

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)	Refugee maternal and perinatal health in Ontario, Canada: a retrospective population-based study
AUTHORS	Wanigaratne, Susitha; Shakya, Yogendra; Gagnon, Anita; Cole, Donald; Rashid, Meb; Blake, Jennifer; Dastoori, Parisa; Moineddin, Rahim; Ray, Joel; Urquia, Marcelo

VERSION 1 – REVIEW

REVIEWER	Hannah Dahlen Hannah Dahlen Professor of Midwifery Higher Degree Research Director School of Nursing and Midwifery Building EB/LG Room 34, Parramatta South Campus P: 9685 9118 F: 9685 99599 E: h.dahlen@westernsydney.edu.au Western Sydney University Locked Bag 1797 Penrith NSW 2751 Australia http://www.uws.edu.au/staff_profiles/uws_profiles/professor_hannah_dahlen Twitter @hannahdahlen
REVIEW RETURNED	25-Aug-2017

GENERAL COMMENTS	<p>Thank you for the opportunity to review this paper. We are green with envy over the quality of routinely collected data the Canadians have available to them.</p> <p>My queries are as follows:</p> <p>I would be more cautious about saying this is the largest study of refugee maternal health in the literature and say it is either one of the largest or the largest for Canada (Page 2 line 49)</p> <p>The last paragraph of the background is a bit messy. I would put all the literature about Canada and what other studies have attempted to do first and then end with the research question</p> <p>Have any validation studies been done on the accuracy of the IRCC-PRD?</p> <p>Page 4 line 13 it does not make sense to say analysis of preterm birth is restricted to 22-41 weeks as this is not the definition of preterm</p> <p>You do not mention BMI in your adjustments. Is this data available? If not it is a limitation</p> <p>Write logistic in full (page 5 line 36)</p>
-------------------------	--

	<p>You report the Figures in a forest plot style. While this is not a common approach it appears to work.</p> <p>There are a lot of supplementary tables and figures which the Editor can decide on.</p>
--	--

REVIEWER	Jane Yelland Murdoch Children's Research Institute Melbourne, Australia
REVIEW RETURNED	12-Sep-2017

GENERAL COMMENTS	<p>This nicely written paper reports a retrospective study of adverse maternal and perinatal outcomes for refugee women, other immigrant women and Canadian-born women in Ontario over a 13 year period. Several studies conducted in other high income countries have examined maternal and perinatal outcomes using maternal country of birth to determine likely refugee background and grouping of countries of birth into world regions. The authors have had the advantage of being able to link official immigration data to hospital and physician billing data. This has provided the first analysis of refugee 'status' and perinatal outcomes compared to other immigrants matched on country of birth and Canadian born women. This is a unique study. Other high income countries are unlikely to be in a position to undertake a similar analysis given differing definitions of refugee and refugee status, and diversity of immigration data and potential for linkage with health data systems. The statistical analysis has been carefully conducted and reported.</p> <p>I found it a little difficult to fully appreciate the migration system in Canada including at what point people attain permanent residency and what this may mean in interpreting the study data. The authors may wish to consider the following comments and suggestions in reviewing their paper.</p> <ol style="list-style-type: none"> 1. A limitation of the study is the definition of refugee status - a limitation articulated by the authors. It would be helpful to have the administrative data more fully outlined in the methods. The refugee cohort is taken from the permanent resident group of refugees. Is this a particularly privileged group? Of all people seeking asylum or coming to Canada as refugees, what proportion are granted permanent residency? Is it likely that newly arrived or asylum seeking women giving birth are not in the data at all? Further clarification would be helpful. Should reference to refugee immigrant women be changed to 'refugees of permanent residence' or the like? 2. Significant proportion of immigrant women giving birth had a duration of residency (since granted permanent resident status) in Canada of 10 years or more - 42% of refugee immigrant women and 28% of non-refugee immigrant women. Is it possible that some of these women came to Canada as young children? If immigrant women arrived as young children what was their status? Were data collected re language and education at immigration for children? How have the authors considered this in statistical analyses? 3. The comparison of adverse maternal and perinatal outcomes by the 5 top refugee source countries is of interest. In some other high income countries (e.g. Australia) over a similar time period to the study, the vast majority of immigrants from Afghanistan
-------------------------	--

	<p>are humanitarian entrants. Is it possible that people with a refugee experience such as the majority of Afghan immigrants (but for whatever reason not identified as having refugee status in the Canadian data) are in the immigrant non-refugee group? Additional contextual information may help the reader understand some of this complexity.</p> <p>4. Were Indigenous women included in the Canadian-born cohort?</p> <p>5. The sections re implications and future research (page 8) identify a number of issues. It would be helpful to have some contextual information around these issues (Canada's migration and refugee program; national efforts to tackle health and health care inequalities) in the manuscript.</p> <p>6. The final statement "Overall, based on an administrative definition of refugee, we do not find a strong need to enhance the health care for refugee mothers and their infants in Canada" even with the caveats, requires a re-think. Other aspects of health care can be compromised for refugee women - experience of trauma & torture and impacts on mental health; poor health literacy limiting access and engagement with services. Measuring these outcomes was obviously not within the scope of the study. The authors may wish to consider a possible re-phrasing as it seems difficult to conclude from the research what the health care outcomes are for refugees and immigrants with permanent resident status in Canada?</p> <p>7. This reviewer's preference is for the language of 'births' to be used as an alternative to 'deliveries'. This is also accepted language of several international perinatal journals with multidisciplinary readership.</p>
--	--

REVIEWER	John S. Preisser University of North Carolina, U.S.A
REVIEW RETURNED	12-Oct-2017

GENERAL COMMENTS	<p style="text-align: center;">Statistical Review</p> <p style="text-align: center;">"Refugee maternal and perinatal health in Ontario Canada: a retrospective population-based study." bmjopen-2017-018979</p> <p style="text-align: center;">By John S. Preisser</p> <p>This statistical review considers three themes: record linkage, matching, and statistical models.</p> <p>Record linkage: In this epidemiological study on refugee and non-refugee immigrant health, the authors have linked five datasets pertaining to immigration, health care, childbirth, death, and HIV, respectively, to study maternal and perinatal health outcomes in Ontario Canada. This is an impressive feat and one that the authors document with several citations that assessed the validity of various aspects of the linkage. This is undoubtedly a complex exercise involving a litany of rules; understandably not all the details are</p>
-------------------------	--

provided. Yet, few details of the linking process are stated, only that “unique encoded identifiers” are used. While the full details are well beyond the scope of this article, the reader is left with some basic questions: (1) what are the identifiers used for linkage and 2) what is the general approach to linking datasets, e.g., exact matching on named identifiers or probabilistic matching.

Matching: The comparison of refugee to non-refugee immigrant health is based on 1:1 matching of first deliveries in Canada based on country of birth, thus producing paired data. Unfortunately, the authors do not explain how the pairs are formed considering that there could be hundreds or even thousands of mothers from a given country and perhaps millions of possible different pairings of mothers. Also, why not also match on other variables besides country of birth such as time (in years) elapsed between the date of becoming a permanent resident and the date of delivery? While matching exactly on variables is one option, it may not be feasible when matching on multiple variables; matching based on propensity score algorithms would identify “best matches” based on practically any desired subset of the study variables. The propensity scores are commonly the predicted probability from a logistic regression for the binary outcome of refugee status as a function of the covariates used for matching. Pairs are formed based on minimizing within-paired differences in predicted probabilities.

Statistical Models: Multivariable log-binomial regression estimated with Generalized Estimating Equations with a working exchangeable correlation for matched pairs is applied to the paired data. Even if the pairs are well formed and the criteria for 1:1 matching is described, the GEE analysis does not provide a matched analysis, but merely accounts for within-pair correlation in the outcome. Matched-pair statistical analyses for the several binary outcomes can be conducted with generalized linear mixed models (with matched pair as the random intercept) or using conditional logistic regression.

A minor comment is that I would recommend splitting the final paragraph of the introduction section into two paragraphs, the first containing the rest of the literature review, the second stating what will be done in the manuscript.

In summary, I found the paper quite interesting and I was impressed by the scope of work undertaken, as it relates to the unique compiled dataset and the totality of statistical analyses. Nonetheless, I would like to know a few more details about the linkage and the matching. Finally, I am curious as to the results that a true matched pair analysis would give. Otherwise, the matching and the analysis seem only to accomplish matching on a group (country of birth) level in terms of matching the sample size of the two groups of women within each country, which is perhaps all the authors intended.

REVIEWER	Mari Palta University of Wisconsin- Madison Madison, WI USA
REVIEW RETURNED	22-Oct-2017

GENERAL COMMENTS	<p>Review: Refugee maternal and perinatal health in Ontario Canada: a retrospective population-based study.</p> <p>This is an interesting and mostly well done study. The manuscript and analyses have some room for improvement in language and clarity. Some specific comments are below:</p> <ol style="list-style-type: none"> 1. First of all, this is not really a “matched” study, in the usual sense of matching individuals. Unless I missed something, women were matched only on country of birth, which is shared by groups of women. Essentially the investigators eliminated women from the analysis to make the two groups balanced on country of birth. The matching is handled by GEE, which would not have worked on unbalanced data- hence the elimination of subjects I assume. However, simply entering indicator variables for countries of birth and keeping the whole sample would have been preferable and a more explicit approach to matching (or rather "stratification"). This could also have been done via a conditional logistic analysis. These so called “fixed effect” approaches would also have allowed testing for the hypothesis stated by the authors, (see 3. Below) but not fully addressed. 2. The paper referenced regarding the matched analysis does not seem publicly available and the reference does not include publication venue. The title seems to imply problems analyzing matched data by SAS. It is unclear why this would be without access to the paper. Besides, as noted above the sample here was not really in need of matched analysis. 3. The authors state as one of their objectives to investigate whether effect is independent of country. This seems addressed by sub-study of specific countries of birth only. It would have been interesting to see a more complete analysis that testes interaction effects, as the objective seems to anticipate. 4. The abstract needs to be more explicit about the findings. I was confused by the statements that refugees and no-refugees differed only on HIV, versus differed on most complications. Before reading the paper, I also did not understand what “in the same direction” refers to. 5. The writing needs some general editing for language and clarity. In particular, there are too many acronyms. I do not see a reason for why most could not be spelled out. It is especially difficult to read the conclusions with all these acronyms, as many readers may wish to do.
-------------------------	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Hannah Dahlen

Institution and Country: Professor of Midwifery, Higher Degree Research Director School of Nursing and Midwifery, Western Sydney University, Australia Competing Interests: Nil

Thank you for the opportunity to review this paper. We are green with envy over the quality of routinely collected data the Canadians have available to them.

My queries are as follows:

I would be more cautious about saying this is the largest study of refugee maternal health in the literature and say it is either one of the largest or the largest for Canada (Page 2 line 49)

This sentence has been edited to "This is one of the largest studies of refugee maternal and perinatal health in the literature."

The last paragraph of the background is a bit messy. I would put all the literature about Canada and what other studies have attempted to do first and then end with the research question.

This paragraph has been reorganized into two paragraphs. The first paragraph summarizes the background literature and the second paragraph outlines the research questions.

Have any validation studies been done on the accuracy of the IRCC-PRD?

The IRCC-PRD is an official database of the Canadian government and is based on legal documentation potential permanent residents submit to the government at the time of their application for permanent residency. It is considered a gold standard for many of the immigration (e.g., date of landing, country of birth etc) and socio-demographic (e.g., official language ability, education at arrival) variables that were used in this study since they were used to assess an applicant's legal eligibility for immigration to Canada. As such, validation studies have not been undertaken. We have clarified in the methods section that this database is used for legal purposes.

Page 4 line 13 it does not make sense to say analysis of preterm birth is restricted to 22-41 weeks as this is not the definition of preterm

Sorry for the confusion. I was attempting to describe the populations that are included and excluded when determining the risk of the outcome, which changes with the perinatal health outcome examined and not the definitions of the outcomes themselves. I have revised this section to read "Since many of the outcomes in this study are commonly used in epidemiologic surveillance, specifications based on gestational age and/or birthweight used by the Canadian Perinatal Surveillance System [20] were implemented where possible. These specifications relate to including births that are reasonably expected to be at risk for the outcome; e.g., births < 500 grams and/or <20 weeks gestation are less likely to be viable."

You do not mention BMI in your adjustments. Is this data available? If not it is a limitation

Unfortunately, BMI was not collected so we could not adjust for it. This has been added to the limitations section in the discussion as a fourth limitation.

Write logistic in full (page 5 line 36)

The type of regression we conducted is commonly referred to as "log-binomial" regression. We did not use logistic regression, as the reviewer may be implying, so we have left this terminology as is in the manuscript.

You report the Figures in a forest plot style. While this is not a common approach it appears to work.

There are a lot of supplementary tables and figures which the Editor can decide on.

Reviewer: 2

Reviewer Name: Jane Yelland

Institution and Country: Murdoch Children's Research Institute, Melbourne, Australia

Competing Interests: None declared

This nicely written paper reports a retrospective study of adverse maternal and perinatal outcomes for refugee women, other immigrant women and Canadian-born women in Ontario over a 13 year period. Several studies conducted in other high income countries have examined maternal and perinatal outcomes using maternal country of birth to determine likely refugee background and grouping of countries of birth into world regions. The authors have had the advantage of being able to link official immigration data to hospital and physician billing data. This has provided the first analysis of refugee 'status' and perinatal outcomes compared to other immigrants matched on country of birth and Canadian born women. This is a unique study. Other high income countries are unlikely to be in a position to undertake a similar analysis given differing definitions of refugee and refugee status, and diversity of immigration data and potential for linkage with health data systems.

The statistical analysis has been carefully conducted and reported.

I found it a little difficult to fully appreciate the migration system in Canada including at what point people attain permanent residency and what this may mean in interpreting the study data. The authors may wish to consider the following comments and suggestions in reviewing their paper.

1. A limitation of the study is the definition of refugee status - a limitation articulated by the authors. It would be helpful to have the administrative data more fully outlined in the methods. The refugee cohort is taken from the permanent resident group of refugees. Is this a particularly privileged group? Of all people seeking asylum or coming to Canada as refugees, what proportion are granted permanent residency? Is it likely that newly arrived or asylum seeking women giving birth are not in the data at all? Further clarification would be helpful. Should reference to refugee immigrant women be changed to 'refugees of permanent residence' or the like?

Thank you for this suggestion to add additional details regarding the refugees included in the IRCC-PRD. Additional details have been added to the Methods section, under "Variables" – please see pg 5. The main purpose of this added section is to describe categories of refugees in Canada and what services and supports are available to them during the immigration process. Within the categories of refugees who are permanent residents, benefits and services vary within the first year of arrival. After the first year, if financial and social supports are still required, they are eligible for the same supports that are available to other (low-income) Canadians (i.e., income support, additional health care benefits such as dental). Those who do not become permanent residents (i.e., failed refugee claimants) are at a disadvantage compared to those who do become permanent residents. However, those without permanent residency status are not included in the IRCC-PRD and unfortunately we cannot draw any conclusions about them from these data. In Canada, understandably minimal research is available looking at maternal and infant health for failed refugee claimants – however if the reviewer is interested I'd recommend: Wilson-Mitchell K. Increasing Access to Prenatal Care: Disease Prevention and Sound Business Practice. *Health Care Women Int.* 2014 Feb;35(2):120–6.

The new section on page 5 in the methods section reads as follows:

"Refugee status was defined using the IRCC-PRD. There are four categories of refugees in the database – i) government sponsored refugees, who are provided with financial and settlement

assistance during their 1st year in Canada by the federal government; ii) privately sponsored refugees, who are provided with financial and settlement assistance during their 1st year in Canada by a group of Canadians; iii) refugee claimants, who arrive to Canada unsupported and make a legal claim to refugee status; and iv) refugees who are dependents of a primary refugee applicant. Prior to arrival, the two groups of sponsored refugees were registered with the UN High Commissioner for Refugees (UNHCR) and are chosen for immigration to Canada based on vulnerability. Sponsored refugees become permanent residents and are eligible for provincial health care upon arrival to Canada. Non-sponsored refugees (i.e., refugee claimants) [4,24] are eligible for federally funded health care (administered by the provinces) while they wait for their refugee determination hearing. The proportion of refugee claims approved during the time span of the IRCC-PRD is unknown but recent data indicate approvals have risen from 38.1% in 2013 to 66.1% in 2016 [25]. Successful refugee claimants, who make up the remaining 50% of permanent residents who are refugees, become eligible for permanent residency and for provincial health care once their claim is approved. Unsuccessful refugee claimants are not included in the IRCC-PRD.

Non-refugee immigrants in the IRCC-PRD are predominately skilled immigrants or their family members. Skilled immigrants are selected based on high levels of education, official language fluency, and work experience. Family class immigrants must be related to a permanent resident or Canadian citizen able to provide financial support. Soon after arrival in Canada both groups become permanent residents and are eligible for universal, provincially funded health care.

All immigrants in the IRCC-PRD were subject to an immigration medical exam (IME) during the application process. Prior to 2002 immigration applicants could be rejected if they placed “excessive demand” on health and social services [26]. However, in 2002 the Immigration & Refugee Protection Act (IRPA) [27] came into effect which changed this “excessive demand” criteria so it only applied to skilled immigrants and not family class immigrants or refugees.”

A sentence has also been added to the “Strengths and Limitations” section on page 9 – “Finally, our findings are not generalizable to unsuccessful refugee claimants (since our study was limited to permanent residents) who may be more representative of refugees and asylum seekers in other countries.”

2. Significant proportion of immigrant women giving birth had a duration of residency (since granted permanent resident status) in Canada of 10 years or more - 42% of refugee immigrant women and 28% of non-refugee immigrant women. Is it possible that some of these women came to Canada as young children? If immigrant women arrived as young children what was their status? Were data collected re language and education at immigration for children? How have the authors considered this in statistical analyses?

Yes, some of the refugee and non-refugee mothers arrived as children (<15 years of age) to Canada – in the matched cohort (revised main analyses, now matched for country of birth, year of arrival and age at arrival) about 18% of refugees and 15% of non-refugees.

Of the refugees who arrived <15 years of age, 29% were government assisted refugees, 38% were privately sponsored refugees, 21% were refugee claimants and the remaining 12% were refugee dependents. Of the non-refugees who arrived <15 years of age, 72% were equal parts “spouses/partners” or “spousal dependents of skilled workers”, and another 15% were “sons and daughters” and the remainder entered Canada in various other official immigration categories.

Yes, language and education were collected at arrival for these mothers. We recognize that since these mothers entered Canada as children, they likely enrolled in schools and improved both their language ability and their education therefore what is captured in the IRCC database was likely not accurate at the time they give birth in Canada

However, as indicated above, both refugees and non-refugees have roughly equal proportions of those arriving in Canada before the age of 15. Therefore any misclassification of their education and language ability will affect both groups and wouldn't be expected to change the results. To ensure that this was the case, a sensitivity analysis was run excluding mothers under 15 upon arrival. The maternal and perinatal outcomes that were statistically significant remained the same as those when mothers arriving at all ages were included, with virtually no difference in the effect estimates.

3. The comparison of adverse maternal and perinatal outcomes by the 5 top refugee source countries is of interest. In some other high-income countries (e.g. Australia) over a similar time period to the study, the vast majority of immigrants from Afghanistan are humanitarian entrants. Is it possible that people with a refugee experience such as the majority of Afghan immigrants (but for whatever reason not identified as having refugee status in the Canadian data) are in the immigrant non-refugee group? Additional contextual information may help the reader understand some of this complexity.

Thank you for this comment – unfortunately it does not have a straight-forward answer. At the request of this reviewer, the differences between refugee and non-refugee immigrants in Canada have been explained in greater detail in the methods section (see question 1 above). In this section, we have attempted to explain that refugees and non-refugees are selected for immigration using different criteria. However, this does not mean that these criteria are mutually exclusive. Refugees are chosen based on experiences of persecution but it is also possible they are skilled or have family members in Canada such that they could apply to immigrate as a non-refugee immigrant. The same applies for non-refugee immigrants – they have applied for immigration based on their skill level or the relationship to family members in Canada but this does not mean that they had no experiences of persecution in their country of origin. There is likely a spectrum of both skill levels and experiences of persecution in both the refugee and non-refugee groups. What is most certainly different between refugee and non-refugee immigrants is that sponsored refugees (50% of all refugees) would have been exposed to living in a refugee camp or under refugee-like conditions for some period before coming to Canada (as this is where they would have registered with the UN). Non-refugee immigrants would be much less likely to have these experiences (since most apply for immigration from their country of birth), but again, we cannot rule out the possibility. In the original version of this paper, we have mentioned that non-refugee immigrants may have had “refugee experiences” (see pg 9 in the “implications” section of the discussion). This sentence reads “Secondly, non-refugee immigrants from refugee-source countries may be just as likely to experience pre-departure health risks (related to persecution) as their refugee counterparts, reducing specificity and minimizing any differences between the groups.”

The simple answer to this question is that we need to know more about the experiences of both “refugees” and “non-refugee immigrants” to determine what migration experiences are harmful to health. In the original version of this paper (in the “Future Directions” section, pg 10), this is also mentioned - “Further refining refugee status based on detailed migration experiences would also be beneficial.”

4. Were Indigenous women included in the Canadian-born cohort?

Yes, Indigenous women were included in the “Canadian-born” cohort. We were unable to exclude Indigenous women since the data set which would have made their exclusion possible was not available during the time of the linkage. A report by the Canadian Perinatal Surveillance System in 2008 suggests that Indigenous peoples in Canada have sub-optimal maternal and perinatal health status therefore including Indigenous mothers in this group likely increases the risk of outcomes in the Canadian-born group and potentially biases the risk ratios we report towards the null. However, since ~ 1/6th of Canadian mothers are Indigenous their inclusion is unlikely to change our results or our interpretation substantially. We now mention this in the paper (in “Study design and

inclusion/exclusion criteria”, pg 3) – “Births not linked to the IRCC PRD were attributed to Canadian-born mothers (Indigenous mothers could not be excluded at the time of the linkage).”

5. The sections re implications and future research (page 8) identify a number of issues. It would be helpful to have some contextual information around these issues (Canada’s migration and refugee program; national efforts to tackle health and health care inequalities) in the manuscript.

Thank you for this suggestion. We have added some additional detail in this section of the paper that provides additional context. This section (pg 9) now reads as follows – “Thirdly, all refugees to Canada receive financial, social supports (e.g., housing, resettlement), particularly in the first year after arrival (as described in the method section) as well as universal health care. Specialized primary health care centres catering to the unique health needs of refugees are available [52,53]. There are also national efforts to highlight the need for a focus on equity in the quality of care received and migrant friendly maternity care [54]. These specialized efforts may be helping to minimize potential health inequities experienced by refugees. Lastly, despite official immigration policies, such as the Immigration & Refugee Protection Act (IRPA, 2002)[26,27] (see methods section for more detail), it is possible that unofficial processes select refugees based on factors such as skill level and language fluency (i.e., similar to non-refugee immigrants), effectively selecting for healthy refugees.”

6. The final statement “Overall, based on an administrative definition of refugee, we do not find a strong need to enhance the health care for refugee mothers and their infants in Canada” even with the caveats, requires a re-think. Other aspects of health care can be compromised for refugee women - experience of trauma & torture and impacts on mental health; poor health literacy limiting access and engagement with services. Measuring these outcomes was obviously not within the scope of the study. The authors may wish to consider a possible re-phrasing as it seems difficult to conclude from the research what the health care outcomes are for refugees and immigrants with permanent resident status in Canada?

Thank you for this comment. We agree that a rephrasing is necessary. This section now reads as follows.

“Research has described that refugees and other immigrants in Canada experience barriers to accessing health care[50], had unaddressed health concerns after birth [51] and experienced culturally insensitive policies [50]. Indeed, such health care deficiencies may have contributed to the 1/3rd of outcomes where refugee and non-refugee immigrant mothers experienced greater risk when compared to Canadian-born mothers. By the same token, it is surprising that refugee mothers did not experience an excess of maternal and infant health risks compared to non-refugee immigrants since these health care deficiencies are likely experienced more acutely by refugee mothers. There are a few important caveats to our findings...”

7. This reviewer’s preference is for the language of ‘births’ to be used as an alternative to ‘deliveries’. This is also accepted language of several international perinatal journals with multidisciplinary readership.

> This suggestion has been incorporated throughout the paper.

Reviewer: 3

Reviewer Name: John S. Preisser

Institution and Country: University of North Carolina, U.S.A Competing Interests: I have no competing interests.

Further details on the 1:1 matching are requested by this reviewer.

Record linkage: In this epidemiological study on refugee and non-refugee immigrant health, the authors have linked five datasets pertaining to immigration, health care, childbirth, death, and HIV, respectively, to study maternal and perinatal health outcomes in Ontario Canada. This is an impressive feat and one that the authors document with several citations that assessed the validity of various aspects of the linkage. This is undoubtedly a complex exercise involving a litany of rules; understandably not all the details are provided. Yet, few details of the linking process are stated, only that “unique encoded identifiers” are used. While the full details are well beyond the scope of this article, the reader is left with some basic questions: (1) what are the identifiers used for linkage and 2) what is the general approach to linking datasets, e.g., exact matching on named identifiers or probabilistic matching.

The organization which holds the health care data used in this study (the Institute for Clinical Evaluative Sciences – ICES - in Toronto) was responsible for conducting the linkage between the immigration database and Ontario’s health care registry (known as the Registered Persons Database). Details of this linkage have been published by Chiu et al (in BMC Medical Informatics and Decision Making, 2016). A short summary of the linkage process and the basic questions posed above have now been added to the methods section of the paper on page 4. An expanded version of the explanation added to the paper, is provided here for the reviewer –

“In summary, the Automatch probabilistic record linkage program was used to link the two datasets. Three rounds of deterministic linkages were undertaken using several personal identifiers (i.e., sex, last name, given name, birth date) resulting in 68.2% of all immigrants being deterministically linked to the health care registry. Unmatched records from the deterministic process then were submitted to a probabilistic process which generated definite matches and probable matches, the latter which were reviewed manually. The probabilistic and manual processes resulted in an additional 18.2% records being linked while the remaining 13.6% could not be linked. Bias in the linkage process was investigated by comparing immigration variables between matched and unmatched individual and little bias was detected. ”

For immigrants in the IRCC-PRD, successful linkage to the RPDB permitted the assignment of their unique encrypted health care number (aka - “unique encoded identifier”, which is the language preferred by ICES but this has been further clarified in the text – see page 4). Every Ontarian eligible for universal health care coverage is given a unique health care number which is encrypted for use at ICES and used to link across health care databases held at ICES (healthcare, HIV etc). An additional sentence has been added to clarify this – “The linkage between the IRCC-PRD and the health care registry permitted assignment of an individual’s unique encrypted health care number which facilitated linkage to health care databases used to identify outcomes.”

Matching: The comparison of refugee to non-refugee immigrant health is based on 1:1 matching of first deliveries in Canada based on country of birth, thus producing paired data. Unfortunately, the authors do not explain how the pairs are formed considering that there could be hundreds or even thousands of mothers from a given country and perhaps millions of possible different pairings of mothers. Also, why not also match on other variables besides country of birth such as time (in years) elapsed between the date of becoming a permanent resident and the date of delivery? While matching exactly on variables is one option, it may not be feasible when matching on multiple variables; matching based on propensity score algorithms would identify “best matches” based on practically any desired subset of the study variables. The propensity scores are commonly the predicted probability from a logistic regression for the binary outcome of refugee status as a function of the covariates used for matching. Pairs are formed based on minimizing within-paired differences in predicted probabilities.

Thank you for this comment. Our original approach was very simple, we matched deterministically on country of birth, so the reviewer is correct – for many countries there were hundreds and sometimes thousands of possible matches. Given that this approach was overly simplistic, we took the advice of the reviewer and revised our analyses (objective 1) to match on additional variables. We matched on country of birth, arrival year +/- 5 years and age at arrival +/- 5 years.

It was decided to match on the additional variables of arrival year and age at arrival because matching on these variables serves the purpose of identifying a refugee and non-refugee mother that are as similar as possible in terms of their pre-migration (i.e., contextual) circumstances except their refugee status. In addition to being from the same country, the women would have been a similar age when they left their country (so they would have been exposed to their home country for the same number of years before coming to Canada) and they would have left their home country at approximately the same time (so are more likely to be exposed to similar political and social exposures in their country of birth prior to arriving in Canada).

Matching on these three variables resulted in approximately 85% of refugee mothers being successfully matched (1:1) to a non-refugee mother. This is a very similar proportion to our approach where we just matched on just country of birth. Socio-demographic variables for refugees who were matched and those who were unmatched were compared. For the majority of variables, the frequencies were approximately the same. However, we did find that over half of the unmatched refugees were from Sub Saharan Africa. Given the importance of this region in contributing to refugee populations world-wide, we were concerned that we may be potentially losing information. As a sensitivity analysis, we did a second round of matching where we submitted unmatched refugees and non-refugees from the first process to an additional round of matching where we matched on world sub-regions (as defined by the United Nations – e.g., East Africa, West Africa etc) as well as arrival year and age at arrival (same parameters). With this second round of matching, we matched 99% of all refugees to non-refugees. Comparing risk ratios estimated from the cohort from the first round of matching to those from the cohort from both the first and second round of matching we found the same outcomes to be important with similar magnitudes for the risk ratios. Since the first approach to matching is cleaner and addresses the objective more effectively (i.e., comparing two mothers that are approximately equal other than refugee status) we chose to present the main findings using the first round of matching only in the newly submitted version of this paper. A description of this sensitivity analysis has been included in the Strengths and Limitations section of the discussion (page 8).

Statistical Models: Multivariable log-binomial regression estimated with Generalized Estimating Equations with a working exchangeable correlation for matched pairs is applied to the paired data. Even if the pairs are well formed and the criteria for 1:1 matching is described, the GEE analysis does not provide a matched analysis, but merely accounts for within-pair correlation in the outcome. Matched pair statistical analyses for the several binary outcomes can be conducted with generalized linear mixed models (with matched pair as the random intercept) or using conditional logistic regression.

Thank you for this comment. We have more fully explained our decision to use log-binomial regression with GEE with some additional supporting references in the Analysis portion of the manuscript. This section now reads as follows (pg 6) – “With this matched cohort design, estimating risk ratios is preferred over odds ratios since it is more appropriate to model the risk in a cohort study and because risk ratios do not exaggerate relative risks for common outcomes (which odds ratios tend to do). For these reasons we opted to use log-binomial regression instead of conditional logistic regression which estimates odds ratios. In addition, to compare refugee and non-refugee immigrants within matched pairs (or clusters) we used fixed-effects Generalized Estimating Equations (GEE) with an exchangeable correlation matrix.[31] This was deemed an acceptable approach since analysis of

clustered data using conditional logistic regression and logistic regression with GEE was shown to provide similar estimates and standard errors.[32]"

A minor comment is that I would recommend splitting the final paragraph of the introduction section into two paragraphs, the first containing the rest of the literature review, the second stating what will be done in the manuscript.

Thank you for this suggestion. The paragraph has been divided into two as per this reviewer's suggestion.

In summary, I found the paper quite interesting and I was impressed by the scope of work undertaken, as it relates to the unique compiled dataset and the totality of statistical analyses. Nonetheless, I would like to know a few more details about the linkage and the matching. Finally, I am curious as to the results that a true matched pair analysis would give. Otherwise, the matching and the analysis seem only to accomplish matching on a group (country of birth) level in terms of matching the sample size of the two groups of women within each country, which is perhaps all the authors intended.

Thank you for the kind review, we are pleased that you found the paper interesting. We have taken your suggestion to do a true matched pair analysis by matching refugee mothers to non-refugee mothers on country of birth as well as year and age at arrival (+/- 5 years). We have presented these results in the revised manuscript. The results and interpretation using this revised approach compared to what was initially reported in the manuscript are very similar with some marginally significant outcomes dropping out of significance (very preterm birth) or becoming significant (caesarean section). Gestational diabetes and HIV remain statistically significant while moderate preterm birth became marginally non-significant.

Reviewer: 4

Reviewer Name: Mari Palta

Institution and Country: University of Wisconsin- Madison, Madison, WI, USA Competing Interests: None declared.

Review: Refugee maternal and perinatal health in Ontario Canada: a retrospective population-based study.

This is an interesting and mostly well done study.

>Thank you, we are happy to hear you found the paper interesting!

The manuscript and analyses have some room for improvement in language and clarity. Some specific comments are below:

1. First of all, this is not really a "matched" study, in the usual sense of matching individuals. Unless I missed something, women were matched only on country of birth, which is shared by groups of women. Essentially the investigators eliminated women from the analysis to make the two groups balanced on country of birth. The matching is handled by GEE, which would not have worked on unbalanced data- hence the elimination of subjects I assume. However, simply entering indicator variables for countries of birth and keeping the whole sample would have been preferable and a more explicit approach to matching (or rather "stratification"). This could also have been done via a

conditional logistic analysis. These so called “fixed effect” approaches would also have allowed testing for the hypothesis stated by the authors, (see 3. Below) but not fully addressed.

Thank you for these comments/questions.

Re: matching

This reviewer is correct in the statement that we matched only on country of birth which made the two groups of immigrants balanced on country of birth. Our objective was not merely to adjust for country of birth (which I think is what is being suggested by the suggestion to enter indicator variables for countries of birth) but to identify whether a refugee mother experiences greater risk of adverse outcomes compared to a similar non-refugee mother from the same country of birth. In other words, we aimed to identify whether refugee status itself and not circumstances related to the country of birth, confers greater risk of outcomes. Adjusting for country of birth does not explicitly do this, which is why matching was undertaken. We recognize that our initial approach to matching could be improved so we conducted a new analysis in which we matched refugee mothers and non-refugee mothers to country of birth as well as landing year and age at arrival within 5 years.

This revised round of matching allowed 85% of all refugee mothers to be matched to a non-refugee mother which was very similar to our initial approach. While this proportion of matching is high enough such that the matched refugees are unlikely to be biased, we conducted an additional round of matching where the unmatched refugees and non-refugees from the first round of matching were subject to matching on sub-region of birth (e.g., East Africa). The second round of matching matched an additional 15% of all refugees, leading to 99% of all refugee mothers being matched on COB or sub-region. Analyzing the first set of matches and the first and second set of matches together yielded similar risk ratios and the same interpretation for maternal and perinatal outcomes. Since results were similar but the first set of matches more explicitly addressed our objective, we opted to present only the first set of matches in the revised paper. The second round of matching and the similarity of the results are described as a sensitivity analysis in the discussion section (pg 8).

RE: analytic approach

We now more fully explain our rationale for using log-binomial regression with fixed-effects GEE rather than conditional logistic regression to analyze matched pairs in the Analysis section of the paper (pg 6 of the revised manuscript) – “With this matched cohort design, estimating risk ratios is preferred over odds ratios since it is more appropriate to model the risk in a cohort study and because risk ratios do not exaggerate relative risks for common outcomes (which odds ratios tend to do). For these reasons we opted to use log-binomial regression instead of conditional logistic regression which estimates odds ratios. In addition, to compare refugee and non-refugee immigrants within matched pairs (or clusters) we used fixed-effects Generalized Estimating Equations (GEE) with an exchangeable correlation matrix.[31] This was deemed an acceptable approach since analysis of clustered data using conditional logistic regression and logistic regression with GEE was shown to provide similar estimates and standard errors.[32]”

2. The paper referenced regarding the matched analysis does not seem publicly available and the reference does not include publication venue. The title seems to imply problems analyzing matched data by SAS. It is unclear why this would be without access to the paper. Besides, as noted above the sample here was not really in need of matched analysis.

The explanation as to why a matched analysis was necessary is described in detail in response to question 1 (above). A web address to the SAS paper in question has now been included as part of its

reference (lexjansen.com/nesug/nesug07/sa/sa01.pdf). This SAS paper describes the approach for the analysis of matched cohort data using SAS.

3. The authors state as one of their objectives to investigate whether effect is independent of country. This seems addressed by sub-study of specific countries of birth only. It would have been interesting to see a more complete analysis that testes interaction effects, as the objective seems to anticipate.

The objective of whether refugee status is independent of country of birth was addressed by creating a cohort of refugee mothers and non-refugee mothers matched on country of birth, landing year and age at arrival as well as the sub-study of specific countries. However, we recognize that the statement of the first objective may not be clear enough. Therefore, in the abstract and the paper's introduction the first objective has been reworded as follows:

In the abstract: "Herein, we examined whether: i) a refugee mother experiences greater risk of adverse maternal and perinatal health outcomes compared to a similar non-refugee mother from the same COB, and ii)..."

In the introduction: "Given this background, our first objective was to determine whether a refugee immigrant mother experiences greater risk of adverse maternal and perinatal health outcomes compared to a similar non-refugee immigrant mother from the same country of birth."

The first objective was not intended to imply an interest in interaction effects. Presumably the reviewer is referring to examining an interaction between refugee status and country of birth which was not within the scope of this study. Given that there are well over 100 countries of birth represented in the data, examining interaction effects would be a time-consuming exercise and not necessarily meaningful unless there was a hypothesis related to specific countries of birth.

4. The abstract needs to be more explicit about the findings. I was confused by the statements that refugees and no-refugees differed only on HIV, versus differed on most complications. Before reading the paper, I also did not understand what "in the same direction" refers to.

The abstract results and conclusions have been edited to be more explicit about the findings. These two sections now read as follows:

Results: Refugees differed from non-refugee immigrants most notably for HIV, with respective rates of 0.39% and 0.20% and an ARR of 1.51 (95% CI 1.07-2.11). Other elevated outcomes included gestational diabetes mellitus (ARR 1.07, 95% CI 1.00-1.14) caesarean section (ARR 1.03, 95% 1.00-1.05) and moderate preterm birth (ARR 1.06, 95% CI 0.99-1.15). For most outcomes, refugee and non-refugee immigrants experienced similar ARRs when compared with Canadian-born mothers.

Conclusions: Refugee status was associated with a few adverse maternal and perinatal health outcomes but associations were not strong with the exception of HIV. The definition of refugee status used herein may not sensitively identify refugees at highest risk. Future research would benefit from further refining refugee status based on migration experiences.

5. The writing needs some general editing for language and clarity. In particular, there are too many acronyms. I do not see a reason for why most could not be spelled out. It is especially difficult to read the conclusions with all these acronyms, as many readers may wish to do.

Thank you for this suggestion, we apologize for the confusion regarding the use of many acronyms. The manuscript has been edited throughout to spell out terms not used commonly. If a term is used

frequently in certain parts of the text only, the acronym for this term was reintroduced to improve clarity.

VERSION 2 – REVIEW

REVIEWER	Hannah Dahlen Western Sydney University, Australia
REVIEW RETURNED	06-Dec-2017

GENERAL COMMENTS	Thanks I am satisfied the authors have addressed the suggested changes
-------------------------	--

REVIEWER	John S. Preisser University of North Carolina, U.S A.
REVIEW RETURNED	22-Dec-2017

GENERAL COMMENTS	<p>Record Linkage: In their revised paper, the authors have satisfactorily addressed my questions by giving additional details regarding record linkage.</p> <p>Matching: I like the authors revised approach to form matches based on three variables: country of birth, arrival year +/- 5 years and age at arrival +/- 5 years.</p> <p>Statistical Models: The revised analysis section states “In addition, to compare refugee and non-refugee immigrants within matched pairs (or clusters) we used fixed effects Generalized Estimating Equations (GEE) ...” This is an incorrect statement because GEE applies to regression models where parameters have population-averaged interpretations (not conditional upon the matched pair) regardless of whether you adjust for the correlation with an exchangeable correlation matrix or not at all (i.e., independence structure). So the phrase “within matched pairs (or clusters)” in the aforementioned sentence should be removed.</p> <p>Generally speaking, matching may be used in the design only or in both the design and analysis. In statistical analysis, matched sets (with matched pairs as a special case) are often treated as strata that are specified as fixed or random effects in regression models. Conditional logistic regression is a stratified logistic regression model that includes dummy indicators as fixed effects for the matched sets, rendering interpretations of other regression parameters to be conditional upon the matched set to give within-matched pair comparisons. Because there are many indicator variables (or intercepts), they are eliminated from the conditional likelihood function through conditioning arguments. Nonetheless, the conditional interpretation of regression coefficients for the within-pair covariates remains.</p> <p>Conversely, GEE is a method of estimation of regression models for clustered data where the models are unstratified, i.e., they don't include dummy indicators for matched pairs in the marginal mean (probability, for binary outcomes) model. As I wrote in my original review “GEE analysis does not provide a matched analysis”, however the authors continue to make the erroneous assertion that</p>
-------------------------	---

	<p>they have conducted a matched pair analysis with GEE. They have cited a SAS paper that is a useful resource for applying many SAS procedures to clustered data problems, but a deeper understanding of statistical motivation, models, methodology and interpretation (presumably not the intention of the SAS paper) should be sought elsewhere; there are many sources – two are given below.</p> <p>The authors' GEE analysis of unstratified log-binomial models doesn't take advantage of the extra effort the authors have undertaken to refine their pair matching procedures. I am not referring to question of odds ratios versus risk ratios, which the authors have addressed. I am simply referring to the fact that GEE does not give a matched pairs (or pair-specific) analysis (for within-pair comparisons) but rather it gives a population averaged analysis. On the other hand, a matched pair analysis such as given by conditional logistic regression would have addressed the authors objective as stated in their reply to Reviewer 4: "Our objective was ... to identify whether a refugee mother experiences greater risk of adverse outcomes compared to a similar non-refugee mother from the same country of birth" [and of similar arrival year and age at arrival].</p> <p>Finally, I recommend omitting the last sentence of the first paragraph in the analysis section stating that GEE and conditional logistic regression gave similar estimates and standard errors; the basis of my recommendation is the distinct differences in interpretation of the estimated effects in unstratified and stratified logistic regression models, respectively. While the estimates may be similar numerically, they are estimating different regression parameters nonetheless.</p> <p>Preisser JS, Koch GG (1997). "Categorical data analysis in public health", Annual Review of Public Health, 18:51-82.</p> <p>Stokes ME, Davis CS, Koch GG (2012). Categorical data analysis using the SAS system, 3rd edition, Cary, NC: SAS Institute Inc.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Hannah Dahlen

Institution and Country: Western Sydney University, Australia

Competing Interests: None declared

Thanks I am satisfied the authors have addressed the suggested changes

Reviewer: 3

Reviewer Name: John S. Preisser

Institution and Country: University of North Carolina, U.S A.

Competing Interests: None declared

Record Linkage: In their revised paper, the authors have satisfactorily addressed my questions by giving additional details regarding record linkage.

Matching: I like the authors revised approach to form matches based on three variables: country of birth, arrival year +/- 5 years and age at arrival +/- 5 years.

>We are happy to hear this reviewer likes our revised approach to matching.

Statistical Models: The revised analysis section states “In addition, to compare refugee and non-refugee immigrants within matched pairs (or clusters) we used fixed effects Generalized Estimating Equations (GEE) ...” This is an incorrect statement because GEE applies to regression models where parameters have population-averaged interpretations (not conditional upon the matched pair) regardless of whether you adjust for the correlation with an exchangeable correlation matrix or not at all (i.e., independence structure). So the phrase “within matched pairs (or clusters)” in the aforementioned sentence should be removed.

>As per this reviewer’s suggestion, the phrase mentioned has been removed from the methods section.

Generally speaking, matching may be used in the design only or in both the design and analysis. In statistical analysis, matched sets (with matched pairs as a special case) are often treated as strata that are specified as fixed or random effects in regression models. Conditional logistic regression is a stratified logistic regression model that includes dummy indicators as fixed effects for the matched sets, rendering interpretations of other regression parameters to be conditional upon the matched set to give within-matched pair comparisons. Because there are many indicator variables (or intercepts), they are eliminated from the conditional likelihood function through conditioning arguments. Nonetheless, the conditional interpretation of regression coefficients for the within-pair covariates remains.

Conversely, GEE is a method of estimation of regression models for clustered data where the models are unstratified, i.e., they don’t include dummy indicators for matched pairs in the marginal mean (probability, for binary outcomes) model. As I wrote in my original review “GEE analysis does not provide a matched analysis”, however the authors continue to make the erroneous assertion that they have conducted a matched pair analysis with GEE. They have cited a SAS paper that is a useful resource for applying many SAS procedures to clustered data problems, but a deeper understanding of statistical motivation, models, methodology and interpretation (presumably not the intention of the SAS paper) should be sought elsewhere; there are many sources – two are given below.

>> We thank this reviewer for taking the time to provide this detailed explanation (and provide additional resources) as to why our previous analyses using log-binomial using GEE were not

appropriate for conducting a matched analysis. We appreciate the opportunity to gain a deeper understanding of analysis of matched data.

The authors' GEE analysis of unstratified log-binomial models doesn't take advantage of the extra effort the authors have undertaken to refine their pair matching procedures. I am not referring to question of odds ratios versus risk ratios, which the authors have addressed. I am simply referring to the fact that GEE does not give a matched pairs (or pair-specific) analysis (for within-pair comparisons) but rather it gives a population averaged analysis. On the other hand, a matched pair analysis such as given by conditional logistic regression would have addressed the authors objective as stated in their reply to Reviewer 4: "Our objective was ... to identify whether a refugee mother experiences greater risk of adverse outcomes compared to a similar non-refugee mother from the same country of birth" [and of similar arrival year and age at arrival].

>> We completely agree with this reviewer's explanation of the important differences between log-binomial regression with Generalized Estimating Equations (GEE) and conditional logistic regression to analyze matched data. We have revised our matched analysis (summarized in revised Figures 1 and 2) to instead use conditional logistic regression to address our objective. We trust the use of conditional logistic regression to analyse the matched cohort addresses this reviewer's concerns with our earlier approach.

Finally, I recommend omitting the last sentence of the first paragraph in the analysis section stating that GEE and conditional logistic regression gave similar estimates and standard errors; the basis of my recommendation is the distinct differences in interpretation of the estimated effects in unstratified and stratified logistic regression models, respectively. While the estimates may be similar numerically, they are estimating different regression parameters nonetheless.

>> We have removed the statement which states that GEE and conditional logistic regression give similar estimates and standard errors, as suggested by this reviewer. This paragraph in the methods section now reads – "To estimate whether refugee status increases the risk of adverse outcomes between a refugee mother and a non-refugee mother with a similar pre-migration circumstances (objective 1), we 1:1 matched first births in Canada among refugees to non-refugee immigrants on COB, year of arrival (+/- 5 years) and age at arrival (+/- 5 years). Analyses were restricted to the first delivery in the hospitalization database to prevent matching several births from a single refugee mother to births to more than one non-refugee immigrant mother. We conducted a matched pair analysis using conditional logistic regression."

Preisser JS, Koch GG (1997). "Categorical data analysis in public health", Annual Review of Public Health, 18:51-82.

Stokes ME, Davis CS, Koch GG (2012). Categorical data analysis using the SAS system, 3rd edition, Cary, NC: SAS Institute Inc.

>>For consistency and to ease comparability between matched results now reported using odds ratios (Figures 1 & 2) and unmatched results previously reported using risk ratios (Figures 3 & 4 and all supplementary figures), all regression analyses in the paper have been changed from log-binomial regression to logistic regression so that odds ratios are estimated throughout the paper. The methods, results and discussion sections and all relevant tables and figures reflect this change to estimating odds ratios rather than risk ratios.

VERSION 3 – REVIEW

REVIEWER	John S. Preisser University of North Carolina, U.S.A.
REVIEW RETURNED	06-Feb-2018
GENERAL COMMENTS	The authors have satisfactorily addressed all my concerns