were data on pollen counts for 5/7 days, the average of these 5 days was used. What if there were pollen data only for 2 or 3 days during some weeks? 6. Table 2: please show results (RR and 95% CI) for the adjusting variables **Other comments**

	1. Abstract, in outcome measures: "… adjusted the data for location, seasonality and meteorological conditions". How did you adjust for seasonality apart from including the meteorological variables? 2. p.5, third paragraph: " Psychiatric hospital or ward with a diagnosis…". Please add that diagnoses from outpatient settings were also included. 3. p.8, first paragraph: "…an increase of pollen counts…" and "Pollen counts arising from…". Please rephrase so that a temporal change of pollen counts is not implied. 4. p.8, third paragraph: Please rewrite, so that the striking findings with the highest RR for suicide among males with mood disorders are presented. 5. P.9, third paragraph: "Etiological study" refers probably to "ecological study"? 6. P.11, The second paragraph is somehow speculative. The proposed explanations can merely explain the increase in suicide within one week. Decreased quality of life and social interaction can lead to depression and feelings of hopelessness but it is highly unlikely that these effects would take place within one week. 7. P. 12, second paragraph: "did not noted" → "did not note" 8. P. 12, second paragraph: The lack of effect of a very high level of air pollen is indeed intriguing. A dose-dependent relationship would strengthen the argument of a causal association between pollen and suicide. These results are difficult to interpret and the explanation provided is not sufficient, as one would expect that high levels of pollen (more than 100) would have at least the same suicide risk increase with the "30-100" level of pollen. Unless, susceptible individuals had already committed suicide before the even higher increase of pollen, as in reality there is a temporal relationship and pollen levels succeed temporally each other. 9. Table 4. Females <35 years old, RR = 0.738 for pollen "<10". It is not mentioned in the text. Possible explanation? Chance finding? 10. Table 5. The findings on males with mood disorders should be emphasized as the most striking increase in suicide risk.

VERSION 1 – AUTHOR RESPONSE

Reviewer:Jean-Pierre Besancenot, PhD
National Network of Aerobiological Monitoring France The reviewer declares that he has no relevant conflicts of interest.

[Comment 1_overall]:
The aim of this paper was to test the hypothesis that airborne pollen, and more especially tree pollen, may be a significant immune trigger for suicidal behavior. The study was about a population of 2.86 million (what account for more than half the national population of Denmark) from January, 1982 to December, 2006 (1304 consecutive weeks). A so extensive sample size has the advantage ensuring a strong statistical power in the analysis; however, it may find significant connections that are not dominantly present in the real world.

According to the authors, which are most likely right, it is the first attempt to test the association between atmospheric pollen and suicide risk using empirical and individually linked data. The statistical analysis used a generalized additive Poisson model and was rigorously conducted. The data were correctly adjusted for possible confounding effects of location, calendar time and weather conditions as temperature, cloud cover and humidity – factors that have been suggested to be possible bioclimatic mediators of seasonal peaks of suicide through changes in neurotransmitters and hormones. Lastly, the stratified analyses were wisely performed by sex, age group and history of mood disorder of the subjects. Results indicate a significant association between suicide risk and air pollen counts. In the main the observed association remained after controlling for effects of region, calendar time, and weather conditions. The influence of pollen counts on suicide differed significantly by sex. So, males appeared to be more sensitive to lower concentrations of pollen, while females tended to show a stronger effect but only when the pollen counts reach to a certain level. Interestingly, the authors did not note any additional effect of a very high level of air pollen (more than 100 grains per cubic meter of air and per day). The personal history of mood disorders only had slight influence on suicide rate in the population, regardless of sex and age. The discussion is rather relevant and well managed. The paper is supported by a substantial bibliography, which includes 44 references; 35 of them are after 2000, and 23 after 2005, the most recent ones being dated 2012.

I only have a couple of quick points.

[Response 1_overall]: We are pleased with the reviewer's positive comments on this study.

[Comment 1_1]: First of all, only data of grand pollen counts, i.e. all taxa confounded, were taken into account. The authors try to justify this practice on the basis of the limited availability of the data. Such an assertion is surprising: if the Danmarks Meteorologiske Institut has not correctly stored all the files, these could probably be obtained from the European Aeroallergen Network Pollen Database. Additional information on that point would be welcome. In particular, birch pollen really deserved to be considered separately.
[Response 1_1]: We obtained the data on a few taxa including birch, but such data were only for the years of later study period and had a lot of missing values. Birch pollen is a major type of allergic pollen in Denmark and normally peaks for about 2-3 weeks in April-May. The main reason of using the grand pollen counts for final analyses is that suicide incidents in Denmark exhibit two peaks, a major one in the Spring and a minor one in the early Fall, which coincident well with peaks of tree pollen (in Spring, major peak) and mugwort (in early Fall, minor). We feel it is not easy to incorporate separate examinations on the specific effect of birch pollen into the present study, but agree with the reviewer that it may be worthwhile to have a future study focusing on this taxa.

[Comment 1_2]: Second, it would have been helpful to indicate why the study closes in 2006, while the paper is only submitted in December, 2012.
[Response 1_2]: There was a heavy delay of data on cause-of-death for the years 2007-20010 to be included in the national Cause-of-Death Register in Denmark. We were only able to get the cause of death data up to the year 2006 at the time when enquiring the data for the present study. It would be of course ideal to have an up-to-date data on cause of death, but at this point we feel the effort of this enterprise would exceed benefits considering the laborious re-application for approvals required by Danish legislations.

[Comment 1_3]: Finally, among the possible mechanisms underlying the link between pollen and suicide, the authors emphasize, and rightly so, that aeroallergens may precipitate suicide via mediators of allergic inflammation. But it might not have been too much to devote a few lines to the side effects of some antiallergic drugs (pseudoephedrine, antihistamines and corticosteroids), which may worsen prosuicidal factors such as night-insomnia, day-somnolence, agitation, anxiety, impulsivity, and cognitive disturbance.

[Response 1_3]: Thanks for the elaboration. We have now added a few lines about this issue in the discussion (page 11)

[Comment 1_4]: In my view, with these minor additions, the paper could be published.
[Response 1_4]: Thank you.

Reviewer: Ho Kim, Ph.D.
PROFESSOR of BIOSTATISTICS & EPIDEMIOLOGY Associate Dean,
SCHOOL OF PUBLIC HEALTH, SEOUL NATIONAL UNIVERSITY
1 KWANAK-RO KWANAK-GU, SEOUL 151-742, KOREA

[Comment 2_overall]:
This manuscript aimed to report an association between suicide and pollen exposure with individual level data from Danish registers. I think, as they said, this is the first paper successfully showing the association. The paper is well written in general. My concern is about the statistical model. It is well kwon among air pollution epidemiologists that the estimated effects are very sensitive to the choices of variables and other options of the model. I don't think the authors provided enough information on their models. Detailed information of the statistical model should be included to show the validity of this study. I strongly recommend to include a professional statistician as a co-author of the manuscript. They used "none-parametric" which is an unusual terminology. "non-parametric" or "nonparametric" is a proper word. Please google "none parametric" and see what happens. Please refer to the comments below.
[Response 2_overall]: This study is carried at a research center with a number of highly qualified statisticians including one of the co-authors. While we are confident about the statistical model being properly applied in this study, we realize that our description may not be sufficient enough to allow people understand it well. We have now provided more description about it in the manuscript and also included an appendix (page 6 and page 26).

We regret misspelling “non-parametric” as “none parametric” and have made the correction throughout the manuscript.

[Comment 2_1]: 1. I'm not quite sure that the final model is valid. At Fig 1, CI's for large value of total pollen covers 0 to 1 which means 100% uncertainty. And I'm not sure about the meaning of "Pr>Chisq: 0.0126". Please specify the exact model at the bottom of Fig 1.

[Response 2_1]: Fig. 1 shows the untransformed non-linear non-parametric effect of pollen counts on the weekly number of suicide incidents. The confidence interval ranging from 0 to 1 DOES NOT mean 100% uncertainty, it means that for a large pollen count there is a wide confidence interval due to a small number of weeks with very high levels of pollen.

The Chi-square test for the hypothesis that the non-linear non-parametric effect of the pollen counts is constant at 0 results in a p-value of 0.0126, which, on a statistical significance level, rejects the hypothesis. In other words, the result indicates that the non-linear non-parametric effect of pollen on suicide risk is NOT constant, i.e., there is a significant effect of pollen counts on the number of suicides. We agree with the reviewer, that this has not been clearly explained, so further explanations has been added (page 7) “(p=0.0126), meaning that the effect of weekly average pollen counts on the weekly number of suicides is not constantly at 0.”

The specific model used for the analysis in fig. 1 is
$\log(\lambda)=\mu+\llbracket \beta x+s \rrbracket_1 (v_1,\llbracket DF \rrbracket_1)+s_2 (v_2,\llbracket DF \rrbracket_2)$
Where β is the untransformed parametric effect of x, that is the region (Copenhagen or Viborg), and $s_i (v_i,\llbracket DF \rrbracket_i)$ is the non-linear and non-parametric effect of v_i estimated using the local regression

method (LOESS) with 〖DF〗_i degrees of freedom. In this model v_1 is the weekly average pollen count and v_2 is the date of the Monday of the week. In order to make this more understandable, we have now provided more detailed description to the legend of figure 1 (page 21). An appendix has also been added to this manuscript to further explain the models (page 26).

[Comment 2_2]: 2. It is not quite sure how they control for other factors like temperature, cloud cover? I guess they were added linearly into the model. It is well kwon that temperature shows nonlinear association with mortality. Please be more specific on describing adjustment methods for other variables.
[Response 2_2]: The effects of temperature, cloud cover, humidity and calendar week were handled in the same non-linear non-parametric way using local regression in a model as described above. This is explained in the section about statistical methods (page 5-6) and also in the appendix (page 26).

[Comment 2_3]: 3. The authors should compare the effect of pollen to other variables like temperature, season, etc. This comparisons will be helpful to discuss the effect of pollen with known risk factors.
[Response 2_3]: We are not quite sure why we should compare the effect of pollen with that of other variables like temperature, season, etc., while the focus of this study is to assess the effect of pollen exposure? We don't think such comparison makes particular sense to us.

[Comment 2_4]: 4. How to control time is not that simple. We have to control long-term trend, year-to-year variation, seasonal and monthly trend. Please be more specific on time trend adjustment.

[Response 2_4]: For the analysis, we are only interested in removing the confounding effect "time" might have on the effect of pollen counts on suicide counts. With our analytic approach, it should be sufficient to have a finely categorized time variable included for adjustment in the model. We don't see the particular need to decompose the time effect into elements as long-term trend, year-to-year variation, seasonal and monthly trend and so on. In our analyses, the variable "time" is defined as date of the Monday of each given week, and the estimated effect of this variable leaves an effect of each week throughout the study period of 1304 consecutive weeks. So we used a more general model than the one leaving an additive effect of the multiple variables as year and month. By doing so, the estimate has the capacity to account for the general effects from e.g., year trend, season trend, etc., that may possibly exist along the actual time.

To further clarify the variable holding the week information, we have now rephrased the description (page 6) "calendar time (defined in week as a continuous variable)" to "calendar week (defined as the date of the Monday of the week)".

Reviewer: Fotios Papadopoulos, MD
Associate Professor
Dep of Neuroscience, Psychiatry
Uppsala University

[Comment 3_overall]:
Only completed suicides are included in this study. There is no control group. Results cannot be generalized for the whole population or even the population of patients with depression or allergy.

This is a well-written paper with an interesting research question, interesting statistical approach and engaging discussion of the results. The combination of individual-level data in a time-series analysis is undoubtedly advantageous and superior over simpler ecological approaches. Sub-analyses by

gender, age groups and history of mood disorders are also adding value to the paper.

[Response 3_overall]: We thank the reviewer for the positive comments. About the absence of control group in the study, we want to clarify that this study does not model the personal risk of suicide associated with pollen counts, but models the effect of pollen counts per week on the number of suicides per week in a population. The reference is the number of suicides of this population in the weeks when air pollen counts are 0. No control population is needed.

We understand the reviewer's concern about the generalizability of our findings, although the study uses data from two major regions that covers about 52% of the national population in Denmark. We have discussed this as limitations that "the results from this study may not apply to countries or regions where the climate and therefore plants producing allergenic pollen are very different from that in Denmark" (page 13).

Major concerns
[Comment 3_1]: 1. pollen counts as the exposure measure are obtained from two stations covering the two included regions, the Middle Jutland Region with an area of 13.000 km² and Copenhagen. No data is provided on the variation of pollen counts within regions. I am not so confident with Denmark's geography, but I believe that the including regions have a quite large area and include both rural and urban/semiurban areas, something that may also influence pollen counts. Thus these pollen counts should be seen as rough approximations of the actual exposure in individual level. This should be discussed in the paper.
[Response 3_1]: We agree with the reviewer, that these pollen counts are rough approximations of the actual exposure of individuals. Along with the suggestion from this reviewer, we have now extended our discussion to cover this limitation of our study (page 13).

[Comment 3_2]: 2. Please provide description of the raw data, preferably as a time-series graph of suicide and pollen counts, temperature etc.
[Response 3_2]: We have tried to use a time-series graph to illustrate the distribution of the raw data on study variables, but the graphs look too busy and dense to give a clear overview. This is because our data spans over 1304 weeks and the interval spaces are too small to allow proper visualization of the data differences over time. We have therefor chosen to not include these graphs in the manuscript, but would be happy to provide upon request from the Editor.

[Comment 3_3]: 3. Suicide is a very complex phenomenon. It would be very interesting to quantify the potential of pollen-associated pathophysiological mechanisms in suicide prevention. How much of the suicide variance can be explained by pollen counts versus by other climatic variables?
[Response 3_3]: This is an important suggestion, and we understand that many people are interested in knowing such estimates. Unfortunately, the present study was not designed for answering this question. To assess how much of suicide variance can be explained by pollen counts versus by other climatic factors would require another study in its own right.

[Comment 3_4]: 4. The authors control for climatic variables in their analyses. I wonder how cloud cover is defined. If it takes into account day length or not. If not then sunlight would be a better measure of sunlight exposure, which is associated with suicide risk in the literature.
[Response 3_4]: The cloud coverage is defined as the percentage of the sky covered by clouds (viewed from an observation spot) at a given time point of the day (page 4) — a standard measurement in Meteorology. Cloud coverage of 0 indicates a clear sky while coverage of 100 means a totally clouded sky. The measure does not take into account the length of the day nor the daily sunlight, but works as a crude proxy for the amount of sunlight that the cohort is exposed to. We agree that it might be better to consider actual sunlight time in the model; unfortunately, we did not investigate the availability and quality of data about actual sunlight and did not enquire this data when

planning the study. This limitation is now stated in the discussion (page 13)

[Comment 3_5]: 5. Please provide more information on missing values for pollen counts and how those were managed. It is stated eg that if there were data on pollen counts for 5/7 days, the average of these 5 days was used. What if there were pollen data only for 2 or 3 days during some weeks?
[Response 3_5]: The reviewer is correct. If we had data of pollen counts for only 3 days in a week, the average level of the week is set to the average of these 3 days. In winter time most weeks have missing values for all weekdays, and these are set to have a pollen count of 0. To make this clear, we have now added a sentence "If there were no information on pollen counts for all 7 weekdays of a week (mostly Winter time), then the pollen count of that week was set at 0" (page 6).

[Comment 3_6]: 6. Table 2: please show results (RR and 95% CI) for the adjusting variables
[Response 3_6]: We are not so sure if it is a good idea to include the results for the adjusting variables along with the table 2 (normally we don't include). However, in response to this suggestion, we are presenting the results for the adjusting variables as an appendix in this point-to-point response letter. We are open to the Editor to decide whether or not to include such information in the report.

Other comments
[Comment 3_7]: 1. Abstract, in outcome measures: "… adjusted the data for location, seasonality and meteorological conditions". How did you adjust for seasonality apart from including the meteorological variables?
[Response 3_7]: The local regression method was used to estimate the non-linear non-parametric effect of calendar weeks. We have described in details in our responses to comment 2_1 and comment 2-4, and provided additional descriptions on this issue in the manuscript (page 6) and the enclosed appendix (page 26).

[Comment 3_8]: 2. p.5, third paragraph: " Psychiatric hospital or ward with a diagnosis…". Please add that diagnoses from outpatient settings were also included.
[Response 3_8]: Thank you, we have now added this information at the place (page 5).

[Comment 3_9]: 3. p.8, first paragraph: "…an increase of pollen counts…" and "Pollen counts arising from…". Please rephrase so that a temporal change of pollen counts is not implied.
[Response 3_9]: We have rephrased our description using the phrase 'a change of pollen level' (pages 2 and 8).

[Comment 3_10]: 4. p.8, third paragraph: Please rewrite, so that the striking findings with the highest RR for suicide among males with mood disorders are presented.
[Response 3_10]: We have made some revision describing this finding, which, hopefully, makes e better sense (page 8-9).

[Comment 3_11]: 5. P.9, third paragraph: "Etiological study" refers probably to "ecological study"?
[Response 3_11]: Yes, this cited study is an ecological study. We have made the correction (page 10). Thanks for pointing out.

[Comment 3_12]: 6. P.11, The second paragraph is somehow speculative. The proposed explanations can merely explain the increase in suicide within one week. Decreased quality of life and social interaction can lead to depression and feelings of hopelessness but it is highly unlikely that these effects would take place within one week.
[Response 3_12]: We agree with the reviewer that this section is somehow speculative and have chosen to remove this piece of information.

[Comment 3_13]: 7. P. 12, second paragraph: "did not noted" à "did not note"

[Response 3_13]: We have now made the correction, thank you.

[Comment 3_14]: 8. P. 12, second paragraph: The lack of effect of a very high level of air pollen is indeed intriguing. A dose-dependent relationship would strengthen the argument of a causal association between pollen and suicide. These results are difficult to interpret and the explanation provided is not sufficient, as one would expect that high levels of pollen (more than 100) would have at least the same suicide risk increase with the “30-100” level of pollen. Unless, susceptible individuals had already committed suicide before the even higher increase of pollen, as in reality there is a temporal relationship and pollen levels succeed temporally each other.
[Response 3_14]: It is intriguing that our study did not show an additional effect of a very high level of air pollen, i.e., > 100 pollen grains per cubic meter air. There may be a selection process existing here that susceptible individuals had already committed suicide before the even higher increase of pollen. We have now extended our discussion to include this possibility amongst other explanations (page 12).

[Comment 3_15]: 9. Table 4. Females <35 years old, RR = 0.738 for pollen “<10”. It is not mentioned in the text. Possible explanation? Chance finding?
[Response 3_15]: We appreciate the reviewer’s exceptional focus and attention to details. We do not know how to explain this result. While age differences in sex-specific hormones, mood regulation, and allergic sensitivity, etc., may contribute to the observation, it may be due to chance, or to an unmeasured concurring environmental factor. We have now included this in the discussion (page 12).

[Comment 3_16]: 10. Table 5. The findings on males with mood disorders should be emphasized as the most striking increase in suicide risk.
[Response 3_16]: We agree that the finding on males with mood disorder is interesting in this study. We have emphasized this finding in the text, but we don’t think it is the most striking one among other results. In general, we have chosen to modestly present what we have found because our findings are still preliminary and repetition studies are needed in order to draw a firm conclusion.

In addition, we have made minor corrections on a number of places for clarification and sincerely hope you will find this revised version suitable for publication.

We look forward to hear from you soon.

Sincerely,

Ping Qin
Corresponding author of the study

Appendix (In response to the comment 3_6 from the third reviewer)
Effect of adjusting factors (calendar time, temperature, humidity and cloud cover) in the adjusted model presented in Table 2.

(Since formula and pictures can not be displayed in this text box, we therefore have chosen to submit our response letter as a pdf file.)

VERSION 2 – REVIEW

REVIEWER	Ho Kim, PROFESSOR of BIOSTATISTICS & EPIDEMIOLOGY Associate Dean, SCHOOL OF PUBLIC HEALTH, SEOUL NATIONAL UNIVERSITY 1 KWANAK-RO KWANAK-GU, SEOUL 151, South Korea I declare that I have no relevant conflicts of interest.
REVIEW RETURNED	04-Apr-2013

GENERAL COMMENTS	The manuscript has been improved significantly. I think it is ready to be published.

REVIEWER	Fotios C Papadopoulos, MD Associate Professor Department of Neuroscience, Psychiatry Uppsala University
REVIEW RETURNED	04-Apr-2013

THE STUDY	The title and the abstract strongly imply a longitudinal analysis with follow up of 2.86 million people, while this is not the case. See also the attached document.
GENERAL COMMENTS	The authors have revised their manuscript with emphasis on better explaining their statistical methods, something very positive. The extra appendix is very welcome and pedagogical. The other smaller revisions are also in the right direction and the article is improved. I would like to thank the reviewers for substantially addressing my comments. I have however a couple remaining major issues to discuss, before, in my opinion, the manuscript can be accepted for publication. These are: properly adjusting for seasonality, changing the language to become less causal and describe the missing values in more detail. 1) Adequately *adjusting for seasonality*, which is also mentioned in R2's question no4. The problem here is that both pollen and suicide are seasonal phenomena and co-vary. Both increase during spring. The finding of an association between higher pollen levels and suicide can simply mean that suicide increases when pollen increases, which is in spring. "Spring" is the big confounding factor here which if properly adjusted, we can better approach a possible real effect of pollen in suicide. Otherwise, we can also refer to the finding as an association finding of ecological level of evidence (that is low). The authors try to adjust for seasonality by controlling for

	"time", temperature, cloud cover and humidity. I do not really know the statistical consequences of inserting these three intercorrelated variables in the model simultaneously; however the GAM model should be superior when it comes to such problems. As it is clearly shown in the supplementary figures (thanks for those!) with the effects of the covariates, time has an impressing effect with lower suicide counts over time. It is my impression, that time the way it is inserted in the model, accounts mostly for the decreasing trend of suicide rates in Denmark over time. I doubt if this is an appropriate adjustment for season, but I am not a statistician. The next thing one thinks when it comes to controlling for season is of course meteorological parameters, and the authors have included temp, cloud cover and humidity. Regarding suicide and meteorological factors, research mostly supports an association between temperature and sunshine, even humidity. The role of cloud cover in suicide is much less studied, if studied at all. I suggested the use of sunshine data as a more accurate seasonality confounding variable and the authors replied that they did not investigate the availability and quality of data about actual sunlight and did not enquire this data when planning the study. DMI (www.dmi.dk) should be able to promptly reply about such availability and it is probable that the data on solar radiance are available. Of course this is up to the authors to decide. Another suggestion to approach this issue is to do sensitivity analyses, running the same models for each season, or to be accurate for each season there is enough variability in pollen counts. If the effect of pollen in suicide counts remains significant in spring or in summer then the findings and their impact get a lot stronger. Otherwise one can assume that the effect was simply due to unmeasured confounding. I would strongly recommend to do this. 2. In author's response 3overall, they state "…this study does not model the personal risk of suicide associated with pollen counts, …". I fully agree but such a purpose is implied in the manuscript. I would recommend adding your clarifying reply also in the manuscript. I would also suggest changing the language in order be less "causal". Some examples follow, but I would like to ask the authors to check for such inconsistencies throughout the manuscript: - In the title "A study based on Danish longitudinal data". The word longitudinal data suggest a study on personal risk of suicide. Actually, the analysis is on aggregated weekly data and not individual-based data. Individual-

	based information on mood disorder is used for stratifying analyses. - In the abstract, background: “with individual level data”, same problem as in previous point. - In the abstract, methods: “we followed a population of 2.86 million…”. You actually did not. You do not report data on controls, you analyze suicide counts. This formulation strongly implies a longitudinal design with individual based data. - Abstract, results: “a change of… RESULTED in a relative risk of…” Better to use less causal language as “was associated with”. - Abstract, results: “…suicide risk in males started to rise when…” . A temporal causal association is strongly implied here. Similarly when writing “elevate risk” - Introduction, 2nd paragraph: again the word “longitudinal” and “in a population of 2.86 million over…” can mislead the reader. Actually the size of population is not of great interest for the purpose of this study, makes more harm than good (by implying that those whose did not commit suicide were also studied by some way) 3. Missing values. Thank you for your reply, but I would like to ask some more questions on this important issue – the exposure variable of the study. Missing values mean that pollen counts were not registered for some days due to a variety of reasons (technical or other). They cannot be assumed to be 0 (=no pollen), except if you received such instructions from DMI. When you say “if <u>we had data on pollen</u> counts for only 3 day in a week, the average level of the week is set to the average of these 3 days” do you mean that the other 4 days had missing values or 0 values? Please explain more the definition of missing values and their data cleaning/management. I would also like to know how many missing values there were in %.

VERSION 2 – AUTHOR RESPONSE

Reviewer: Fotios C Papadopoulos, MD
Associate Professor
Department of Neuroscience, Psychiatry
Uppsala University

The authors have revised their manuscript with emphasis on better explaining their statistical methods, something very positive. The extra appendix is very welcome and pedagogical. The other smaller revisions are also in the right direction and the article is improved. I would like to thank the reviewers for substantially addressing my comments.

I have however a couple remaining major issues to discuss, before, in my opinion, the manuscript can be accepted for publication. These are:
properly adjusting for seasonality, changing the language to become less causal and describe the missing values in more detail.

[Comment] 1. Adequately adjusting for seasonality, which is also mentioned in R2's question no4. The problem here is that both pollen and suicide are seasonal phenomena and co-vary. Both increase during spring. The finding of an association between higher pollen levels and suicide can simply mean that suicide increases when pollen increases, which is in spring.
"Spring" is the big confounding factor here which if properly adjusted, we can better approach a possible real effect of pollen in suicide.
Otherwise, we can also refer to the finding as an association finding of ecological level of evidence (that is low).

The authors try to adjust for seasonality by controlling for "time", temperature, cloud cover and humidity. I do not really know the statistical consequences of inserting these three intercorrelated variables in the model simultaneously; however the GAM model should be superior when it comes to such problems.

As it is clearly shown in the supplementary figures (thanks for those!) with the effects of the covariates, time has an impressing effect with lower suicide counts over time. It is my impression, that time the way it is inserted in the model, accounts mostly for the decreasing trend of suicide rates in Denmark over time. I doubt if this is an appropriate adjustment for season, but I am not a statistician.

The next thing one thinks when it comes to controlling for season is of course meteorological parameters, and the authors have included temp, cloud cover and humidity. Regarding suicide and meteorological factors, research mostly supports an association between temperature and sunshine, even humidity. The role of cloud cover in suicide is much less studied, if studied at all. I suggested the use of sunshine data as a more accurate seasonality confounding variable and the authors replied that they did not investigate the availability and quality of data about actual sunlight and did not enquire this data when planning the study.
DMI (www.dmi.dk) should be able to promptly reply about such availability and it is probable that the data on solar radiance are available. Of course this is up to the authors to decide.

Another suggestion to approach this issue is to do sensitivity analyses, running the same models for each season, or to be accurate for each season there is enough variability in pollen counts. If the effect of pollen in suicide counts remains significant in spring or in summer then the findings and their impact get a lot stronger. Otherwise one can assume that the effect was simply due to unmeasured confounding. I would strongly recommend to do this.

[Authors' Response]:
Regarding how we adjusted the data for possible effect of seasonality, we expanded our description in the text and also added an appendix in the previous revision. Because the local regression (LOESS) method estimates, for each point, a regression depending on the points in a chosen neighborhood of the specific point, the estimated effect of 'calendar week', introduced in the model with this method, has the capacity to account for the general effects from e.g., year trend, season trend, etc., that may possibly exist along the actual time. Yet one can always argue whether such adjustment is sufficient, we believe it is sufficient enough for our interesting in removing the confounding effect 'time' might have on the association between pollen level and suicide risk.

Following the suggestion from this reviewer, we have performed a sensitive analysis looking the effect of pollen on the number of suicides that occurred in the months of March, April, May and June. We got similar results with the adjusted RR=1.08, 1.05, 1.12 and 1.07, respectively, associated with level of pollen counts at <10, 10-<30, 30-100 and >100, respectively. These estimates were at less statistical significance level because of the loss of power when discarding more than half of the suicide cases, but the effect size as well the general trend of the effect are pretty comparable to our

results presented in the paper. We have now mentioned in the result section that our sensitive analysis supports the observed association. See page 8.

[Comment] 2. In author's response 3overall, they state "…this study does not model the personal risk of suicide associated with pollen counts, …". I fully agree but such a purpose is implied in the manuscript. I would recommend adding your clarifying reply also in the manuscript. I would also suggest changing the language in order be less "causal". Some examples follow, but I would like to ask the authors to check for such inconsistencies throughout the manuscript:

- In the title "A study based on Danish longitudinal data". The word
longitudinal data suggest a study on personal risk of suicide. Actually, the analysis is on aggregated weekly data and not individual-based data.
Individual-based information on mood disorder is used for stratifying analyses.
- In the abstract, background: "with individual level data", same
problem as in previous point.
- In the abstract, methods: "we followed a population of 2.86 million…".
You actually did not. You do not report data on controls, you analyze suicide counts. This formulation strongly implies a longitudinal design with individual based data.
- Abstract, results: "a change of… RESULTED in a relative risk of…"
Better to use less causal language as "was associated with".
- Abstract, results: "…suicide risk in males started to rise when…" . A
temporal causal association is strongly implied here. Similarly when writing "elevate risk"
- Introduction, 2nd paragraph: again the word "longitudinal" and "in a
population of 2.86 million over…" can mislead the reader. Actually the size of population is not of great interest for the purpose of this study, makes more harm than good (by implying that those whose did not commit suicide were also studied by some way)

[Authors' Response]:
We think it is perfectly correct to call our data "longitudinal data" because they were collected longitudinally for the registers. This does not mean, however, that our study design is a longitudinal one. In order to avoid possible confusion to the readers, as pointed by the reviewer, we have tried to reduce the use of this term when describing our data. Please see the changes at multiple places in the manuscript. See pages 1, 2 and 14.

Taking into account the suggestions from this reviewer, we have now rephrased the language at multiple places, making our finding of the association reads less a causal relationship. See pages 2, 3, 8, 9 and 14.

[Comment] 3. Missing values. Thank you for your reply, but I would like to ask some more questions on this important issue – the exposure variable of the study. Missing values mean that pollen counts were not registered for some days due to a variety of reasons (technical or other). They cannot be assumed to be 0 (=no pollen), except if you received such instructions from DMI.
When you say "if we had data on pollen counts for only 3 day in a week, the average level of the week is set to the average of these 3 days" do you mean that the other 4 days had missing values or 0 values?
Please explain more the definition of missing values and their data cleaning/management. I would also like to know how many missing values there were in %.

[Authors' Response]:
We fully agree that missing values should not be set to 0 by default, which we have taken care of this

carefully and provided further explanations in the present revision.

We want to point out, however, that the reviewer incorrectly interpreted what we say “if we had data on pollen counts for only 3 day in a week, the average level of the week is set to the average of these 3 days”. In this case, we mean that the other 4 days had missing values and therefore we only used the average value of the 3 days for which the pollen data were recorded as the average pollen level for that specific week.

In our data, the missing values on pollen counts had a nice distribution. The weeks with missing values on all 7 days ONLY occurred in the off season of pollen monitoring, that is, the winter season. During the period monitoring air pollen counts (usually from the middle of February until the beginning of October, varying slightly from year to year), there were occasionally a missing value for 1 or 2 days of a week. The proportion of days with missing value on pollen counts by month is 4.8% in April, 3.4% in May, 2.7% in June, 5.9% in August, 83.7% in October, and almost 100% in November, December and January. We have now included more description on this issue in the method section of the manuscript. See pages 4 and 6.

VERSION 3 - REVIEW

REVIEWER	Fotios Papadopoulos, MD, PhD Associate Professor Dep of Neuroscience, Psychiatry Uppsala University
REVIEW RETURNED	19-Apr-2013

THE STUDY	The first paragraph in the discussion and a sentence in the Summary panel should be rephrased.
GENERAL COMMENTS	The paper is even more improved in this revision and I would like to thank the authors for their replies to my comments. I won't raise any new issues, but I would like to comment on the following authors' replies: 1. Sensitivity analysis. Thank you for following my suggestion. This is indeed interesting and although I share your view of decreased statistical power when looking only at four months, I would suggest to write in results that the findings followed the same trend, but did not reach statistical significance, probably due to lack of statistical power. 2. Rephrasing. Thanks for all the changes, I do believe that the current presentation will help the reader follow your methods and avoid confusion. The following points were missed and should also be rephrased: i) the first paragraph of the discussion "...the association between air pollen counts and residents' risk for suicide with empirical and individually linked data that covered a population of 2.86 million over 1304 consecutive weeks", in the presence form is misleading. I would suggest not to refer to "residents' risk", "individually linked data" as well as to the size of the population. You could instead refer to the 13700 suicides over 1304 weeks, which is more accurate description. ii) in the article summary, in the strengths and limitations, I would suggest to replace "...covers 2.86 million residents" with "...covers 13700 suicides;" iii) in the abstract's methods I would also suggest to replace "We

	studied a population of 2.86 million over 1304 consecutive weeks for which data on suicides and air pollen counts were recorded in Danish registries" with i.e. "We studied 13700 suicides over 1304 consecutive weeks for which data on air pollen..." I do believe that it more appropriate to refer to the total number of suicides that were actually analysed than the population base of that suicide number. A minor comment. The ref 48 is listed twice (also as no3)

VERSION 3 – AUTHOR RESPONSE

Reviewer: Fotios Papadopoulos, MD, PhD
Associate Professor
Dep of Neuroscience, Psychiatry
Uppsala University

The paper is even more improved in this revision and I would like to thank the authors for their replies to my comments. I won't raise any new issues, but I would like to comment on the following authors' replies:

[Comment 1]
1. Sensitivity analysis. Thank you for following my suggestion. This is indeed interesting and although I share your view of decreased statistical power when looking only at four months, I would suggest to write in results that the findings followed the same trend, but did not reach statistical significance, probably due to lack of statistical power.

[Authors' response] We have now added the findings of sensitive analysis in the result section (see page 8)

[Comment 2]
2. Rephrasing. Thanks for all the changes, I do believe that the current presentation will help the reader follow your methods and avoid confusion. The following points were missed and should also be rephrased:
i) the first paragraph of the discussion "...the association between air pollen counts and residents' risk for suicide with empirical and individually linked data that covered a population of 2.86 million over 1304 consecutive weeks", in the presence form is misleading. I would suggest not to refer to "residents' risk", "individually linked data" as well as to the size of the population. You could instead refer to the 13700 suicides over 1304 weeks, which is more accurate description.
ii) in the article summary, in the strengths and limitations, I would suggest to replace "...covers 2.86 million residents" with "...covers 13700 suicides;"
iii) in the abstract's methods I would also suggest to replace "We studied a population of 2.86 million over 1304 consecutive weeks for which data on suicides and air pollen counts were recorded in Danish registries" with i.e. "We studied 13700 suicides over 1304 consecutive weeks for which data on air pollen..."

I do believe that it more appropriate to refer to the total number of suicides that were actually analysed than the population base of that suicide number.

[Authors' response] We have now rephrased the description at these 3 places as suggested by the reviewer (see page 2, 9 and 14, respectively).

[Comment 3]
A minor comment. The ref 48 is listed twice (also as no3)

[Authors' response] We would like to thank the reviewer drawing our attention on the duplication of this reference in our RM database. We have now made the correction in the manuscript (see page 13 and page 19 of the reference list).