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)	The use of infectious disease surveillance reports to monitor the Zika virus epidemic in Latin America and the Caribbean from 2015-2017 : strengths and deficiencies
AUTHORS	Morris, Joan; Dolk, Helen; Durán, Pablo; Orioli, Ieda Maria

VERSION 1 – REVIEW

REVIEWER	Renata Artimos de Oliveira Vianna Universidade Federal Fluminense, Brazil
REVIEW RETURNED	06-Aug-2020

GENERAL COMMENTS	My overall impression is that the paper showed important facts for the interpretation of Congenital Zika Syndrome surveillance in Latin America and the Caribbean, inferring different limitations and complexities of each country. The authors clarified very well all these particularities during the results and discussion. One small suggestion is to abbreviate "Zika virus", as ZIKV. If the authors prefer to not abbreviate, please write Zika with a capital "Z", as it is the name of a forest in Uganda.

REVIEWER	HELOISA VISCAINO F.S. PEREIRA RIO DE JANEIRO STATE UNIVERSITY -UERJ
REVIEW RETURNED	17-Aug-2020

GENERAL COMMENTS	The study is of great relevance since it addresses a method for better comparing population data free of some of the bias frequently present in countries with poor resources. In order to have the vaidation of the mathematical approach of these results and to have them proposed as an index for Zika and other disorders I would propose the review of the statistical method, which seems appropriate to me but in which I am not a specialist.

REVIEWER	Jimena Barbeito-Andrés CONICET, Argentina
REVIEW RETURNED	25-Aug-2020

GENERAL COMMENTS	The study by Morris and cols is aimed to provide new insights on Congenital Zika Syndrome (CZS) prevalence across Latin America and the Caribbean based on available epidemiological data. Zika virus appeared in the Americas recently and it disseminated in many regions causing birth defects that were unknown. The determination of the scientific community to understand Zika virus dynamics at several levels has been noticeable and, in this line, retrospective analyses that seek to re-interpret available epidemiological data are crucial. However, I have some concerns to be addressed that could

potentially improve the current manuscript.

- My main concern is related to how the objectives of this study are presented. I think the authors should revise the manuscript to highlight the specific contributions of their analyses. The authors presented the number of CZS in relation to number of births (Variable 1) and to the number of (estimated) infected women during pregnancy (Variable 2). As far as I know, Variable 1 has been reported in previous studies for specific countries (for instance, de Oliveira et al 2017, doi: 10.1016/S0140-6736(17)31368-5) Is it the first time that it is compared at the continental scale? Or is there other original contribution here in relation to this variable?
Variable 2 is very interesting to discuss the potential incidence of Zika virus infection in pregnant women that were undiagnosed and that had children with birth defects. In fact, the authors speculate that this "might give an indication of the quality of the publicly available data." Is there any variable related to health system investment, number of performed PCR/serological test or any other parameter that could be studied to contrast this hypothesis? Since almost all the discussion as well as the comparison with the prediction by Zhang et al 2017 are focused on the idea of under-reporting ZIKV infections and CZS cases in some countries, I think an effort to explicitly test this hypothesis should be made.
As well, I understand that the authors are considering a similar probability to have a CZS case when infected with Zika virus during pregnancy along the territories. However, several studies suggest that there are factors that could predispose or prevent the development of CZS when infected during pregnancy and that they could explain the asymmetrical CZS geographical distribution (a short and incomplete list includes: doi: 10.1038/s41593-020-0664-0, doi: 10.1038/s41467-018-03497-1, doi: 10.1101/2020.06.03.132878, doi: 10.1126/sciadv.aaw6284, doi: 10.1371/journal.pntd.0008060). How does this evidence fit with your interpretations?
In sum, I believe that there are potentially interesting results that need to be clearly put in the context of what is known and to be explored in depth to dissect confounding or contributing factors.

- In the title, the period 2015-2018 is presented as the focus of the article. But, as far as I noticed, only 2015-2017 information was analyzed. Please, clarify this point in order to preserve the coherence along the article and reflect the actual aim of the study in the title.

- The authors suggest that mainland and island territories have different patterns and some results are presented in relation to this point. I think it is an interesting aspect of the problem but that it is approached without a systematic perspective. It would be important to state some hypotheses about it, or at least expectations, based on previous background. As well, these expectations and your results should be comprehensively discussed. By the way, in the Material and Methods section the authors stated that "A random effects model was fitted to determine if this measure is related to whether the country is an island or part of the mainland". It is not clear to me where the results for this analysis are presented.

- In the Introduction the authors stated: "A study of women in America who were pregnant and had evidence of Zika virus infection during pregnancy (both symptomatic and asymptomatic) found that 60 per 1,000 pregnancies were diagnosed with CZS." A reference should follow this sentence.

2

	- Results presented in Table 2 are not described in the Result section but only introduced in the Discussion. I believe it is essential to introduce these results properly.
	- In the Result section (page 6) the authors claimed that there is a significant linear correlation between the number of CZS cases per 1,000 births and the incidence rate of Zika virus infections. After that, they continued discussing on "weak" and "significant" correlations. However, no indication on the actual correlation coefficients as well as the kind of coefficient used (Pearson, Spearman, etc) and other important values such as R2 are provided. This information is essential.
	- In the section for "Limitations and Strengths" I could not find an explicit statement about the limitations of the study. I think that this part of the manuscript should be reformulated.
	- Minor: Some typos and writing inaccuracies were detected. I think a process of language revision could largely improve this manuscript. Some examples: "Columbia" for "Colombia", "date" for "data", zika/Zika virus, etc

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1	
My overall impression is that the paper showed important facts for the interpretation of Congenital Zika Syndrome surveillance in Latin America and the Caribbean, inferring different limitations and complexities of each country. The authors clarified very well all these particularities during the results and discussion. One small suggestion is to abbreviate "Zika virus", as ZIKV. If the authors prefer to not abbreviate, please write Zika with a capital "Z", as it is the name of a forest in Uganda.	Done - We have capitalised all mentions of Zika
Reviewer: 2	
The study is of great relevance since it addresses a method for better comparing population data free of some of the bias frequently present in countries with poor resources. In order to have the vaidation of the mathematical approach of these results and to have them proposed as an index for Zika and other disorders I would propose the review of the statistical method, which seems appropriate to me but in which I am not a specialist.	One of the co-authors is a statistician and we are confident the method is correct.

Reviewer: 3	
The study by Morris and cols is aimed to provide new insights on Congenital Zika Syndrome (CZS) prevalence across Latin America and the Caribbean based on available epidemiological data. Zika virus appeared in the Americas recently and it disseminated in many regions causing birth defects that were unknown. The determination of the scientific community to understand Zika virus dynamics at several levels has been noticeable and, in this line, retrospective analyses that seek to re-interpret available epidemiological data are crucial. However, I have some concerns to be addressed that could potentially improve the current manuscript.	
- My main concern is related to how the objectives of this study are presented. I think the authors should revise the manuscript to highlight the specific contributions of their analyses. The authors presented the number of CZS in relation to number of births (Variable 1) and to the number of (estimated) infected women during pregnancy (Variable 2). As far as I know, Variable 1 has been reported in previous studies for specific countries (for instance, de Oliveira et al 2017, doi: 10.1016/S0140-6736(17)31368-5) Is it the first time that it is compared at the continental scale? Or is there other original contribution here in relation to this variable?	The reviewer is correct – variable 1 has been used by other authors. However, to our knowledge, we are the first to compare across countries. We have added this to the manuscript in the following places : Objective in abstract 2nd paragraph in Introduction
Variable 2 is very interesting to discuss the potential incidence of Zika virus infection in pregnant women that were undiagnosed and that had children with birth defects. In fact, the authors speculate that this "might give an indication of the quality of the publicly available data." Is there any variable related to health system investment, number of performed PCR/serological test or any other parameter that could be studied to contrast this hypothesis? Since almost all the discussion as well as the comparison with the prediction by Zhang et al 2017 are focused on the idea of under-reporting ZIKV infections and CZS cases in some countries, I think an effort to explicitly test this hypothesis should be made.	We agree with the reviewer that it would be interesting to identify women who were infected during pregnancy and whose children were undiagnosed. However, as we have already stated in our paper, there is no information on the numbers of people tested in each country which makes interpreting incidence of Zika virus infections very difficult. There are individual country reports from 25 September 2017 which provide additional information on the number of pregnant women who were reported as being infected with Zika virus. We have included this data in a new analysis and additional text that identifies the issues about under-reporting more clearly, there is an additional table. We hope that this satisfies the reviewers questions.
As well, I understand that the authors are considering a similar probability to have a CZS case when infected with Zika virus during pregnancy along the territories.	We thank the reviewer for raising the issue of predisposing factors for developing CZS when infected during pregnancy and have added text and references to the discussion

However, several studies suggest that there are factors that could predispose or prevent the development of CZS when infected during pregnancy and that they could explain the asymmetrical CZS geographical distribution (a short and incomplete list includes: doi: 10.1038/s41593-020-0664-0, doi: 10.1038/s41467-018-03497-1, doi: 10.1101/2020.06.03.132878, doi: 10.1126/sciadv.aaw6284, doi: 10.1371/journal.pntd.0008060). How does this evidence fit with your interpretations? In sum, I believe that there are potentially interesting results that need to be clearly put in the context of what is known and to be explored in depth to dissect confounding or contributing factors.	about this point. However, in this study, the order of magnitude of the effect of under-reporting of Zika virus infections is much greater than the effect of an imbalance of risk factors and therefore we wish to leave the emphasis in the study to be on under-reporting as we believe that under-reporting is the most likely explanation for most of this variation
- In the title, the period 2015-2018 is presented as the focus of the article. But, as far as I noticed, only 2015-2017 information was analyzed. Please, clarify this point in order to preserve the coherence along the article and reflect the actual aim of the study in the title.	Corrected
- The authors suggest that mainland and island territories have different patterns and some results are presented in relation to this point. I think it is an interesting aspect of the problem but that it is approached without a systematic perspective. It would be important to state some hypotheses about it, or at least expectations, based on previous background. As well, these expectations and your results should be comprehensively discussed. By the way, in the Material and Methods section the authors stated that "A random effects model was fitted to determine if this measure is related to whether the country is an island or part of the mainland". It is not clear to me where the results for this analysis are presented.	We have added text to the statistical analysis describing the hypothesis and have also expanded the results of the model fitting (2nd para of results) and the discussion of these results (1st para discussion)
- In the Introduction the authors stated: "A study of women in America who were pregnant and had evidence of Zika virus infection during pregnancy (both symptomatic and asymptomatic) found that 60 per 1,000 pregnancies were diagnosed with CZS." A reference should follow this sentence.	Done
- Results presented in Table 2 are not described in the Result section but only introduced in the Discussion. I believe it is essential to introduce these results properly.	Done -We have moved the text comparing our data to that predicted by reference 20 (old table 2 – new table 3) into the results section.

- In the Result section (page 6) the authors claimed that there is a significant linear correlation between the number of CZS cases per 1,000 births and the incidence rate of Zika virus infections. After that, they continued discussing on "weak" and "significant" correlations. However, no indication on the actual correlation coefficients as well as the kind of coefficient used (Pearson, Spearman, etc) and other important values such as R2 are provided. This information is essential.	Spearman's rho correlation coefficient is provided in the results and the details given in the methods section.
- In the section for "Limitations and Strengths" I could not find an explicit statement about the limitations of the study. I think that this part of the manuscript should be reformulated.	This has been rewritten.
- Minor: Some typos and writing inaccuracies were detected. I think a process of language revision could largely improve this manuscript. Some examples: "Columbia" for "Colombia", "date" for "data", zika/Zika virus, etc	Apologies – these have been corrected.

VERSION 2 – REVIEW

REVIEWER	Jimena Barbeito CONICET
REVIEW RETURNED	30-Oct-2020

| GENERAL COMMENTS | The revised version of the manuscript has addressed large part of the issues raised in the previous revision. In my opinion, the topic of the article is relevant and the author's contribution is very valuable. However, I still have some comments that might help to improve the manuscript.
- The objective, as stated in the Abstract and Introduction, is still not very clear. Why do the authors refer to "two outcomes" instead of specifically introducing the aspects that were analyzed?

- In table 2, the last column for "Number of pregnant women with Zika virus suspected or confirmed per 1,000 births" is obtained by C/E*1000? Or should be G/E *1000?

- CZS occurs only in part of the cases of ZIKV infected mothers, as the authors stated. As the authors also indicated there are variable estimations for this incidence but range from 4 to 60 cases for 1,000 infected pregnancies. Then, I think that for those countries with smaller populations or with less ZIKV infections, it is difficult to derive problems in the reports from the presented data. To illustrate the point, in countries with small ZIKV outbreaks, for instance less than 500 ZIKV infections (suspected + confirmed), the chance to have no CZS is really high, as well as the chance to have a relatively large number of CZS. How do you think this can impact in your |

interpretations?

- I understand that since this study has a continental scale for comparisons, it is expectable to find some generalizations. However, in my opinion to assume similar susceptibility to ZIKV infection and CZS in all countries is not sustained by current data. It is not clear to me why the authors conclude that the effect of under-reporting is in a different order of magnitude that other factors that have been linked to differential distribution of ZIKV infections and CZS.

Minor:
- In the Abstract it is confusing to me the use of the figures "25/48" or "27/48" without an explanation on what "48" means.

- There are some typos (for instance, flavirus) and specific terms such as Zika virus should be used with the same style throughout the text. I am not a native English speaker but I am sure that a revision on the written English will largely improve the manuscript.

VERSION 2 – AUTHOR RESPONSE

Response to Reviewers' Comments

Reviewers' Comments	Our responses
- The objective, as stated in the Abstract and Introduction, is still not very clear. Why do the authors refer to "two outcomes" instead of specifically introducing the aspects that were analyzed?	We have replaced the reference to "two outcomes" with the actual outcomes measured as suggested
- In table 2, the last column for "Number of pregnant women with Zika virus suspected or confirmed per 1,000 births" is obtained by C/E*1000? Or should be G/E *1000?	This was a typing error and has been corrected
- CZS occurs only in part of the cases of ZIKV infected mothers, as the authors stated. As the authors also indicated there are variable estimations for this incidence but range from 4 to 60 cases for 1,000 infected pregnancies. Then, I think that for those countries with smaller populations or with less ZIKV infections, it is difficult to derive problems in the reports from the presented data. To illustrate the point, in countries with small ZIKV outbreaks, for instance less than 500 ZIKV infections (suspected + confirmed), the chance to have no CZS is really high, as well as the chance to have a relatively large number of CZS. How do you think this can impact in your interpretations?	We agree that for countries with small ZIKV outbreaks the incidence estimates will be highly influenced by random error. We have added the four countries (reporting more than 200 infected pregnancies but no CZS cases) to figure 2 where we compared the incidence in each country with the range 4 to 60 per 1,000. We display the 95% confidencecentervals for each country to enable the reader to see how the smaller countries have much wider confidence intervals. This resulted in us realising that no CZS cases for both Peru and Jamaica was not an unexpected finding and therefore did not indicate under-reporting of CZS . So we have removed this statement from the abstract and revised the text in the results (page 8)
- I understand that since this study has a continental scale for comparisons, it is expectable to find some generalizations. However, in my opinion to assume similar susceptibility to ZIKV infection and CZS in all countries is not sustained by current data. It is not clear to me why the authors conclude that	We agree that susceptibility to Zika infections varies hugely according to climatic, environmental and social factors. In this study, we are only commenting on the chances that an infected pregnancy results in a CZS outcome.Estimates from studies of the birth prevalence of CZS varied from 4 to 60

7

the effect of under-reporting is in a different order of magnitude that other factors that have been linked to differential distribution of ZIKV infections and CZS.	cases for 1,000 infected pregnancies. Yet the estimates from the reported data varied from over 100 to below 0.005. We have revised the text (2nd para discussion to make this clearer)
Minor: - In the Abstract it is confusing to me the use of the figures "25/48" or "27/48" without an explanation on what "48" means.	We have redrafted this to make it clearer
- There are some typos (for instance, flavirus) and specific terms such as Zika virus should be used with the same style throughout the text. I am not a native English speaker but I am sure that a revision on the written English will largely improve the manuscript.	We have checked the manuscript for typos